OXFORD

Applying Al to digital archives: trust, collaboration and shared professional ethics

Lise Jaillant (D) 1,*, Arran Rees (D) 2

- ¹School of Social Sciences and Humanities, Loughborough University, Loughborough, UK
- ²School of Fine Art, History of Art and Cultural Studies, University of Leeds, Leeds, UK
- *Correspondence: Lise Jaillant, School of Social Sciences and Humanities, Loughborough University, Loughborough, UK. Email: I.jaillant@lboro.ac.uk

Abstract

Policy makers produce digital records on a daily basis. A selection of records is then preserved in archival repositories. However, getting access to these archival materials is extremely complicated for many reasons—including data protection, sensitivity, national security, and copyright. Artificial Intelligence (AI) can be applied to archives to make them more accessible, but it is still at an experimental stage. While skills gaps contribute to keeping archives 'dark', it is also essential to examine issues of mistrust and miscommunication. This article argues that although civil servants, archivists, and academics have similar professional principles articulated through professional codes of ethics, these are not often communicated to each other. This lack of communication leads to feelings of mistrust between stakeholders. Mistrust of technology also contributes to the barriers to effective implementation of AI tools. Therefore, we propose that surfacing the shared professional ethics between stakeholders can contribute to deeper collaborations between humans. In turn, these collaborations can lead to the building of trust in AI systems and tools. The research is informed by semi-structured interviews with thirty government professionals, archivists, historians, digital humanists, and computer scientists. Previous research has largely focused on preservation of digital records, rather than access to these records, and on archivists rather than records creators such as government professionals. This article is the first to examine the application of AI to digital archives as an issue that requires trust and collaboration across the entire archival circle (from record creators to archivists, and from archivists to users).

For there to be betrayal, there would have to have been trust first.

— Suzanne Collins, The Hunger Games

1 Introduction

Born-digital archives¹ have been acquired and preserved for the best part of thirty years. Yet, access to these records remains a complex challenge for the institutions that hold and manage them, as well as the researchers seeking to use them. Obstacles to access are numerous and include issues with data protection, sensitivity, and copyright. In the case of government archival collections, releasing potentially sensitive and private information could threaten national security and embarrass foreign partners. However, not all collections present the same level of risk, and giving access to digital records is essential to make government accountable and enable the writing of history.

As born-digital records increasingly represent the largest part of new accessions, finding solutions to these obstacles is becoming a key priority. Unlocking vast amounts of digital data is not a task that can be done manually. Artificial Intelligence² (AI) has the potential to make born-digital archives more accessible and usable. Sensitive information can be automatically identified, making possible the release of non-sensitive data. AI can also be used when keyword search is not effective (for example in the case of web archives, which include terabytes of information). An AI-driven process of recommendation—similar to the functionality 'customers who bought this item also bought'—could enable discovery of previously inaccessible archival materials.

Several projects have recently sought to identify key challenges and find solutions to the problem of locked born-digital archives, using AI and other advanced technologies.³ Yet, AI applied to archives remains at the experimental stage; rarely implemented beyond small data sets and collections identified for

experimentation. Expanding the application of AI to vast amounts of archival materials is complicated for many reasons—including the difficulty to recruit specialists with the right skills, and the relatively low salaries in the archive sector. These obstacles could be overcome with additional funding to attract talent, as well as training opportunities for staff who already work in the sector. A thornier problem is the frequent lack of trust and shared professional ethics across the archival circle (from record creators to archivists, and from archivists to researchers).

The Oxford English Dictionary defines 'trust' as the 'firm belief in the reliability, truth, or ability of someone or something' and the 'confidence or faith in a person or thing, or in an attribute of a person or thing.' The word was inherited from Germanic languages, and is often associated with protection, shelter, and safety. We trust if we feel safe. On the opposite, we lack trust in a situation that seems risky and unsafe. Building trust is a requirement for many professionals, who need to show that their work is beneficial, or at least innocuous. This is why academics must obtain approval from ethics committees to conduct certain kinds of research (including research with human participants). For the OED, 'ethics' is defined as 'the codes of conduct or moral principles recognized in a particular profession' and other aspects of life. Professionals do not always agree on what is or is not ethical. In the case of new technologies such as AI, the lack of trust and shared professional ethics can lead to a deadlock. For example, archivists will hesitate to release data for computational research if they do not trust that academic researchers will make good use of these data. In this context, AI can do nothing to make digital archives more accessible: the first step is to address the lack of trust and common codes of conduct. Once the trust between humans has been restored. AI can come in as the next step to unlock digital archives.

In this article, we investigate these issues of trust and shared professional ethics using qualitative data from the project 'Unlocking our Digital Past' which brought together our team of Digital Humanists with government professionals and archivists. The project was organized in collaboration with the UK Cabinet Office, and although its main focus was Britain, it also involved professionals from North America and Continental Europe. We sought to better understand the barriers to implementing AI-driven tools to make digital archival collections more accessible, and identify key concerns from stakeholders in the entire archival process; from the civil servants and policy makers creating the documents and records that end up in The National Archives, to the archivists and GLAM and (Galleries, Libraries. Archives Museums) professionals preserving the digital collections, and the academic researchers attempting to access and use them.

The research is informed by semi-structured interviews with 30 government professionals, archivists, historians, digital humanists, and computer scientists, and was driven by two key research questions: What are the potential applications of AI to make born-digital archives more accessible and usable? And what are the barriers to implementing them? The project found that many of the concerns about the application of AI are shared between stakeholders in the civil service, GLAM sector, and academia. Yet, collaborative conversations, events and projects between the three groups are rare.

This article argues that although civil servants, GLAM professionals and academics have similar professional principles articulated through professional codes of ethics, these are not often communicated to each other. This lack of communication leads to feelings of mistrust between stakeholders (for example, when archivists do not trust that researchers will make the right decision when confronted with sensitive information). Mistrust of technology also contributes to the barriers to effective implementation of AI tools. Therefore, we propose that surfacing the shared professional ethics between stakeholders can contribute to deeper collaborations between humans. In turn, these collaborations can lead to the building of trust in AI systems and tools. If record creators and archivists trust that researchers will treat data in an ethical way, they will be more eager to share these data for research (including computational research). Trust in other humans can lead to trust in technology.

The three sections of this article develop the argument that shared professional ethics can contribute to closer collaborations between record creators, archivists and researchers, leading to the building of trust in AI tools. The first section looks at the general context and related work on the key barriers to accessing borndigital collections, and the potential applications of AI that could unlock these collections. It also explains our methodology and approach. Drawing on our series of interviews, the second section then focuses on the issues of mistrust of other stakeholders and mistrust of technology, which contribute to the lack of access to digital collections. The third section explores our recommendation to build up trust through collaboration and shared professional ethics, in order to step up the application of AI and new technologies to digital archives. Unquestioning trust in AI is neither desirable nor possible. Yet, the building of trust in collectively developed and explainable AI tools across the civil service, GLAM sector and academia will help make digital archives more accessible and usable.

2 Context, related work, and methodology

Substantial work has been undertaken over the past thirty years or so to establish effective practices for preserving born-digital materials in the GLAM sector (Deegan and Tanner, 2006; Delve and Anderson, 2014; Corrado and Sandy, 2017; Harvey and Weatherburn, 2018; Owens, 2018). Influential support organizations like the Digital Preservation Coalition have helped embed and sustain the work of digital preservation practitioners across a range of cultural heritage organizations. However, while preservation processes are becoming ever more integrated into the core work of GLAM institutions, processes for providing access to those same born-digital collections has been slower to develop (Jaillant, 2019). There are multiple reasons for this, from content-related legal issues such as privacy concerns and copyright, to technology issues that make it difficult to provide access to materials in appropriate formats.

The legal concerns mostly relate to the General Data Protection Regulations (GDPR) in the European Union, the Data Protection Act 2018 in the UK and Copyright legislation. The GDPR and Data Protection legislation place a responsibility to build in 'privacy by design' on organizations and institutions handling personal data, but also provide exemptions to aspects of the regulations when personal data is being archived in the public interest (Lomas, 2019). Whilst The National Archive's Guide to Archiving Personal Data (2018) makes clear that the updated data protection does not prevent archiving, it does influence how archival institutions need to articulate the purposes of archiving personal data and how they provide access to collections containing it (The National Archives, 2018). Records deemed to contain personal data are carefully managed and normally kept closed from public access during the lifetime of the individual the data is relating to (The National Archives suggests that a lifespan of 100 years should be presumed, unless proven otherwise).4 However, in some circumstances access to these records can be granted, and decisions need to be taken with respect to the safeguarding measures defined in the Data Protection legislation, and fully documented by the archive (The National Archives, 2018).

Additional concerns relate to Copyright legislation which has an impact on how archival material can be processed and made accessible. The application of copyright law on born-digital materials in GLAM institutions is an ongoing and complex issue. Many born-digital collections contain content from multiple copyright holders, described at varying levels of detail and often without proof of consent. For example, an archived website may have personal information, images, creative writing, or videos embedded within it. It is not possible to assume that all the content belongs to the

person or organisation that created it, and it is often not possible to trace who the copyright holder is. Many GLAM institutions with web archives face difficult decisions about the amount of their collections they can make available whilst managing the potential risks of making it available (Hockx-Yu, 2014; Vlassenroot et al., 2019). Copyright legislation in the UK and in the EU has acknowledged the digitization of physical material as a form of preservation in GLAM institutions and has developed exemptions to support this. However, the management of born-digital material is largely ignored in the legal context. Creating preservation and access copies of born-digital content remains a grey area (Koščík and Myška, 2019).

With so many digital collections closed to users due to these issues with data protection and copyright, AI has been presented as a possible solution to the problem of locked archives (Jaillant and Caputo, 2022; Jaillant 2022a, 2022b). For the purposes of this article, we refer to AI in a broad sense. Instead of focusing on particular subsets of AI such as machine learning, computer vision, or natural language processing, we consider AI as a broad suit of computational approaches used to make decisions and complete tasks. Although often discussed through the lens of new and emerging technologies, AI has a longer history, with waves of excitement and periods of disinterest (Jackson, 2019; Agar, 2020). The current period is characterized by sustained development, and AI technologies are increasingly featuring in all aspects of our everyday lives. The role of AI technologies in making born-digital archives more accessible and usable is increasingly acknowledged as an important direction for record creators and managers (including civil servants responsible for knowledge, information, and records management), GLAM professionals, and researchers interested in using the archival material.

Examples of AI tools used to increase the accessibility and usability of digital archives include: the automatic sensitivity classification of government records before being released through Freedom of Information requests or transferred to national archives (The National Archives UK, 2016; Mcdonald et al., 2020a; Baron et al., 2020); experimentations with computer vision to enhance the discovery of visual archival collections (Angelova et al., 2020); and algorithmic search tools (Nix and Decker, 2021). Additionally, scholarship and research networks, such as AURA, AEOLIAN and AI4LAM bring together multiple stakeholders interested in the application of AI to archival collections.

As awareness of AI technologies increase, so does public scrutiny, and ethical questions, including ones on transparency and bias, begin entering public discourses (Fast and Horvitz, 2017; Cave *et al.*, 2019; Kelley *et al.*, 2021). Calls to develop more widely

agreed standards and processes of enforcement for the algorithms that underpin a lot of AI technologies continue to grow. Projects such as *Assembling Accountability* have sought to develop frameworks to support the assessment of algorithmic impacts when utilising AI-drive tools. This project argues that impact assessments are 'a promising model of algorithmic governance because it bundles an account of potential and actual harms of a system with a means for identifying who is responsible for their remedy' (Moss *et al.*, 2021, p. 1).

Civil servants, GLAM sector professionals and academic researchers increasingly engage with ethical issues related to AI systems. In 2018, the UK government established the Centre for Data Ethics and Innovation to provide independent advice on measures needed to ensure safe, ethical, and innovative uses of AI. One of its core missions is to help public sector partners to use data and AI in a way that commands and retains public trust.⁵ And in October 2020, the UK Parliament's Office of Science and Technology released a note on interpretable machine learning, to ensure that ML systems are designed and deployed in an ethical and responsible way (Christie, 2020). In the GLAM sector, an emerging body of work addresses transparency, accountability, and bias in libraries and archives (Padilla, 2019; Cordell 2020), as well as museums (Murphy and Villaespesa, 2020). This work explores the public discourse around ethical uses of AI. However, it is not sufficiently connected to considerations of professional ethics across the civil service, GLAM sector, and academic researchers.

Ethical questions in relation to new technologies begin to intersect with the codes of ethics that govern professional practice in those sectors. Civil servants, GLAM professionals, and academics all subscribe to some form of ethical practice, whether it be through established codes of practice, like the Civil Service Code⁶ and the International Council on Archives' Code of Ethics, or processes of ethical approval seen in university-based academic research. As well as the ethical practices embedded within the civil service, GLAM sector and academic professions, the application of AI technologies is also subject to ethical codes and guidelines, although less established and not formally agreed amongst private companies, research institutions, and public sector bodies (Jobin et al., 2019; Mittelstadt, 2019; Hagendorff, 2020). One study found that there were eighty-four sets of ethical principles or guidelines for AI published globally (Jobin et al., 2019). These existing frameworks lack specific actions and coordination across professional sectors.

Although there is no single agreed set of ethical principles in relation to AI, the notions of transparency,

justice, and fairness, non-maleficence, responsibility, and privacy are the most commonly discussed (Jobin et al., 2019, p. 391).⁸ These five principles can also be found in the ethical codes and professional guidelines followed by our interviewees. Evaluating the effectiveness of these ethical codes and guidelines is outside the scope of this article. However, as Iason Gabriel notes, we should be looking 'more closely at principles that would be supported by a global overlapping consensus of opinion' to support the use of AI (Gabriel, 2020, p. 433). Working through shared principles, expressed in codes of ethics, presents itself as a productive place to start building trust in the application of AI to born-digital archives.

The common ethical notions of transparency, justice and fairness, non-maleficence, responsibility, and privacy noted by Jobin *et al.* do not inherently conflict with the principles expressed in the professional codes and practices seen in the civil service, GLAM sector, and academic researchers. The Civil Service Code specifies integrity, honesty, objectivity, and impartiality as the key values for civil servants, and these principles permeate into every facet of how civil servants approach their professional roles. A number of these principles are underpinned through laws and regulations like the Freedom of Information Act and Public Records Act that build in a culture of accountability into the documents that civil servants and public office holders create every day.

Moreover, codes of ethics and practice have also been instrumental in the professionalisation of the GLAM sector. In 1986, the International Council of Museums (ICOM) released their code of ethics. Ten years later, the International Council on Archives (ICA) introduced their own ethical framework. The International Federation of Library Associations (IFLA) approved a code in 2012. These codes are often supplemented by country specific codes, and the United Kingdom also has codes of ethics from the UK Museums Association and CILIP (the UK's library and information association). Integrity, impartiality, objectivity, privacy, and co-operative working feature as principles for archivists to work with.

As well as the codes that govern civil servants and GLAM sector professionals, academic researchers also work within ethical frameworks. Ethics within academic research is governed by a set code that is integrated into individual university strategies and research approval processes. The UK Research Integrity Office (UKROI) is thus responsible for the UK's Code of Practice for Researchers. Loughborough University, home to the 'Unlocking our Digital Past' project, has an Ethical Policy Framework informed by the UKRIO code of practice, designed to build ethical considerations into research practices. This article does not claim

that all professionals working in the civil service, GLAM sector and academia consult their codes of practice or ethics before they make new decisions or undertake a new task, but that these codes broadly influence the principles and areas of concern for the sectors.

Concepts of trust, transparency, and professional ethics emerged as a recurring theme throughout 'Unlocking our Digital Past'. The project was part of an ongoing collaborative relationship between Loughborough University and the Cabinet Office of the UK Government that focuses on the application of AI to born-digital archives. The project facilitated a series of workshops aimed at increasing interdisciplinary dialogue between civil servants, GLAM sector professionals, and academics. After obtaining approval from the Ethics Committee at our institution, our team of Digital Humanists conducted thirty interviews with key thinkers and practitioners who have a stake in the accessibility and usability of born-digital archives now and in the future. Our background is in cultural history, literary studies and museum studies, rather than computer science. This has informed the nature of our questions: instead of focusing on the technical aspects of AI, the questions we asked centred on the key obstacles to making born-digital and digitised collections more accessible, and on the possible solutions to these issues. We were interested in the day-to-day practice of professionals-from record creators who need to prepare archival records ahead of transfer to The National Archives and other repositories, to archivists who need to preserve and make accessible these records. We also interviewed historians who need access to archives (including born-digital archives) for their research. A common thread of these interviews was the lack of trust in other stakeholders and in technology. In turn, this mistrust has an impact on the accessibility and usability of digital archives.

3 Mistrust of other stakeholders and of technology

Collaborative discussions between professionals in government, GLAMs and academia are the exception rather than the norm. This, in turn, has an impact on the way professionals see other professionals, often leading to misunderstandings. James Baker, who studied history and worked as a curator at the British Library before moving to academia, gave an example of this lack of dialogue between archivists and researchers. While archivists see appraisal and selection of records as a central aspect of their role, researchers often expect complete archives. In other words, academics view archivists as record keepers, rather than professionals who must make a selection and throw

away records that lack lasting value. This task of appraisal is particularly complicated due to the scale of born-digital archives. Baker said:

I always feel very sorry for my archivist friends every time historians turn round and say, well, you must collect everything, and why aren't you trying, you know, we need all this stuff and it's going to be a digital dark age. And they're like, it's not going to be a digital dark age, we always do selection, you know, it's just that selection issues are kind of way more complicated now in the age of the born digital. ¹³

There was a consensus among our interviewees who engage closely with born-digital records: keeping and describing everything is not a viable option. Clifford Lynch, Executive Director of the Coalition for Networked Information, pointed out that in the very early days of the internet, there were attempts to apply library-style cataloguing to internet resources. But these 'were rapidly abandoned because the scale was just unmanageable'. ¹⁴ The Google-style, computational approach to records became the norm to search records at scale.

The problem of scale is particularly acute for government digital records. Jason Baron, who served for 13 years as Director of Litigation for the US National Archives and Records Administration (NARA) before moving to the private sector and then academia, gives the example of US presidents' email records:

The National Archives in the US today has 500 million emails from the Regan administration to the Obama administration. The Trump emails haven't been counted yet or haven't been fully processed. But it has 500 million distinct emails, and then more than a billion pages because those emails have attachments. ¹⁵

In this role at NARA, Baron pushed for an approach that would diminish the number of government emails transferred to archival repositories. This CAPSTONE approach adopted in 2013 led to an email preservation policy based on the seniority of the record creator. Individuals at the top or near the top of an agency have all their emails permanently preserved, while others see their emails destroyed after a certain period (generally seven years).

CAPSTONE addressed the issue of email exponential increase, but emails are of course not the only born-digital records produced by US government. Leslie Johnston (Director of Digital Preservation at NARA) notes that the Trump administration produced about 500 terabytes of archival records—compared to

200 terabytes for the two terms of the Obama administration.¹⁷ Libraries and archival institutions lack the staff necessary to handle enormous amounts of data. Baron gives the example of the Clinton Presidential Library, which has six archivists and 'a 10 million document queue estimate for Freedom of Information Act requests'.¹⁸ This leads to a lack of accountability in the short term, and risk impacting the cultural memory in the long term.

Jason Baron and others have argued that automation is a necessity rather than a choice. How do you search a billion objects?', asks Baron. 'Well, you don't do it manually, as an archivist, and you don't do it with keyword searches, although you can try, but those are terribly inefficient'. 19 Indeed, an approach based on keyword search does not work well with collections characterized by their huge size and lack of metadata. For instance, a search for 'Brexit' on the UK web archive gives 63,968,885 results.²⁰ As Leontien Talboom (Doctoral Researcher at The National Archives UK and University College London) told us, analysing each record manually is not an option.²¹ Instead, a computational approach is the only way to make sense of this mass of results. 'How can they have any success of opening a substantial amount of history for the 21st century and before, if they don't use machine learning techniques to perform searches?' Baron inquires.²² It should be noted that AI-assisted searches are not perfect solutions, since they might also miss relevant results.23

The application of AI and machine learning to archival records is not a new thing. In the early 2000s, Jason Baron was part of the government legal team involved in a giant tobacco suit against Philip Morris and other tobacco companies. Key information was scattered in the Clinton email archive, which included 30 million emails. Baron decided to seek out information scientists and computer scientists to figure out a better way for lawyers to search, using other techniques than keyword searching.²⁴ At around the same time, commercial firms started marketing eDiscovery software that drew on machine learning to search and find information in vast amounts of data. The archive sector on both sides of the Atlantic then investigated the potential of eDiscovery in solving the challenges of born-digital records. In 2016, a report from The National Archives (TNA) in the UK concluded that 'technology-assisted review using eDiscovery software can support government departments during appraisal, selection and sensitivity review as part of a born-digital records transfer' to TNA (The National Archives UK, 2016, p. 5).

Researchers such as Graham McDonald (Lecturer in Information Retrieval at the University of Glasgow) have shown that AI and machine learning can indeed unlock government archives, using two main approaches: 'protect then search' or 'search then

protect'.²⁵ McDonald's work has used mostly the first approach. Tools that his team developed can help humans identify the sensitivities in a particular collection, and then make informed decisions before transferring records to archival institutions. For example, record creators can choose to protect sensitive information by redacting it, ahead of transfer to archives. This process lowers the level of risk in providing access to that collection.

The second approach consists in making a collection entirely accessible, with a search framework that hides sensitive information when it comes across it. The machine is trained using a subset of data identified as sensitive or confidential and is then expected to identify sensitivity in larger datasets. The problem is that there is no universal definition of sensitivity. It largely depends on the context, and the machine is not always able to correctly identify contextual information that makes a record sensitive. ²⁶

While Graham McDonald recognizes that sensitivity review is still an imperfect process, the risk of giving access to archival records is worth taking, according to Jason Baron. AI 'will get a tremendous amount of return' even though it comes with a risk of bias.²⁷ The most important thing is to allow researchers, journalists, historians, and other users to make discoveries that will serve the public good. The stakes are lower than in the criminal justice system, where AI has in the past led to disastrous results [e.g. algorithms used in the USA to evaluate the risk of prisoners to re-offend were shown to be racially biased (Wadsworth et al., 2018)]. Despite the risk that comes with AI, the police and justice system on both sides of the Atlantic see advanced technologies as a necessary tool in the fight against crime. In 2014, the UK Home Office released a report on eDiscovery in digital forensic investigations (Lawton et al., 2014). Four years later, it announced the development of new technology to automatically detect terrorist content on online platforms.²⁸

If AI is routinely deployed by police officers and judges to analyse large datasets, why is it rarely applied to government archives? It is still exceptional to use AI for sensitivity review, or to find results in vast digital collections. In order to train AI systems, access to data is needed. And this notion of access (even limited access to training datasets) is problematic for many record creators and archivists. When users are not trusted to make the right decisions, it seems logical to close entire collections—or to restrict access only to groups who are seen as trustworthy. For example, web archives generally require users to travel to the library due to copyright reasons. Content is available onsite via browsers, rather than offsite via downloading data. Jane Winters, Professor of Digital Humanities at the University of London, notes that Denmark has a

researcher exception to this restrictive access policy. Researchers attached to a Danish university, as well as overseas collaborators, can get remote access to web archive data anywhere in the world. 'You have to sign waivers and so on that you will treat the data appropriately', Winters said. This can be seen as a first step 'to start to think about opening things up'.²⁹

The risk of releasing potentially sensitive data should be balanced against the risk of keeping records dark and inaccessible. Indeed, access is at the centre of ethical codes in the GLAM sector. The Museums Association puts 'public engagement and public benefit' as the first principle in its Code of Ethics. More specifically, the code states that museums and other cultural heritage organisations should 'provide public access to, and meaningful engagement with, museums, collections, and information about collections without discrimination'. Likewise, the UK's library and information association CILIP declares that 'preservation and continuity of access to knowledge' is central to libraries' mission. Similarly, the International Council on Archives asserts that 'archivists should promote the widest possible access to archival material and provide an impartial service to all users'. Of course, access is not the only aspect that GLAM professionals need to consider. Codes of ethics include references to privacy and copyright as limits to access. For the ICA, 'archivists should respect both access and privacy'. Yet, the balance is often skewed towards privacy and data protection.

While archivists must respect legal frameworks and protect the private data of individuals, they could also take limited risks and make some collections available without necessarily checking all materials. The need for trust and shared codes of practice and conduct was highlighted by our interviewees. Jane Winters (University of London) gave the example of personal archives, rather than web archives and other large-scale collections. 'Relationships of trusts between depositing authors, archivists, and researchers' are central. However, these relationships are also 'time-consuming' and 'excluding because if you happen to know the librarians it's much easier to have those conversations than if you're coming to it cold'. In other words, access is often decided on a one-to-one basis based on preliminary discussions with archivists. 'Having those negotiations and building relationships of trust around the use of this material' can unlock previously closed collections.³⁰ Winters's comments echo the experience of literary scholars who have had access to previously closed collections. Lise Jaillant has thus written about her experience of accessing the email archive of the British writer Ian McEwan at the Harry Ransom Center (HRC) in Texas, which is officially inaccessible to researchers. Access was granted through personal contacts, and the fact that the HRC archivist was participating in the same research project (Jaillant 2019).

Providing ad hoc access to known users is not a longstanding solution since it excludes entire groups—for example, independent scholars or family historians who lack a university affiliation and professional networks. As we have seen, access without discrimination is central to GLAM codes of ethics. Adam Nix, Lecturer in Responsible Business at the University of Birmingham, favours a systematic policy of access for those who respect certain codes of behaviour:

I would like to see more focus on sharing that duty of care between the archives and the users. I think users need to take a far more active and conscious role in maintaining ... the integrity of the archival discovery process. And I think that can be done by having well developed and well-respected codes of conduct that the user very consciously agrees with beforehand, which means that even if they see sensitive information, that sensitive information goes no further than that individual researcher.³¹

This proposed process of opening up archives to users who behave in an ethical manner is based on trust. But it is also based on the possibility of sanctions (including legal sanctions) for those who transgress codes of conduct. A similar system of trust and possible sanction is described in Frances Harris and Fergus Lyon's research on collaborations across professional cultures. Their central findings are that first, trust is a vital ingredient when collaborating across disciplines and sectors. Second, maintaining collaboration across professional cultures presents particular challenges for building trust. Third, trust in interdisciplinary and transdisciplinary teams is based on norms, information, sanctions, and controls (Harris and Lyon, 2013).

John Sheridan, Digital Director at The National Archives UK, is in favour of quantifying risk and therefore trust. 'Looking at our processes around digitization, all tend to lean much more heavily on managing the risks through expert knowledge, than on systems', Sheridan said.³² He advocates for a computational approach using statistical models to assess risk. To this purpose, TNA has explored a Bayesian network, i.e. a graphical model that represents a set of variables and conditions, often used for probability analysis (Barons et al., 2021). For Sheridan, archival institutions should strike the right balance between risk and access to potentially sensitive materials. It is not a case of 'transparency above everything', as:

archives are not Wikileaks, and we're not in the WikiLeaks business. ... It's not responsible to data subjects; it's not responsible to other people's intellectual property rights; it's not lawful. So, we then

need to build the techniques to provide access responsibly.

Until systematic risk models are an inherent part of archives' digital practice, Sheridan recommends what he calls 'gradating access'. A process that 'maximises use but manages the risks of publication', gradating access can lead to a more flexible system of publishing potentially sensitive records (but not in breach of data protection laws), by identifying varying levels of risk, and determining necessary exemptions.³³

As the number of born-digital records continues to increase, and the capacity of expert assessment becomes overstretched, AI technologies can enable risks to be managed and reduced. Several interviewees told us that whilst AI tools could support the access and use of digital archives, there were still barriers to scaling up experiments and widespread acceptance of its usage. These concerns can be roughly categorised as accountability, control, and bias. We suggest that these concerns could be addressed through closer collaboration between civil servants, GLAM professionals and academics (both in the humanities and computer science).

3.1 Accountability

The need to be accountable and transparent about the work the government does is a fundamental part of the codes of practice and professional ethics civil servants work within. Accountability and transparency are also embedded within the Public Records Act. The tools used to aid decisions about the government's records sent to The National Archives are subject to this need for accountability. James Lappin, Digital Era lead for KIRM³⁴ at the Cabinet Office, explained that if a decision is made to use an AI tool, then 'the people who deploy it are accountable for the outcomes' and 'you've got to defend your algorithm ... the process by which you trained your algorithm, and the process by which you monitored your algorithm'. ³⁵

Reticence over the implementation of AI technologies is often connected to the fear of unexpected or unintended results, and where the responsibility for those lies. Clifford Lynch (Coalition for Networked Information) noted that the algorithm often gets blamed, but incorrectly: 'I don't get mad at the algorithm, I don't get mad at the person who wrote the algorithm, I get mad at the people who are applying the algorithm believing that it's doing the right thing and not understanding what they've got there'. Lynch's comments emphasise both the need for those who implement AI tools to understand and be accountable for them, and the common misconception that the algorithm can and should be blamed. Andrew Blick, Head of the Department of Political Economy at King's

College London, reminded us that the controversy surrounding the AI-driven A-Level grade assignment undertaken in 2020³⁷ was initially blamed on the algorithm. But public discourse rejected that and demanded accountability for the implementation of the algorithm. This requirement for accountability was articulated in other interviews through the need for explainable AI.

Explainable AI was discussed by seven of our interviewees as a way of ensuring accountability and transparency in the tools being used by the civil service, as well as for the institutions holding digital archives and the researchers trying to access them. Jane Winters (University of London) was explicit that explainable AI was part of building relationships of trust, and that archives were generally good at being aware of issues like trust and transparency.³⁹ This again is linked the professional codes of ethics and frameworks that GLAM professionals work within. The ICA Code of Ethics specifically references the need to be able to record and justify actions, provide impartial services to users and 'the special trust given to them' (International Council on Archives, 1996). Laura Millar explains how archivists work within the codes of ethics, and that impartiality and transparency are of particular importance when considering ethical archival work (Millar, 2017, p. 95).

A real tension emerges between the professional ethics of archivists and the application of AI where there is no scope for explaining how it works. These forms of unexplainable AI are often referred to as 'black box' due to their opacity. The unexplainable nature of black box AI was also a problem for the historians we interviewed. Adam Nix (University of Birmingham) explained that with black box AI systems 'we have something that we can't fully critique as a researcher and we can't fully understand the results we got or the findings we generate, it becomes an aspect of that process that we can't understand'. Nix's point is supported by the requirement for accountability highlighted in the UKROI's Code of Practice for Researchers. 41

Whilst the nature of some AI tools conflicts with key ethical principles, computer scientists and programmers developing the AI technology are working to make tools more transparent and explainable. Graham McDonald (University of Glasgow) explained how although a lot of the tools are currently black boxes, work is being done on 'transparency, explainability and ensuring the fairness of predications'. ⁴² McDonald went on to explain that transparency will be one of the most important factors in giving users of AI systems the confidence that they understand what the system is doing. ⁴³

An example of this can be seen through the Digital Sensitivity Review project being undertaken between the Foreign, Commonwealth and Development Office (FCDO), and the consultancy company SVGC in partnership with the University of Glasgow. 44 Andrew Dixon, Managing Director of SVGC, explained that in using AI tools to make decisions about public records to be archives, the public has a right to know how those selections are being made. This is of course derived from the civil service's professional code requirement for accountability and transparency. Dixon explained that this influenced the tools they use in the project: 'we're not using neural networks which are unexplainable... we've avoided technologies that create random outcomes'-instead opting to use the kind that are 'predictable and repeatable technologies . . . so you can determine the outcome'. 45 Here we see how the principles and professional codes of ethics that govern the civil service, GLAM sector and academic researchers are influencing the technology being used.

McDonald, who also works on this project, pointed out that intuitive explanations of what is happening when you apply an AI tool is crucial for building trust—especially when those using the tool are not the computer scientists who have written the algorithms, tested, and evaluated them. Successful AI projects like these are a product of collaborations between the civil servants who manage the sensitivity review process, with archival partners, academic researchers, and computer scientists who provide expertise on the technology. Stakeholders effectively communicated professional principles and codes, and through that, AI tools that are congruent with the notions of accountability and transparency were implemented.

3.2 Control

The second tension that emerged through our interviews focused on the lack of control and trust in AI tools. Eleven of our interviewees brought up the idea of control in discussing the application of AI tools to digital archives. Although we do not want to present colleagues in the civil service, GLAM sector and academia as control obsessed, it should be acknowledged that in these professions, control over what is created, released, archived, described, included, and excluded is important. For civil servants and GLAM professionals, there is a professional responsibility toward maintaining privacy and managing sensitive information. James Lappin (Cabinet Office) explained how once a decision is made to release a document, there is no going back: 'If you use AI to determine access permissions, to open up access ... it's an irrevocable decision, because once you've given me that access, there isn't much point in then taking it away from me, and the horse has bolted, if you like'. 47 Similarly, Leontien Talboom (The National Archives UK/University College London) discussed how archivists tend to default to wanting material to only be available in reading rooms where there can be more control over the type of access to it.⁴⁸ Likewise, Andrew Riley, Senior Archivist at the Churchill Archival Centre, explained how archivists value control and that the loss of that through the use of AI with respect to sensitive data was one of his concerns.⁴⁹

These responsibilities towards protecting privacy and sensitive information connects to the code of practice and professional ethics that govern the sectors. The Civil Service's principle of integrity is particularly relevant here as it references the obligation to undertaking duties responsibly, professionally, and also taking into account the ethical standards that govern other professionals. The ICA Code of Ethics is more explicit in the archivist's responsibilities towards privacy, specifically noting that archivist should 'take care that corporate and personal privacy as well as national security are protected without destroying information, especially in the case of electronic records where updating and erasure are common practice' (International Council on Archives, 1996).

The fear over a loss of control was articulated differently by the Humanities academics interviewed during the project. In most cases, the lack of control came down to not knowing what was being missed. Historians highlighted the importance of serendipitous findings when accessing a box of physical archival items. Lindsay Aqui, Research Fellow at the University of Westminster, explained how she felt an 'immediate discomfort' around AI-assisted searches because she 'wouldn't want there to be something in the process that meant 10% was being excluded because it was deemed irrelevant or didn't match the search criteria in a way that I thought it would'. 51 This sentiment was echoed by Emily Robinson, Senior Lecturer in Politics at the University of Sussex, who noted that using AI-driven search tools may prove a barrier to finding the things you did not know you were trying to find in the archive.⁵² Helen McCarthy, Professor in Modern and Contemporary British History at the University of Cambridge, described her feelings towards the application of AI to digital archival collections as ambivalent: simultaneously open to the potentials, but nervous about the loss of control for civil servants and archivist, and the potential losses in material that may come of this.⁵³

One way in which computer scientists are attempting to address this concern over a loss of control is by flipping the narrative on the use of AI. The term 'humanin-the-loop' is often used to refer to the way in which AI tools are trained with human interaction. Graham McDonald explained how 'if you're putting across a

narrative that it's human in the loop, there's an expectation that the Artificial Intelligence is going to be the one that's making the decisions and doing the job. But that's not right—that's not actually what it is expected to be in practice'. ⁵⁴ Instead, McDonald argued that we should focus more on 'computer in the loop processes where it's the AI that is there to help and assist the human in the decision-making process'. ⁵⁵ He suggested that 'these little changes in narrative can help to change the perspectives of how people actually see it'. ⁵⁶ Shifting perspectives in this way may offer some alleviation to the fears that as responsible civil servants, archivists, or researchers might no longer have control over their own work remit.

3.3 Bias

The third tension that emerged in our interviews was that of bias. Twelve of our interviewees highlighted that algorithmic bias was a concern for them or others around them. The notion of bias is particularly pertinent to the codes of practice and professional ethics across the civil service, GLAM sector, and academia. Impartiality is a key feature of the Civil Service code and specifically noted that a civil servant must not 'act in a way that unjustifiably favours or discriminates against particular individuals or interests'. 57 This sentiment is echoed in the ICA Code of Ethics which states that archivists should 'provide an impartial service' (International Council on Archives, 1996, p. 2) and UKRIO Code of Practice emphasizes the importance of recognizing the dignity, rights, and safety of people in research.58

Oonagh Murphy, Lecturer in Arts Management at Goldsmiths University, described how in working with AI and the GLAM sector there is scope for what she framed as bias squared. Murphy explained how many galleries, libraries, museums, and archives already have biases built into their collections through hundreds of years of colonialism, and the forms and structures that emerged from that and influenced how collections were put together and described. The use of AI in institutions with such collections risked adding additional elements of bias—compounding the issue.⁵⁹ The danger of this happening was recognized not just by those who work in or research about the GLAM sector, but also by Sébastien A. Krier (AI Policy and Governance Advisor) who acknowledged that training AI tools on cultural collections would encounter issues such as sexism. 60 Likewise, historians like Lindsay Aqui explained that they had 'a bit of suspicion around what might end up excluded from search results ... because archives are already curated and people's voices, stories, experiences do get excluded from the archive, so there is a risk that some AI tools may further marginalise already marginalised people'. 61 The potential to reinforce bias is a genuine concern for many stakeholders.

Whilst bias should not be embedded as normative practice, several interviewees recognized that bias can never be fully removed. Jane Winters noted how Humanities researchers tend to understand that bias is always present, and that acknowledging bias and understanding what it means for your research is important: 'you can't get rid of biases, you just have to acknowledge that they're there, and understand what that means for the results that you get... if you try and get rid of bias, you just introduce a new one, so it's acknowledging that it's there and working with that, rather than trying to get rid of it'. 62 Gareth Millward, Research Fellow at the University of Birmingham, was also keen to note that researchers need to be 'hyperaware of the biases that can be baked into some of the technologies that we use ... and at least be able to contextualise them if not be able to eliminate those biases altogether'.63

4 Collaborations and building trust

This section now seeks to explore how the deeper collaborations between civil servants, GLAM sector professionals, and both humanities and computer science academics can help establish appropriate levels of trust in AI tools and support their use in making digital archives more accessible and usable. As professions, the civil service, GLAM sector and academia all value collaborative working. Within the Civil Service Code, there are references to co-operation and respect with colleagues and partners, as well as to taking into account the ethical standards governing other professions. 64 The ICA Code of Ethics states that archivists 'should cooperate with members of related professions on the basis of mutual respect and understanding' (International Council on Archives, 1996, p. 3), and the UKRIO Code of Practice advises that researchers working collaboratively do so with awareness of the ethical guidelines and standards of collaborators. 65 In practice, however, civil servants, GLAM professionals and academics seldom collaborate to unlock digital archives.

Wider, cross-disciplinary discussions are required: discussions about shared concerns, shared professional and ethical principles, as well as solutions with the computer scientists developing the tools, rather than a focus purely on the barriers to implementing AI. The three main tensions discussed in the previous section contribute to a general mistrust of AI throughout the professionals working in the civil service, GLAM sector and academia, but within each of the tensions, there were interviewees who identified potential solutions

that were compliant with the different codes of practice and professional ethics.

A central takeaway is that interdisciplinary partnerships are vitally important in developing trustable AI tools. John Sheridan (The National Archives UK) noted:

It's all about the knowledge transfer into our sector from outside. No one is building anything specifically for us. It is all about things that are being built in the world, and then someone's going, 'Oh, I could apply this to archives.' And I think all we can really do is equip ourselves with learning to understand what that kind of technology transfer process looks like and understand what the implications are.

Sheridan went on to explain how 'it's quite hard for us to know how to navigate this world where, whilst we have collections, we're effectively, certainly in terms of budgets, and then in terms of capability, we're fleas around elephants. But we're not on our own. These issues are pervasive, and again, I think it's about not just talking to each other, but talking to other people who share concerns, is going to be very important, too'. 66

Daniel van Strien, Digital Curator at the British Library also emphasized that the GLAM sector needs to be actively involved in how AI tools are applied to collections. If these tools are developed without engaging with GLAM professionals, they may end up not being useful for the sector in practice.⁶⁷

Whilst some interviewees expressed mistrust over the idea of AI as a tool, there were also expressions of trust in the archive's decisions around potential uses of AI. Gareth Millward (University of Birmingham) explained how 'if The National Archive is providing me with the tool to search its own archive, I trust that they have thought about that, and that is part of their archival process... just as I have to put my faith in their cataloguing technique'. ⁶⁸

In practice, however, not many users have the knowledge and ability to use advanced technologies such as AI to explore archives. In turn, this skills gap hinders collaborative discussions. One professional in a large archival institution in London told us:

we have had a few attempts at trying to run events for researchers to find out what they might be expecting, or what they might want. And ... the majority of our existing core research audience is not yet working with born-digital material.

This is particularly frustrating for archivists and digital preservation specialists who have been working on these kinds of records for years. 'The events turn into a training event on what the archive profession has learnt from 10 years of trying to do digital preservation, rather than really extracting expectations from researchers'. For this professional, it was surprising to see little demand from researchers wanting to use digital tools to explore a digital archive. 'A bit like the depositors, [researchers] transfer their paper world and their paper research methods to the digital world'. 69

Andrew Riley (Churchill Archival Centre) also mentioned the lack of skills which prevents users from productively working in digital collections. He mentioned his experience of attending a workshop at the Foreign Office organised by historian Helen McCarthy on born-digital records. Riley discovered that like many 'traditional' archivists trained to process paper records, historians at the workshop felt insecure when it came to using born-digital materials:

I thought they'd be much more self-confident, but they all seemed to doubt their essential skill sets as historians in using this kind of material and, for similar reasons, hoping to be retired too before they had to do too much of this, which really surprised me.⁷⁰

In this context, it is not surprising that many researchers turn to more easily accessible and usable collections: paper collections but also digitized collections searchable via keyword search. Concerns about big data and Artificial Intelligence could be allayed through a conceptual understanding of the techniques used. Yet, humanities scholars have few institutional opportunities to fill the gaps in their knowledge around AI. For example, the 2022 'Digital Humanities at Oxford' Summer School did not include any course on AI.

Without significant pressure from researchers to open up born-digital collections, and without any easy solutions to the issues of sensitivity and copyright, it can seem rational to keep archives closed for now. James Lappin thus praised NARA's decision to preserve the Whitehouse digital archive, even if it is still "dark." Preservation was presented as the most important thing, since it makes possible access at a later stage. The Breandán Knowlton (who works in a government team) pushed for openness and told us:

When you're talking about the public sector, the work we do is funded by public budgets or scrutinised by Parliament and by journalists and by everybody else. Most of our work should be public by default. . . I think if more of our work was public, I think more people would find it, kind of, boring. But also, it wouldn't be [in case of data leaks] 'Ooh, secret trove of Cabinet Office emails released by aide'. ⁷²

As we have seen, accountability and transparency are central to civil servants' codes of ethics, but this should be balanced against the need to protect national security and avoid leaking sensitive records. Having more collaborative discussions with other stakeholders (including archivists and academics) would offer government professionals new perspectives and allow them to make informed decisions on records to transfer to archival repositories.

5 Conclusion

This article has shown that the lack of communication between civil servants, GLAM professionals and academics often leads to mistrust of other stakeholders. Interviewees anticipated situations where data would leak, reputations would be damaged, and in the worst cases, national security would be at risk. But there can be no betrayal without trust. And it is precisely the lack of trust in other stakeholders, and in new technologies, that was highlighted in our interviews. Mistrust of technology makes it difficult to implement AI tools. The concerns highlighted by interviewees—in terms of accountability, control, and bias-make it difficult to accelerate the development of AI applied to archives. The article recommends more cross-sector and crossdisciplinary dialogue to surface the shared professional ethics between stakeholders. Deeper understanding between humans is an essential step to apply AI to born-digital records. More research is also needed across disciplinary boundaries. Indeed, the issues of innovative technologies applied to digital archives have too often been explored within single disciplines such as archival studies or computer science. Yet, applying AI to archives requires a variety of skillsets: not only computational skills, but also domain knowledge, i.e. knowledge about the entities that are being treated by the algorithms. Thinking within disciplinary boxes hinders our collective ability to apply technology to digital records. Digital Humanities-which is in essence an interdisciplinary field—has an important role to play in bridging the gap between record creators, archivists, researchers, and other users.

Notes

- Born-digital archives refer to archival materials that originated in a digital format (for examples: emails, PDFs, Word documents, audio, and video digital files). These born-digital records differ from "digitised" materials created through processes such as scanning and photographing.
- Artificial intelligence (AI) refers to the use of computational processes to learn, make decision, and solve problems.
 Machine learning (ML), which is often referenced in

discussions about AI, is an application of it. ML is the process by which a computer system is able to continue learning and improving on its own based on previous processes it has undertaken. ML is considered a subset of AI.

- For example, AURA (Archives in the UK/ Republic of Ireland and AI) www.aura-network.net; AEOLIAN (Artificial Intelligence for Cultural Organisations) www.aeo lian-network.net; and AI4LAM https://sites.google.com/ view/ai4lam It should be noted that AI is also enthusiastically embraced in relation to digitized collections (not only born-digital collections).
- 4. The National Archives has a lifespan assumption that governs their approach to personal data when the subject is not personally known to the archive. It states that: 'Given the large number of individuals commonly featuring in archive collections, archive services will not be in a position to ascertain whether they are still alive. If it is not known whether a data subject is alive or dead, the following working assumptions can be used:
 - Assume a lifespan of 100 years;
 - If the age of an adult data subject is not known, assume that they were 16 at the time of the records;
 - If the age of a child data subject is not known, assume person was less than 1 at the time of the records.

If the individual is known to be more than 100 years old and still living then compliance with data protection law is still required. They are entitled to make a subject access request or to exercise any of their other rights' (The National Archives, 2018, p. 34).

- https://www.gov.uk/government/organisations/centre-fordata-ethics-and-innovation/about (accessed 23 April 2022).
- 6. The Civil Service Code can be found at https://www.gov.uk/government/publications/civil-service-code/the-civil-service-code (accessed 23 April 2022).
- The International Council on Archives (ICA) Code of Ethics was formally adopted at the General Assembly of the ICA in 1996. The code is available at https://www.ica.org/en/ica-code-ethics (accessed 23 April 2022).
- Related guidelines include the FAIR principles on Findability, Accessibility, Interoperability, and Reuse to facilitate the reuse of data (Wilkinson et al., 2016); and the CARE principles (Collective Benefit, Authority to Control, Responsibility, Ethics) for Indigenous Data Governance: https://www.gida-global.org/care (accessed 2 September 2022).
- The ICOM Code of Ethics is available via https://icom.mu seum/en/resources/standards-guidelines/code-of-ethics/ (accessed 23 April 2022).
- The IFLA Code of Ethics is available via https://www.ifla. org/faife/professional-codes-of-ethics-for-librarians (accessed 25 April 2022).
- 11. The Museum Association's Code of Ethics is available via https://www.museumsassociation.org/campaigns/ethics/code-ofethics/ (accessed 23 April 2022), and CILIP's code is available via https://www.cilip.org.uk/page/ethics (accessed 25 April 2022).
- The UKROI's Code of Practice for Researchers is available via https://ukrio.org/publications/code-of-practice-for-research/1-0introduction/ (accessed 25 April 2022).

- 13. Interview with James Baker, 22 June 2022.
- 14. Interview with Clifford Lynch, 17 June 2021.
- 15. Interview with Jason Baron, 22 June 2021.
- https://www.archives.gov/records-mgmt/bulletins/2013/2013-02.html (accessed 23 April 2022).
- 17. Interview with Leslie Johnston, 18 June 2021.
- 18. Interview with Jason Baron, 22 June 2021.
- 19.Ibid.
- 20. https://www.webarchive.org.uk/ (accessed 23 April 2022).
- 21. Interview with Leontien Talboom, 22 June 2021.
- 22. Interview with Jason Baron, 22 June 2021.
- 23. For more information on AI-assisted search, see Winters and Prescott (2019).
- 24.Ibid
- 25. Interview with Graham McDonald, 23 June 2021.
- 26. For more on sensitivity review, see McDonald (2020a, 2020b).
- 27. Interview with Jason Baron, 22 June 2021.
- https://www.gov.uk/government/news/new-technology-revealed-to-help-fight-terrorist-content-online (accessed 23 April 2022).
- 29. Interview with Jane Winters, 24 June 2021.
- 30. Interview with Jane Winters, 24 June 2021.
- 31. Interview with Adam Nix, 22 June 2021.
- 32. Interview with John Sheridan, 29 June 2021.
- 33.Ibid.
- 34.KIM refers to Knowledge, Information and Records Management.
- 35. Interview with James Lappin, 29 June 2021.
- 36. Interview with Clifford Lynch, 17 June 2021.
- 37. In 2020 due to the COVID-19 pandemic, A-Level examinations were cancelled in the UK and grades were assigned by the country's exam regulator, Ofqual. They used past data and teacher's predictions to produce computationally derived grades which resulted in 40% of grades being downgraded from teacher predictions. This was met with public outcry and the government was forced to abandon the results and use only the teacher's predicted results. See https://inews.co.uk/news/technology/a-level-results-2020-trust-algorithms-exams-scandal-635222 for more detail (accessed 25 April 2022). The use of the word algorithm created certain pre-conceptions in the mind of the public and the media that it was an example of an AI 'black box'. In fact, it was based on human driven statistical analysis and was fully documented.
- 38. Interview with Andrew Blick, 16 July 2021.
- 39. Interview with Jane Winters, 24 June 2021.
- 40. Interview with Adam Nix, 22 June 2021.
- 41.Principle 2.5 of the UKROI Code of Practice for Researchers in on Accountability. https://ukrio.org/publica tions/code-of-practice-for-research/2-0-principles/ (accessed 25 April 2022).
- 42. Interview with Graham McDonald, 23 June 2021.
- 43.Ibid.
- 44. Digital Sensitivity Review is the review of digital files to check whether they contain sensitive information. The project at the FCDO began in 2018 and is ongoing. More information can be found on the project at https://svgc.co.uk/fcdo-services-consultancy/ (accessed 25 April 2022) and https://www.fcdoservices.gov.uk/what-we-offer/digital-records-sensitivity-review/ (accessed 25 April 2022).

- 45. Interview with Andrew Dixon, 1 July 2021. As explained earlier in this article, we are using a broad definition of AI as a set of computational approaches (rather than focusing on AI in relation to deep learning and neural networks).
- 46. Interview with Graham McDonald, 23 June 2021.
- 47. Interview with James Lappin, 29 June 2021.
- 48. Interview with Leontien Talboom, 22 June 2021.
- Interview with Andrew Riley and Christopher Knowles, 7 July 2021.
- 50. "Integrity," The Civil Service Code. Available via: https://www.gov.uk/government/publications/civil-service-code/the-civil-service-code (accessed 25 April 2022).
- 51. Interview with Lindsay Aqui, 1 July 2021.
- 52. Interview with Emily Robinson, 15 July 2021.
- 53. Interview with Helen McCarthy, 1 July 2021.
- 54. Interview with Graham McDonald, 23 June 2021.
- 55.Ibid.
- 56.Ibid.
- 57. 'Impartiality' in the Civil Service Code, available via https://www.gov.uk/government/publications/civil-service-code/the-civil-service-code (accessed 25 April 2022).
- 58. 'Safety' in the UKRIO Code of Practice, available via https:// ukrio.org/publications/code-of-practice-for-research/2-0-princi ples/ (accessed 25 April 2022).
- 59. Interview with Oonagh Murphy, 30 June 2021.
- 60. Interview with Sébastien A. Krier, 2 July 2021.
- 61. Interview with Lindsay Aqui, 1 July 2021.
- 62. Interview with Jane Winters, 24 June 2021.
- 63. Interview with Gareth Millward, 7 July 2021.
- 64. 'Integrity' in the Civil Service Code, available via https://www.gov.uk/government/publications/civil-service-code/the-civil-service-code (accessed 25 April 2022).
- 65.UKRIO Code of Practice. '3.5 Collaborative Working.' Available via https://ukrio.org/publications/code-of-practice-for-research/3-0-standards-for-organisations-and-researchers/3-5-collaborative-working/ (accessed 25 April 2022).
- 66. Interview with John Sheridan, 29 June 2021.
- 67. Interview with Daniel van Strien, 2 July 2021.
- 68. Interview with Gareth Millward, 7 July 2021.
- 69. Interview with [name withheld], 27 May 2021.
- Interview with Andrew Riley and Christopher Knowles, 7 July 2021.
- 71. Interview with James Lappin, 29 June 2021.
- 72. Interview with Breandán Knowlton, 1 July 2021.

Funding

This research was funded by the Enterprise Projects Group (EPG) at Loughborough University. The EPG funding comes from the Higher Education Innovation Fund.

References

Agar, J. (2020) What is science for? The Lighthill report on artificial intelligence reinterpreted. The British Journal for the History of Science, 53(3): 289–310. https://doi.org/10.1017/S0007087420000230

- Angelova, L., Collomosse, J., Smillie, R., Cubey, R., and Norman, J. (2020) Deep discoveries: Interim Report. Foundation Projects, Towards a National Collection, Arts and Humanities Research Council. https://www.nationalcollection.org.uk/sites/default/files/2021-02/Deep%20 Discoveries.pdf (accessed 4 November 2022).
- Baron, J. R., Sayed, M. F., and Oard, D. W. (2020) Providing more efficient access to government records: a use case involving application of machine learning to improve FOIA Review for the deliberative process privilege. Journal on Computing and Cultural Heritage, 15(1):1–19. https://doi. org/10.1145/3481045
- Barons, M., Bhatia, S., Double, J., et al. (2021) Safeguarding the nation's digital memory: towards a Bayesian model of digital preservation risk. Archives and Records, 42(1): 58–78. https://doi.org/10.1080/23257962.2021.1873121
- Cave, S., Coughlan, K., and Dihal, K. (2019) 'Scary robots': Examining public responses to AI. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Association for Computing Machinery Honolulu, HI, USA. https://doi.org/10.1145/3306618.3314232
- Christie, L. (2020) Interpretable machine learning (POSTnote No. 633), UK Parliament, London. https://post.parliament. uk/research-briefings/post-pn-0633/ (accessed 4 November 2022).
- Cordell, R. (2020) Machine Learning + Libraries, Library of Congress, Washington, D.C. https://labs.loc.gov/static/labs/ work/reports/Cordell-LOC-ML-report.pdf?loclr=blogsig (accessed 4 November 2022).
- Corrado, E. M. and Sandy, H. M. (2017) Digital Preservation for Libraries, Archives, and Museums. London: Rowman & Littlefield Publishers.
- Deegan, M. and Tanner, S. (eds). (2006) Digital Preservation. London: Facet.
- Delve, J. and Anderson, D. (eds) (2014) Preserving Complex Digital Objects. London: Facet.
- Fast, E. and Horvitz, E. (2017) Long-term trends in the public perception of artificial intelligence. AAAI (Association for the Advancement of Artificial Intelligence) Conference, pp. 963–69.
- Gabriel, I. (2020) Artificial intelligence, values, and alignment. Minds and Machines, 30: 411–37. https://doi.org/10.1007/ s11023-020-09539-2
- Hagendorff, T. (2020) The ethics of AI ethics: an evaluation of guidelines. *Minds and Machines*, 30: 99–120. https://doi.org/ 10.1007/s11023-020-09517-8
- Harris, F. and Lyon, F. (2013) Transdisciplinary environmental research: building trust across professional cultures. *Environmental Science & Policy*, 31: 109–19. https://doi.org/10.1016/j.envsci.2013.02.006
- Harvey, R. and Weatherburn, J. (2018) Preserving Digital Materials. Washington DC: Rowman & Littlefield.
- Hockx-Yu, H. (2014) Access and scholarly use of web archives. Alexandria: The Journal of National and International Library and Information Issues, 25(1–2): 113–27. https://doi.org/10.7227/ALX.0023
- International Council On Archives. (1996) Code of Ethics, International Council on Archives.

- Jackson, P. C. (2019) Introduction to Artificial Intelligence. Mineola, NY: Dover Publications.
- Jaillant, L. (2022a) How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions. Archival Science. https://doi.org/10.1007/s10502-022-09390-7
- Jaillant, L. (ed.) (2022b) Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections. Bielefeld, Germany: Transcript. https:// doi.org/10.14361/9783839455845
- Jaillant, L. (2019) After the digital revolution: working with emails and born-digital records in literary and publishers' archives. Archives and Manuscripts, 47(3): 285–304. https:// doi.org/10.1080/01576895.2019.1640555
- Jaillant, L. and Caputo, A. (2022) Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data, AI & Society, https://doi.org/10.1007/s00146-021-01367-x
- Jobin, A., Ienca, M., and Vayena, E. (2019), Artificial Intelligence: the global landscape of ethics guidelines. *Nature Machine Intelligence*, 1: 389–99. https://doi.org/10.1038/s42256-019-0088-2
- Kelley, P. G., Yang, Y., Heldreth, C., et al. (2021) Exciting, useful, worrying, futuristic: public perception of artificial intelligence in 8 countries. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, Association for Computing Machinery. https://doi.org/10.1145/3461702.3462605
- Koščík, M. and Myška, M. (2019) Copyright law challenges of preservation of born-digital digital content as cultural heritage. European Journal of Law and Technology, 10(1): 1–21
- Lawton, D., Stacey, R., and Dodd, G. (2014) eDiscovery in digital forensic investigations (No. 32/14), CAST Publication, Home Office UK, London. https://www.gov.uk/government/publications/ediscovery-in-digital-forensic-investigations (accessed 4 November 2022).
- Lomas, E. J. (2017) Navigating confidentiality, legal privilege and GDPR to maintain legal records for future generations: the case for archiving. Presented at Legal Records at Risk, London. Available at: https://discovery.ucl.ac.uk/id/eprint/10026237
- Mcdonald, G., Macdonald, C., and Ounis, I. (2020a) How the accuracy and confidence of sensitivity classification affects digital sensitivity review. ACM Transactions on Information Systems (TOIS), 39(1): 1–34. https://doi.org/10.1145/3417334
- McDonald, G., Macdonald, C., and Ounis, I. (2020b) 'Active Learning Stopping Strategies for Technology-Assisted Sensitivity Review', in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: Association for Computing Machinery, 2053–56. Available at: https://doi.org/10.1145/3397271.3401267 (Accessed: 9 August 2021).
- Millar, L. (2017) Archives: Principles and Practices. London:
- Mittelstadt, B. (2019) Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1: 501–07. https://doi.org/ 10.1038/s42256-019-0114-4
- Moss, E., Watkins, E. A., Singh, R., Elish, M. C., and Metcalf, J. (2021) Assembling Accountability: Algorithmic Impact Assessment for the Public Interest. Available at SSRN:

- https://ssrn.com/abstract=3877437 or http://dx.doi.org/10. 2139/ssrn.3877437 (accessed 4 November 2022).
- Murphy, O. and Villaespesa, E. (2020) AI: A Museum Planning Toolkit, *The Museums* + *AI Network*, Goldsmiths, University of London. https://themuseumsai.network/toolkit/(accessed 4 November 2022).
- Nix, A. and Decker, S. (2021) Using digital sources: the future of business history? *Business History*, 1–24. https://doi.org/10. 1080/00076791.2021.1909572
- Owens, T. (2018) The Theory and Craft of Digital Preservation. Baltimore: Johns Hopkins University Press,.
- Padilla, T. (2019) Responsible Operations: Data Science, Machine Learning, and AI in Libraries, OCLC Research, Dublin, Ohio. https://www.oclc.org/content/dam/research/ publications/2019/oclcresearch-responsible-operations-datascience-machine-learning-ai.pdf (accessed 4 November 2022).
- The National Archives UK. (2016) The application of technology-assisted review to bron-digital records transfer, inquiries and beyond. www.nationalarchives.gov.uk/docu

- ments/technology-assisted-review-to-born-digital-records-transfer.pdf (accessed 4 November 2022).
- The National Archives UK. (2018), A guide to archiving personal data. https://cdn.nationalarchives.gov.uk/documents/information-management/guide-to-archiving-personal-data.pdf (accessed 4 November 2022).
- Vlassenroot, E., Chambers, S., Di Pretoro, E., *et al.* (2019) Web archives as a data resource for digital scholars. *International Journal of Digital Humanities*, 1: 85–111. https://doi.org/10.1007/s42803-019-00007-7
- Wadsworth, C., Vera, F., and Piech, C. (2018) Achieving fairness through adversarial learning: an application to recidivism prediction https://doi.org/10.48550/arXiv.1807.00199
- Wilkinson, M.D. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1): 160018. https://doi.org/10.1038/sdata.2016.18
- Winters, J. and Prescott, A. (2019) Negotiating the born-digital: a problem of search. *Archives and Manuscripts*, 47(3): 391–403. https://doi.org/10.1080/01576895.2019.1640753