



AI and the Archives programme Thursday 27th April 2023



Venue: Sussex Humanities Lab, Silverstone Building, University of Sussex
<https://www.sussex.ac.uk/about/campus/map>

- 09.30-10.00 Conference registration and coffee
- 10.00 - 10.20 Introduction and Welcome, Sharon Webb
- 10.20 - 11.15 **Sensitivity and Access: Unlocking Colonial Visual Archives with Artificial Intelligence**
Lise Jaillant
- 11.15 - 11.45 Break and Coffee
- 11.45 - 12.45 Panel 1: Critical Computation and Archives (Chair James Baker)
- Critical feminist screenshotting: capturing search results as partial perspectives of situated knowledge** Renée Ridgway (in-person)
- Confronting bias: Setting approaches from artificial intelligence and archiving into dialogue** Sarah Burkhardt (in-person)
- 12.45 - 13.30 Lunch
- 13.30 - 14.30 Panel 2: Audio-visual Archives and AI (Chair Sarah Arnold)
- Co-creating anti-racist datasets in AI workflows utilising films as data.** Amanda Egbe (in-person)
- Exploring the Methodological Possibilities of Automatic Speech Transcripts of Archived TV News** Jasmijn Van Gorp, Mary-Joy van der Duere (on-line)
- 14.30 - 15.00 Break and coffee
- 15.00 - 16.00 Panel 3: Critical Museum Interventions (Chair: Sharon Webb)
- Benin 1897 and the Edo Empire - two views of one subject** Alex Morrison (in-person)
- Innovative tech and invisible labor: lessons from early museum computing**
Alexandria Rayburn (online)
- 16:30 Drinks reception

Abstracts:

Sensitivity and Access: Unlocking Colonial Visual Archives with Artificial Intelligence Lise Jaillant

This talk will present the recent work of the EyCon project on AI applied to historical colonial photographs. It is part of a new co-authored article where the EyCon team explores issues of sensitivity and access to colonial visual archives, using new technologies such as Artificial Intelligence.

In recent decades, archival institutions have digitised ever-increasing quantities of material under the rubric of open access, including from colonial archives. However, large quantities of material have been digitised without sufficient metadata, which makes it difficult to discover and use these archives. When metadata exists, it often uses problematic language that replicates the perspectives of their colonial creators. Manual improvement of huge amounts of records and metadata is impossible due to lack of resources and funding. Moreover, much of the most sensitive material from these collections remains undigitised and hidden due in large part to ethical concerns. This has led to a situation in which the colonial archive risks becoming overly sanitised, as well as difficult to navigate and analyse.

Critical and transparent multimodal AI – that is to say, working on text and image at the same time – offers a way to improve access to colonial archives, without losing sight of the sensitivity of the visual materials. Indeed, computer vision using convolutional neural networks (CNNs) makes it possible to analyse visual material in digital archives at scale. Our project EyCon is thus relying on AI techniques to analyse a database of sensitive visual materials from colonial conflicts, in order to improve access to these records.

Taking EyCon as a case study, the talk explains how AI projects could be built with sensitivity and equity in mind, improving access to colonial archives while enabling more ethical approaches to potentially sensitive material. This includes involving domain experts and source communities; using explainable AI techniques and multimodal AI for identification of sensitive material and metadata enrichment; and combining “distant” and “close” reading of archival materials. These approaches can unlock colonial visual archives at scale, and lead to new historical knowledge and interpretations.

Critical feminist screenshotting: capturing search results as partial perspectives of situated knowledge Renée Ridgway

‘Ubiquitous googling’ (Ridgway 2021) with keywords is a new media habit (Chun 2016) and Google serves a window into users’ thoughts, interests and desires. Instead of receiving 10 hyperlinks on a page, increasing AI chatbots (Bard) deliver search results yet Google’s proprietary algorithms are shrouded in patents and ‘guarded, like missile codes’ (Noble 2018). Transient as well as opaque, what are the criteria determining search results and how can they be captured? Latanya Sweeney’s screenshotting showed how Google Ads with a greater percentage of the word ‘arrest’ appeared more often for black identifying first names in searches’ (2013:34), including her own. Safiya Noble unearthed how Google’s ‘algorithmically crafted web search’ delivered racism and sexism as the first results with ‘black girls’ through paid advertising (2018:5), based on her own searching and screenshotting. These examples of discrimination within machine learning search algorithms (Chun 2019:64) reflect how data sets are attached to bodies as ‘biopolitical implications’ related to ‘gender, race, sexuality, class, disability’ (Cifor et al., 2019).

In this presentation I show results of an auto-ethnography that visualises the black-boxed algorithms of search results through ‘critical feminist screenshotting’, which engenders the ‘partial perspective’

of 'situated knowledge' (Haraway 1988:587). Documentary, evidentiary but also empowering, screenshots capture the 'gaze of the search engine' (Noble 2018:71,116) yet an intersectional lens of analysis also exposes some of their workings and effects. The method contributes to the scholarship 'Feminist Data Manifesto No' and subversive movements informed by feminist, critical data and critical archival practices (D'Ignazio and Klein, 2020).

Confronting bias: Setting approaches from artificial intelligence and archiving into dialogue

Sarah Burkhardt

Archival practices and their narrative powers have always emerged in dialogue with particular technical media, their affordances and logics, carrying a historical and cultural momentum of their own. Following this logic, the seemingly unstoppable rise and power of AI technologies must shape not only the material to archive but also the archival practices and tools. However, the sparse amount of research on the role of AI in archiving shows that rather than being widely applied, infrastructural, legal and institutional barriers prevent a smooth fusion. But even if AI was not widespread in archival practices, its ubiquity in underpinning today's socio-technical fabrics and digital cultures reconfigures epistemological approaches to narrate and understand societal or cultural phenomena. Rather than being a mere technical "tool", the specific technicity and issues related to AI should or already do indirectly stimulate archival practices and concerns as part of a shared zeitgeist. Here, bias poses one of the most characteristic and concerning problems to AI's narrative power. But what does bias even mean when approached as a technical concept in AI, and how do computer scientific approaches to mitigate bias relate to or conflict with archival approaches? Referring to bias as an issue that gained growing attention through the prevalence of AI, the research seeks to approach and discuss this "computational concern" from an archival perspective. It does so by building on practical insights from ongoing cross-media research on the Dutch #MeToo debate that attends to a variety of different media outlets and archived data, from Twitter to television to websites or news. In particular, it discusses how reflections on bias from both perspectives-AI and archiving-could inform feminist research practices and methodological design across multiple layers and actors, ranging from data, institutions, algorithms, media affordances, to the researchers themselves.

Co-creating anti-racist datasets in AI workflows utilising films as data. Amanda Egbe

Considering the concern for racial bias within AI algorithms, could creative responses within moving image archival practice and critical film theory foreground possibilities for intersectional approaches?

This paper stems from artist practices with moving image archival materials that consider the materiality, historiography and ontology of film, such as *Tom Tom the Pipers Son* (Jacobs, 1969). Coupled with critical race theory, feminist and intersectional approaches (Anthias and Yuval-Davis, 1992), (Yuval-Davis, 2011), we can consider the broader applications of producing datasets for and with film analysis to resist bias in AI and computer vision workflows.

Situated within the context of practice, drawing in broader image archives to produce critical moving image datasets, this approach aims to consider strategies for AI and cinema workflows that are independent of uncritical responses to the film canon and responsive to the notion of the coded gaze (Buolamwini, 2016).

The paper reflects on the author's artistic practice of utilising image archives to supplement the analysis of the representation of race by creating datasets. These datasets reflect the context of on-screen racial depiction and the situated context of the broader cultural environment of the film's making.

References:

Anthias, F. and Yuval-Davis, N. (1992) *Racialized Boundaries: Race, Nation, Gender, Colour and Class and the Anti-Racist Struggle*, London: Routledge.

Buolamwini, J. (2016) *The Algorithmic Justice League*. medium.com December 14.
<https://medium.com/mit-media-lab/the-algorithmic-justice-league-3cc4131c5148>

Jacobs, K. (1969) *Tom Tom the Piper's Son*

Yuval-Davis, N. (2011), *The Politics of Belonging: Intersectional Contestation*, Sage.

Exploring the Methodological Possibilities of Automatic Speech Transcripts of Archived TV News

Jasmijn Van Gorp, Mary-Joy van der Duere

In recent decades, broadcast archives opened up their collections with automatic speech recognition (ASR). For archives, ASR is predominantly used as metadata enrichment to increase the chance of retrieval: ASR translates the spoken dialogues to text and enriches the metadata to increase the chance of discovery (e.g. Ordelman and van Hessen, 2018). In this paper, we shift perspectives to the methodological opportunities of ASR-data for media studies research.

In our research project, we consider ASR-data as 'locally inscribed data' (Loukissas, 2019). Each step - from preservation to processing of broadcasts through an ASR-pipeline - requires a myriad of choices made by archivists, who in turn depend on the constraints of the technology used. What archivists keep and what they delete depends on an archive's policy and archivists' practices in a specific historical context (cf. e.g. Cooke, 2011). We contend, then, that it is important to take local context into account when assessing available and missing ASR-files as research data.

More specifically, we formulate and test a method based on Dutch archived TV-news on the Chernobyl nuclear disaster. As first part of the method, we conduct a data-analysis in which we compare ASR-files with other metadata fields. This comparison helps us to discover patterns, to come into grips with ASR's specificities and to understand the gaps in the data. As second part, we share and discuss the results of the data-analysis with archivists in order to better grasp ASR's local context.

Benin 1897 and the Edo Empire - two views of one subject Alex Morrison

We have recently been working on a research project for the Edo Museum of West African Art aimed at finding out what material is held in UK archives that relates to the Edo Empire, Benin, the notorious raid of 1897 and the artefacts that then made their way into Western collections. Meanwhile, a parallel project has been approaching the same subject from the African side. A team in Germany has been working with experts and community members in Benin to produce an online resource called Digital Benin. This recently launched digital resource contains a rich combination of object records, oral history, bibliographic references and interpretative material. This short talk will present and reflect on the differences between the two views of what is ostensibly the same subject. Questions that arise include: what can we learn from this case study about the possibilities for decolonisation of archival material? What are the possibilities and limits of digital resources?

Innovative tech and invisible labor: lessons from early museum computing Alexandria Rayburn
(online)

This presentation shares findings from a recent literature review on computing practices in museum collections from the 1960s to present. I argue that computing work in museums has historically been a feminized profession, drawing first on the ‘unskilled’ labor of key punchers, which was a precarious and invisible position in collections in the 1960s. This trend continues into the present, leading museum computing work to be precarious, underfunded, and not well understood in the larger field of museums.

Similar to the introduction of the mainframe computer in the 1960s, many practitioners in museums today are questioning the role that artificial intelligence (AI) can play in our collections (Moriarty 2019, Ciecko 2022, Wu 2022). However, in this literature there is little focus on the current state of AI labor, as gig based, underpaid, and supported by an invisible workforce (Gray and Suri 2019). It seems as scholarship surrounding AI, at least in museums, tends to fall into similar traps of previous computing work, which assumes the implementation will be done by low paid tech workers, not skilled museum professionals. The lack of understanding surrounding the nuances of computing work in collections contributes to its professional undervaluing. This presentation aims to contribute to conversations questioning whether AI can be integrated into the work practices of memory institutions in an ethical way, or if it will just reinforce precarious labor practices we already witness in AI work, and within memory institutions.

Ciecko, Brendan. 2020. “AI Sees What? The Good, the Bad, and the Ugly of Machine Vision for Museum Collections.” In MW20.

Gray, Mary L., and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston New York NY: Houghton Mifflin Harcourt.

Moriarty, Adam. 2019. “A Crisis of Capacity: How Can Museums Use Machine Learning, the Gig Economy and the Power of the Crowd to Tackle Our Backlogs.” In *Museums on the Web 2019*. <https://www.museweb.net/bibliography/?bib=5799>.

Wu, Mingfang, Hans Brandhorst, Maria-Cristina Marinescu, Joaquim More Lopez, Margorie Hlava, and Joseph Busch. 2022. “Automated Metadata Annotation: What Is and Is Not Possible with Machine Learning.” *Data Intelligence*, September, 1–17. https://doi.org/10.1162/dint_a_00162.

Speakers



Amanda Egbe

Amanda Egbe is an artist, filmmaker, researcher, and senior lecturer in Media Production at the University of the West of England (UWE). Her research and practice focus on archives, digital technologies, the moving image, activism and race. She has exhibited work nationally and internationally at festivals, conferences and galleries.

Sarah Burkhardt

Sarah Burkhardt (she/her) is a PhD candidate in media studies at the University of Amsterdam with a background in computer science, fine arts, and new media studies. Her PhD is part of a research infrastructure project called Twi-XL that is aimed at developing computational cross-media methods for social sciences and humanities with Dutch archives and institutions. Being interested in the question of how (new) media transform power relations, her PhD research investigates the role of different media (social media, newspapers, television, websites) for feminist activism and #MeToo in the Netherlands. Her primary approach to locating power dynamics at the level of technical design requires a continuous critical reflection and theoretical grounding of digital research methods and infrastructures. For this, she engages with media theory and feminist theories to understand computational research methods not in isolation but as entangled practices and interventions within the issue spaces at hand.



Lisa Jaillant



Dr Lise Jaillant is Reader in Digital Humanities at Loughborough University. She was awarded a British Academy Rising Star Engagement Award (2017-18) for her project “After the Digital Revolution,” which was followed by an AHRC Leadership Fellowship (2018-20) to work on the born-digital records of the poetry publisher Carcanet Press. Jaillant has extensive experience of international research networks. Since 2020, she has led four AHRC-funded projects:

- (1) AURA: a UK/ Irish network to connect Digital Humanists, Computer Scientists and archivists to unlock cultural assets;
- (2) AEOLIAN (UK/ US: AI for Cultural Organisations);
- (3) EyCon (Visual AI and Early Conflict Photography) in partnership with French researchers;
- (4) LUSTRE (Unlocking our Digital Past with AI) in partnership with the Cabinet Office.

Recent publications include the edited collection *Archives, Access and AI* (published open access, 2022), and articles in *AI & Society*, *Archival Science* and *American Archivist*.

Alex Morrison

Alex wrote his first program, an online game for an IBM mainframe, on a deck of punched cards. After a spell with Arthur Andersen as a consultant, he became involved in artificial intelligence. This led, in 1985, to a move to Brighton to found a company associated with the University of Sussex's Cognitive Studies programme, Cognitive Applications; now Cogapp.



Alex and colleagues at the new company started developing applications using hypertext and multimedia. Projects in the cultural sector followed including a development for the National Gallery's Sainsbury Wing. This formative three-year project gave birth to the Micro Gallery. Opened in 1991, the Micro Gallery was the first large-scale application of digital media in a museum.

Since 1991, Alex and his colleagues have been applying digital media technologies for museums, galleries and archives around the world.

Alex has a degree in Mathematics and Philosophy from Oxford University.

Alexandra Rayburn



Alexandria Rayburn is a Doctoral Candidate at the University of Michigan School of Information where she also received her Master of Science in Information. Her research centers around computing in museums and Knowledge Infrastructures. She is particularly interested in the maintenance of digital systems within collections, and the labor practices associated with that maintenance. Prior to graduate school, Alexandria worked in cultural heritage museum collections as a cataloger.

Renée Ridgway

Renée Ridgway is a researcher, educator and media artist, having studied at the Rhode Island School of Design (BFA)/Brown University, Piet Zwart Institute/ Plymouth University (MA) and Copenhagen Business School in their Management, Philosophy and Politics department (PhD). Ridgway was also a research associate at Leuphana University's Digital Cultures Research Lab (DCRL) (2014-2017), in 2018 and 2019 a fellow at CAIS (Centre for Advanced Internet Studies) in Bochum, DE and in 2022, a Digital Cultures fellow at KU (Katholische Universität), DE. Presently she is a Post-doctoral researcher in the newly created SHAPE (Shaping Digital Citizenship) centre at the department of Digital Design and Information Studies, Aarhus University, DK where she addresses the problematics and politics of Google search, along with researching its alternatives (European public index) and the future of search (chatbots).

Mary-Joy van der Deure

Mary-Joy van der Deure has completed the Research Master programme Media, Arts and Performance Studies at Utrecht University (2022) with a thesis on the materiality of digital, audio-visual heritage. She is currently a junior researcher at Utrecht University, working at the CLICK-NL Project "Re-Frame" and the infrastructure project "CLARIAH WP5: Media Studies and Audio-visual Data".





Jasmijn Van Gorp

Jasmijn Van Gorp is assistant professor in Digital Television History at Utrecht University. She is co-lead of “CLARIAH WP5: Media Studies and Audio-visual Data”, which develops a digital research infrastructure for audio-visual archives. In addition, she leads the CLICK-NL Project “Re-Frame” that focuses on the use of AI and audio-visual archives in journalistic practices.