# arbido

Suderman Jim,

Tags: Archiv Big Data,

A seismic shift for archival ethics?

Exploring how archival concepts and principles can inform the development of responsible artificial intelligence (AI) is one of the goals of InterPARES Trust AI resarch project. This article looks at commonalities and differences of the ethical principles of the archival and AI communities.

The «Principled Artificial Intelligence» study by Jessica Fjeld, et al, provides a snapshot of the ethical landscape of the AI field in early 2020. The study examines 36 sets of principles, dating from 2016 to 2018, selected for their perceived influence, geographical representation, approach, and source (categorized as governments, intergovernmental organizations, companies, professional associations, advocacy groups, and multi-stakeholder initiatives). The study's three main findings are:

1. The principles contained fall into eight common themes;

2. More recent principles tended to cover all eight themes

3. 64% of the codes reference human rights, with five (14%) explicitly using a human rights framework.
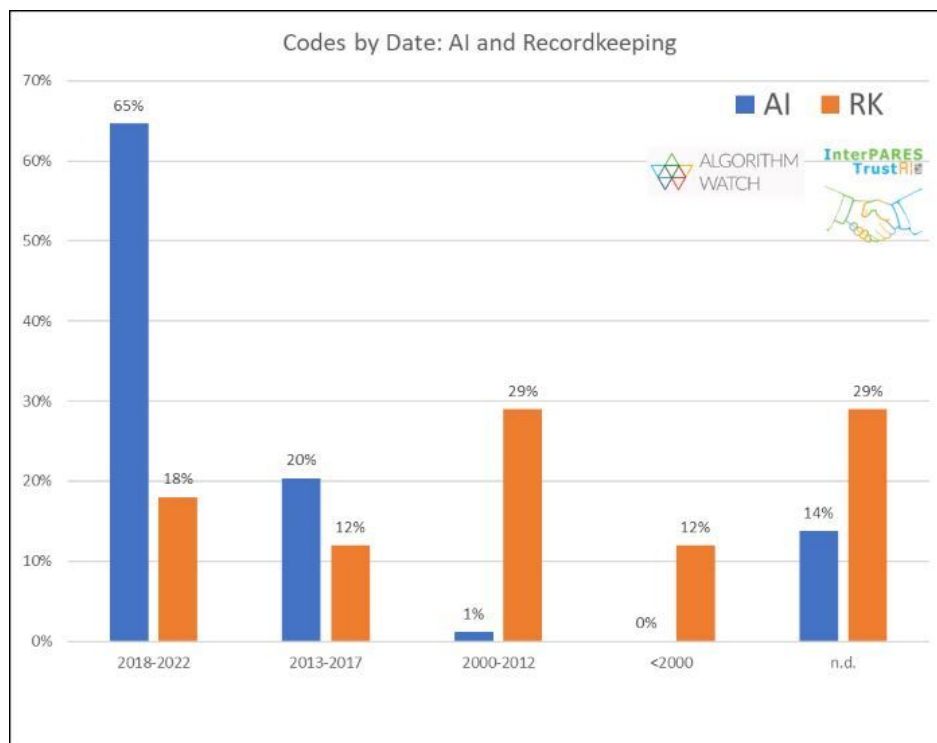
Codes by Date

An overview of archival ethical principles was taken by InterPARES researchers for the purpose of supporting a comparative study. Ten archival ethical codes, dating from 1993 to 2020 were selected to establish a baseline. All are products of national or international professional associations. This study identified themes similar to those of the Fjeld study, including accountability, privacy, and transparency. Two additional themes, labelled 'Social Justice' and 'Trustworthiness of Recordkeepers', have much in common with the themes of 'Promotion of Human Values' and 'Professional Responsibility' identified by the Fjeld study.

The implementation of Machine Learning (ML) and other Artificial Intelligence (AI) tools and technologies will exert a significant impact on archivists and record keepers. Statistics show adoption of AI systems is increasing and is widespread, with ISO's technical report on AI use cases identifying no less than 24 domains for categorizing AI use cases, including automated vehicles, fraud detection, medical diagnosis, and judicial recommendations. AI systems consume and produce a great deal of data, often with unclear governance, fostering uncertainty in terms of what existing laws are applicable, which of those can be enforced in this context, and what additional risks are being run.

Continuing value of authentic records

Core archival values, such as the value of authentic records, will likely remain constant. For example, the accountability principle described by the IEEE's «Ethically Aligned Design" states: «Systems for registration and record-keeping should be established so that it is always possible to find out who is legally responsible for a particular A/IS [autonomous and intelligent systems].» But fulfilling such a requirement will be challenging when «it's not always clear when an AI system has been implemented in a given context, and for what task.»
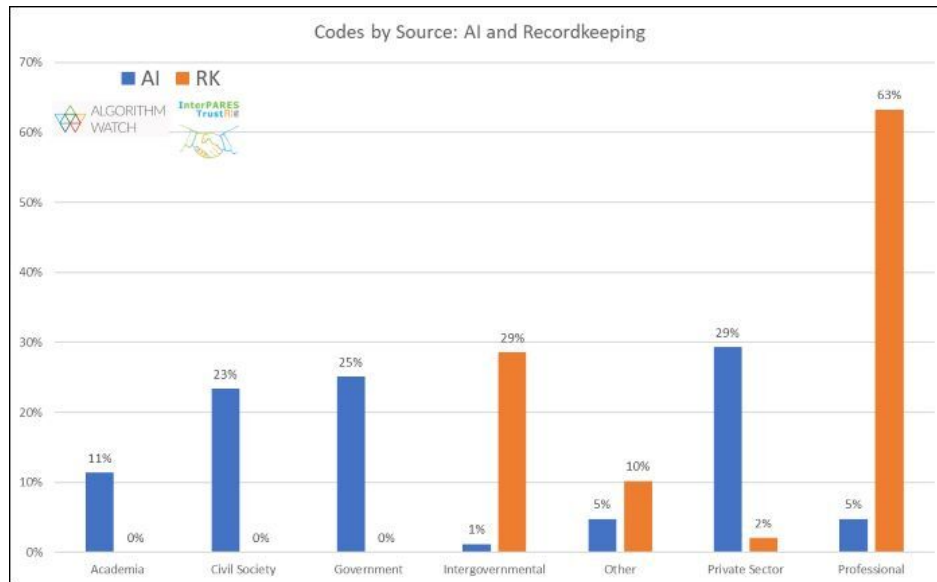
Codes by Source

Trust is fundamental to the ethical principles of both communities. In the Fjeld study it relates to human control of AI-based technology, transparency, accountability, and the protection of human rights. The scope and scale of impact of AI systems is evident in documents such as the OECD's «Recommendation of the Council on Artificial Intelligence» (OECD/LEGAL/0449, 2022), which states:

Alongside benefits, AI also raises challenges for our societies and economies, notably regarding economic shifts and inequalities, competition, transitions in the labour market, and implications for democracy and human rights. (p 3)

The range of actors identified by the AI Risk Management Framework (NIST, 2023), which draws from the OECD's framework for classifying AI systems, is similarly inclusive: «…trade associations, standards developing organizations, researchers, advocacy groups, environmental groups, civil society organizations, end users, and potentially impacted individuals and communities.» (p. 10)

Compare the foregoing with the ICA's «Universal Declaration on Archives» (2010):

> [Archives] play an essential role in the development of societies by safeguarding and contributing to individual and community memory. Open access to archives enriches our knowledge of human society, promotes democracy, protects citizens' rights and enhances the quality of life.

Trust is fundamental to the ethical principles of both communities. In the Fjeld study it relates to human control of AI-based technology, transparency, accountability, and the protection of human rights. The scope and scale of impact of AI systems is evident in documents such as the OECD's «Recommendation of the Council on Artificial Intelligence» (OECD/LEGAL/0449, 2022), which states:

Alongside benefits, AI also raises challenges for our societies and economies, notably regarding economic shifts and inequalities, competition, transitions in the labour market, and implications for democracy and human rights. (p 3)

The range of actors identified by the AI Risk Management Framework (NIST, 2023), which draws from the OECD's framework for classifying AI systems, is similarly inclusive: «…trade associations, standards developing organizations, researchers, advocacy groups, environmental groups, civil society organizations, end users, and potentially impacted individuals and communities.» (p. 10)

Compare the foregoing with the ICA's «Universal Declaration on Archives» (2010):

[Archives] play an essential role in the development of societies by safeguarding and contributing to individual and community memory. Open access to archives enriches our knowledge of human society, promotes democracy, protects citizens' rights and enhances the quality of life.

The UDA asserts that managing archives is « [t]collective responsibility of all – citizens, public administrators and decision-makers, owners or holders of public or private archives, and archivists and other information specialists.» So, from the perspectives of both AI and archival ethics, archivists might very well be considered 'AI actors' – as contributors of data for model training, as members of civil society organizations, as end users, etc., where principles require us to avoid or minimize significant injustices and any destructive impact "on the individual and collective rights and the welfare of our fellow citizens.»

Ethical accountability

Many archival principles are complemented by practice guidance. However, such guidance is often absent when the principle statement is vague, e.g., admonitions to respect privacy. Just over half of the AI codes examined by the Fjeld set out using tools such as human rights impact assessments. This tool, while imperfect, will at least contribute to transparency by estimating the risks or harms that might result from implementation of an AI system and outlining how they are mitigated.
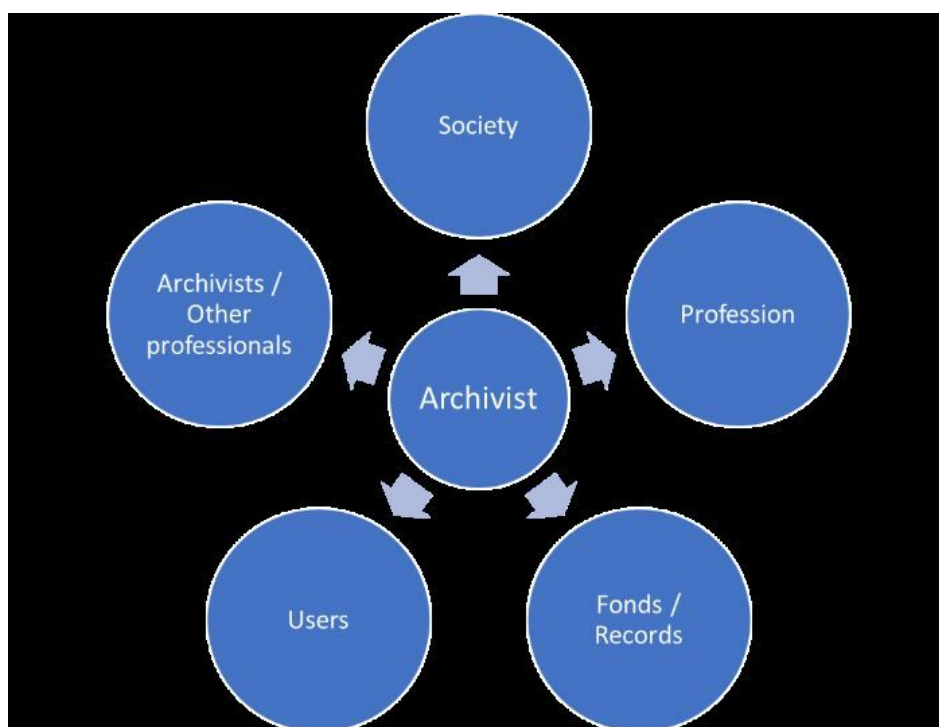
Catalan Code

Any consideration of accountability principles must encompass those to whom practitioners are accountable. The code of the Archivists Association of Catalonia has accountability embedded into its very structure; its principles grouped into five relationships: i) towards society; ii) the profession; iii) fonds and records; iv) users; and v) archivists and other professionals. While these accountabilities are clear, the consequences of failing to live up to them are not. By contrast, accountability principles in the AI field emphasize evaluation and auditing of AI systems by independent monitors internal or external to the organization, with some including outcomes such as terminating an AI system that is found to violate international conventions or human rights or establishing remedies for automated decisions.

Another interesting contrast between the principles of the archival and AI communities is the scope and duration of responsibility.

Archivists must perform their duties and functions in accordance with archival principles, with regard to the creation, maintenance and disposition of current and semi-current records, including electronic and multimedia records, the selection and acquisition of records for archival custody, the safeguarding, preservation and conservation of archives in their care, and the arrangement, description, publication and making available for use of those documents.

There is not much that falls outside this scope, especially when «making available» encompasses respecting the privacy of record creators and subjects. The Fjeld study observes that some codes «point out that existing systems may be sufficient to guarantee legal responsibility for AI harms, with actors including Microsoft and the Indian AI strategy looking to tort law and specifically negligence as a sufficient solution.» (p. 34) Others, such as the Chinese White Paper on AI Standardization, advocates distinguishing between system development and application, with transparency being the primary consideration for systems in development and equal rights and responsibility for system implementations.

Few would disagree that the field of AI is large and dynamic. The archival community also seems to be in a state of transition, the driving force perhaps having more to do with human rights protection and social justice. If that is correct, then the values these are driving may generate a convergence with similar priorities for the ethics of the AI community. If so, then the scope responsibility set out in archival ethical principles may be considerably extended by the values and risks of the AI field.

Jim Suderman

Jim Suderman recently retired from the position of Director of Information Access at the City of Toronto where he oversaw the operations of the City's records management, archives, and information and privacy protection programs. He is a researcher with the InterPARES Trust AI Project, based at the University of British Columbia.

Abstract

Français
Cet article reflète l'état actuel d'une étude d'éthique comparative en cours dans le cadre du projet de recherche InterPARES Trust AI. Au-delà des premiers résultats présentés par l'auteur lors de la journée professionnelle du VSA-AAS en 2022, il conclut que les archivistes devront peut-être se considérer comme des acteurs de l'IA, s'appuyant sur des principes des deux domaines, ce qui entraînerait un élargissement de la portée des responsabilités archivistiques.

Deutsch
Dieser Artikel spiegelt den aktuellen Stand einer vergleichenden Ethikstudie wider, die im Rahmen des KI-Forschungsprojekts InterPARES Trust durchgeführt wird. Ausgehend von den ersten Ergebnissen, die der Autor auf der VSA-AAS-Fachtagung 2022 vorgestellt hat, kommt er zu dem Schluss, dass sich Archivare möglicherweise als KI-Akteure betrachten müssen, die sich auf Prinzipien aus beiden Bereichen stützen, was zu einer Erweiterung des archivarischen Verantwortungsbereichs führt.

English
This article reflects the current state of a comparative ethics study underway within the InterPARES Trust AI research project. Going beyond the early findings presented by the author at the VSA-AAS Symposium last year, it concludes that archivists may need to consider themselves as AI actors based on principles from both fields, with a resulting expansion of scope of archival responsibilities.