Teachable AI for the Archival Professions – Module 2: **Critical AI/ML, Indigenous AI/ML, and AI/ML Ethics for Archival Professionals**

Kaila Fewster[1] and Richard Arias-Hernandez[2]

This educational module is part of a series of learning materials developed by InterPARES Trust AI[3] researchers and educators to train archival professionals and students to effectively leverage artificial intelligence in their archival work. The current version was completed on December 11, 2024.

---

[1] InterPARES Trust AI Graduate Academic Assistant, University of British Columbia.

[2] Associate Professor of Teaching, School of Information, University of British Columbia and InterPARES AI Trust Researcher. richard.arias@ubc.ca

[3] This case study is an outcome of InterPARES Trust AI, an international research partnership led by Drs. Luciana Duranti and Muhammad Abdul-Mageed, University of British Columbia, and funded by the Social Sciences and Humanities Research Council of Canada (SSHRC). https://interparestrustai.org/

Module 2: Critical AI/ML, Indigenous AI/ML, and AI/ML Ethics for Archival Professionals

READ FOR THIS MODULE

Required:

- Cameron, S., Franks, P., & Hamidzadeh, B. (2023). Positioning Paradata: A Conceptual Frame for AI Processual Documentation in Archives and Recordkeeping Contexts. *Journal on Computing and Cultural Heritage*. https://doi.org/10.1145/3594728

- Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, *8*(1), 108. Operationalizing the CARE and FAIR Principles for Indigenous data futures | Scientific Data

- Hervieux, S. & Wheatley, A. (2020). *The ROBOT test* [Evaluation tool]. The LibrAIry. https://thelibrairy.wordpress.com/2020/03/11/the-robot-test

- Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*, 1–22. https://doi.org/10.1080/0960085X.2021.1927212

- Lewis, J. E., Abdilla, A., Arista, N., Baker, K., Benesiinaabandan, S., Brown, M., Cheung, M., Coleman, M., Cordes, A., Davison, J., Duncan, K., Garzon, S., Harrell, D. F., Jones, P.-L., Kealiikanakaoleohaililani, K., Kelleher, M., Kite, S., Lagon, O., Leigh, J., … Whaanga, H. (2020). *Indigenous Protocol and Artificial Intelligence Position Paper*[Monograph]. Indigenous Protocol and Artificial Intelligence Working Group and the Canadian Institute for Advanced Research. p. 3-24. https://doi.org/10.11573/spectrum.library.concordia.ca.00986506

- Memarian, B., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, *5*, 100152. https://doi.org/10.1016/j.caeai.2023.100152

Recommended:

- Chapman University. (n.d.). *Bias in AI.* Artificial Intelligence (AI) Hub. Retrieved September 19, 2024, from *Bias in AI | Chapman University*

- Duranti, L., & Rogers, C. (2024). *Artificial Intelligence and Documentary Heritage* (pp. 1–99). UNESCO.Artificial intelligence and documentary heritage

- European Commission. (n.d.). *AI Act*. Shaping Europe's Digital Future. Retrieved November 14, 2024, from AI Act | Shaping Europe's digital future

- Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, *30*(1), 99–120. The Ethics of AI Ethics: An Evaluation of Guidelines | Minds and Machines

- Rana, V. (2024). Indigenous Data Sovereignty: A Catalyst for Ethical AI in Business. *Business & Society*, 00076503241271143. https://doi.org/10.1177/00076503241271143

- Siddik, M. A. B., Shehabi, A., & Marston, L. (2021). The environmental footprint of data centers in the United States. *Environmental Research Letters*, *16*(6), 064017. https://doi.org/10.1088/1748-9326/abfba1

- Tapu, Ian Falefuafua, & Fa'agau, Terina Kamailelauli'i. (2022). new age indigenous instrument: artificial intelligence & its potential for (de)colonialized data. Harvard Civil Rights-Civil Liberties Law Review, 57(2), 715-754.

OVERVIEW

This module provides an overview of the ethical challenges that artificial intelligence (AI) and machine learning (ML) can pose to archives and records management, and illustrates how archival professionals can more critically engage with these technologies in their work to mitigate issues of bias, privacy, and transparency. Regarding Indigenous data sovereignty and AI/ML, this module emphasizes that Indigenous data should only be used with appropriate permissions and in ways that respect community ownership, Indigenous protocols, and cultural values. Furthermore, it also discusses how to critically assess AI tools using different evaluative frameworks, explores the environmental impact of AI, and highlights the importance of documenting AI processes using paradata.

LEARNING OBJECTIVES

By the end of this lesson, students will be able to:

- Explain the importance of ethics, decolonization, and critical theory in regard to applications of AI in archival practice.
- Critically evaluate AI models and datasets for their potential social, ethical, and environmental impacts.
- Understand the importance of Indigenous data sovereignty and community partnerships when working with Indigenous data and AI/ML.
- Identify opportunities to use paradata to improve on the accountability of archival institutions using AI for their workflows.

Introduction

When working with AI applications and models, it is necessary to be aware of the potential harms and challenges that come along with these technologies. Not only is it important for archivists and records managers to evaluate how and why they are using these tools in their work, but it is also relevant to consider the social and ethical implications of integrating AI into archival workflows.

Outside of the archives, conversations around AI ethics have become widespread with the emergence of big data, automated decision-making systems (ADMS), and large language models (LLMs), which have transformed the way information is created and shared across personal and professional settings (Chalmers, 2023; Floridi, 2023; UNESCO, 2022). Given this transformation, understanding and reflecting on the ethical considerations of using AI is essential in any field, but even more so in archives and records management, where concerns of accuracy, bias, compliance, and privacy are already at the top of mind.

AI ethics is a quickly emerging branch of applied ethics which is primarily concerned with the ethical issues that arise from understanding AI systems as objects (i.e. bias, privacy concerns) and the moral questions raised by recognizing AI systems as subjects (i.e. general artificial intelligence) (Waelen, 2022). More broadly, applied ethics is a subfield of the more extensive study of moral philosophy (or ethics) that considers the practical aspects of right and wrong concerning real-life actions and behaviours. In other words, applied ethics is not concerned with purely theoretical issues but is instead grounded in practical normative challenges (Søbirk & Ryberg,

2019). For records managers and archivists, understanding the ethical issues raised by AI systems *as objects* is most important when looking at the ethics of AI.

It is important to note that there are distinct differences between ideal ethical AI guidelines and the legal and regulatory frameworks currently governing AI systems. While ethics and the law are both systems of rules and norms humans, or in this case, AI systems and their developers, are expected to follow when conducting themselves in society, ethics concerns internal sets of controls, whereas the law refers to external mechanisms of control (Gundugurti, 2022). In other words, ethics is a form of governance which, although non-binding, can shape behaviour and actions through social norms (Koniakou, 2023). On the other hand, the law governs behaviour by imposing binding rules enforced through most often government institutions. Both of these governance modalities are necessary for meaningfully regulating the development and use of AI; however, the extent to which one shows more promise in regulating AI is still up for debate (see Black & Murray, 2019; Hagendorff, 2020).

Despite this, a good example of ethical and legal AI governance comes in the form of the General Data Protection Regulation, passed by the EU Parliament in 2016. While not specifically about AI, specific GDPR provisions introduce the principles of transparency, explainability and enhanced accountability for personal data processing and decision-making by AI (Koniakou, 2023). More recently, in 2024, the EU Parliament adopted the EU Artificial Intelligence Act, a first-of-its-kind piece of legislation focused on promoting safety, transparency, and traceability when using AI systems (European Parliament, 2023). The Act defines three risk-level categories for AI applications through which they are regulated differently and serves as a preliminary legal framework for other governing bodies to follow suit (Future of Life Institute, 2024). This legislation mainly outlines regulatory obligations for high-risk AI

developers, which may not cover all the challenges of using lower-risk AI relevant to archivists and records managers, including concerns with the authenticity and accuracy of AI outcomes.

Therefore, from a critical theory perspective, AI ethics still have a practical goal in attempting to empower individuals and protect them from systems of power (Waelen, 2022). Critical theory, as a school of thought, is rooted in critiques and social movements that have organized against the unequal power relationships entrenched in our current societal structures (Ryoo & McLaren, 2010). Understanding AI ethics from this lens can help pinpoint other ethically relevant issues that may have been missed without a power analysis (Waelan, 2022). As archives themselves are often embedded into pre-existing institutions of power, it is necessary to consider how these existing power relations can impact the use of AI tools and their outcomes in these spaces.

Analyzing the challenges of using AI in archives and records management from different critical perspectives helps highlight potential concerns that arise from embedding these technologies into archival and records management workflows. It also provides the opportunity to address these concerns before integration to avoid perpetuating harm through biases and black-box algorithms.

ACTIVITY #1

- In groups, students will roleplay through a predetermined scenario in which an information organization (e.g. archive, museum, cultural centre, etc.) is considering using Indigenous visual records of arts and culture as training data for a Generative AI model. Students will then be asked to discuss the role of ethics and law in this situation from an

Indigenous relationality-focused perspective and from a Western copyright and moral rights perspective. Students address the question: what role do ethics and law play in this situation?

Critical Artificial Intelligence & Machine Learning

Working with AI in archives and records management requires understanding and critically evaluating the systems used to ensure that they are fit for purpose in the archive and that their outcomes are trustworthy and authentic. There are several useful frameworks for critically evaluating AI models when it comes to their accuracy, fairness, and ethicality, which involve reflecting on the training data, the algorithm, and the uses of the model.

One of these frameworks is called the ROBOT test, which was designed by the LibrAIry team at McGill University and requires reflecting on sets of questions about the AI model being used (Hervieux & Wheatley, 2020). This test primarily assesses the system's Reliability, Objective, Bias, Ownership, and Type to evaluate its legitimacy. For instance, when looking at the Reliability of the model, it is relevant to consider how much information is available about the model itself, including the parties responsible for developing and training the algorithms and any possible biases in this provided information. Similarly, it is important to look at the Objective of the model, and consider the goals of using this technology and sharing information about it. As mentioned, it is also necessary to look for potential Bias in models and be aware of any related ethical implications, which will be further discussed below. Another measure of evaluation in the ROBOT test focuses on Ownership of the model, including reflecting on who developed it, who is responsible for it, and the potential restrictions on who can access and

use the model. Finally, the last part of the ROBOT test requires an evaluation of the Type of system being used, including the subtype of AI, the type of information it relies on, and whether it needs human intervention.

While the ROBOT test is useful for evaluating external forces which may impact AI models and their usage, it is also relevant to examine the models from a more holistic perspective to evaluate how well they respond to human and environmental needs and limitations. The Fairness, Accountability, Transparency and Ethics, or FATE, framework is a loose and flexible collection of relevant principles to be considered and implemented into AI development and education. The principle of fairness is the most often referenced framework principle in AI literature, and in general, refers to the landscape, culture, situation or practices in development that attempt to mitigate bias in models' outcomes so that benefits and burdens are equally distributed among impacted stakeholders (Memarian & Doleck, 2023). In other words, looking for fairness in AI algorithms requires consideration of both external and internal forces that could impact the system's outcomes. Similarly, the principle of accountability focuses on examining the different preventative or mitigation strategies designed to hold those who own, design, sell and use AI algorithms responsible for the system's outcomes (Memarian & Doleck, 2023). When it comes to evaluating AI algorithmic transparency, it can be understood through several different lenses. From a high-level perspective, transparency concerns making black box model design more apparent, and ensuring AI use is clearly outlined in institutional or organizational policy (Memarian & Doleck, 2023). On the other hand, at a lower level, AI algorithms should be transparent in how they work, either mathematically or in layman's terms, to their users (Memarian & Doleck, 2023). Finally, the principle of ethics is more broad and encompassing than the others and generally considers the need to raise awareness around AI's various ethical issues like bias or misinformation, and makes the case for stronger governance measures, support systems and organizational

structures to minimize risk and ensure AI development and use is legally compliant.

Following an evaluative framework like the ROBOT test or the FATE Framework makes it easier to assess the quality of data used to train AI models, as well as understand how the algorithms are being used in the model and determine whether they produce unbiased, accurate and trustworthy results. These frameworks are also valuable for ensuring the model complies with legal regulations and ethical guidelines and for prompting archivists and records managers using these tools to reflect on their individual roles in ensuring high-quality results when working with AI models.

ACTIVITY #2

- In groups or alone, evaluate the DataWorks Plus program and its use by the Government to support decision-making using two different frameworks (ROBOT, VALID-AI). Then compare their results from both frameworks to critically evaluate their usefulness.

Complying with legal regulations when using AI is essential for upholding the principles of data privacy and protection. This is especially important for archivists and records managers when using AI tools with sensitive documents or records containing personal information. As mentioned in the introduction, frameworks like the EU's GDPR and AI Act have been implemented to protect the collection and processing of individuals' personal information by third parties, whether using AI or not (Wolford, 2018). However, no legislation as robust as the GDPR has been introduced in Canada or the United States yet.

In Canada, the Office of the Privacy Commissioner did release a set of privacy principles, primarily aimed at generative AI, in late 2023. Essentially, the document suggests key privacy principles for both AI developers and organizational users to keep in mind when working with the AI system. These principles include following all existing legal regulations around personal information collection, considering the appropriate purposes and necessity of collecting personal information, and ensuring transparency and accountability around how the information is processed, stored, and disposed of (Office of the Privacy Commissioner of Canada, 2023).

Similar to Canada, the US has no comprehensive federal legislation regulating the development and use of AI. However, in 2020, the National Artificial Intelligence Act was passed, which guides AI research and development undertaken at federal science agencies and established the National AI Initiatives Office, which is responsible for overseeing the US' AI strategies (White & Case LLP, 2024). More recently, an executive order directed at federal agencies as well as major players in AI development like Google, OpenAI and Meta was signed by President Joe Biden in October 2023 and mandated the development of federal standards along with open dissemination of safety testing results among AI developers (Szczepański, 2024). The order also calls for accelerating the development of privacy-protection techniques across public and private organizations using AI to protect citizens from AI-related risks; however, this also speaks to a broader need for more robust data privacy legislation federally in the US.

When working within these legal frameworks, regardless of the use of AI, archivists and records managers need to clearly articulate their purposes for storing, keeping, and subsequently providing access to personal data. Once AI is integrated into archival workflows, maintaining transparent decision-making audit trails becomes paramount. In this case, engaging with these technologies in the archives requires a framework of AI governance that is

informed by "well developed language and procedures of consent, power, inclusivity [and] transparency" (Jaillant & Caputo, 2022; Sadler, 2024). Therefore, archivists and records managers must pay close attention to the ethical principles of AI development and usage to ensure these benchmarks are meaningfully met.

Beyond the broad governance frameworks regulating AI use and development, there are also copyright concerns when using AI models, especially those that are trained on exceptionally large datasets. For instance, should an archive choose to use an off-the-shelf AI image recognition model, it is possible that the model was trained on copyrighted data, which naturally affects its outcomes. In particular, generative AI models face significant copyright issues as their outcomes can often be, in whole or in part, directly generated from copyrighted content. When training AI models, the quantity of data is essential, but it is also important to consider the data quality; many developers decide to use datasets that, knowingly or not, include copyrighted materials for training their models (Levine & Bolton, 2023). Furthermore, data related to Indigenous traditional knowledge, which is not currently protected within the Western copyright system, is used for training models without the appropriate permissions from such communities (Lewis et al., 2020). In this sense, archivists and records managers must consider the source of AI models' training data to ensure that copyrighted data does not influence its outcomes.

With the growing importance of recognizing Indigenous data sovereignty, it is relevant to reflect on how culturally sensitive data is being integrated into AI models, appropriately or not, with or without due permission from originating communities. When data is scraped from the internet and used to train AI models, it can often exacerbate existing social biases and deepen inequalities by reinforcing existing social issues within the models' algorithms (Lewis et al., 2020). As research on and about Indigenous Peoples has historically

been extremely exploitative and extractive, there is general distrust among communities in non-native control and use of this data (Tapu & Fa'agau, 2022). Additionally, current copyright frameworks fail to properly recognize Indigenous ways of knowing and recordkeeping like oral histories, rituals, chants, and artefacts, which hold the traditions and customs that form cultural knowledge and guide Indigenous Peoples' lives (Tapu & Fa'agau, 2022). In this sense, Western liberal copyright laws privilege individual intellectual property rights and authorship, while Indigenous conceptions of cultural and intellectual property more often privilege community relationships and consider the responsibilities and obligations associated with different types of knowledge (Mills 2017). Thus, in our current system, this kind of knowledge becomes unprotected and open to extraction by non-Indigenous individuals and organizations. In response to these challenges, the International Indigenous Data Sovereignty Interest Group developed the CARE principles, a framework for Indigenous data governance which grounds the collection, use, and dissemination of Indigenous data in Indigenous worldviews (Indigenous Data Sovereignty Interest Group, 2023). Designed to work in tandem with other open data movements, the framework focuses on the **C**ollective benefit, **A**uthority to control, **R**esponsibility and **E**thics of working with Indigenous data (Carroll et al., 2021). Records managers and archivists should take care to implement the CARE principles and other Indigenous data sovereignty frameworks when working with different AI models to ensure culturally sensitive data is being handled appropriately and with the correct permissions from communities.

Aside from misuse of culturally sensitive data in training AI models, online data scraping for AI training has caused issues in other industries, such as the stock photo company Getty Images' lawsuit against StabilityAI, a popular text-to-image AI model developer, which accuses them of misusing over 12 million copyrighted photos for training their models (Brittain, 2023). Furthermore, several high-profile creatives, like comedian Sarah Silverman,

have also joined lawsuits against Meta and OpenAI, alleging protected works were also copied and ingested to train these companies' AI programs (Small, 2023). Beyond training the models, there has also been concern about AI models taking on celebrity likeness, for instance, when OpenAI and ChatGPT came under fire for releasing a voice chat option for the GPT engine that sounded eerily like actress Scarlett Johansson, who had previously voiced an AI in the 2013 movie, Her (Tenbarge, 2024). Johansson came out against OpenAI, stating that CEO Sam Altman had previously asked her to provide her voice for the model, but she declined, indicating that the company may have partially used her likeness without her consent, violating her personality rights (Tenbarge, 2024). In this sense, there is a general concern across AI model creation, development, and dissemination about the illegal use of copyrighted data and even personal likeness in training models, which archivists and records managers must consider when working with third-party models. Additionally, archivists should also reflect on how uploading collections online in the name of accessibility may also lead to inadvertently contributing data to AI model training sets.

ACTIVITY #3 (requires reading Lewis et al. (2020))

- Warning: this activity exposes students to cultural appropriation, lack of sensitivity to Indigenous production of cultural works, and an image that it was created without following proper Indigenous protocols, with the goal of provisioning for a teaching moment that brings these topics of discussion to the core.
- Students are presented with an image generated by commercial GenAI that exemplifies cultural appropriation of Indigenous art.
- In groups, students will then discuss the potential ethical issues around the creation and use of these types of

images. Trigger question: What is wrong with this? (training data, algorithm, commercial product, company behind the product, prompt, the generator of the prompt/user of GenAI).

As mentioned throughout the module so far, biases in AI models and algorithms have proven to be a substantial issue as outputs are highly dependent on the training data they receive, which often reflects existing social inequalities and entrenches them within the algorithms. Using these technologies in archives and records management could cause further harm by perpetuating these biases if applications are not properly evaluated beforehand. However, there is also growing concern about misinformation being spread by AI models, either through algorithmic biases or 'hallucinations,' an issue most common with Large Language Models, or LLMs. AI hallucinations can roughly be understood as the model failing to produce outputs based on the training data or following an identifiable pattern, essentially "making answers up," which leads to inaccurate or nonsensical outputs (Metz, 2023). Reliable and non-reliable sources are indistinguishable to AI models, and given that much of the training data for these models comes from Western-centric internet sources, the system's knowledge base is inherently skewed and thus more prone to misinformation (Sweetman & Djerbal, 2023). Furthermore, these hallucinations can be exacerbated by the LLMs' inability to recall long-term information in complex scenarios and tendency to capture spurious correlations as causal relationships (Huang et al., 2023). Archivists and records managers should be aware of the possibility of their AI applications hallucinating and should be prepared to mitigate them by clearly defining the application's purpose and use cases, putting limits on potential response outputs, testing the system continually, and using high-quality training data. Since there is always a risk of perpetuating bias and misinformation when using LLMs and other AI applications, it becomes the responsibility of the human user to double-check

responses and ensure AI project outcomes are realistic and meaningfully applicable.

ACTIVITY #4

- Some Harm Considerations of LLMs: Visit this link to engage with an interactive image that outlines some of the potential harms to consider when engaging with AI models, specifically Large Language Models, or LLMs. Choose one harm that can potentially be caused by LLMs, read the corresponding vignette and skim through the additional resources. Write a short summary of this harm in 250 words or less and share it with the class.

While AI is rapidly being integrated into applications across industries, less thought is often given to the environmental impacts of developing and using such energy-heavy technologies. Massive data centres have already been identified as meaningful contributors to climate change, requiring over 1% of the world's total energy production to store our digital information (Siddik et al., 2021). When it comes to AI models, training a single model can produce hundreds of tons of carbon, equivalent to the annual carbon emissions of hundreds of North American households combined (Ren & Wierman, 2024). Training and running AI models requires massive amounts of data, which is often stored in data centres, plagued with their own myriad of adverse environmental impacts, including massive consumption of energy, noise pollution, massive freshwater consumption, and negative physiological and psychological effects on nearby communities (Siddik et al., 2021). Data centres are also becoming increasingly naturalized into our environment in the sense that they are becoming integral to parts of local industrial and natural landscapes and community narratives (Hogan & Vonderau, 2019). The concept of the cloud as immaterial and omnipresent has contributed to how data centres are perceived as an inevitable aspect of nature in today's

connected society (Hogan & Vonderau, 2019). This omnipresent attitude towards massive amounts of data stored in the cloud has also contributed to AI being perceived as an essential component of all applications. As archivists and records managers increasingly work with digital resources, it is worth considering how the storage of that data and the use of AI applications to manipulate that data may be negatively contributing to environmental degradation.

Critical Algorithms & Critical Data

As previously mentioned, all algorithms have some type of bias. There are two commonly known types of bias: implicit and explicit. Implicit bias refers to attitudes or internalized feelings that unconsciously affect actions and decisions, whereas explicit bias refers to attitudes which consciously and directly influence beliefs, values and actions (Stoneybrook et al., 2008). When it comes to algorithms, they necessarily reflect both the external and internal biases of the developer unless steps are taken to mitigate the integration of these attitudes. Most commonly, algorithms are affected by systemic and implicit biases. There have been several cases of large organizations like Amazon and Apple having issues with gender discrimination in their recruitment algorithms, and in 2016, ProPublica released an expose illustrating how risk assessment algorithms used in the US criminal justice system had a significant bias against African Americans and other people of colour (Kordazeh et al., 2021; Angwin et al., 2016). However, legislation like the EU's GDPR has encouraged AI and ML developers to contend with these biases and the algorithms they use in interaction with the public (Kordazeh et al., 2021). Nonetheless, computational techniques alone are not enough to prevent bias, and systems must be improved to include transparency, auditability, and control features to encourage bias detection and mitigation (Kordazeh et al., 2021). When working with algorithms, archivists and records managers should take an

active role in identifying biases in their applications and be proactive in mitigating their potential negative effects on users.

When it comes to working with data in AI applications, especially for archivists and records managers, it is essential to consider the data itself as records of the processes taking place within the applications. Whether its external datasets used to train models or the datafied records held by the organization being processed by AI and ML, it is relevant to evaluate the authenticity, integrity and trustworthiness of this data to ensure it is well-suited to providing the desired outcomes (Cameron and Hamidzadeh, 2024). It is also important to obtain and maintain custody over these datasets as they were used originally in AI-related processes, as evidence (Cameron et al., 2023). Archivists and records managers are well-situated to engage with these issues based on the existing interests in the field in promoting transparent, consensual and trustworthy records.

Thinking of data(sets) as records also brings metadata into the discussion as a method of tracking data provenance and continually recording processes in which the data is involved (Gilliland-Swetland, 2016). Metadata is used to arrange, describe, document, preserve, and manage digital resources, which is essential for identifying provenance. Understanding data(sets) as records then means including information like ownership, rights permissions, data source, and other administrative metadata with datasets to ensure maximum transparency in articulating the original context of creation and use for the data(set). Initiatives like the Data Documentation Initiative (DDI) offer a suite of standards and products, including metadata schemas, which are developed to better document data(sets) and improve discoverability in line with the FAIR principles discussed above (Data Documentation Initiative, n.d.). These schemas are interoperable with other common recordkeeping standards and provide an easier way for developers, AI or otherwise, to begin documenting how their data was collected and what happens with it

over time. To document how data(sets) were initially used in AI processes and to capture the necessary information to ensure accountability and transparency of these processes, it is worth considering how paradata can be employed. While paradata will be discussed in more depth below, it can essentially be understood as "information about the procedure(s) and tools used to create and process information resources, along with information about the persons carrying out those procedures." (Cameron et al., 2023). In other words, paradata captures information, or in an archival sense, evidence, of the processes in which the dataset(s) and the algorithm(s) were used, and information about the individuals carrying out these processes. Recording this information when working with AI applications is essential because it enables procedural transparency and accountability, unlike the more traditional black-box AI model. With this in mind, both metadata and paradata are necessary components to understanding data(sets) as records as they provide the contextual information needed to meaningfully evaluate the integrity and authenticity of AI applications and the outcomes they produce. As such, archivists and records managers should continue to understand data(sets), algorithms, and related documentation as records and push to properly document AI processes using metadata and paradata to preserve the evidence of these transactions.

As data and datasets are key components of all AI models, it is worth briefly discussing how this data can be biased or manipulated to often dramatically influence the outcomes of the model. Ideally, the hope is that the data used in all applications is free from biases and manipulation, but this is often not the case. Data manipulation techniques like cherry-picking, data dredging, and data falsification can all have drastic impacts on an AI model's outcomes by misrepresenting the data inputs and changing the paths of causality. For archivists and records managers using AI applications, this misrepresented causality can dramatically affect the outcomes produced by the model, thus impacting the results' authenticity, integrity, and evidential value. For

example, cherry-picking is a common way of manipulating data where only the most favourable results are considered, and those results which do not support the desired results are ignored. In the context of AI, cherry-picking data involves highlighting the most desirable outcomes from a set of possibilities outlined by the system and downplaying less favourable ones, which leads to a biased representation of the algorithm's performance capabilities and impacts the system's overall trustworthiness and integrity ("*Cherry Picking*," n.d.). Thus, to minimize the risk of cherry-picking, it is necessary for the datasets used by AI models and applications to have extensive metadata records detailing their provenances and how the data was collected.

A similar issue to cherry-picking that can negatively affect AI model outcomes is data dredging, also sometimes known as data fishing or p-hacking. Data dredging is a misuse of statistical analysis to find patterns that are presented as statistically significant in large volumes of data, like those used for training AI models (Awati, 2022). This can lead to a higher rate of false positives, where the 'statistically significant' data was cherry-picked or manipulated in some way (Awati, 2022). When it comes to working with AI and ML applications, data dredging, while making the model appear very robust, actually can lead to a model that does not generalize or accept new data well, is difficult to reproduce, and may perpetuate additional biases ("*P-Hacking a Statistical Pitfall of Machine Learning," 2020*). Like cherry-picking, data-dredging can be avoided by having high-quality metadata and paradata records for both the datasets and AI/ML models to transparently define how outcomes are determined to maintain result authenticity and integrity and how the specific AI application was evaluated.

Another potential issue is data falsification, also known as data fabrication, which broadly refers to data manipulation with the intent of misrepresenting the results (*Data Fabrication/data falsification,* n.d.). Data falsification can

take various forms, such as changing or adding data points or even removing 'inconvenient' results (*Data Fabrication/data falsification,* n.d.). There are already concerns about how some LLMs have the ability to generate realistic but entirely fake datasets, which can then spread as misinformation or be used to pose serious threats to research integrity (Chen et al., 2024). When it comes to the data ingested by an AI model, if these datasets have been previously falsified, then they can skew the model's training and, therefore, its results.

In the world of AI, there are some gray areas, though, with situations in which data can be fabricated without being falsified. This is the case of synthetic data. Synthetic data has been growing in popularity for training AI models in an attempt to mitigate the privacy and copyright concerns of using actual human data. Synthetic data is based on real-world data, where AI models are trained to recognize the patterns, correlations, and statistical properties of the real data (*What is Synthetic Data,* 2021). Once trained, the models can reproduce these characteristics in statistically identical but synthetic data (*What is Synthetic Data,* 2021). There are potential benefits for using synthetic data in AI models, like more effective protection of people's data and improved fairness by attempting not to replicate societal biases (Riemann, 2024). However, the quality of synthetic data is highly correlated to its original dataset, and inaccuracies in the initial training stages may cause broader systemic issues in the entire synthetic dataset. Moreover, if synthetic data is repeatedly used to train AI models, it is likely that the data itself will become increasingly disconnected from reality, leading to potentially inaccurate or even completely false results. In this sense, then, it is important to know the provenances of the datasets used to train AI models to determine if there has been any data manipulation, either intentional or otherwise, and to ensure the data is reflective of reality.

ACTIVITY #5

Read this summary of the 2023 New York Times v. OpenAI lawsuit and explore how understanding data(sets) as records is relevant in this case. In small groups, discuss how paradata could be beneficial in addressing AI lawsuits.

Finally, as discussed above, bias is a major issue when it comes to properly training AI models. Datasets can often replicate systemic and societal biases, which, when used to train AI models, can reproduce those biases in their outcomes. There are many types of bias which can impact AI/ML models and their outcomes, including societal biases (e.g., systemic racism, patriarchalism, homophobia, etc.), selection bias, confirmation bias, and measurement bias, which can be introduced to the system in all stages of development, from data collection and data labelling, to model training and deployment (Chapman University, n.d.). Societal bias refers to when AI models are trained, unintentionally or otherwise, to reflect social intolerance or systemic discrimination (*Bias in AI and Machine Learning*, 2022). Although the datasets and algorithms may appear unbiased, their outputs may still reinforce societal biases if not properly screened, and even then, societal biases are hard to trace (*Bias in AI and Machine Learning*, 2022). Since these biases are already ingrained in everyday life, and therefore real-world datasets, it can be challenging to recognize and address them. Selection bias occurs when the data used to train the model is not representative of the reality intended to be modeled, and can be caused by incomplete data, biased sampling or any other reason which could lead to unrepresentative data (Chapman University, n.d.). In an archival or records management context, selection bias could occur due to incomplete metadata records or incorrect classifications for different training documents. Confirmation bias happens when the system relies too much on existing trends in the data, which reinforces the bias within the model and fails to highlight other potentially new and meaningful patterns. This can cause the model to be

unrepresentative of reality and exacerbate other biases within the system. Lastly, measurement bias refers to when the data collected and used to train the model is systematically different from the actual variables of interest, thus leading to inaccurate or incomplete outcomes (Chapman University, n.d.). This can lead to models missing statistically significant relationships and potentially highlighting irrelevant patterns. To mitigate concerns of measurement bias, it is important to make sure that models are highly attuned to the type of work they're completing and the training datasets reflect the topics of interest.

ACTIVITY #6

Data Fallacies to Avoid: Visit this link to explore different types of data fallacies, which all involve statistical data being misused, misrepresented or misapplied. In small groups of 2-3, discuss how each fallacy may impact AI/ML models and/or datasets, and consider how these issues could be mitigated from an archival perspective.

Given the amount of resources needed to train and run AI models, their development is primarily overseen by large, pre-existing technology companies or governments with the capacity to run complex models. However, there is a power dependence in these systems between those who develop and shape the system and those who use the system (Maas, 2023). Algorithmic systems like AI models have power over individuals in the sense that as people's lives become increasingly datafied, this data is used to 'reveal reality' and thus cannot lie (Pop Stefanija, 2023). As such, the decisions made through the algorithms based on our data end up guiding who is provided services, who is refused them, and ultimately, how individuals are perceived (Pop Stefanija, 2023). Even if properly screened for data fallacies, AI models can still show signs of systemic societal biases, which are hard to trace and mitigate. Smaller organizations developing AI

models may face difficulties in screening their algorithms for biases due to a lack of resources or expertise and, in this sense, may be more prone to externalizing power asymmetries. As a result, it is crucial to embed accountability measures within the model to track its behaviour and determine how it produces its outcomes.

Indigenous data and Indigenous AI/ML in the Archives

When considering how AI can be implemented into archival and records management practice, it is relevant to consider the type and content of the materials held in the archives, and how AI could improve (or detract) from the arrangement, description, and discoverability of a collection. Many archives hold sensitive data which may not be appropriate for ingesting into a model. For example, archival institutions in settler-colonial states like the U.S. and Canada have thousands of records about Indigenous Peoples, the atrocities committed against them, and their ongoing disenfranchisement. These records are often made inaccessible to communities. Archives in Canada, in particular, have faced a reckoning with the purpose of their collections after the release of The Truth and Reconciliation Commission of Canada's Calls to Action in 2015, which called for interrogation around archival treatment of Indigenous records and data (Truth and Reconciliation Commission of Canada, 2015; Steering Committee on Canada's Archives, 2022). Furthermore, with the adoption of the United Nations Declaration on the Rights on Indigenous Peoples, even more focus has been drawn to Indigenous data sovereignty and the protection of Indigenous People's right to their traditional knowledge (United Nations, 2007). In this sense, there is an inherent tension between an archive's use of AI models and its obligation to protect Indigenous data rights.

The term Indigenous data is often used shorthand for describing Indigenous knowledge, information, and materials, as it can encompass data by and

about Indigenous Peoples, their cultures and customs, and their relationships with the land (Mukunda, 2023). Similarly, Indigenous data sovereignty refers to the right of Indigenous people to have ownership of and stewardship over their own data (whether analogue or digital) and information pertaining to their distinct societies. This means that Indigenous Peoples, as independent nations, have the ability to manage information in ways that are consistent with their ways of life, cultures, and customs (Kukutai et al., 2016). As archivists and records managers who may hold or steward Indigenous records, it is important to consider how to create opportunities for Indigenous Peoples to claim greater control over the data connected to them held in institutional repositories, whether these repositories are directly owned and managed by them or not (Kukutai et al., 2016). Working towards data decolonization requires that Indigenous peoples hold the power to determine who is considered Indigenous and what records and data are considered Indigenous rather than settler states or other non-Indigenous organizations. It also requires that data collected by or about Indigenous Peoples reflects their ontology, interests, and priorities. Indigenous communities should not only have power over the content of the data collected about them, but also over who can access it (Kukutai et al., 2016).

Following this line of reasoning, it holds that using Indigenous data held in archival institutions for training and working with AI models may be inappropriate if the proper permissions for these types of data uses are not explicitly granted by the communities represented in the data. Meaningful engagement between Indigenous data and AI requires a shift away from an extractive model of AI and a focus instead on developing models from a more accountable, collaborative, and culturally responsive approach derived from sustained partnerships between Indigenous communities and archives (Rana, 2024). There is worthwhile potential in using AI to benefit Indigenous communities, like language revitalization work and environmental monitoring. However, archivists and records managers considering working

with these technologies must go beyond superficial relationships with communities and instead develop "deep, sustained partnerships that center the voices, knowledge and priorities of Indigenous Peoples" (Rana, 2024; Walter & Kukutai, 2018). Furthermore, as discussed throughout this module, more transparency and accountability is needed during AI model development and deployment to ensure Indigenous communities influence how their information is being used in these systems and can evaluate the convenience of these information developments for their community. In particular, there needs to be clear mechanisms for accountability and reparations for when AI does cause harm to Indigenous communities through misusing or misappropriating Indigenous data (Rana, 2024). Finally, archivists and records managers can lend their support to Indigenous-based digital records initiatives by acknowledging Indigenous data sovereignty and providing access to the records by or about Indigenous Peoples on their own terms. Additionally, when developing digital records solutions and working with Indigenous data, it is necessary that the focus of the project remains grounded in Indigenous knowledge and value systems, and ultimately focus on the communities' priorities over the organizations. This can only be ensured by creating, growing, and sustaining deep and meaningful partnerships with Indigenous communities.

Indigenous knowledge systems are "holistic, dynamic, and generative system[s] that are embedded in lived experience" (Lewis et al., 2020). However, Indigenous knowledge systems are also particularly vulnerable to the impacts of globalization and the underlying goal of creating a "global village based on cultural, social, political, and economic homogenization" (Lewis et al., 2020). With homogenization can come loss, and the loss of languages, histories, cultures, and ecosystems contributes to the dissolution of identities and, ultimately, a loss of power (Lewis et al., 2020). In this sense, there are valid concerns among marginalized communities that if AI

applications accelerate these homogenizing changes, they will potentially also exacerbate the losses.

Reasonably, many Indigenous communities will choose not to engage with AI, and therefore also have the right to refuse institutional use of their data for training and working with AI models. Tuck illustrates that research and the pursuit of knowledge within the academy can often be exploitative and represent an ongoing form of extractive settler colonialism (Tuck and Yang, 2014). As such, refusal is more than just a 'no,' but a broader rejection of colonization as an inevitable and monopolizing force (Tuck and Yang, 2014). Furthermore, Simpson highlights that refusal can in itself be generative, as a way of telling archivists when to stop, what not to do, why not to do it, and how to strengthen relationships with Indigenous communities (Simpson, 2007). As institutions with settler-colonial roots, it is necessary for archives to consider how their projects may contribute negatively to the exploitation and extraction of knowledge from Indigenous communities and in meaningful relationships with these communities open the space for refusal. Archivists must respect and abide by Indigenous communities' decisions, especially when communities refuse to have their records to be part of projects they do not feel will meaningfully serve them. In the context of AI, this becomes paramount as data extraction from third-party AI models is already a concern for continuing and unauthorized extraction of Indigenous data and knowledge. In this sense, when working with AI, it is relevant to consider whose realities the models reflect and what underlying assumptions inform their algorithmic rules.

Furthermore, it is also necessary to recognize that algorithms will not recognize social and historical contexts and may also ignore or contravene Indigenous protocols unless they are taught. As such, when it comes to working with Indigenous data and AI, there must be a concerted effort during the algorithmic training process to ensure the model can recognize the social,

historical, cultural, and political contexts of the records and reflect them in its outputs (Walter and Kukutai, 2018). Generative AI presents severe challenges to protecting Indigenous data, knowledge and intellectual property rights, as many freely available models do not adequately consider the cultural contexts of the information they ingest, and thus, many outputs appropriate Indigenous knowledge without proper community involvement and due permissions (Cardona-Rivera et al., 2024). This poses serious threats to Indigenous data sovereignty, and as stewards of the Indigenous records which exist throughout colonial institutions, archivists and records managers have a responsibility to use these technologies cautiously through partnership with, and under the guidance of, Indigenous communities. AI has the potential for use in both archives and Indigenous communities; however, its meaningful implementation requires detailed evaluation and tailoring of models to ensure they meet the community's needs and uphold the principles of Indigenous data sovereignty.

AI/ML Records and Paradata

As previously discussed in this module, there is a rapidly growing need to document the processes and procedures behind using dataset(s) and algorithm(s) in AI applications and information about the individuals carrying out these processes. Collecting this information, called paradata, is essential for enabling procedural transparency and audit-trail-like accountability for AI outcomes. For archives in particular, paradata can be understood as "information recorded and preserved about records' processing with AI tools," which provides archivists and records managers with a framework for articulating their assessment, application, and documentation needs when working with AI applications for archival purposes (Cameron et al., 2023). With this data, it is necessary for archivists to develop AI records to provide evidence of the decisions that went into choosing and implementing AI tools (Cameron et al., 2023). It is also worth noting here that AI records, or

paradata more broadly, cannot only document the application's algorithms and the dataset(s) but must also provide information on the full scope of the tool's use and context to illustrate its potential impacts on the collection. Moreover, many AI tools do not naturally produce the appropriate or necessary documentation, and thus archivists and records managers must take an active and deliberate role in ensuring the correct paradata is created (Cameron et al., 2023). As such, it is worth considering how paradata and AI records are valuable, not just to archivists and records managers, and how they can be used to inform critical AI and ML practice in the archive.

Paradata, while a newer concept in the context of archives and AI, has previously been applied in statistical sciences, virtual heritage visualization, and research dataset documentation, with the common definition focusing on providing information about the processes of creation, curation and management of other information resources (Cameron et al., 2023.) Paradata is an opportunity for archivists and records managers to not only make AI processes more transparent and accountable but also use computerized tools to develop better and more effective ways of transparently documenting archival decisions made throughout a record's lifecycle (Duranti & Rogers, 2024). Therefore, applying the concept of paradata in archival contexts also includes creating processual documentation, or evidence, of the different actions taken by actors, human or otherwise, on materials and the impact of these decisions on accessibility, authenticity and trustworthiness of the records throughout their life cycles.

Since there is extreme variability in AI tools and the circumstances in which they are applied, there is no standard method of documenting their actions and decisions, and these processes can become even more complex when considering the specific needs of stakeholders in different contexts. Nonetheless, one way to conceptualize how paradata can be used is through the Machine Learning Lifecycle, which illustrates different types of actions

taken during the AI model training process (Franks, 2024a). The ML lifecycle can be broken into six broad categories: obtaining and formatting the dataset, developing or obtaining the ML model, training the model with the dataset that was prepared, evaluating the model performance, implementing the model, and monitoring and possibly continuously improving the model with new data (Franks, 2024a). Through each of these steps, paradata like design plans, generated computer logs, model training parameters, and impact assessments can be collected and aggregated to be referenced along with the information resource modified or produced by the AI tools (Franks, 2024a). However, it is often not as simple as collecting all the documentation from each step and making it freely available, as many algorithms, training datasets and models are restricted by legal contracts and security measures, making creating or obtaining transparent and comprehensive processual AI documentation difficult.

Nonetheless, collecting paradata is still important as AI applications become more prevalent in high-risk sectors like immigration, education, and employment, where documentation about the design, development and deployment of the application is a necessary mechanism for transparency and accountability. For instance, the proposed Artificial Intelligence and Data Act (AIDA) in Canada, if passed, will establish measures to mitigate risks of harm and bias outputs, including publishing plain-language descriptions of systems and the risk reduction actions taken, and requirements for assessing, monitoring and documenting possible harms and risk mitigation measures (Franks, 2024b; Parliament of Canada, 2022). Similarly, the EU's AI Act also considers how to document AI processes based on their assessed risk level (Franks, 2024b). In particular, high-risk applications would be subject to strict regulations requiring activity logs, detailed documentation about the system, developer, and deployer, and high-quality datasets (European Commission, n.d.). In this sense, paradata is an essential component to meaningfully governing AI and ensuring its decision-making

processes are transparent, fair and equitable. From an archival perspective, capturing this information is also necessary to establish authoritative records about AI processes and preserve the records' authenticity, reliability, integrity and usability (Franks, 2024b).

To get a good sense of how paradata works practically, it is helpful to look at a real-world example. iTrust AI researchers Alex Richmond and Mario Beauchamp, working with the Bank of Canada, developed a proof of concept illustrating what constitutes paradata and how it is captured (Richmond, 2023). Recall that paradata is information about the model, the data used, and the processes which provide the desired results. With this in mind, Richmond and Beauchamp highlight the type of paradata that should be collected at the Bank of Canada in Table 1 below.

Table 1: Information Paradata Should Capture

Value(s)	Details it provides
Identification Metadata	Name of the data sets, types, associations, pipelines
Data-set Metadata	Ablation method, training/validation/test split ratios, size, date, source
Model-Related Metadata	Learning rate, parameters such as weights & biases, hyperparameters, quality & performance metrics
Experiment or Project Metadata	What has been used to capture data processing or model training runs, number of epochs, optimization algorithm

Pipeline Metadata	Details on how to execute the ML workflows
Operationalization Metadata	Audit logs details, result statistics

Following this framework, they also tested their proof of concept on a real algorithm used at the Bank of Canada. The goal of the AI application was to improve their existing program to review external content about international oil market rates. It used an EconBERT model developed by the researchers. The model used training data from research papers, federal reserve reports, Bank of Canada documents and the Daily Oil Bulletin. The results, which illustrate real, captured paradata, are shown in Table 2 below.

Table 2: Paradata Examples from BoC Algorithm

Value(s)	Details it provides
Identification Metadata	**Name**: Daily Oil Bulletin
Data-set Metadata	**Training/validation/test split ratios:** training 90 %/ validation 10 %
Model-Related Metadata	**Hyperparameters:** •Learning Rate 3e-1 •Epoch 2 •Batch Size: 16 •weight decay of 0.001 **Performance Metrics;** Accuracy 88%
Experiment or Project Metadata	**Optimization Algorithm**: Adam W

Pipeline Metadata	**Validates models:** •Oil Bulletin from current day •Separate by sentence •Score each sentence •Aggregation of relevancy •Send results to economists

This case study is merely an example of how paradata can be captured, and as previously mentioned, every AI algorithm and model is different; therefore, it is challenging to develop a standardized way of capturing paradata. Nonetheless, proof of concept examples like that of Richmond and Beauchamp illustrate that paradata can be operationalized in a specific application domain, and it is valuable and necessary to capture to ensure AI and algorithmic accountability.

While AI and ML present many challenges to archives and records management, they also have significant potential for automating manual recordkeeping processes and improving discoverability, among other even more innovative uses. Still, a lack of transparency and explainability among models can pose issues in these contexts where trustworthiness and authenticity are essential. In this sense, paradata is proposed to help uncover issues in the current archival documentation of AI processes and potentially contribute to the development of archival AI documentation standards in the future (Cameron et al., 2023). As such, paradata is relevant to consider when investigating how to ethically use AI in archival contexts, as it provides an opportunity to better document the model's agency and impact as an actor working with the records. Additionally, working with paradata marks a shift towards more critical practice in the archive, where the agency and contexts of those working with the records are just as important to document as that of the records themselves. In this sense, integrating AI into these spaces is also an opportunity to more critically evaluate existing

archival and records management practices and establish more ethical ways of working with records using AI, ML, or otherwise.

INTERACT/ACTIVITY #7

Visit this link to access the results of an AI impact assessment on Canada's Federal Access to Information and Privacy (ATIP) Online Request Service using Canada's Algorithmic Impact Assessment Tool. Hit the download bottom at the top of the page to get the report. Look at the 2022 updated results and identify questions which are asking about forms of paradata. Then, discuss in groups whether enough paradata is being collected and made available about the ATIP Online Request Service Model, based on the results of the assessment.

MODULE COMPREHENSION ACTIVITY

Imagine a real or fictionalized relationship between AI, records, and a specific community (e.g., racialized, Indigenous, intellectual property holders, etc.) and develop a creative assignment (e.g., comic, design of a board game, short story, children illustrated story, 5-min podcast, 3-min video, etc.) that investigates this relationship from a critical perspective, based on the concepts covered throughout this module.

SUMMARY

In summary, while AI has the potential to bring meaningful change to archival and records management practice, it is also important to remember the current ethical challenges

with AI/ML, particularly when it comes to privacy, transparency, and bias. The integration of AI across industries has been rapid, and governance and regulation has yet to catch up when it comes to developing and enforcing ethical use guidelines and laws. In this sense, archivists and records managers should make use of practical tools like the ROBOT test and the FATE framework to critically evaluate AI tools and determine how they can be used to produce trustworthy and unbiased outcomes. Furthermore, respecting Indigenous data sovereignty is essential when using AI tools in the archive, and community partnerships are necessary to ensure ethical and responsible handling of culturally sensitive information. It is also important to consider the environmental impacts of using AI, as the activity of training models is energy-intensive and has been shown to have negative environmental effects. Finally, it is necessary to better document AI processes using paradata to maintain transparency and accountability in archival work while using these technologies. Although AI holds significant potential for archival practice, it is ultimately the responsibility of archivists and records managers to ensure that these technologies are used in contextually appropriate, culturally sensitive, and ethically responsible ways.

REFERENCES

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Awati, R. (2022, April). *What is Data Dredging (data fishing)?* TechTarget. https://www.techtarget.com/searchdatamanagement/definition/data-dredging

Bias in AI and Machine Learning: Sources and Solutions. (2022, December 7). Lexalytics. https://www.lexalytics.com/blog/bias-in-ai-machine-learning/

Black, J., & Murray, A. D. (2019). Regulating AI and machine learning: Setting the regulatory agenda. *European Journal of Law and Technology*, *10*(3), 1–17. https://www.ejlt.org/index.php/ejlt/article/view/722

Brittain, B. (2023, February 6). Getty Images lawsuit says Stability AI misused photos to train AI. *Reuters*. https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/

Cameron, S., Franks, P., & Hamidzadeh, B. (2023). Positioning Paradata: A Conceptual Frame for AI Processual Documentation in Archives and Recordkeeping Contexts. *Journal on Computing and Cultural Heritage*. https://doi.org/10.1145/3594728

Cameron, S., & Hamidzadeh, B. (2024). *Preserving Paradata for Accountability of Semi-Autonomous Ai Agents in Dynamic Environments: An Archival Perspective*. Elsevier BV. https://doi.org/10.2139/ssrn.4681230

Cardona-Rivera, R. E., Alladin, J. K., Litts, B. K., & Tehee, M. (2024). *Indigenous Futures in Generative Artificial Intelligence: The Paradox of Participation*. https://uen.pressbooks.pub/teachingandgenerativeai/chapter/indigenous-futures-in-generative-artificial-intelligence-the-paradox-of-participation/

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, *8*(1), 108. https://doi.org/10.1038/s41597-021-00892-0

Chalmers, D. J. (2023). *Could a Large Language Model be Conscious?* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2303.07103\

Chapman University. (n.d.). *Bias in AI.* Artificial Intelligence (AI) Hub. Retrieved September 19, 2024, from https://www.chapman.edu/ai/bias-in-ai.aspx

Chen, Z., Chen, C., Yang, G., He, X., Chi, X., Zeng, Z., & Chen, X. (2024). Research integrity in the era of artificial intelligence: Challenges and responses. Medicine, 103(27), e38811. https://doi.org/10.1097/MD.0000000000038811

Cherry Picking. (n.d.). Envisioning. Retrieved September 19, 2024, from https://www.envisioning.io/work/cherry-picking

Data Documentation Initiative. (n.d.). *Why Use DDI?* Data Documentation Initiative. Retrieved September 19, 2024, from https://ddialliance.org/learn/why-use-ddi

Data fabrication / data falsification. (n.d.). Springer - International Publisher. Retrieved September 25, 2024, from https://www.springer.com/gp/authors-editors/editors/data-fabrication-data-falsification/4170?srsltid=AfmBOoqzZmlWz8TgLT97rub5lOU47yP2dwuHlND6iXgVFroSfKD11Pke

Duranti, L., & Rogers, C. (2024). *Artificial Intelligence and Documentary Heritage* (pp. 1–99). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000389844

European Commission. (n.d.). *AI Act*. Shaping Europe's Digital Future. Retrieved November 14, 2024, from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Parliament. (2023, June 8). *EU AI Act: First regulation on artificial intelligence*. Topics | European Parliament. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities* (1st ed.). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198883098.001.0001

Franks, P. (2024a). In the Pursuit of Archival Accountability: Positioning Paradata as AI Processual Documentation. *Society of American Archivists*. https://www2.archivists.org/sites/all/files/Franks_In%20the%20Pursuit%20of%20Archival%20Accountability.pdf

Franks, P. (2024b). *Paradata: What is it and why do we care?* 5th International Symposium, ITrust AI, Honolulu, HI. https://interparestrustai.org/assets/public/dissemination/6-FRANKSPARADATATrustAISymposium_pptx.pdf

Future of Life Institute. (2024). *High-level summary of the AI Act | EU Artificial Intelligence Act*. https://artificialintelligenceact.eu/high-level-summary/

Gilliland-Sweland, A. (2016). Introduction to Metadata: Setting the Stage. In *Introduction to Metadata* (3rd ed.). Getty Research Institute. https://www.getty.edu/publications/intrometadata/setting-the-stage/

Gundugurti, P. R., Bhattacharyya, R., Kondepi, S., Chakraborty, K., & Mukherjee, A. (2022). Ethics and Law. *Indian Journal of Psychiatry*, *64*(Suppl 1), S7–S15. https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry_726_21

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

Hervieux, S. & Wheatley, A. (2020). *The ROBOT test* [Evaluation tool]. The LibrAIry. https://thelibrairy.wordpress.com/2020/03/11/the-robot-test

Hogan, M., & Vonderau, A. (2019). The Nature of Data Centers. *Culture Machine*, *18*. https://culturemachine.net/vol-18-the-nature-of-data-centers/the-nature-of-data-centers/

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2023). *A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions*(Version 1). arXiv. https://doi.org/10.48550/ARXIV.2311.05232

Indigenous Data Sovereignty Interest Group. (2023). *CARE Principles*. Global Indigenous Data Alliance. https://www.gida-global.org/care

Jaillant, L., & Caputo, A. (2022). Unlocking digital archives: Cross-disciplinary perspectives on AI and born-digital data. *AI & SOCIETY*, *37*(3), 823–835. https://doi.org/10.1007/s00146-021-01367-x

Koniakou, V. (2023). From the "rush to ethics" to the "race for governance" in Artificial Intelligence. *Information Systems Frontiers*, *25*(1), 71–102. https://doi.org/10.1007/s10796-022-10300-6

Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*, 1–22. https://doi.org/10.1080/0960085X.2021.1927212

Levine, J., & Bolton, J. (2023, November 14). *Primer: Training AI Models with Copyrighted Work*. AAF. https://www.americanactionforum.org/insight/primer-training-ai-models-with-copyrighted-work/

Levy, S. R., & Killen, M. (2008). *Intergroup Attitudes and Relations in Childhood Through Adulthood*. Oxford University Press, USA.

Lewis, J. E., Abdilla, A., Arista, N., Baker, K., Benesiinaabandan, S., Brown, M., Cheung, M., Coleman, M., Cordes, A., Davison, J., Duncan, K., Garzon, S., Harrell, D. F., Jones, P.-L., Kealiikanakaoleohaililani, K., Kelleher, M., Kite, S., Lagon, O., Leigh, J., … Whaanga, H. (2020). *Indigenous Protocol and Artificial Intelligence Position Paper*[Monograph]. Indigenous Protocol and Artificial Intelligence Working Group and the Canadian Institute for Advanced Research. https://doi.org/10.11573/spectrum.library.concordia.ca.00986506

Maas, J. (2023). Machine learning and power relations. *AI & SOCIETY*, *38*(4), 1493–1500. https://doi.org/10.1007/s00146-022-01400-7

Memarian, B., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, *5*, 100152. https://doi.org/10.1016/j.caeai.2023.100152

Metz, C. (2023, March 29). What Makes A.I. Chatbots Go Wrong? *The New York Times*. https://www.nytimes.com/2023/03/29/technology/ai-chatbots-hallucinations.html

Mills, A. (2017). Learning to Listen: Archival Sound Recordings and Indigenous Cultural and Intellectual Property. Archivaria, 83, 109-124. Retrieved from https://archivaria.ca/index.php/archivaria/article/view/13602

Mukunda, K. (2023, August 4). *Indigenous data sovereignty | SFU Library*. https://www.lib.sfu.ca/help/publish/research-data-management/indigenous-data-sovereignty

Office of the Privacy Commissioner of Canada. (2023, December 7). *Principles for responsible, trustworthy, and privacy-protective generative AI technologies*. https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/

Parliament of Canada. (2022, June 16). *C-27 (44-1)—First Reading—Digital Charter Implementation Act, 2022.* https://www.parl.ca/documentviewer/en/44-1/bill/C-27/first-reading

P-Hacking a Statistical Pitfall of Machine Learning. (2020, February 10). Wovenware. https://www.wovenware.com/blog/2020/02/p-hack-ml-pitfall/

Pop Stefanija, A. (2023). Power asymmetries, epistemic imbalances and barriers to knowledge: The (im)possibility of knowing algorithms. In S. Lindgren

(Ed.), *Handbook of Critical Studies of Artificial Intelligence* (pp. 563–572). Edward Elgar Publishing. https://doi.org/10.4337/9781803928562.00058

Rana, V. (2024). Indigenous Data Sovereignty: A Catalyst for Ethical AI in Business. *Business & Society*, 00076503241271143. https://doi.org/10.1177/00076503241271143

Ren, S., & Wierman, A. (2024, July 15). *The Uneven Distribution of AI's Environmental Impacts*. Harvard Business Review. https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts

Richmond, A., & Beauchamp, M. (2023, October 28). *Architecting Accountability in AI: Paradata and AI in Commodity Markets at the Bank of Canada*. 4th International Symposium, iTrustAI, Vancouver British Columbia Canada. https://interparestrustai.org/assets/public/dissemination/6-Richmond-ArchitectingAccountabilityInterPARESUBCPlenaryandSymposiumPresentation.pptx

Riemann, R. (2024, September 24). *Synthetic Data.* European Data Protection Supervisor. https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data

Ryoo, J. J., & McLaren, P. (2010). Critical Theory. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education (Third Edition)* (pp. 348–353). Elsevier. https://doi.org/10.1016/B978-0-08-044894-7.01388-9

Sadler, J. K. (2024, May 16). Analyzing Artificial Intelligence Methods in Digital Preservation Workflow. *Association of Canadian Archivists*. https://www.archivists.ca/Blog/13358000

Small, Z. (2023, July 10). *Sarah Silverman Sues OpenAI and Meta Over Copyright Infringement*. The New York Times. https://www.nytimes.com/2023/07/10/arts/sarah-silverman-lawsuit-openai-meta.html

Siddik, M. A. B., Shehabi, A., & Marston, L. (2021). The environmental footprint of data centers in the United States. *Environmental Research Letters*, *16*(6), 064017. https://doi.org/10.1088/1748-9326/abfba1

Simpson, A. (2007). On Ethnographic Refusal: Indigeneity, 'Voice' and Colonial Citizenship. *Junctures: The Journal for Thematic Dialogue*, *9*, Article 9. https://junctures.org/index.php/junctures/article/view/66

Søbirk Petersen, T., & Ryberg, J. (2019, September 25). *Applied Ethics*. OxfordBibliographies.

https://www.oxfordbibliographies.com/display/document/obo-9780195396577/obo-9780195396577-0006.xml

Steering Committee on Canada's Archives. (2022). *Reconciliation Framework: The Response to the Report of the Truth and Reconciliation Commission Taskforce* (pp. 1–117). Steering Committee on Canada's Archives. https://archives2026.com/wp-content/uploads/2022/02/reconciliationframeworkreport_en.pdf

Sweetman, R., & Djerbal, Y. (2023, May 25). *ChatGPT? We need to talk about LLMs*. University Affairs. https://universityaffairs.ca/opinion/in-my-opinion/chatgpt-we-need-to-talk-about-llms/

Szczepański, M. (2024). *United States approach to artificial intelligence*. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/757605/EPRS_ATA(2024)757605_EN.pdf

Tapu, Ian Falefuafua, & Fa'agau, Terina Kamailelauli'i. (2022). new age indigenous instrument: artificial intelligence & its potential for (de)colonialized data. Harvard Civil Rights-Civil Liberties Law Review, 57(2), 715-754.

Tuck, E., & Yang, K. W. (2014). R-Words: Refusing Research. In *Humanizing Research: Decolonizing Qualitative Inquiry with Youth and Communities* (pp. 223–248). SAGE Publications.

Truth and Reconciliation Commission of Canada. (2015). *Truth and Reconciliation Commission: Calls to Action*. Truth and Reconciliation Commission of Canada. http://nctr.ca/ assets/reports/Calls_to_Action_English+.pdf.

United Nations. (2007). *United Nations Declaration on the Rights of Indigenous Peoples*. https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2018/11/UNDRIP_E_web.pdf

UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence* (1-43). https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

Waelen, R. (2022). Why AI Ethics Is a Critical Theory. *Philosophy & Technology*, *35*(1). https://doi.org/10.1007/s13347-022-00507-5

Walter, M., & Kukutai, T. (2018). *Artificial Intelligence and Indigenous Data Sovereignty* [Input Paper for the Horizon Scanning Project "The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing" on behalf of the Australian Council of Learned Academies].

https://acola.org/wp-content/uploads/2019/07/acola-ai-input-paper_indigenous-data-sovereignty_walter-kukutai.pdf

What is synthetic data? (2021, September 24). *Mostly AI*. https://mostly.ai/what-is-synthetic-data

White & Case LLP. (2024, May 13). *AI Watch: Global regulatory tracker - United States*. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wolford, B. (2018, November 7). *What is GDPR, the EU's new data protection law?* GDPR.Eu. https://gdpr.eu/what-is-gdpr/