

Article

Automatic Classification of Records and Archives as Data: A Survey of Experiments Using Machine Learning

Eduardo Watanabe^{1,*}, Renato Tarciso Barbosa de Sousa¹¹Faculty of Information Science, Universidade de Brasília, 70910-900 Brasília, Distrito Federal, Brazil*Correspondence: edwatanabe@protonmail.com (Eduardo Watanabe)

Academic Editor: Natália Tognoli

Submitted: 15 February 2025 Revised: 27 August 2025 Accepted: 5 January 2026 Published: 22 April 2026



Eduardo Watanabe is a PhD candidate in Information Science at the University of Brasília (UnB). He holds a Master's degree from the same program, a specialization in Labor Law and Procedure, and a Bachelor's degree in Law from the University of São Paulo (USP). He is a Union Attorney since 2003 and has held leadership positions in judicial litigation, knowledge, information and strategic management at the Attorney General of Union (Advocacia-Geral da União – AGU). Currently, he is on the development team for the SAPIENS system (Sistema AGU de Inteligência Jurídica). His work and research focus primarily on knowledge and information organization, artificial intelligence, public administration, jurimetrics, and results-based management.



Renato Tarciso Barbosa de Sousa is a Full Professor at the University of Brasília (UnB). He holds a PhD in Social History from the University of São Paulo (USP) and a Master's in Library and Information Science from UnB. Currently, he serves as the Director of the Faculty of Information Science, is the Vice-leader of the Research Group on Information and Knowledge Representation and Organization Studies, and teaches in the Archival Science and Information Science graduate programs. With extensive experience in Information Science, his work focuses on archival science, the organization and retrieval of archives, record management, professional training, and public archive policies.

Abstract

This paper investigates the application of machine learning (ML) to the automatic classification of records and archives, framing it as a critical challenge in Knowledge Organization (KO). As digitization creates massive volumes of uncategorized data, the following research question arises: how can fundamental archival principles—such as provenance, original order, and hierarchical description—be translated into this new computational paradigm? This study first synthesizes, based on a multidisciplinary review of archival science, classification theory, KO, computer science, and information science, a proposal of six fundamental guidelines for the responsible application of artificial intelligence (AI) in records and archives. These guidelines connect traditional archival theory with the modern imperatives of trustworthy and explainable AI. Second, we conduct a comparative analysis of 24 published ML experiments, assessing their adherence to these guidelines. Our analysis reveals a significant and troubling disconnect. While most experiments acknowledge the principle of provenance (75.0%), they demonstrate profound neglect of guidelines related to diverse perspectives (25.0%), explainability (16.7%), and, most critically, algorithmic accountability (0.0%). The results indicate that current practices often succeed in basic content categorization but fail in the more sophisticated archival task of preserving archives' evidentiary and relational integrity by treating records as decontextualized data. The study calls for urgently developing an Archival AI Lifecycle—a framework that weaves archival principles, classification theory, and knowledge organization into AI development, safeguarding archival practice's intellectual and ethical integrity in the digital age.

Keywords: machine learning; archival science; automatic classification; knowledge organization; trustworthy AI



1. Introduction

Digitizing records transforms them from texts to be read to data to be mined (Moss et al., 2018; Mordell, 2019). The increasing volume of unstructured, uncategorized, and highly diverse data has challenged manually performed archival classification and evaluation tasks (Sousa, 2022; Vellino and Alberts, 2016). For Greene and Meissner (2005), the accumulation of unprocessed records is a problem exacerbated by traditional approaches to handling collections. After reviewing the literature and archival practices in the United States and the United Kingdom, they conclude that a lack of professional consensus on the minimum components and work metrics results in a waste of resources and an accumulation of unprocessed records. Thus, they propose the concept of “more product, less process” (MPLP) to allocate resources most efficiently for the benefit of the end users of the collection. Trace (2022) points out that the loss of backlog accumulation interrupts the distribution and final consumption of the research process, which constitutes a knowledge infrastructure problem with clear harm to society. Problems include the obstruction of research, the loss of contextual information from records, a limited understanding of the past, increased preservation costs, a loss of public trust, and reduced visibility of the social and cultural impact of collections.

The challenges we identified can be categorized into three distinct orders. The first challenge is the lack of engagement by archivists with the conceptual principles of new information technology (Bunn, 2019), which has consequently limited the application of computational analysis in support of archival practices. The second challenge lies in the technical ignorance among archivists; this lack of knowledge often results in computational tools being perceived as opaque and insufficiently transparent regarding their actual functioning (Mordell, 2019; Amozurrutia et al., 2023). The third challenge involves promoting the proper use of computational thinking within archives to ensure that archival requirements are met and that such endeavors contribute to the emerging transdisciplinary field known as Computational Archival Science (Marciano et al., 2018).

In the literature, several researchers advocate for the use of computational analysis tools, such as machine learning (ML) and natural language processing (NLP), which identify patterns in data, make predictions, automate repetitive tasks, and speed up processing (Rolan et al., 2019; Colavizza et al., 2021; Bunn, 2019; Jaillant, 2022a; Sousa, 2022; Alaoui, 2024). As presented, automatic records classification appears to help manage the accumulation of records processing tasks, with additional benefits such as identifying and signaling confidential and sensitive information and improving data quality. Additionally, Chabin (2020) demonstrates how the use of archival knowledge and diplomatic analysis improves the performance of artificial intelligence (AI) models and thus enriched the corpus of records from the large-scale “yellow vest” social move-

ment in France in 2019. Shabou et al. (2020) combines an archival model with a data mining model to create an automated method for evaluating the relevance of data recorded in different formats and content. In this context, we identified a lack of research that comparatively analyzes the results and methodologies used in experiments that apply ML methods to the automatic classification of records, especially if they use knowledge and instruments from archival science and information science. Therefore, we formulated two research questions:

(1) What are the guidelines outlined in the literature for developing artificial intelligence projects in records and archives?

(2) Do the automatic records and archives classification experiments reported in the literature meet the guidelines identified in objective 1?

This research is fundamentally rooted in Knowledge Organization (KO). We argue that automatic record classification is not merely a technical data processing challenge but a complex KO problem. Traditional archival practices, such as establishing provenance, maintaining original order, and creating hierarchical descriptions, are sophisticated, human-driven methods of classifying and organizing knowledge, that impose intellectual order on chaos, reveal context, and facilitate access. This paper examines how well the fundamental KO principles of archival science are being translated, preserved, or potentially threatened with the intrusion of data-driven machine learning algorithmic classification. Our goal is to provide a critical lens for the KO community to assess the impact of AI on one of its oldest domains of practice.

This article addresses archival principles and their articulation with classification in section 2, describes and defines artificial intelligence and its subdisciplines in section 3, section 4 addresses the application of AI in records management, and section 5 summarizes the methodological procedures and the literature. The results are discussed in section 6 and the conclusions in section 7.

2. Foundational Archival Principles and Their Articulation With Classification Theory

The management, preservation, and accessibility of records and archival rely on core principles developed over centuries of practice. Before examining how Computational Archival Science (CAS) transforms archival paradigms, we must establish how fundamental archival practices embody Knowledge Organization (KO) principles. This chapter explores key archival tenets—provenance, original order, record group conceptualization, and archival description—demonstrating their intrinsic connections to classification theory and practice.

Contemporary archival theory distinguishes between records and archives. Records are documents created or received during organizational activities and preserved as ev-

idence, characterized by their organicity and primary value (administrative, legal, fiscal). Archives comprise records that, having fulfilled their immediate administrative purpose, are selected for permanent preservation due to their secondary value—evidential, informational, and historical (Pearce-Moses, 2005). This distinction is crucial for automatic classification, as context and metadata differ substantially between active records (associated with workflows) and historical archives (associated with series and fonds). This study encompasses records and archives, applying experiments across the continuum—from current records management to permanent archive organization—demonstrating how Knowledge Organization principles apply distinctly at each stage.

Since the nineteenth century, provenance has been archival science's theoretical and practical foundation (Tognoli and Guimarães, 2020). The International Council on Archives defines provenance as the relationship between records and the organizations or individuals that created, accumulated, maintained, and used them during personal or corporate activities (ICA, 2007). The principle's codification through the French *respect des fonds* (in 1841) established that documents sharing common origin must remain together and be organized systematically (Tognoli and Guimarães, 2020).

Provenance operates as the primary classification mechanism in archival knowledge organization. By establishing that records from a single creator constitute a distinct fond, this principle creates the fundamental classificatory level within archives (Tognoli and Guimarães, 2020). Based on creator-record relationships, this grouping exemplifies core KO practice, delineating knowledge spheres corresponding to creators' functions and activities. The principle's classificatory power generates meaningful intellectual boundaries reflecting the administrative realities of record creation and the epistemological frameworks structuring original knowledge production.

Original order (*Registraturprinzip* in German, *respect de l'ordre primitif* in French) is intrinsically linked to provenance and emerged from European archival traditions (Tognoli and Guimarães, 2020). This principle mandates maintaining records in the sequence and establishing configuration creators during operational activities, revealing how creators actually operated—their working methods, administrative priorities, decision-making processes, and information management approaches.

The principle of the original order is articulated with the organization of knowledge by being responsible for validating the classification system of its creator, in which it recognizes the original classification scheme as a legitimate and significant structure essential to delimit the context and integrity of the documents. These arrangements represent indigenous classification schemes and filing methodologies developed to meet specific operational requirements (Tognoli and Guimarães, 2020). Archival respect

for original order acknowledges existing knowledge organization systems emerging from particular functional contexts. Analyzing creator-imposed structures yields insights into organizational activities, decision-making frameworks, and intended intellectual relationships among documents—central concerns in KO theory and practice.

The scale and complexity of fonds often necessitate subdivision for effective management and access. Archivists typically organize fonds into series, which are records structured according to a filing system or maintained as a unit due to originating from the same accumulation, filing process, or activity (ICA, 2007). The series formation may reflect the creators' original arrangements or archivists' subsequent organization based on function, subject, or format.

Delineating fonds and series constitutes explicit classification and hierarchical knowledge organization. The progression from fonds through series (and potentially sub-series, files, and items) creates a classification system organizing records into intellectually coherent units. Classification criteria—administrative function, operational activity, subject matter, or documentary form—derive from the creation context and functional use. Subdividing fonds requires systematic analysis to identify logical groupings and structural relationships, representing fundamental KO activity that reveals internal architecture while maintaining provenance-based integrity.

Archival description creates comprehensive representations that facilitate record identification, management, understanding, and use (ICA, 2007). Duranti (1993) characterizes description as analysis, identification, and organization for “control, retrieval and access”, producing representations that illuminate “archival material, its provenance and documentary context, interrelationships and how it can be identified and used”.

According to Vital and Brascher (2016), archival description and classification are central and inextricably linked activities within information organization and representation. Description structures and presents documentary information and acts as a mediating mechanism between the holdings and their users, enabling efficient management and retrieval. The researchers contend that organizing archival materials fundamentally depends on systematic observation, analytical examination, and synthesizing activities designed to discern commonalities and distinctions among documentary records. At its core, this methodology represents a classification-based approach that creates significant connections and meaningful differentiations throughout archival collections.

Description is the primary mechanism for communicating archival classification and organization to users, demonstrating profound KO integration. Finding aids—description's primary output—typically mirror provenance and original order principles, reflecting hierarchical classification while articulating each unit's intellectual and func-

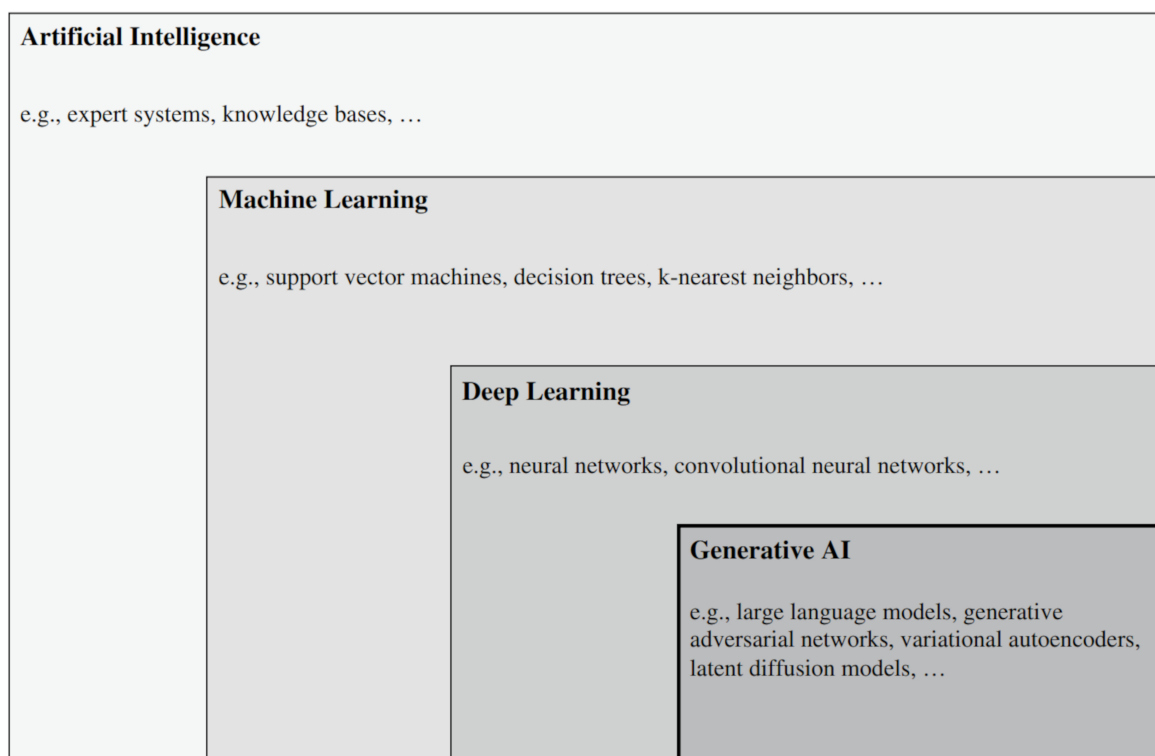


Fig. 1. AI concepts. Reprinted from [Banh and Strobel, 2023](#), p. 63. AI, Artificial intelligence.

tional basis. Complex knowledge organization systems embedded within archives become navigable and comprehensible through descriptive apparatus.

These foundational principles reveal deep integration with knowledge organization theory. Provenance establishes primary classificatory boundaries; original order preserves organization systems; hierarchical structuring creates manageable intellectual units; description renders organizational systems accessible. This intrinsic relationship between archival principles and KO provides theoretical grounding for understanding how computational approaches simultaneously build upon and transform traditional organizational frameworks—a transformation examined in subsequent sections.

3. Computer Science and Artificial Intelligence

AI is a discipline that researches and develops mechanisms and applications of AI systems, considered those that generate content outputs, predictions, recommendations or decisions to meet human-defined objectives ([ISO/IEC, 2022](#)). The evolution of AI has been marked by the emergence of the subdisciplines illustrated in [Fig. 1](#).

The first AI systems were expert systems and knowledge bases that depended on humans to provide established rules to generate their results ([Banh and Strobel, 2023](#)). The advent of ML as a subdiscipline of AI has profoundly changed the development of algorithms capable of solving

tasks autonomously, as they rely on exposure to data without the need for explicit programming, as shown in [Fig. 2](#).

For automatic text classification, model training data is collected and undergoes cleaning and pre-processing steps to then be submitted to feature engineering. Feature engineering (1) processes the raw data into features that can be understood by the algorithms and (2) transforms them into resources that help algorithms achieve better results in terms of prediction and/or interpretability ([Verdonck et al., 2021](#)). [Fig. 3](#) illustrates the feature engineering for classification.

[Mosqueira-Rey et al. \(2023\)](#) reviewed the literature on new interactions between humans and machine learning algorithms, referred to as Human-in-the-loop machine learning (HITL-ML), where the goal is not solely to achieve higher accuracy or speed, but also to make humans more effective and efficient. The main approaches of Human-in-the-loop machine learning share interactivity as a key factor but have different degrees and purposes of this interactivity ([Mosqueira-Rey et al., 2023](#)). Active Learning, Interactive Machine Learning, and Machine Teaching focus on controlling the learning process, while Curriculum Learning and Explainable AI concentrate on improving the efficiency and transparency of learning.

3.1 Trustworthy AI and Explainable AI

The rapid rise of AI has raised concerns about its potential risks and negative impacts. The increasing complexity of AI systems with deep learning algorithms and more

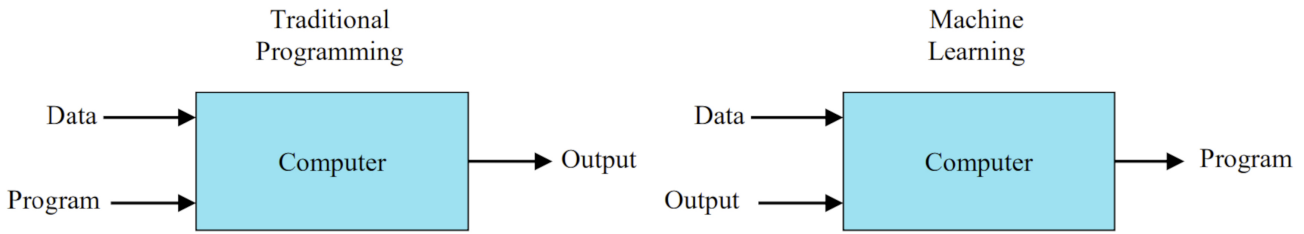


Fig. 2. Differences between Traditional Programming and Machine Learning. Reprinted from Shyam and Singh, 2021, p.19.

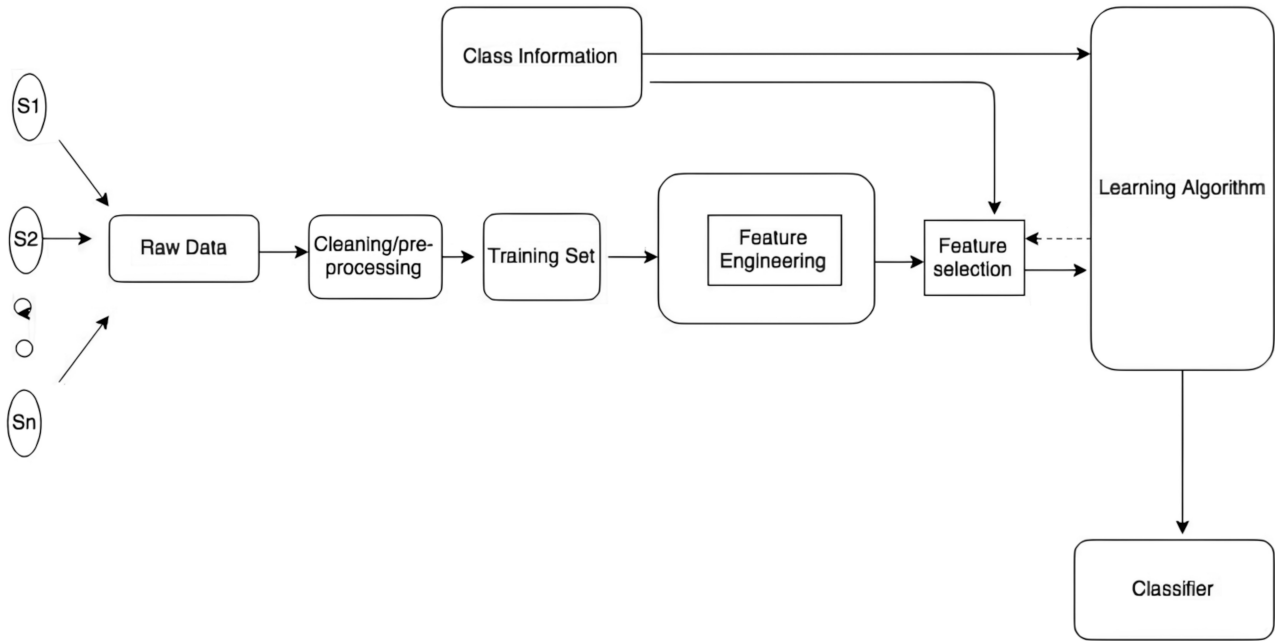


Fig. 3. A general framework of feature engineering for classification. Reprinted from Rawat and Khemchandani, 2017, p. 170.

recently Generative AI makes it difficult to understand their internal workings and ensure their reliability. Cases like the investigation surrounding the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), which assessed the risk of reoffending and was used by judges in decision-making, revealed that the software generated higher false positive rates for African Americans compared to Caucasian offenders (Mehrabi et al., 2022).

It is in this context that Trustworthy AI emerges, a framework created to ensure that an AI system is worthy of trust based on evidence related to its declared requirements, so that user and stakeholder expectations are verifiably met, as defined by the ISO/IEC 24027:2021 standard (ISO/IEC, 2021). Trustworthy AI studies encompass aspects such as human agency and oversight, robustness and safety, privacy and data governance, transparency, diversity and fairness, social and environmental well-being, and accountability (Chamola et al., 2023).

To ensure Trustworthy AI, human participation is essential at all stages of the AI lifecycle, in what can be called a “Human-Centered Approach to Trustworthy AI (Human + AI)” (Kaur et al., 2023). This collaboration between hu-

mans and machines aims to combine human cognitive capacity with the computational power of machines, creating more robust, ethical, and efficient systems.

The explainability of an AI system consists of its ability to express the important factors influencing the outcomes it produces in a way that humans can understand (ISO/IEC, 2022). It is linked to reliability (the property of consistent behavior and results), robustness (the system’s ability to maintain its performance level under any circumstance), transparency (the property of the system to provide due information to stakeholders), and predictability (which should allow stakeholders to make reliable assumptions about the outputs) (Kale et al., 2023).

Provenance in the AI context provides a detailed record of the origin, processing, and transformations of data in a system, answering the “who, what, when, where” questions about each step of a process (Kale et al., 2023). In Fig. 4, the topics of provenance, Explainable AI and Trustworthy AI are presented and their reciprocal relationships.

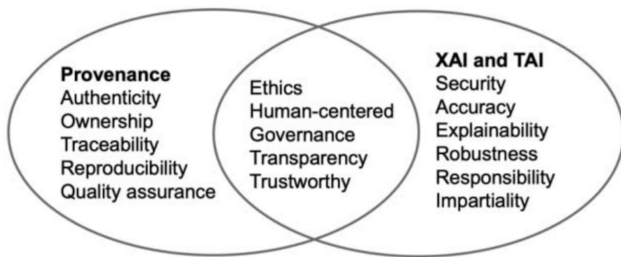


Fig. 4. Similarity of topics involving AI provenance, explainable AI, and trustworthy AI. Reprinted from [Kale et al., 2023](#), p. 151. XAI, Explainable AI.

3.2 Methodologies for Developing Projects Using Data Science and AI

The development of projects to extract useful knowledge from large volumes of data had its initial landmark with the creation of the term Knowledge Discovery in Databases by [Piatetsky-Shapiro \(1990\)](#). However, it was only five years later that the Knowledge Discovery in Databases Process was published, a non-trivial and iterative process with the objective of identifying valid, new, potentially useful, and understandable patterns in data. The process converts raw data into knowledge by selecting relevant data, preprocessing it for quality, transforming it into suitable formats, applying data-mining techniques to extract patterns, and finally evaluating these patterns to derive meaningful insights.

[Martinez-Plumed et al. \(2019\)](#) developed a graphical representation of the evolution of data science methodologies in [Fig. 5](#).

According to [Fig. 5](#), the KDD Process had a very relevant role, but with the advent of the Cross-Industry Standard Process for Data Mining (CRISP-DM), most of the major methodologies since then have been directly influenced by the latter. CRISP-DM organizes the process into six phases.

The main contributions to CRISP-DM were based on practical experience, as it was developed by industry leaders from real projects, which is why it became a robust and effective model in the real world ([Shearer, 2000](#)), as illustrated in [Fig. 6](#). The model also encourages the adoption of best practices by describing the tasks and subtasks within each phase, highlighting the importance of business understanding, data quality, and evaluation of results.

On the other hand, the TDSP (Team Data Science Process) is an agile and iterative methodology created by Microsoft from CRISP-DM to support solutions for predictive analysis and intelligent applications ([Microsoft, 2024](#)). Its main components are the Data Science Lifecycle, the standardized project structure, the recommended infrastructure and resources for data science projects, and recommended tools and utilities for project execution, as illustrated in [Fig. 7](#).

The Domino DS Lifecycle is an agile methodology developed by Domino Data Lab in 2017 based on CRISP-DM, which adopts a holistic approach covering the entire lifecycle from ideation to delivery and monitoring ([Domino Data Lab, 2017](#)).

Finally, the Agile Data Science Lifecycle is a framework created by [Jurney \(2017\)](#) for rapid prototyping, exploratory data analysis, interactive visualization, and applied machine learning, from the perspective of utilizing data science through web applications.

The author elaborates the Agile Data Science Manifesto with principles that support the framework. The first is “Iterate, iterate, iterate”, which emphasizes the need for data to be analyzed, formatted, classified, aggregated, and summarized before being understood ([Jurney, 2017](#)).

[Silva and Alahakoon \(2022\)](#), researchers at the Centre for Data Analytics and Cognition, created the CDAC AI Life Cycle to cover all stages of AI development, from conception to production. Due to the increasing sophistication and incorporation of AI in all forms of digital systems and services, CDAC aims to address risk assessment issues related to: Privacy; Cybersecurity; Trust, interpretability, explainability, and robustness; Usability; and Social Implications.

The CDAC is composed of three main phases that unfold into 19 stages, as shown in [Fig. 8](#).

4. Records Management and Artificial Intelligence

The high volume of digital records makes so-called “close reading” unfeasible, which requires archivists to adopt a macroscopic approach supported by computational methods to automate the processing, organization, and analysis of digital records ([Ranade, 2016](#); [Moss et al., 2018](#)).

ML automatically associates records with archival categories, and for this the archivist does not need to specifically define the characteristics of each category was done in a “capstone project” in which emails were categorized according to functional rules and user accounts ([NARA, 2014](#)).

However, the statistical modeling inherent in ML, while presenting the advantage of simplifying relationships between input variables and responses ([Edmond et al., 2022](#)), is not an exact emulation of nature and remains far from objective. Furthermore, the application of NLP in the organization and description of digital files, as noted by [Mordell \(Mordell, 2019\)](#), often involves opaque interfaces in computational tools, which may lead users to perceive these tools as neutral, disregarding their historically conditioned nature.

The judicious selection of computational tools, considering their functionality, scope, customization options, and data types, is crucial ([Underwood and Marciano, 2019](#)). Problems must be reformulated to align with existing computational tools, involving decomposition, rephrasing into

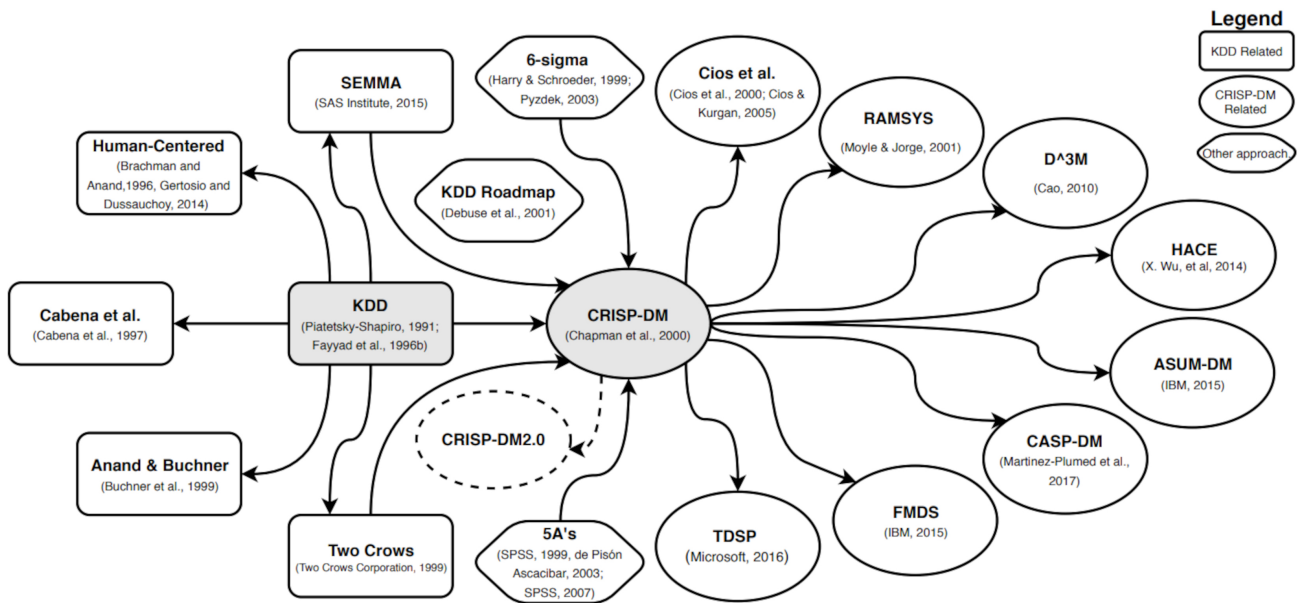


Fig. 5. Evolution of the most important models and methodologies of Data Mining and Data Science. Reprinted from [Martinez-Plumed et al., 2019](#), p. 4. KDD, Knowledge Discovery in Databases. CRISP-DM, CRoss-Industry Standard Process for Data Mining. SEMMA, Sampling, Exploring, Modifying, Modeling and Assessing. RAMSYS, RAPid collaborative data Mining SYstem. D^3M , Domain-Driven Data Mining. ASUM-DM, Analytics Solutions Unified Method for Data Mining. CASP-DM, Context-Aware Standard Process for Data Mining. FMDS, Foundational Methodology for Data Science. TDSP, Team Data Science Process. 5A's, Assess, Access, Analyze, Act and Automate.

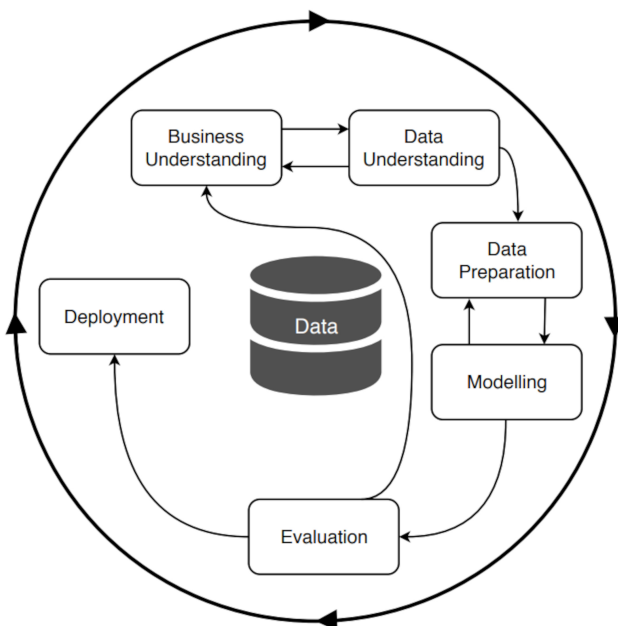


Fig. 6. Phases of CRISP-DM. Reprinted from [Martinez-Plumed et al., 2019](#), p. 2. CRISP-DM, Cross-Industry Standard Process for Data Mining.

known problems, and simplification. A fundamental understanding of how a computational model represents a phenomenon is essential, including the aspects faithfully modeled, those simplified, and the underlying assump-

tions made. It is also critical to acknowledge that algorithms are formalized opinions in mathematics; they reflect the choices, biases, and intentions of their creators, which can perpetuate and even amplify social injustices ([O'Neil, 2020](#)). Consequently, increased transparency and accountability in their development and implementation are paramount.

The consensus, as highlighted by [Amozurrutia et al. \(2023\)](#), is that automated tools should not replace human activity, but rather support it. While AI systems may be capable of generating administrative metadata, as suggested by [Cushing and Osti \(2023\)](#), the necessity for human intervention remains in providing context and understanding events. Therefore, the role of AI lies in supporting human work through a human-machine partnership, where humans retain the ability to verify AI technology outputs for potential issues before making collections available to users. Furthermore, initiatives such as crowdsourcing, as proposed by [Ranade \(2018\)](#), are important for accommodating different, and potentially conflicting, interpretations of documents. This innovative descriptive practice aims to present raw data alongside derived data, qualified with essential confidence measures, promoting a shift in archival thinking based on technical advancements.

The "Traces Through Time" project at The National Archives of the United Kingdom demonstrates how computational techniques can transform access to documents, by recognizing connections between individuals in the col-

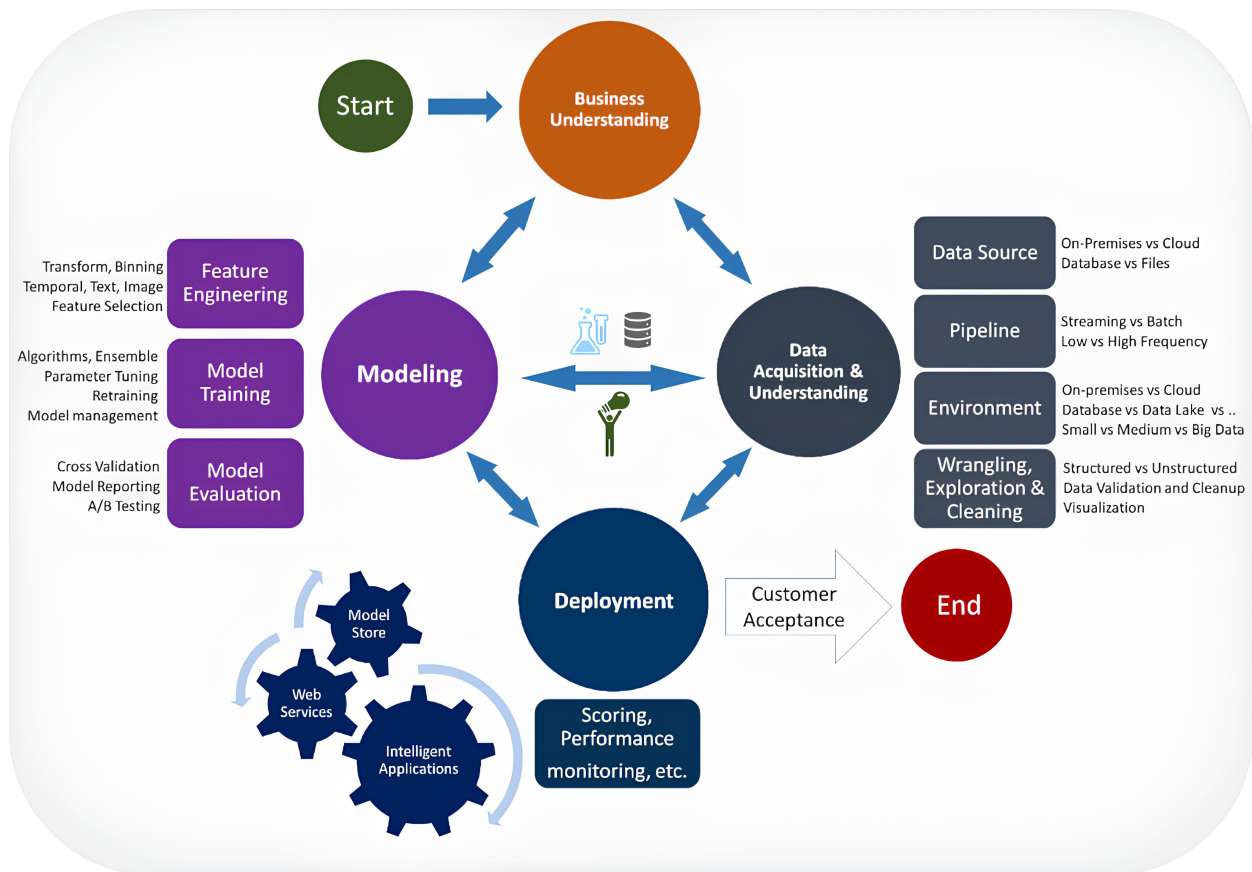


Fig. 7. TDSP data science lifecycle. Reprinted from Microsoft, 2024. TDSP, Team Data Science Process.

lections and automatically transcribing paper documents (Ranade, 2016). This effort highlights the challenge of enhancing computational techniques and addressing the implications of their probabilistic nature and inherent uncertainty for archives, which requires a fundamental shift in archival thinking.

New approaches should focus on improving communication and transparency. Another critical challenge is to ensure that AI usage aligns with core archival concepts, such as archival bonds (Sullivan, 2023). The French government's use of AI tools to organize information from over 1.5 million contributions in the 2019 "Great National Debate" illustrates how incorporating archival knowledge and diplomatic analysis, particularly the formal elements of contributions, can enrich the document corpus and enhance AI model performance (Chabin, 2020). Indeed, according to Chabin (2020), the diplomatic approach allows for demonstrating the reliability of the material, offering a critical assessment of its representativeness, understanding what is expressed beyond words, and considering the production context including place, time, silences, repetitions, and discourse organization.

The implementation of responsible practices to manage biases is crucial, through symposiums, best practice

exchanges, the formation of committees, and the auditing of methods used (Padilla et al., 2019). Moreover, it is vital to ensure the participation of underrepresented and marginalized sectors of society, as proposed by Punzalan and Caswell (2016) as a key aspect of social justice in archival studies.

Computational Archival Science

The CAS is a transdisciplinary field that applies computational methods and resources, design standards, socio-technical constructs, and man-machine interaction to the processing, analysis, storage, long-term preservation, and access problems of large-scale documents and archives (big data) (Marciano et al., 2018). CAS originated from a series of workshops held between 2013 and 2015 on Big Humanities Data in collaboration with King's College London at the IEEE Big Data Conference (Marciano et al., 2024).

According to Marciano et al. (2018), the definition of CAS presented is still provisional and in constant development due to its recent origin. Transdisciplinarity demands a bidirectional exchange of knowledge among the foundational disciplines, an aspect that the current definition does not yet address satisfactorily. The goals of CAS are to enhance and optimize efficiency, authenticity,



Fig. 8. CDAC AI life cycle. Reprinted from [Silva and Alahakoon, 2022](#), p. 6. CDAC, Centre for Data Analytics and Cognition AI life cycle. ML, Machine Learning.

veracity, provenance, productivity, computation, informational structure and design, accuracy, and man-machine interaction in support of acquisition, appraisal, arrangement and description, preservation, communication, transmission, analysis, and access decisions. As an important contribution, the authors emphasize the importance of examining the “computational theories and methods that dominate document practices” ([Marciano et al., 2018](#), p. 179).

For [Wing \(2006\)](#), computational thinking is a fundamental skill and a form of thought that involves multi-layered abstractions, the decomposition of a complex task into smaller parts, heuristic reasoning in the discovery and resolution of various problems, even outside the field of computer science. Computational thinking is characterized by its reflection on multiple levels of abstraction, in a way that resembles human rather than computer thinking.

Building on the contributions of [Weintrop et al. \(2016\)](#), CAS seeks to integrate computational thinking with

archival thinking in academic teaching and archival practice ([Marciano et al., 2018](#); [Underwood et al., 2018](#)). According to the characteristics of each research project, relevant computational thinking practices illustrated in the taxonomy in [Fig. 9](#) are selected and applied, in an effort to structure computational thinking applied to digital archives ([Marciano, 2022](#)).

[Weintrop et al. \(2016\)](#) consider the growing trend of research in science and mathematics being strongly influenced by computing; they seek to contribute a theoretical foundation for computational thinking to be incorporated into teaching according to the taxonomy presented in [Fig. 7](#).

From the work presented at the 2017 CAS Workshop, [Marciano \(2022\)](#) mapped archival concepts with computational methods in [Table 1](#).

For [Mordell \(2019\)](#), CAS represents a crucial junction that unites two trajectories of archival theory and practice that have so far operated on segregated planes, which are so-

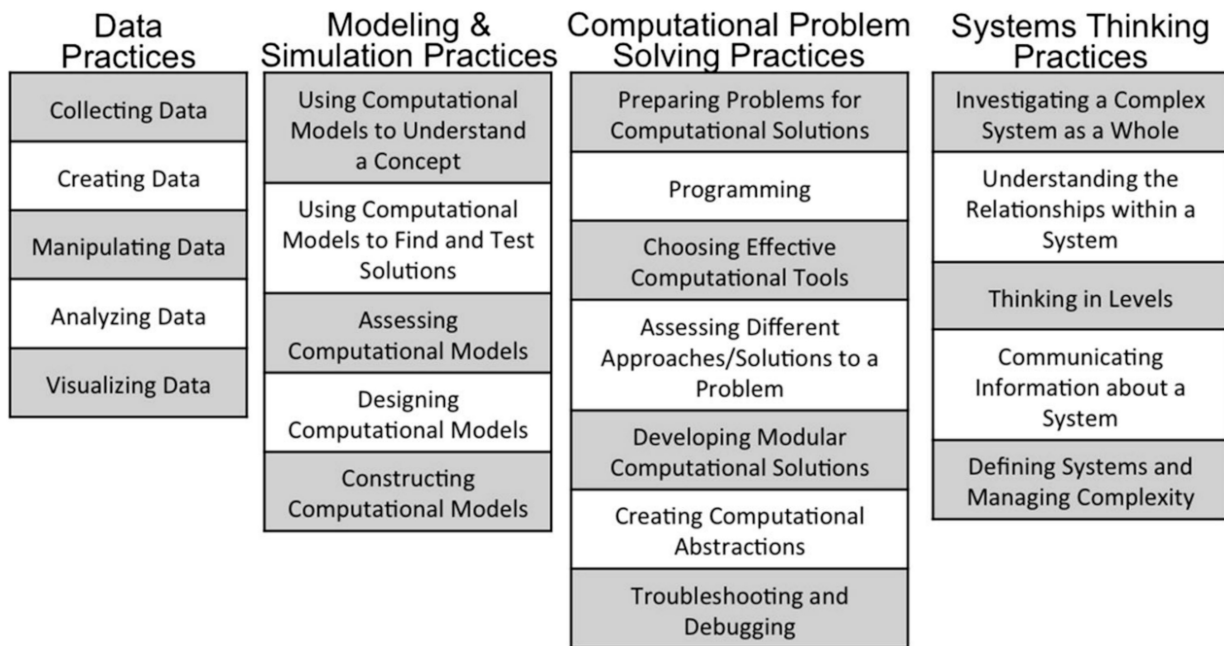


Fig. 9. Taxonomy of computational thinking practices in mathematics and sciences. Reprinted from Weintrop et al., 2016, p. 9.

Table 1. Archival concepts and computational methods.

Archival concepts	Computational methods
Transitioning from paper catalog entries to digital catalogs, matching records in distributed databases	Graph and probabilistic databases
Technology-assisted review accessibility for presidential and federal emails accessed in National Archives	Analysis, predictive coding to address personal data
Provenance in terms of why, who, and how	Abstraction and ontology construction
Appraisal	File format characterization, file format policies, bulk extractor (identifies personal data), content visualization, tagging
Classification of archival images	AI, line detection, image segmentation
Document Management	Auto-categorization, auto-classification, e-discovery, machine learning
Personally Identifiable Information (PII)	NLP, entity recognition, sentiment analysis
Structured data interfaces for archival materials	APIs for cultural heritage materials, graph databases
Decentralized recordkeeping	Blockchain, secure computing, reliability

Source: adapted from Marciano, 2022.

AI, artificial intelligence; NLP, natural language processing; APIs, Application Programming Interfaces.

cial justice issues and technology. And Pajares et al. (2023) highlight the potential of Computational Archival Science (CAS) with the use of NLP, including creating new datasets, detecting problems in digitized documents, enriching data management, producing quality information for decision-making, and improving document management systems.

5. Methodological Procedures

The present study adopts a qualitative research design that combines literature review and comparative analysis to address two research questions: (1) identifying and synthesizing existing guidelines for AI projects in archives from diverse literature sources and (2) assessing adherence to re-

ported automatic records classification experiments to these synthesized guidelines.

Comparative analysis examines relationships between macro-level cases to explain differences, similarities, and contextual relations beyond single instances (Esser and Vliegenthart, 2017). This method enables reflexive understanding through contextual comparison, discovers variations across units, and provides means to test and develop causal theories by assessing empirical limits (Azarian, 2011).

Within this research, comparative analysis serves dual functions:

- Guideline Development (RQ1): A comparative synthesis of literature from computer science, archival science,

and information science identified recurring themes, critical concerns, and best practices. These elements were analyzed and integrated to formulate coherent guidelines (Section 5.1), establishing literature-grounded benchmarks for evaluation.

- Experiment Evaluation (RQ2): The developed guidelines served as an analytical framework to examine ML experiments for automatic record classification. Each experiment was analyzed according to the guidelines and respective verification rules to assess its adherence. This evaluation reveals how the theoretical considerations translate into practice, which allows identifying strengths, limitations and opportunities for improvement.

The methodological choice addresses the research gap highlighted in the literature on comparative evaluations by structuring the investigation, synthesizing diverse information, and, thus, allowing the systematic evaluation of theory and practice (Esser and Vliegenthart, 2017).

5.1 Research Design and Execution

This study employs a two-phase qualitative research design to address its research questions. The first phase focuses on developing a set of guiding principles through a literature synthesis, while the second phase uses these principles as a framework for a comparative analysis of published machine learning experiments.

5.1.1 Phase 1: Guideline Development From Literature Synthesis

The initial step was a comprehensive bibliographic survey to build a corpus of relevant literature. The bibliographic survey included the following databases: Information Science Abstracts databases - LISA and Web of Science (from 1969 to 03/24/2023 exclusively with items reviewed by experts) and the Information Science Database - BRAPCI (from 1972 to 03/24/2023). The search operators used were: (recordkeeping OR “records management” OR archivistic OR “archival science”) AND (“artificial intelligence” OR “machine learning” OR “natural language processing”); (recordkeeping OR “records management”) AND (“auto-categorization” OR “auto-classification” OR “automatic classification”); appraisal AND archival AND (“artificial intelligence” OR “machine learning” OR “natural language processing”); “computational archival science” OR “archival engineering”.

We retrieved $n = 325$ items, of which 20 duplicates were excluded, and 235 were removed because they were not specific, leaving 70 total. Non-specific items include the incidental mention of search terms, for example, articles with the theme of blockchain technology in which record managements and AI are incidentally mentioned without any relationship between the concepts. Through reviewing the articles, we accessed another 131 unique references, totaling 201, of which 57 were used for final analysis.

The 57 selected documents were then subjected to qualitative content analysis to synthesize the guidelines. This synthesis produced six core guidelines for applying AI in automatic records classification.

5.1.2 Phase 2: Comparative Analysis of ML Experiments

The second phase involved a comparative analysis of practical ML applications. From the literature corpus established in Phase 1, we identified 24 published experiments that applied ML to automatic records classification. Selection was based on direct relevance and the provision of sufficient methodological detail to allow for assessment.

Each of the 24 experiments was systematically evaluated against the six guidelines using the corresponding verification rules. Adherence was assessed on a binary basis (i.e., the guideline was either met or not met), based on the information provided in each published article. While this approach does not capture the degree of adherence, it provides a clear macro-level view of current practices. The results of this comparative analysis were then aggregated and analyzed quantitatively to identify overarching patterns and gaps, informing the Discussion and Conclusions of this paper.

6. Results

6.1 Guidelines for Using AI in Automatic Records and Archival Classification

The guidelines, which address Research Question 1 (RQ1), were developed from the literature review and synthesis process detailed in Section 5.1.1 (Phase 1: Guideline Development from Literature Synthesis). This process revealed six recurring thematic areas that emerged consistently across the literature in computer science, archival science, and information science, which are detailed in Table 2.

To synthesize our findings, we refined each theme through several iterations using a systematic approach. We began by coding relevant passages from our selected literature and then compared these codes and concepts across three key disciplines: archival science, computer science, and information science. This analysis across disciplines revealed overlaps and distinctions, uncovering insights that may have been overlooked from a single disciplinary perspective.

Transforming our thematic analysis into practical guidelines required a structured methodology. Let us use “Human Expertise Integration” as a practical example of how this approach works. This theme ultimately became Guideline 1 after synthesizing converging observations from multiple fields. Archival literature repeatedly highlights how archivists’ tacit knowledge and interpretive skills remain essential for truly understanding records. Computer science research has demonstrated promising results with Human-in-the-Loop machine learning approaches, highlighting domain expertise’s impor-

Table 2. Thematic areas identified through literature analysis: frequency and supporting references.

Id	Thematic area	Frequency of occurrence	Key supporting literature
1	Human expertise integration	14	Bell and Bunn (2022), Rolan et al. (2019), Alaoui (2024), Lee (2018), Frendo (2007), Mordell (2019), Edmond et al. (2022), Cushing and Osti (2023), Amozurrutia et al. (2023), Pajares et al. (2023), Ranade (2016; 2018), Chabin (2020), Sullivan (2023) and Jo and Gebru (2020)
2	Contextual preservation	11	Tognoli and Guimarães (2020), Edmond et al. (2022), Jaillant et al. (2022a), Vellino et al. (2016), Anderson (2021), Frendo (2007), Chabin (2020), Ranade (2018), Sousa (2022), Bell and Bunn (2022) and Colavizza et al. (2021)
3	Data quality concerns	10	Shearer (2000), Amozurrutia et al. (2023), Pajares et al. (2023), Rolan et al. (2019), Edmond et al. (2022), Cushing and Osti (2023), Bell and Bunn (2022), Alaoui (2024), Jo and Gebru (2020) and Padilla et al. (2019)
4	Stakeholder diversity	11	Bell and Bunn (2022), Alaoui (2024), Edmond et al. (2022), Chabin (2020), Mordell (2019), Padilla et al. (2019), Marciano et al. (2024), Jaillant (2022a), Jo and Gebru (2020), Cushing and Osti (2023) and Punzalan and Caswell (2016)
5	Ethical accountability	13	Hodel (2022), Alaoui (2024), Edmond et al. (2022), Mordell (2019), Jaillant (2022a), Amozurrutia et al. (2023), Bell and Bunn (2022), Jo and Gebru (2020), Lee (2018), Ranade (2018), Padilla et al. (2019), Edmond et al. (2022) Cushing and Osti (2023)
6	Interpretability requirements	15	Edmond et al. (2022), Marciano (2022), Hodel (2022), Bell and Bunn (2022), Mordell (2019), Pajares et al. (2023), Amozurrutia et al. (2023), Alaoui (2024), Jo and Gebru (2020), Jaillant and Caputo (2022b), Rolan et al. (2019), Underwood et al. (2018), Cushing and Osti (2023), Anderson (2021) and Colavizza et al. (2021)

tance, especially in the model development and validation phases. From the information science perspective, we found strong support for collaborative system design and user-centered approaches involving key stakeholders—particularly archivists.

We applied this same systematic synthesis process to all six themes from our analysis, resulting in the comprehensive guidelines presented in the following sections. Each guideline brings together insights from multiple disciplines, drawing on recurring themes we observed, critical concerns raised by researchers, expert recommendations from the field, and foundational principles from both AI governance frameworks and archival science. We grounded our guidelines in this multidisciplinary literature base to create theoretically sound and practical recommendations for organizations looking to implement AI in archival settings.

6.1.1 Integrate Professionals or Professors in Archives or Records Management

The rise of AI, driven by machine learning models, offers new possibilities for the treatment of archival documents. However, it is crucial to recognize that machine learning should not be viewed as an automated solution that dispenses with human reasoning (Bell and Bunn, 2022). Despite their capacity for data processing and mathematical

calculations, these models lack the understanding of reality and the semantic meaning intrinsic to documents (Bell and Bunn, 2022; Rolan et al., 2019).

Alaoui (2024) argues that while AI can automate specific tasks, the archivist’s role evolves into that of an “archivist-informatician”, essential for continuously validating AI outputs and ensuring contextual integrity, a perception echoed by Lee (2018) who sees AI as a support rather than a replacement for appraisal decisions. Frendo (2007) highlights that the intellectual control offered by traditional file plans, a form of human expertise, is at risk if AI systems focus solely on discrete metadata without archivist guidance. The author considers that planning and contextual reasoning are uniquely human skills that must be considered to give meaning to documents.

Mordell (2019) emphasizes that AI tools are not neutral, and their application to archival description requires acknowledging how they privilege certain records and how assumptions (e.g., based on race and gender) may be encoded into their design, necessitating archivist involvement in adapting or developing these tools. The intellectual work of establishing provenance, respecting original order, and creating hierarchical structures like series represents expert knowledge organization that AI must leverage, requiring deep archivist involvement.

In this context, AI should be seen as a complementary tool to the work of archival specialists, valuing their knowledge and experience in a human-machine collaboration (Edmond et al., 2022; Cushing and Osti, 2023; Amozurrutia et al., 2023; Pajares et al., 2023). The analysis of archival documents, by its nature, involves interpretation and subjectivity, essential skills possessed by professionals in the field (Edmond et al., 2022). The speed with which AI models can generate responses may sometimes obscure more complex issues. Therefore, a deep engagement of the experiment team members is essential to understand the nuances and particularities inherent to archival documents (Ranade, 2016; Ranade, 2018). Chabin (2020) powerfully illustrates this with the French great national debate, where archival expertise was vital to enrich AI-processed data by providing context that algorithms alone would miss. Sullivan (2023) highlights that expert knowledge, particularly in feature selection, can significantly enhance AI and stresses the need for AI tools to conform to archival concepts.

Given this, the tacit knowledge of archivists assumes a fundamental role and must be integrated into all phases of the development and application of the technology (Ranade, 2016). This integration should permeate the entire classification process, from planning to implementation (Edmond et al., 2022), including the design, training, and validation of AI models. The active participation of archivists in these stages is crucial to ensure accurate, contextualized results that are aligned with the specific needs of the collection (Amozurrutia et al., 2023). Colavizza et al. (2021) point out the need for collaboration between researchers in the digital humanities and archivists to create and implement effective archival solutions that make use of AI.

The involvement of archivists in these stages is essential to ensure that outcomes are precise, contextually appropriate, and aligned with specific collection needs (Amozurrutia et al., 2023; Underwood et al., 2018). Jo and Gebru (2020) also affirm that ML data collection can significantly benefit from established archival strategies concerning ethics and inclusivity. On the other hand, computer scientists and archivists need to reflect together on how AI has implications for document management and, conversely, how document management can provide contributions to the advancement and proper use of AI and data in society.

6.1.2 Consider the Provenance and Context of Records and Archives

This guideline directly operationalizes core knowledge organization principles such as provenance and original order, as detailed in Section 2. From a KO perspective, provenance is the primary classificatory act, creating meaningful intellectual boundaries by grouping records based on their creator (Tognoli and Guimarães, 2020). Original order preserves the creator's classification system, a vital source of contextual evidence. Therefore, any machine learning

model that aims to perform archival classification should be designed to categorize content and respect and, ideally, strengthen these preexisting knowledge structures using human curation.

The preservation of authenticity and the adequate understanding of records and archival documents require a careful analysis of their provenance and context. This involves considering the history of creation and preservation, the original structure of the archival fonds, and the organizational criteria established by the creators (Edmond et al., 2022; Jaillant, 2022a). Provenance, the relationship between records and their creators, is the primary classification mechanism in archives, creating distinct *fonds* that delineate knowledge spheres corresponding to creators' functions and activities (Tognoli and Guimarães, 2020; ICA, 2007). The treatment of archives through machine learning systems must go beyond mere classification, and it is essential that the system preserves and explicates information about the context of creation and use, connecting each item to its historical trajectory (Vellino and Alberts, 2016; Anderson, 2021).

Frendo (2007) critiques the decontextualization inherent in treating information as "disembodied" discrete metadata, arguing that the record's significance stems from its context, which traditional classification helps maintain. Chabin (2020) found that AI processing focusing solely on text missed crucial contextual metadata (source, date, arrangement) in the French great national debate. Ranade (2018) argues that digital records in their native form can be richer than we recognize, containing embedded information (e.g., timestamps, geo-references, audit trails) that AI could help expose, but transfer processes often strip this richness, reducing records to digital facsimiles. Sousa (2022) emphasizes that the archival bond, preserved through classification, gives a document meaning beyond mere information.

Machine learning models, therefore, must be sensitive to the context of production and use, avoiding the treatment of documents as abstract data, disconnected from their origin and trajectory (Edmond et al., 2022). Bell and Bunn (2022) argue that archives are not "raw data" but are products of multiple interpretive acts, making their context indispensable for sense-making. Edmond et al. (2022) warn that "datafication" can strip records of their original narrative and cultural context. To this end, the reconciliation between new technologies and the fundamental principles of archival science (Jaillant, 2022a) is essential, so that the models are temporally and contextually aware. This approach allows users to explore the content from the perspectives of time, place, and identity (Ranade, 2018), enriching the analysis and interpretation of the documents.

6.1.3 Taking Care of Data Quality

Data quality is a fundamental pillar for the development of effective and reliable classification models (Shearer, 2000). Therefore, it is imperative that the data

used in training is rigorously evaluated and selected, ensuring its relevance, representativeness, and absence of biases (Amozurrutia et al., 2023). This process demands meticulous attention to the accuracy, completeness, consistency, and representativeness of the data used both in the training and classification phases (Pajares et al., 2023; Rolan et al., 2019). Human validation and review emerge as crucial steps in this context, acting as a control mechanism to ensure the accuracy of classifications and identify possible errors or inconsistencies that may compromise the model's performance (Edmond et al., 2022).

The importance of high-quality training data is reiterated in the literature, along with the need for rigorous methods to assess the accuracy and interpretability of the resulting models (Amershi et al., 2019). Alaoui (2024) cautions that the “datafication” process can degrade archival quality if not managed. Edmond et al. (2022) discuss “hidden data” and the imperative of “data cleaning”, directly addressing data quality. Jo and Gebru (2020) dedicate much of their work to the biases stemming from poor or unrepresentative training data in ML. Meticulous documentation of dataset creation, including rationale, composition, collection methodologies, preprocessing steps, and usage recommendations, is vital for reproducibility and reliability (Jo and Gebru, 2020; Padilla et al., 2019).

Professionals tasked with dataset creation must rigorously document every aspect of the process, including the rationale behind dataset design, its intricate composition, collection methodologies, preprocessing protocols, and usage recommendations (Jo and Gebru, 2020). This comprehensive documentation facilitates in-depth data assessment and aids in identifying and mitigating potential biases, ultimately enhancing the robustness and equity of the resulting models. Bell and Bunn (2022) suggest a nuanced view of data quality, arguing that archives should not be classified as “raw data” as they inherently involve an element of curation that impacts their integrity and usability.

6.1.4 Include Different Perspectives

Classification, whether conducted by archivists, researchers, or automated systems, is intrinsically a process influenced by a series of non-neutral factors (Bell and Bunn, 2022). Perspectives, intentions, and decisions made in the classification process shape the data and, consequently, impact the training of machine learning algorithms. This subjectivity inherent in classification underscores the importance of adopting an inclusive approach in the development of machine learning systems.

Alaoui (2024), through Terry Cook's paradigms, underlines archives' “communal” role in representing diverse voices and mitigating silences, a principle AI must defend. In order to mitigate the imposition of a single perspective on the collection, it is imperative that the design of machine learning models incorporates different views, originating from diverse user groups (Edmond et al., 2022; Bell

and Bunn, 2022). The inclusion of multiple perspectives ensures that the diversity of interpretations and user needs are duly considered in the classification process.

Edmond et al. (2022) warn that AI can make “minoritized material” invisible if not designed inclusively. Chabin (2020) critiqued the AI processing of the French debate for potentially homogenizing citizen expression, emphasizing the need for AI to capture diverse inputs. Mordell (2019) argues that archivists should acknowledge how AI tools privilege certain records and creators and how biases (race, gender, and class) can be encoded, calling for involvement in developing more inclusive systems.

In this sense, the active involvement of user communities in the development and implementation of classification systems becomes a fundamental requirement (Padilla et al., 2019; Marciano et al., 2024). By involving users from the initial stages, it is possible to ensure that their needs and perspectives are considered, resulting in fairer, more effective systems that are representative of the diversity of experiences.

Jaillant (2022a) focuses on user-centered design and design thinking for “dark archives”, meaning understanding diverse user access needs and involving end-users in the design process from inspiration to iteration. Jo and Gebru (2020) highlight community archives and participatory methods as crucial for inclusive ML data collection. Cushing and Osti (2023) found practitioners deeply concerned about AI bias affecting marginalized communities, calling for diverse perspectives in AI design. Punzalan and Caswell (2016) advocate for archival pluralism, acknowledging multiple coexisting archival realities and diverse ways of knowing.

6.1.5 Ensuring Algorithmic Accountability

The application of ML techniques in the analysis and classification of archives presents transformative potential, but also raises crucial ethical and methodological questions. It is essential to recognize that ML does not constitute a neutral tool and may, inadvertently, perpetuate and even amplify pre-existing biases in the training data (Hodel, 2022). Therefore, a rigorous analysis of both the algorithms and the data used is essential to mitigate the risks of biased interpretations and unfair results. Alaoui (2024) highlights the need for robust ethical frameworks, privacy safeguards, and strong governance for AI in archival contexts. Edmond et al. (2022) critique the “black box” nature of AI and the risks of “surveillance capitalism”, implicitly demanding greater accountability. Mordell (2019) warns that the resurgence of claims of objectivity or neutrality in datafied contexts can undermine efforts to address power relations in archives.

One of the main challenges lies in ensuring transparency and accountability in the use of ML in records and archives. Institutions that adopt these techniques must ensure that all decisions and results are traceable, auditable, and attributable (Jaillant, 2022a). The absence of these

mechanisms can compromise the credibility and legitimacy of the processes, undermining public trust. In this sense, the implementation of ML in archives must be guided by data protection laws and robust ethical principles, which ensure transparency, fairness, and non-discrimination (Amozurrutia et al., 2023).

Bell and Bunn (2022) call for “reasoning over archives”, which implies accountability in how this reasoning, whether human or algorithmic, is conducted and how its outcomes are justified. Jo and Gebru (2020) argue that ML should adopt ethical archival strategies, and Ranade (2018) emphasizes the need for transparency regarding processes for capturing, preserving, and enriching digital records to build trust. Padilla et al. (2019) state that collections as data stewards are guided by ongoing ethical commitments that should be formally documented and publicly available, working transparently to develop trustworthy, long-lived collections.

It is also imperative that the ML models used for classification do not perpetuate or aggravate existing prejudices (Edmond et al., 2022). The careful selection of training data, as well as the proper adjustment of the model parameters, are crucial steps to ensure fairness and avoid the exclusion of minority groups or perspectives. Human expertise, in this context, plays a fundamental role in guaranteeing the ethics and quality of the process (Cushing and Osti, 2023). The collaboration between humans and machines, therefore, should not relegate human experience and judgment to a secondary plane but rather integrate them synergistically and complementarily.

6.1.6 Provide Explainability for Predictions

The increasing adoption of ML techniques for classifying records, archives and other memory institutions brings with it challenges related to transparency and explainability. Complex models, such as deep neural networks (deep learning), often operate as “black boxes”, making it difficult to understand the criteria that led to a given classification (Edmond et al., 2022). In this sense, the need to go beyond the simple evaluation of the accuracy of the models arises, seeking an understanding of the mechanisms that guide their decisions (Marciano, 2022). Algorithmic opacity raises legitimate concerns that need to be addressed by developing and adopting more transparent and interpretable models (Hodel, 2022).

In the context of KO, this guideline is the modern computational equivalent of archival description. As established in Section 2, description makes the archivist’s classification system—the organization of fonds, series, and files—intelligible and navigable to users (Duranti, 1993; Vital and Brascher, 2016). A search engine explains why records and files are organized a certain way. Likewise, an explainable AI model must articulate the logic behind its classifications. Without explainability, an automated system becomes an opaque and undescribed knowledge struc-

ture, violating the archival and KO imperative for transparent and accessible organization.

Bell and Bunn (2022) advocate for making archivists’ “implicit processes more explicit”, a human-centered parallel to the XAI goal in computational systems. Mordell (2019) uses the “Devil’s Bridge” metaphor to describe how new technologies often obscure the material conditions of their production and render human decision-making invisible. Consequently, he advocates for plain language explanations of what AI tools do and the human choices embedded in their design.

The prioritization of models that enable understanding by experts is crucial, especially in sensitive contexts such as access to knowledge (Edmond et al., 2022). In other words, the choice of interpretable models and the clear documentation of the classification process become imperative to ensure the confidence and acceptance of algorithmic decisions by professionals in the field (Pajares et al., 2023; Amozurrutia et al., 2023). The focus should be on the development and use of models that not only perform well in the classification task but also enable an understanding of how the classification was performed and that are replicable (Edmond et al., 2022).

Alaoui (2024) suggests that the “archivist-informatician” must understand and articulate AI’s logic, demanding explainable systems. Jo and Gebru (2020) propose archival documentation as a model for enhancing ML dataset transparency. Explainable AI emerges as a fundamental area of study, with the goal of developing techniques that allow us to understand the reasoning behind the decisions made by ML algorithms (Chamola et al., 2023). Complex systems, such as deep learning systems, present additional challenges in this regard, as they are more difficult to interpret and debug (Rolan et al., 2019). Underwood et al. (2018) have developed lesson plans that integrate computational thinking into archival science teaching, including activities requiring students to understand and edit applications to open the “black box” of computational tools.

In line with the search for explainability, review by human experts becomes indispensable. ML-based classification processes must be open to evaluation by professionals who can verify whether algorithmic decisions are aligned with ethical principles and the needs of the user community (Edmond et al., 2022). Human verification of ML results is essential to ensure the accuracy, sensitivity, and ethical adequacy of classifications (Cushing and Osti, 2023). It is essential to recognize that ML systems are not neutral (Anderson, 2021) and that biases present in the training data can lead to biased results. In this context, critical evaluation and explainability become essential tools for mitigating these risks.

Table 3. Guidelines for using AI in automatic records and archival classification experiments: related specific AI methodology steps.

Id	Guidelines	Related specific AI methodology steps
1	Integrate professionals or professors in archives or records management	Interpretation/evaluation (KDD Process), Business understanding, Data understanding and Evaluation (CRISP-DM, RAMSYS and DST), Business understanding and Data Acquisition & Understanding (TDSP) and Identify and formulate the problem, Review data and AI ethics, Post-deployment review and Monitor and evaluate performance (CDAC).
2	Consider the provenance and context of records and archives	Business understanding, Data understanding and Data preparation (CRISP-DM, RAMSYS and DST), Business understanding and Data Acquisition & Understanding (TDSP) and Data preparation and Data exploration (CDAC).
3	Taking care of data quality	Data understanding and Data preparation (CRISP-DM, RAMSYS and DST), Data Acquisition & Understanding (TDSP).
4	Include different perspectives	-
5	Ensuring algorithmic accountability	Review data and AI ethics (CDAC)
6	Provide explainability for predictions	AI model explainability (CDAC)

Source: [Fayyad et al., 1996](#); [Shearer, 2000](#); [Moyle and Jorge, 2001](#); [Microsoft, 2024](#); [Domino Data Lab, 2017](#); [Jurney, 2017](#); [Martinez-Plumed et al., 2019](#); [Silva and Alahakoon, 2022](#).

Table 4. Guidelines for using AI in automatic records classification experiments: verification rules.

Id	Guidelines	Verification rules
1	Integrate professionals or professors in archives or records management	Check whether professional or professors in archives or records management are involved in the experiment steps.
2	Consider the provenance and context of records and archives	Verify whether machine learning systems preserve and respect the original structure of archival funds, the context of production and use of documents.
3	Taking care of data quality	Implement a human validation and review stage of the data for relevance, representativeness, accuracy and completeness.
4	Include different perspectives	Verify whether different user groups with different profiles and perspectives were identified and involved in the process of developing the classification system.
5	Ensuring algorithmic accountability	Verify that implementation is guided by robust data protection laws and ethical principles, and decisions and outcomes are traceable and auditable.
6	Provide explainability for predictions	Verify the application of practices that favor explainability, such as the use of models that are interpretable by nature, Explainable AI techniques, explanatory reports, and peer review of results.

6.1.7 Consolidation of Results

To consolidate these findings, Table 3 maps each guideline to the corresponding steps within the main data science and artificial intelligence frameworks and methodologies analyzed in Section 3.2. Each listed guideline was evaluated in relation to its practical and theoretical implications for the development of AI models for the automatic classification of archival documents.

Next, we present Table 4, which details the verification rules for each of the guidelines in order to articulate useful criteria so that the developed classification system is reliable, explainable, and appropriate to the archival context.

6.2 Experiments using Machine Learning

In the literature search, we identified 24 experiments on the practical application of ML in the automatic records and archives classification listed in Tables 5,6.

Table 5 shows that the origins of the records and archives used in research vary widely, ranging from public institutions, universities, public archives, personal collections, and data sets from public projects. Some experiments were limited to emails only, and the others cover all types of records. In a temporal analysis, in the period 2001–2016, surveys on emails dominated with 8 experiments (80%), a situation that was practically reversed in the period 2017–2023 with only 3 experiments (21%).

With the objective of assessing the practical adherence to the proposed guidelines, we developed Table 7, which maps each of the automatic classification experiments carried out to the guidelines defined in Table 4. This assessment allows us to attest, systematically, which guidelines were incorporated in each experiment, identifying gaps and areas for improvement in the application of AI in records and archives. One limitation of the assessment made is that the verification is binary, as it only allows the answer that

Table 5. Quantitative experiments: basic characteristics.

Reference	Origin of records	Description	Year
Marcus, 2002; Shinkle, 2017	National Archives and Records Administration (NARA)	Automatically classify records and file them in the system according to the assigned classification.	2001
Cohen et al., 2004	Carnegie Mellon University	Automatically classify emails according to one of five possible action types.	2004
Bennett and Carbonell, 2005	Carnegie Mellon University	Classification of emails and their sentences that have action items.	2005
Carvalho and Cohen, 2005	Carnegie Mellon University	Automatically classify emails according to one of seven possible action types.	2005
Goldstein and Evans Sabin, 2006	Personal email collections	Automatically classify emails into act and genre categories.	2006
Lampert et al., 2010	Academic research with Enron emails	Automatically classify emails that contain requests.	2010
Public Record Office Victoria, 2018	Canadian government bodies	Automatically classify emails into one of the relevant subcategories in the sorting step	2012
Alberts and Vellino, 2013	University of Ottawa	Automatically recognize whether an email has business value	2013
Esteva et al., 2013	United States Department of State	Automatically classify telegrams into secrecy classes	2013
Vellino and Alberts, 2016	Academic research with volunteer information management consultants	Automatically classify emails in the process of evaluating whether they have value for the business	2016
Liu et al., 2017	Academic research with archival records from Gansu Province in China	Automatically classify records with classification code application.	2017
Kim et al., 2017	Seoul Metropolitan Government	Automatically classify records into different classification schemes.	2017
Rolan et al., 2019	New South Wales State Archives (NSWSA)	Automate the evaluation of records according to the institution's temporality table.	2017
Rolan et al., 2019; Alberts and Vellino, 2013	Public Record Office Victoria	Identify the format of emails to reduce the volume of records to be evaluated.	2018
Hutchinson, 2018	University of Saskatchewan Associate Vice President for Information and Communications Technology	Automatically classify human resources records that contain individualized personal information.	2018
Binici, 2019	Çankırı University Karatekin	Classify records according to the Standard File Plan	2019
Shu et al., 2020	Avocado dataset	Automatically classify email intent	2020
Anderson, 2021	Cybernetics Thought Collective project	Sort manually tagged records into four categories	2021
Wang et al., 2021	Archives of Liaoning Province	Classification of data catalog items in accordance with the Chinese File Classification Law	2021
Tkachenko and Denisova, 2022	Siberian State Automobile and Highway University	Automatically classify records from a university	2022
Franks, 2022	Australian Human Rights Commission (AHRC)	Classify records according to the disposal schedule.	2022
Brokensha et al., 2023	National Afrikaans Literary Museum and Research Center	Classify records according to their type.	2023
Payne, 2023	Academic research with Enron emails	Automatically classify emails into administrative or operational	2023
Watanabe and Sousa, 2023	Attorney General of the Union	Classify records according to the temporality table.	2023

Table 6. Quantitative experiments: basic data.

Reference	Record metadata used as independent variables	Classification algorithms	Records used	Classes	Agreement between annotators (Kappa)	Model results*
Marcus, 2002; Shinkle, 2017	Content	AutoRecords proprietary application	n/a	n/a	n/a	A: 96.00%
Cohen et al., 2004	Content, expressions of time and related to pronouns or proper names.	Voted Perceptron, Ada Boost, Support Vector Machine (SVM) and Decision Tree	1135	5	0.72 to 0.82	F1: 85.00%**
Bennett and Carbonell, 2005	Total content and separated into sentences	k-nearest neighbor (kNN), SVM and Naïve Bayes	744	two	0.82 to 0.85	A: 81.73%** F1: 77.90%**
Carvalho and Cohen, 2005	Content, actual and estimated relationship tags between inline emails	Own algorithm based on dependency network	721	7	n/a	F1: 82.96%**
Goldstein and Evans Sabin, 2006	Content and 16 characteristics of speech acts	SVM and Random Forest	280	5	0.89	P: 63.00%**
Lampert et al., 2010	Content and 9 characteristics of speech acts	SVM	505	two	0.68	A: 83.76%** P: 84.90%** A: 83.90%** F1: 84.30%**
Public Record Office Victoria, 2018	Content and non-lexical characteristics (+22)	kNN	1703	13	n/a	A: 88.70%** P: 84.18%** A: 91.09%**
Alberts and Vellino, 2013	Content and non-lexical characteristics (+22)	SVM	173	n/a	n/a	P: 77.30%**
Esteva et al., 2013	Content, Traffic Analysis by Geography and Subject and keywords	SVM	154,392	4	n/a	A: 92.39%
Vellino and Alberts, 2016	Content and other attributes (+13)	SVM	1023	13	n/a	A: 91.00%** P: 89.00%** A: 94.00%** F1: 91.00%**
Liu et al., 2017	Record content and group according to department	Naïve Bayes	3600	6	n/a	P: 89.20%
Kim et al., 2017	Content, Title and metadata of related records	Proprietary application	25,553	120	n/a	A: 97.81%**
Rolan et al., 2019	Content	Multinomial Naïve Bayes and Multi-Layer Perceptron	8784	n/a	n/a	A: 84.00%
Rolan et al., 2019; Alberts and Vellino, 2013	Content, organizational functions, action verbs/objects and function/activity terms.	Nuix proprietary application	4,600,000	n/a	n/a	F1: 83.50% A: 100%**

Table 6. Continued.

Reference	Record metadata used as independent variables	Classification algorithms	Records used	Classes	Agreement between annotators (Kappa)	Model results*
Hutchinson, 2018	Content	SVM	169	30	n/a	A: 87.72%
Binici, 2019	Content	Multinomial Naïve Bayes	1784	n/a	n/a	A: 90.40%** F1: 98.30%**
Shu et al., 2020	Weak content and role labels from user interactions	Hydra, neural network model	3182	3	0.61	A: 80.40%**
Anderson, 2021	Content	Naïve Bayes	154	n/a	n/a	A: 71.10%
Wang et al., 2021	Content, issuing institution, receiving institution, type of record and keywords	SVM and Network Analysis	2500	11	n/a	F1: 71.60%
Tkachenko and Denisova, 2022	Content	SVM and kNN	1778	4	n/a	F1: 98.30%**
Franks, 2022	Content	SVM, C-LSTM, CNN, LSTM, BERT, RoBERTa and XLNet	6217	29	n/a	A: 87.90%** P: 83.93%** A: 78.64%** F1: 77.10%
Brokensha et al., 2023	Content	SVM and Multi-Layer Perceptron	621	6	n/a	P: 92.80%** A: 93.10%** A: 92.80%** F1: 92.90%**
Payne, 2023	Content and STACC Framework (+13)	Random Forest	214	two	n/a	P: 70.00% A: 60.00%
Watanabe and Sousa, 2023	Content and vocabularies of people, places and time	Ridge, Multi-Layer Perceptron, Complement Naïve Bayes and others (+16)	4800	24	n/a	A: 87.10%** F1: 0.870**

* A, Accuracy; P, Precision; F1, F1 score.

** The best result.

C-LSTM, Convolutional Long Short-Term Memory; CNN, Convolutional Neural Network; LSTM, Long Short-Term Memory; BERT, Bidirectional Encoder Representations from Transformers; RoBERTa, Robustly Optimized BERT Pretraining Approach; XLNet, eXtreme Long Net.

the criteria are met or not met, so it does not allow us to distinguish the levels of compliance with the verification rules. Another limitation of the assessment was that the verifications were limited to the content of the articles reporting the experiments, which do not always present many details of the methodological procedures adopted.

7. Discussion

The analysis of the literature enabled the elaboration of the six guidelines for using AI in automatic records and archives classification, which were selected based on the relevance manifested by their occurrence among the majority of the authors researched. Although categorized in different groups, the guidelines have a strong reciprocal influence, as will be addressed below. Next, the analysis of the experiments made it possible to verify how they were developed in comparison with the elaborated guidelines, which enabled us to discuss the results in the sequential order of the guidelines, with a joint analysis of their incidence or non-incidence in the researched experiments. At the end of this chapter, we will discuss the results in general to formulate broader proposals.

The challenge of enabling the use of machine learning in archives finds its way into Computational Archival Science (CAS), a transdiscipline that seeks not only to apply technology to enhance archival practices, but also to transform the founding disciplines into a movement in which the theories and methods of both are integrated and mutually influence each other (Marciano et al., 2018). Table 3 reveals an initial convergence in some of the guidelines originating from the archival literature with the computation methodologies for the use of AI, since in 5 of the guidelines there is a provision for steps in the analyzed methodologies. It should be noted that on the one hand, the guidelines that we have elaborated are specific (Integrate professionals or professors in archives or records management), while on the other hand, the steps of the methodologies are more generic (Business understanding) due to their enormous scope of application, as is evidenced in the very expression “Cross-Industry” of CRISP-DM, one of the studied methodologies.

Table 8 presents the adherence to the guidelines. For guideline no. 1 of “Integrate professionals or professors in archives or records management”, we identified that most of the experiments meet the criteria (54.1%), of which four were projects within the scope of public archives (NARA, NSWSA, National Archives of Korea and Victoria). The form of this collaboration was quite varied, with the prevalence of authors (8), followed by focus groups and interviews with specialists (4), and with archival institutions as project managers (3). The sum is greater than the number of experiments because in some of them, more than one form of collaboration was observed.

In the computer science literature, domain knowledge appears as a fundamental basis to be captured during the experiment, since the ML paradigm requires quality data

Table 7. Verification of compliance with the guidelines in the researched experiments.

Reference	1	2	3	4	5	6
Marcus, 2002; Shinkle, 2017	✓					
Cohen et al., 2004		✓	✓	✓		
Bennett and Carbonell, 2005			✓			
Carvalho and Cohen, 2005		✓	✓			
Goldstein and Evans Sabin, 2006			✓	✓		
Lampert et al., 2010		✓	✓	✓		
Alberts and Forest, 2012	✓	✓	✓	✓		
Alberts and Vellino, 2013	✓	✓	✓			
Esteva et al., 2013		✓				
Vellino and Alberts, 2016	✓	✓	✓	✓		
Liu et al., 2017		✓				
Kim et al., 2017	✓	✓				
Rolan et al., 2019	✓	✓				
Rolan et al., 2019; Public Record Office Victoria, 2018	✓	✓				
Hutchinson, 2018	✓		✓			
Binici, 2019	✓		✓			
Shu et al., 2020		✓	✓			
Anderson, 2021	✓	✓	✓	✓		✓
Wang et al., 2021		✓				
Tkachenko and Denisova, 2022						
Franks, 2022	✓	✓		✓		
Brokensha et al., 2023		✓	✓			
Payne, 2023	✓	✓		✓		
Watanabe and Sousa, 2023	✓	✓				

Note: The ✓ symbol indicates compliance with the corresponding guideline.

and pre-existing outputs to train the algorithms (Banh and Strobel, 2023). The human participation of professionals or professors in archives or records management appears as essential in the perspective of Human-in-the-loop machine learning (Mosqueira-Rey et al., 2023), which values human interaction in different types:

- Active Learning: values human annotation to label unmarked data and increase the model’s accuracy, which will also have a strong relationship with guideline no. 3;
- Interactive Machine Learning: values human collaboration to guide learning and perform various tasks, where human specialists are better able to explain how tasks should be performed and with what quality;
- Machine Teaching: values human expertise to transfer knowledge to the machine learning model, as human experts can develop strategies on how to teach the machine according to the peculiarities of the way it learns;
- Curriculum Learning: values human organization of learning to guide the machine efficiently, starting with the presentation of examples in increasing levels of difficulty, which only a human specialist is able to elaborate;

Table 8. Guidelines for using AI in automatic records and archives classification experiments: quantitative adherence.

Id	Guidelines	Quantity	Percentage
1	Integrate the knowledge of archivists or record managers	13	54.1%
2	Consider the provenance and context of records and archives	18	75.0%
3	Taking care of data quality	13	54.1%
4	Include different perspectives	6	25.0%
5	Ensuring algorithmic accountability	0	00.0%
6	Provide explainability for predictions	4	16.7%
	Total	57	37.5%

- Explainable AI: values the human capacity to understand and trust AI models based on the generation of explainable models, which depend on the evaluation of human specialists to be improved, which also has a direct relationship with guideline no. 6;
- Usable and Useful AI: values the design of human-computer interaction to create practical and accessible solutions, in which human experts can contribute based on the knowledge they have of how users interact with the documents and thus assist in the design and evaluation of the AI system.

To maximize the contribution of specialists, Marciano (2022) suggests that the development of AI solutions should be iterative, involving the prototyping, testing, and adaptation of the solutions. In this way, the refinement and continuous improvement of the model is allowed based on the feedback from archivists and the analysis of results (Amozurrutia et al., 2023).

As Tognoli and Guimarães (2020) emphasize, provenance operates as the primary classification mechanism in archival knowledge organization, creating meaningful intellectual boundaries that reflect the administrative realities of record creation. However, in the experiments reviewed, provenance consideration frequently manifests as simple metadata extraction rather than the deep contextual understanding that archival theory demands. For instance, while Wang et al. (2021) included “issuing institution” and “receiving institution” as variables, and Kim et al. (2017) incorporated “metadata of related records”, these approaches fail to capture what Duranti (1993) characterizes as the full scope of “archival material, its provenance and documentary context, interrelationships and how it can be identified and used”. The experiments demonstrate technical acknowledgment of provenance without achieving the interpretive depth that Chabin (2020) achieved in the French Great National Debate, where diplomatic analysis enriched the AI model’s understanding of documentary context.

The much lower adherence to guidelines that reflect original order (implicit in preserving context) and the complete absence of algorithmic accountability (the new form of archival description) are telling. This suggests that the ML experiments analyzed succeed at a rudimentary level of content-based categorization but largely fail at the more sophisticated KO task of preserving and representing the com-

plex, evidence-based relationships that define an archive. They classify records as isolated items, not interconnected components of a larger, organic knowledge system. This trend points to a significant risk: the efficiency of automatic classification may come at the expense of the very contextual richness that archival KO practices were designed to protect.

In stark contrast, the work of Payne (2023) with the STACC (State, Temporality, Attachment, Context, and Content) framework exemplifies a deep approach. Rather than merely using authorship metadata, STACC attempts to model archival reasoning itself, operationalizing complex concepts like the “archival bond” and the “transactional/supporting” nature of a document into computationally tractable features. This distinction is fundamental: whereas the superficial approach performs content categorization assisted by metadata, the deep approach aims for true archival classification that preserves context. The prevalence of the former over the latter explains the central paradox of this study: a high formal respect for provenance coexisting with a systemic failure to preserve the archive’s relational integrity, effectively treating it as a mere collection of data to be mined, not an organized knowledge system.

The computer science literature seems very convergent with that of archival science regarding guideline no. 3 “Taking care of data quality”. The data science and AI methodologies CRISP-DM, RAMSYS, DST, and TDSP expressly provide for steps to deal with data quality, as expressed in Table 3. And Trustworthy AI requires reliable data beyond the traditional criteria of precision, completeness, and consistency, to also encompass requirements related to equity and the prevention of biases (Kaur et al., 2023), which reveals a strong relationship with guideline no. 5 to be addressed below.

A little more than half of the experiments meet guideline no.3 (54.1%), which was manifested in different forms, such as the need for human review of the data set used in model training to ensure the production of better results (Binici, 2019). In five experiments (20.8% of the total), the Kappa indicator was used, a statistical measure that measures the degree of agreement or agreement between two or more evaluators in the range of -1 to 1, thus offering a more precise assessment of consistency between assess-

ments. However, the greater investment in data quality has the side effect of reducing the set of documents suitable for use. The experiment with the largest number of documents was Liu et al. (2017) with 3600 documents, which is far from being a representative sample of the diversity of most document collections of institutions today. The need for prior annotation of all documents is linked to the use of ML in supervised learning (Emmert-Streib and Dehmer, 2022), which was the option of practically all the experiments analyzed (95.8%). Therefore, a possible path may be the use of semi-supervised learning, which uses the combination of a small set of annotated documents that will serve as a basis for the automatic annotation of a large volume of documents. This path is advocated by Shu et al. (2020), who consider the need to elaborate large-scale annotated training data for model learning, which has proven to be unfeasible given the high costs and a lot of time to be invested, as well as situations related to the privacy of personal data.

The problems highlighted by Kim et al. (2017) and Binici (2019) in the classification plan regarding hierarchical structure, lack of specific classes, and absence of scope notes would have influenced the quality of the training data due to greater divergence between annotators. And in turn, the problems accumulated in these two groups may end up damaging the results obtained by the classification model. Records classification plans, as knowledge organization systems, can provide an explicit representation of the relationships between records, in order to facilitate the understanding of the context and importance as evidence (Frendo, 2007), which we consider an essential stage of projects. A classification plan with clear and mutually exclusive classes helps train the prediction model, as the algorithms are trained from records that fall into each category and from iterative reviews of additional records (NARA, 2014).

Regarding guideline no. 4 “Include different perspectives”, which was verified in 33.3% of the experiments, the ISO/IEC TR 24027:2021 standard (ISO/IEC, 2022) states the need for the project team to have a diverse composition, including contributions from experts in ethics, social sciences, quality, and privacy. One of the risks is social bias, which consists of the unfair treatment shared in a society, which may be encoded and perpetuated through organizational policies and ML models that learn historical patterns. Hence, the need to establish strategies for treating bias at all stages of the life cycle of an AI system, from the analysis of requirements to deployment and monitoring, which involves consideration of external and internal requirements, acceptance criteria, training data, adjustment, fairness metrics, among other points.

The forms of inclusion of different perspectives in the experiments were done in different ways. We can mention the meeting of a multidisciplinary advisory board composed of cybernetics scholars and technologists and the test with a focus group to understand the needs of different users

done by Anderson (2021). This experiment is in line with Padilla’s suggestion (2019) to form committees and audit the methods used as a proposed measure to mitigate these risks. Alberts and Forest (2012) included the perspective of users of email classification systems, and Franks (2022) also included researchers. In five experiments, groups of annotators were integrated for the same documents as a measure to capture different views and calculate the inter-annotator agreement indicator (Cohen et al., 2004; Goldstein and Evans Sabin, 2006; Lampert et al., 2010; Vellino and Alberts, 2016; Payne, 2023).

For guideline no. 5 of “Ensuring algorithmic accountability”, the literature has advanced significantly, but it is a very recent topic, in which the definition of Trustworthy AI was developed only in the early 2020s (ISO/IEC, 2021). This seems to be a determining reason why none of the experiments managed to meet the minimum requirements of the guideline, which range from data protection principles such as privacy and ethical principles to the tracking and auditing of results. Of all the data science and AI methodologies, only CDAC, which is the most recent of all, contemplates the “Review data and AI ethics” phase (Silva and Alahakoon, 2022). This is a phase that goes far beyond the simple technical verification of the data because it represents a deep commitment to responsibility, inclusion, and the mitigation of risks inherent in the implementation of ethical AI systems that are fair and aligned with the values of society, in addition to being effective.

To face these challenges, Kaur et al. (2023) propose the “Human-Centered Approach to Trustworthy AI (Human + AI)”, which provides for human participation in all stages, from the design of the AI system, through development and use, and to subsequent supervision, which is close to the understanding of Cushing and Osti (2023). As a way to materialize this approach, Kaur et al. (2023) propose several techniques and methods to meet the requirements of Trustworthy AI in relation to equity, explainability, accountability, privacy, and acceptance. Due to the complex and probabilistic nature of many AI systems, the authors understand that traditional software testing techniques may be insufficient and, therefore, consolidated some approaches for the verification and validation of Trustworthy AI which are the Metamorphic Test (verification of the system based on the relationships between its inputs and outputs, without depending on an oracle to define the expected result), Expert Panels (human experts evaluate the performance and safety of AI systems), Benchmarking (comparison of the AI system’s performance with established benchmarks, using standardized data sets to assess their accuracy and robustness), Field Testing (evaluation of the AI system in real environments with the observation of its behavior in practical scenarios and with real users), and Comparison with Human Intelligence (performed in specific tasks to assess the ability of the AI system to replicate or surpass human intelligence).

An important concept for Trustworthy AI is the provenance of AI systems, which seeks to provide a detailed record of the origin, processing, and transformations of data about each step of the process (Kale et al., 2023; Rolan et al., 2019). AI provenance brings as benefits increased transparency, support for explainability, promotion of reproducibility, help in identifying biases, and facilitation of auditing and debugging, which are valued by Jaillant (2022a). The models, languages, and tools indicated to document AI provenance are the PROV (Provenance Data Model) of the W3C, ProvStore, Prov Viewer, and initiatives such as OpenML and ModelDB, in addition to the Renku and WholeTale platforms for practical application in research environments. We can perceive a strong link between guideline no. 5 and guidelines no. 3 (data quality), no. 4 (inclusion of different perspectives), no. 6 (explainability), and no. 7 (privacy), which reveals that this is the guideline with the largest number of dependencies with different work fronts, which helps to understand its high complexity and, on the other hand, the challenges in applying it in experiments.

In guideline no. 4 “Provide explainability for predictions”, 16.7% managed to meet the minimum requirements. Burkart and Huber (2021) indicate that the use of self-interpretable (“interpretable by nature”) supervised machine learning models, such as Decision Tree, Naive Bayes, and kNN, can be considered a research decision that favors explainability. There were nine experiments that used “interpretable by nature” models as shown in Table 6. However, the use of such models was not accompanied by the researchers’ justification that the choice was made to enable the explanation of the generated predictions. It is so certain that only four experiments adopted measures to favor the explainability of the results. In the research project of Anderson (2021), explainability is favored by making the documents publicly available, as well as open access to all the code and application used.

The other three experiments favored explainability through the qualitative evaluation of the generated predictions, which proved to be very important for understanding the functioning of automatic classification, with its gaps to support actions to improve the results. Alberts and Forest (2012) developed rankings of discriminant features with the chi-square value for each of the categories, which help the reader to interpret the model results and analyze the document collection itself. However, it should be noted that this analysis was exploratory since the authors did not use the conclusions of the most important characteristics to eliminate the less important ones and thus optimize the results and the model’s own explainability. Next, the authors discovered that the most discriminating non-lexical characteristics were related to the sender and the addressee (e.g., social network, hierarchy, mailing lists). The presence of formatting (bold, capital letters) and attachments also proved to be relevant. Concerning the lexical characteristics, the keywords specific to each category were identified (e.g.,

“approve” for action-authorization, “meeting” for action-meeting), and the lexical analysis revealed nuances in the language and tone of each category.

Kim et al. (2017) identified the types of errors in the model results as class coverage, classification error due to a lack of correspondence between the content and the class indicated by the AI, classification errors in the learning data, class ambiguity, overfitting, and situations in which it was not possible to determine the cause. Binici (2019) lists reasons related to the classification plan that cause the algorithm to make mistakes, such as the hierarchical and content structure of the plan, including the lack of proper classes. He points out that the uncertainty caused by the simplicity of class titles and the lack of additional explanations in the filing system creates uncertainty, the use of similar subtitles in different hierarchies causes confusion, the presence of more than one subject in the document creates uncertainty in the notation, and carelessness in the annotation can motivate errors. In these two experiments, we can see that the focus was on explaining the reasons for the wrong predictions, which brings important conclusions by detecting errors in the quality of the data and in the classification schemes used, which can be later corrected so that the model can be trained again and thus generate better results. However, the focus of these two experiments only on the errors ends up bringing a partial explanation about the model’s predictions. One problem with this approach is that it does not analyze whether the correct classifications were made based on discriminant features that are not based on archival theory.

As warned by Ranade (2018), we cannot sacrifice the nuances in the confusing data of the archives through its improper cleaning or the imposition of structure or order where they did not exist. This is because archival data are probabilistic in nature, and dealing with this reality should require transparency regarding this uncertainty with the adoption of tools and approaches that embrace it.

Once the individual analysis of each guideline has been completed, from this point we will address general aspects of the research results. A very significant challenge of this research resided in the diversity of methodological procedures used in the analyzed experiments. On the other hand, we did not locate methodologies or frameworks that could help in the comparison of automatic classification experiments of archival documents. Therefore, based on a review of the theoretical literature in the area, it was possible to elaborate the six guidelines based on theories in the area that allowed us to first identify the main research questions related and, second, to serve as solid parameters for evaluating the experiments carried out. Another important challenge to be highlighted is that the technical knowledge of computer science required a dual effort in the research to (i) understand the concepts, methods, and techniques of AI and (ii) identify points and forms of articulation with archival science.

Moving forward, the field requires what Wing (2006) describes as true computational thinking: multi-layered abstractions that reflect human rather than computer thinking. In the archival context, this means developing algorithms that can represent and process the complex relationships Duranti (1993) identifies—not just between documents but between documents and their contexts, creators, and communities. It means creating systems embodying what Marciano et al. (2018) envision for CAS: genuine transdisciplinary integration that transforms archival and computational practice.

It seems to us that a solution to advance studies on automatic classification in records and archives is to structure a methodological framework, which can be defined as a structured practical guideline or a tool to guide the user through a process, which is organized in stages or in a step-by-step approach (McMeekin et al., 2020). This is because the data-formatted documents demand the need to create tools, methodologies, and approaches for using and analyzing files as data, as proposed by Moss et al. (2018). The methodological framework includes a “set of structured principles” and a “sequence of methods” (McMeekin et al., 2020), as well as an “approach to make explicit and structure how a certain task is used” (Andrade et al., 2009), and a “body of methods, rules, and postulates employed by a specific procedure or set of procedures” (Rivera et al., 2017, used by Rivera et al., 2017). The methodological framework would bring as benefits improving the consistency, robustness, and documentation of activities (Rodgers et al., 2016), increasing the quality of research, standardizing approaches (Squires et al., 2016), as well as maximizing the reliability of results (Kallio et al., 2016).

8. Conclusions

This article sought to bridge the gap between classification theory and AI practice by answering two guiding questions: (1) What are the fundamental guidelines for applying AI to records and archives? Furthermore, (2) How closely do current experiments in automatic classification adhere to them? We begin by grounding our investigation in the fundamental archival principles for classifying and organizing knowledge, as detailed in Section 2, establishing that provenance, original order, and hierarchical description are not mere administrative procedures but rather sophisticated forms of knowledge organization (Tognoli and Guimarães, 2020; Duranti, 1993). From this foundation, we synthesize a multidisciplinary literature review into six evidence-based guidelines for implementing AI and conduct a comparative analysis of 24 experiments based on this new framework.

Our analysis reveals a critical paradox. From one perspective, the experiments demonstrated significant respect for the principle of provenance (75.0% adherence to Guideline no. 2), which recognizes the creator-record relationship fundamental to classification within the KO frame-

work. This suggests that the core of archival thinking has permeated computational practice. However, this approach has remained superficial for the most part, failing to capture the deeper contextual integrity preserved by the principles of the original order or the rich, hierarchical relationships articulated by archival description. The result is a “datafication” of records that respects their provenance but often strips them of their organic, narrative, and evidentiary structure. It is as if the treatment accorded to records and archives fails to distinguish them from mere data collection.

This study aimed to synthesize guidelines for developing AI projects in records and archives and to assess whether existing automatic records classification experiments adhere to these guidelines. Through a comprehensive literature review spanning computer science, archival science, and information science, we identified six critical guidelines for the application of ML in the automatic classification of records.

The core findings emphasize the need for a collaborative, context-aware, and ethically grounded approach to AI implementation in archival settings. Our analysis of 24 experiments applying machine learning to automatic records classification revealed a mixed landscape. While some guidelines, such as considering provenance and context (75.0%) and integrating archival expertise (54.1%), were reasonably well-addressed, others, particularly algorithmic accountability (0.0%) and explainability (16.7%), were largely neglected. This highlights a significant gap between the theoretical understanding of ethical AI principles and their practical implementation in records and archival contexts. The relative strength in considering provenance suggests a recognition of the inherent value of archival principles, but the weakness in accountability and explainability points to a need for greater awareness and application of Trustworthy AI frameworks.

These findings have significant implications for both theory and practice. Theoretically, this research contributes to the emerging field of Computational Archival Science (CAS) by providing a structured set of guidelines for developing AI-driven archival systems. The guidelines synthesize knowledge from diverse disciplines and offer a framework for integrating computational methods with core archival principles. Practically, our findings can inform the design and implementation of future AI projects in records and archives, promoting more responsible and effective use of these technologies. They also serve as a checklist for evaluating existing systems and identifying areas for improvement.

This study has limitations. First, the assessment of compliance with the guidelines was based on published articles, which may not fully capture the methodological nuances of each experiment. Second, the binary assessment of compliance did not allow for distinguishing levels of adherence to the verification rules. Future research could address these limitations by developing a comprehensive

methodological framework specifically designed for evaluating AI applications in archival contexts is paramount. This framework must have as elements the objectives, key concepts, principles (the basic requirements for trustworthy AI), guidelines (the standards to follow for developing AI models), work processes (the necessary steps to the implementation of AI projects), tools and models (the AI frameworks and algorithms), and metrics and indicators (quantitative and qualitative measurements for the assessment of results).

Future research should also focus on developing concrete strategies for implementing the guidelines identified in this study. This includes developing tools and techniques for data quality assessment, bias detection and mitigation, and explainable AI. Furthermore, research is needed to explore the ethical implications of AI in archives, including issues of privacy, access, and representation.

In conclusion, our findings underscore the importance of a collaborative, context-aware, and ethically grounded approach to AI, ensuring that these tools are used to enhance, rather than undermine, the core values of archival science and the integrity of the historical record. This work serves as a foundation for future research and practice, promoting the development of more transparent, accountable, and equitable AI systems for records and archives.

Availability of Data and Materials

There is no data and material beyond what is mentioned in the article.

Author Contributions

EW and RTBS designed the research study, performed the research and analyzed the data. EW drafted the manuscript. Both authors contributed to critical revision of the manuscript for important intellectual content. Both authors read and approved the final manuscript. Both authors contributed to editorial changes in the manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

References

Alaoui S. Peut-on parler de l'automatisation comme cinquième paradigme archivistique? *The Canadian Journal of Information and Library Science*. 2024; 47: 18–34. <https://doi.org/10.5206/cjils-rcsib.v47i1.17111>. (In French)

Alberts I, Forest D. Email pragmatics and automatic classification: A study in the organizational context. *Journal of the American Society for Information Science and Technology*. 2012; 63: 904–922.

Alberts I, Vellino A. The importance of context in the automatic classification of email as records of business value: A pilot study. *Proceedings of the American Society for Information Science and Technology*. 2013; 50: 1–2. <https://doi.org/10.1002/meet.14505001112>.

Amozurrutia AB, Bernal Astorga Y, Cuéllar Hidalgo R, Escoto Velázquez CA, García-Velázquez LM. Inteligencia artificial en los archivos. *Tábula*. 2023; 25: 41–59. <https://doi.org/10.51598/tab.936>. (In Spanish)

Anderson BG. On constructing a scientific archives network. *Archivaria*. 2021; 91: 104–147. <https://doi.org/10.7202/1078467ar>.

Andrade J, Rodríguez S, Seoane M, Suárez S. Knowledge management systems procedural development. In Rabuñal Dopico J, Dorado J, Pazos A (eds.) *Encyclopedia of Artificial Intelligence* (pp. 975–981). IGI Global Scientific Publishing: Hershey, PA. 2009. <https://doi.org/10.4018/978-1-59904-849-9.ch144>.

Amershi S, Begel A, Bird C, DeLine R, Gall H, Kamar E, Nagappan N, Nushi B, Zimmermann T. Software engineering for machine learning: A case study. In *Proceedings - 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP) 2019* (pp. 291–300). Institute of Electrical and Electronics Engineers Inc.: Seattle, WA, USA. <https://doi.org/10.1109/ICSE-SEIP.2019.00042>.

Azarian R. Potentials and limitations of comparative method in social science. *International Journal of Humanities and Social Science*. 2011; 1: 113–125.

Banh L, Strobel G. Generative artificial intelligence. *Electronic Markets*. 2023; 33: 63. <https://doi.org/10.1007/s12525-023-00680-1>.

Bell M, Bunn J. Dark archives or a dark age for reasoning over archives? *AI & Society*. 2022; 37: 959–966. <https://doi.org/10.1007/s00146-021-01365-z>.

Bennett PN, Carbonell J. Detecting action-items in e-mail. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval 15-19 August 2005, Salvador, Brazil* (pp. 585–586). Association for Computing Machinery: New York. 2005. <https://doi.org/10.1145/1076034.1076140>.

Binici K. Makine öğrenmesi yaklaşımıyla e-belgelere standart dosya plan numaralarının otomatik olarak atanması üzerine bir çalışma. *Bilgi Yönetimi*. 2019; 2: 116–126. <https://doi.org/10.33721/by.654464>. (In Turkish)

Brokensha S, Kotzé E, Senekal B. Machine learning for document classification in an archive of the National Afrikaans Literary Museum and Research Centre. *Journal of the South African Society of Archivists*. 2023; 56: 134–147. <https://doi.org/10.4314/jsasa.v56i1.9>.

Bunn J. Archival automation in the United Kingdom and

- the relationship between standardization and computerization. *Information and Culture*. 2019; 54: 159–178. <https://doi.org/10.7560/IC54202>.
- Burkart N, Huber MF. A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*. 2021; 70: 245–317.
- Carvalho VR, Cohen WW. On the collective classification of email “speech acts”. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval 15-19 August 2005, Salvador, Brazil* (pp. 345–352). Association for Computing Machinery: New York. 2005. <https://doi.org/10.1145/1076034.1076119>.
- Chabin MA. The potential for collaboration between AI and archival science in processing data from the French Great National Debate. *Records Management Journal*. 2020; 30: 241–252. <https://doi.org/10.1108/RMJ-08-2019-0042>.
- Chamola V, Hassija V, Sulthana AR, Ghosh D, Dhingra D, Sikdar B. A review of trustworthy and explainable artificial intelligence (XAI). *IEEE Access*. 2023; 11: 78994–79015.
- Cohen WW, Carvalho VR, Mitchell TM. Learning to classify email into “speech acts”. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing 25-26 July 2004, Barcelona, Spain* (pp. 309–316). Association for Computational Linguistics: Kerrville. 2004.
- Colavizza G, Blanke T, Jeurgens C, Noordegraaf J. Archives and AI: An overview of current debates and future perspectives. *Journal on Computing and Cultural Heritage*. 2021; 15: 1–15. <https://doi.org/10.1145/3479010>.
- Cushing AL, Osti G. “So how do we balance all of these needs?”: How the concept of AI technology impacts digital archival expertise. *Journal of Documentation*. 2023; 79: 12–29. <https://doi.org/10.1108/JD-08-2022-0170>.
- Domino Data Lab. The practical guide to managing data science at scale. 2017. <https://domino.ai/resources/managing-data-science>.
- Duranti L. Origin and development of the concept of archival description. *Archivaria*. 1993; 35: 47–54.
- Edmond J, Horsley N, Lehmann J, Priddy M. *The trouble with big data*. Bloomsbury Academic: London. 2022.
- Emmert-Streib F, Dehmer M. Taxonomy of machine learning paradigms: A data-centric perspective. *WIREs Data Mining and Knowledge Discovery*. 2022; 12: e1470. <https://doi.org/10.1002/widm.1470>.
- Esser F, Vliegenthart R. Comparative research methods. *The International Encyclopedia of Communication Research Methods* (pp. 1–22). Wiley: London. 2017. <https://doi.org/10.1002/9781118901731.iecrm0035>.
- Esteva M, Tang JF, Xu W, Padmanabhan KA. Data mining for “big archives” analysis: A case study. *Proceedings of the American Society for Information Science and Technology*. 2013; 50: 1–10. <https://doi.org/10.1002/meet.14505001076>.
- Fayyad U, Piattetsky-Shapiro G, Smyth P. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*. 1996; 39: 27–34.
- Franks J. Text classification for records management. *Journal on Computing and Cultural Heritage*. 2022; 15: 1–19. <https://doi.org/10.1145/3485846>.
- Frendo R. Disembodied information. *Records Management Journal*. 2007; 17: 157–168. <https://doi.org/10.1108/09565690710833062>.
- Goldstein J, Evans Sabin R. Using speech acts to categorize email and identify email genres. In *Proceedings of the 39th Hawaii International Conference on System Sciences 4-7 January 2005, Hawaii, USA* (pp. 1–10). IEEE: New York. 2006. <https://doi.org/10.1109/HICSS.2006.511>.
- Greene M, Meissner D. More product, less process: Revamping traditional archival processing. *The American Archivist*. 2005; 68: 208–263. <https://doi.org/10.17723/aarc.68.2.c741823776k65863>.
- Hodel T. Supervised and unsupervised: Approaches machine learning for textual entities. In Jaillant L (ed.) *Archives, Access and Artificial Intelligence* (pp. 157–178). Bielefeld University Press: Bern. 2022.
- Hutchinson T. Protecting privacy in the archives: Supervised machine learning and born-digital records. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 2696–2701). IEEE. 2018. <https://doi.org/10.1109/BigData.2018.8621929>.
- International Council on Archives (ICA). ISAD(G): International standard for archival description. International Council on Archives: Paris. 2007.
- International Organization for Standardization (ISO), International Electrotechnical Commission (IEC). ISO/IEC 22989:2022 Information technology – artificial intelligence – artificial intelligence concepts and terminology. ISO: Geneva. 2022.
- International Organization for Standardization (ISO), International Electrotechnical Commission (IEC). ISO/IEC 24027:2021 Information technology – artificial intelligence – bias in AI systems and AI aided decision making. ISO: Geneva. 2021.
- Jaillant L (ed.). Introduction. *Archives, Access and Artificial Intelligence* (pp. 157–178). Bielefeld University Press: Bern. 2022.
- Jaillant L, Caputo A. Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data. *AI & Society*. 2022; 37: 823–835.
- Jo ES, Gebru T. Lessons from archives: Strategies for collecting sociocultural data in machine learning. In *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency 27-30 January 2020, Barcelona, Spain* (pp. 306–316). Association for Computing Machinery: New York. 2020. <https://doi.org/10.1145/3351095.3372829>.
- Journey R. *Agile Data Science 2.0: Building Full-Stack Data Analytics Applications with Spark*. O’Reilly Media: Sebastopol. 2017.
- Kale A, Nguyen T, Harris FC, Li C, Zhang J, Ma X. Provenance documentation to enable explainable and trustworthy

- AI: A literature review. *Data Intelligence*. 2023; 5: 139–162. https://doi.org/10.1162/dint_a_00119.
- Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*. 2016; 72: 2954–2965. <https://doi.org/10.1111/jan.13031>.
- Kaur D, Uslu S, Rittichier KJ, Durresti A. Trustworthy artificial intelligence: A review. *ACM Computing Surveys*. 2023; 55: 1–38. <https://doi.org/10.1145/3491209>.
- Kim HCS, An DJ, Yim JH, Rieh HY. A study on automatic classification of record text using machine learning. *Journal of the Korean Society for Information Management*. 2017; 34: 321–344. <https://doi.org/10.3743/KOSIM.2017.34.4.321>. (In Korean).
- Lampert A, Dale R, Paris C. Detecting emails containing requests for action. In *HLT '10: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 984–992). Association for Computational Linguistics: Ker-ville. 2010.
- Lee CA. Computer-assisted appraisal and selection of archival materials. In *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018* (pp. 2721–2724). Institute of Electrical and Electronics Engineers Inc.: Seattle, WA, USA. 2018. <https://doi.org/10.1109/BigData.2018.8622267>.
- Liu P, Yu H, Xu T, Lan C. Research on archives text classification based on naive Bayes. In *2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (pp. 187–190). IEEE: Chengdu, China. 2017. <https://doi.org/10.1109/ITNEC.2017.8284934>.
- Marciano R. Towards a New Discipline of Computational Archival Science (CAS). In Jaillant L (ed.) *Archives, Access and Artificial Intelligence - Working with Born-Digital and Digitized Archival Collections*. Bielefeld University Press: Bielefeld. 2022.
- Marciano R, Grant R, Hill A, Nicholas P, Scheer N, Wierdak A, Conrad M, McCoy R, Lee M, Robinson P. Bridging traditional digital humanities and archives through computational archival science. In *The Routledge Companion to Libraries, Archives, and the Digital Humanities*. IEEE: London. 2024. <https://ai-collaboratory.net/cas/>.
- Marciano R, Lemieux V, Hedges M, Esteva M, Underwood W, Kurtz M, et al. Archival records and training in the age of big data. *Re-Envisioning the MLS*. 2018; 44: 179–199. <https://doi.org/10.1108/S0065-28302018000044B010>.
- Marcus RW. NARA: A sneak preview. *Information Management Journal*. 2002; 36: 56–60.
- Martinez-Plumed F, Contreras-Ochando L, Ferri C, Hernandez-Orallo J, Kull M, Lachiche N, et al. CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*. 2021; 33: 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>.
- McMeekin N, Wu O, Germeni E, Briggs A. How methodological frameworks are being developed: Evidence from a scoping review. *BMC Medical Research Methodology*. 2020; 20: 173. <https://doi.org/10.1186/s12874-020-01061-4>.
- Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Computing Surveys*. 2022; 54: 1–35. <https://doi.org/10.1145/3457607>.
- Microsoft. O que é o processo de ciência de dados de equipe? 2024. Available at: <https://learn.microsoft.com/pt-br/azure/architecture/data-science-process/overview> (Accessed: 7 February 2024). (In Portuguese)
- Mordell D. Critical questions for archives as (big) data. *Archivaria*. 2019; 87: 140–161.
- Moss M, Thomas D, Gollins T. The reconfiguration of the archive as data to be mined. *Archivaria*. 2018; 86: 118–151.
- Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, Bobes-Bascarán J, Fernández-Leal Á. Human-in-the-loop machine learning: A state of the art. *Artificial Intelligence Review*. 2023; 56: 3005–3054. <https://doi.org/10.1007/s10462-022-10246-w>.
- Moyle S, Jorge A. RAMSYS: A methodology for supporting rapid remote collaborative data mining projects. In *ECML/PKDD01 Workshop: Integrating Aspects of Data Mining, Decision Support and Meta-Learning (IDDM-2001)*. Fürnkranz J, Wrobel S: Freiburg. 2001. Available at: https://www.researchgate.net/publication/247329752_RAMSYS-A_methodology_for_supporting_rapid_remote_collaborative_data_mining_projects (Accessed: 9 March 2024).
- National Archives and Records Administration (NARA). Managing government records directive: Automated electronic records management report/plan. 2014. Available at: <https://www.archives.gov/files/records-mgmt/prmd/A31report-9-19-14.pdf> (Accessed: 10 November 2024).
- O’Neil C. Algoritmos de destruição em massa. Editora Rua do Sabão: Santo André. 2020. (In Portuguese)
- Padilla T, Allen L, Frost H, Potvin S, Roke ER. Always already computational: Collections as data final report. 2019. Available at: <https://digitalcommons.unl.edu/scholcom/181> (Accessed: 22 December 2024).
- Pajares PR, Hernández Tornero C, Simon Martín M. IA y archivo. *Tábula*. 2023; 25: 83–98. <https://doi.org/10.51598/tab.937>. (In Spanish)
- Payne N. An intelligent class – the sequel: The development of a novel context capturing method for the functional auto classification of records. In *2023 IEEE International Conference on Big Data (BigData)* (pp. 2071–2082). IEEE: Sorrento, Italy. 2023. <https://doi.org/10.1109/BigData59044.2023.10386255>.
- Pearce-Moses R. A glossary of archival and records terminology. Society of American Archivists: Chicago. 2005.
- Piatetsky-Shapiro G. Knowledge discovery in real databases: A report on the IJCAI-89 workshop. *AI Magazine*. 1990; 11: 68. <https://doi.org/10.1609/aimag.v11i4.873>.
- Public Record Office Victoria (PROV). Email Machine Assisted

- Appraisal: Proof of Concept. 2018. Available at: <https://prov.vic.gov.au/sites/default/files/files/Blog/Government%20recordkeeping/Victoria%20Government%20Email%20Machine%20Assisted%20Appraisal%20Final.pdf> (Accessed: 12 June 2022).
- Punzalan RL, Caswell M. Critical directions for archival approaches to social justice. *The Library Quarterly*. 2016; 86: 25–42. <https://doi.org/10.1086/684145>.
- Ranade S. Access technologies for the disruptive digital archive. In Brown C (ed.) *Archival Futures* (pp. 79–98). Facet: location. 2018. <https://doi.org/10.29085/9781783302192.007>.
- Ranade S. Traces through time: A probabilistic approach to connected archival data. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 3260–3265). IEEE: Washington, DC, USA. 2016. <https://doi.org/10.1109/BigData.2016.7840983>.
- Rawat T, Khemchandani V. Feature engineering tools and techniques for better classification performance. *International Journal of Innovations in Engineering and Technology*. 2017; 8: 169–179. <https://doi.org/10.21172/ijiet.82.024>.
- Rivera SC, Kyte DG, Aiyegbusi OL, Keeley TJ, Calvert MJ. Assessing the impact of healthcare research: A systematic review of methodological frameworks. *PLOS Medicine*. 2017; 14: e1002370. <https://doi.org/10.1371/journal.pmed.1002370>.
- Rodgers M, Thomas S, Harden M, Parker G, Street A, Eastwood A. Developing a methodological framework for organisational case studies. *Journal of Epidemiology and Community Health*. 2016; A67–A67. <https://doi.org/10.1136/jech-2016-208064.129>.
- Rolan G, Humphries G, Jeffrey L, Samaras E, Antsouпова T, Stuart K. More human than human? Artificial intelligence in the archive. *Archives and Manuscripts*. 2019; 47: 179–203. <https://doi.org/10.1080/01576895.2018.1502088>.
- Shabou BM, Tièche J, Knafou J, Gaudinat A. Algorithmic methods to explore the automation of the appraisal of structured and unstructured digital data. *Records Management Journal*. 2020; 30: 175–200. <https://doi.org/10.1108/RMJ-09-2019-0049>.
- Shearer C. The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*. 2000; 5: 13–21.
- Shinkle T. Automated electronic records management. *IQ: The RIMPA Quarterly Magazine*. 2017; 33: 36–40.
- Shu K, Mukherjee S, Zheng G, Awadallah AH, Shokouhi M, Dumais S. Learning with weak supervision for email intent detection. In SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval 25 July 2020, virtual event, China (pp. 1051–1060). Association for Computing Machinery: New York. 2020. <https://doi.org/10.1145/3397271.3401121>.
- Shyam R, Singh R. A taxonomy of machine learning techniques. *Journal of Advancements in Robotics*. 2021; 8: 18–25.
- Silva DD, Alahakoon D. An artificial intelligence life cycle: From conception to production. *Patterns*. 2022; 3: 100489. <https://doi.org/10.1016/j.patter.2022.100489>.
- Sousa RTB. A classificação automática de documentos de arquivo é uma solução para os problemas que os usuários encontram com a classificação funcional? In Barros THB, Laipelt RdCF (eds.) *Organização e Representação Do Conhecimento Em Múltiplas Abordagens* (pp. 221–250). Pimenta Cultural: São Paulo. 2022. <https://doi.org/10.31560/pimentacultural/2022.95613.7>. (In Portuguese)
- Squires H, Chilcott J, Akehurst R, Burr J, Kelly MP. A framework for developing the structure of public health economic models. *Value in Health*. 2016; 19: 588–601. <https://doi.org/10.1016/j.jval.2016.02.011>.
- Sullivan P. Applying AI tools in archival functions. 3rd International Symposium ITrustAI: San Benedetto del Tronto. 2023. <https://interparestrustai.org/assets/public/dissemination/6-SullivanPlenary8Symposium.pdf>.
- Tkachenko AL, Denisova LA. Designing an information system for the electronic document management of a university: Automatic classification of documents. *Journal of Physics: Conference Series*. 2022; 2182: 012035. <https://doi.org/10.1088/1742-6596/2182/1/012035>.
- Tognoli N, Guimarães JAC. Provenance as a knowledge organization principle. *Knowledge Organization*. 2020; 46: 558–568. <https://doi.org/10.5771/0943-7444-2019-7-558>.
- Trace CB. Archival infrastructure and the information backlog. *Archival Science*. 2022; 22: 75–93. <https://doi.org/10.1007/s10502-021-09368-x>.
- Underwood W, Weintrop D, Kurtz M, Marciano R. Introducing computational thinking into archival science education. In 2018 IEEE International Conference on Big Data (Big Data). IEEE. 2018.
- Underwood W, Marciano R. Computational thinking in archival science research and education. In 2019 IEEE International Conference on Big Data (Big Data). IEEE. 2019.
- Vellino A, Alberts I. Assisting the appraisal of e-mail records with automatic classification. *Records Management Journal*. 2016; 26: 293–313. <https://doi.org/10.1108/RMJ-02-2016-0006>.
- Verdonck T, Baesens B, Óskarsdóttir M, Vanden Broucke S. Special issue on feature engineering editorial. *Machine Learning*. 2021; 113: 3917–3928. <https://doi.org/10.1007/s10994-021-06042-2>.
- Vital LP, Brascher M. Descrição arquivística: Uma discussão conceitual. *Informação & Informação*. 2016; 21: 213–213. <https://doi.org/10.5433/1981-8920.2016v21n1p213>. (In Portuguese)
- Wang Z, Wu J, Yu G, Song Z. Text analysis and visualization research on the Hetu Dangse during the Qing Dynasty of China. *Information Technology and Libraries*. 2021; 40. <https://doi.org/10.6017/ital.v40i3.13279>.
- Watanabe E, Sousa RTB. Uso do aprendizado de máquina para a classificação automática de documentos de arquivo: Experimento inicial em uma organização pública. *Tendências da Pesquisa Brasileira em Ciência da Informação*. 2023; 16: 1–

27. (In Portuguese)

Weintrop D, Beheshti E, Horn M, Orton K, Jona K, Trouille L, et al. Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*. 2016; 25: 127–147. [https://doi.org/10.1007/s10956-](https://doi.org/10.1007/s10956-015-9581-5)

015-9581-5.

Wing JM. Computational thinking. *Communications of the ACM*. 2006; 49: 33–35. <https://doi.org/10.1145/1118178.1118215>