Uncovering Positionality in data/AI-driven cultural heritage pipelines

Anna-Maria Sichani

December 2023-March 2024

Digital cultural heritage data, AI-driven pipelines and positionality

Over the last couple of years, new computationally-intensive methods, including AI, are introduced and applied to enhance and curate digitised cultural heritage data: from data analysis and enrichment of digital collections using Natural Language Processing (NLP) to advanced, sophisticated searchability and access using Computer Vision and Machine Learning algorithms. Recently, digital collections have also been used as training data for fine-tuning and evaluation of AI methods and ML models, both supervised and unsupervised.

Alongside the emergence of AI-driven pipelines in digital cultural heritage, there have been regular claims around the need to assess and mitigate bias in cultural heritage datasets from a theoretical (Posner 2015), methodological (Kizhner et al. 2021; Nyhan & Ortolja-Baird 2021) and institutional perspective (Sever 2022). However, current bias mitigation strategies are focusing on locating and correcting visible gaps or errors in datasets, mainly in the stage of data collection/capture, gaps that are magnified through digitisation and public access. As these datasets are used as initial training data for the AI models, bias enters AI systems. On the other hand, bias can be found in the algorithm (their programming), or the predictions the algorithm produces. Mapping bias in AI algorithms, models and systems used in digital cultural heritage is a complex exercise, yet to be performed, as the community is still surfacing the area.

Although there have been regular, generic claims around the need to assess and mitigate bias in cultural heritage data and/or in AI systems, there is a significant lack of a systematic, yet flexible, way to uncover **the root of bias, that is positionality, t**hroughout AI-driven pipelines. This is what this investigation aims to offer.

Defining Positionality in AI-driven pipelines

Positionality is the culmination of a person's personal experiences and sociocultural positions that influence how they interact with the world. Social positions, such as race, ethnicity, class, gender, socioeconomic status, political affiliation, educational and work background,(dis)ability, nationality, and sexual orientation, hold meaning in our society because of cultural and historical contexts that have attached certain advantages and disadvantages to these characteristics. Positionality, in other words, serves as a lens through which we interpret and understand the world.

Positionality influences not not just how we see and understand the world, but more importantly how we engage with others, the processes we are designing and systems we're part of. Understanding positionality requires a critical reflection of our positions and recognize the privileges, power dynamics, or disadvantages they may entail.

Reflection on positionality is meant to be an iterative, constant process. Our positions change over time, so it's crucial to consistently reflect on our positions and how they align with our environment and practices.

Positionality, and how it embeds itself in processes and systems produced and performed by individuals plays a significant role in the entire data lifecycle in digital cultural heritage, from the formulation of standards and ontologies, the collection and interpretation of data, to the development of data-focused AI tools, such as ML models. AI-driven pipelines and **AI systems are designed by humans, who inevitably bring in their own positionality. However, positionality is often overlooked in the context of data collection, classification, annotation and algorithm development in AI-driven pipelines.**

Some alternative approaches and terms to 'positionality'

Positionality statements

There is an increasing demand to include a positionality statement in research methods publications in academic journals and in the methodology sections of reports and theses. Positionality statements are quite common in journal articles and scholarly works in the USA and focus on an author's racial, gender, class, or other self-identifications, experiences, and privileges, based on the idea that the author's identity can, intentionally or not, influence the results of their research. A positionality statement stands as a 'disclaimer' making clear how the identities of the authors relate to the research topic and to the identities of the participants, and how these identities are represented.

Positionality and *reflexivity*

Reflexivity is a second-order questioning and challenging of one's own thoughts and beliefs. As such, reflexivity is a process that we engage in throughout our research from conception through to dissemination. Positionality, by contrast, is our understanding of ourselves, of who we are and what we bring to our research. The positionality statement in publications, then, is a summary of that understanding of ourselves to highlight how our very being may have shaped our work.

Positionality and *self-awareness*

The concept of self-awareness encompasses various interpretations, such as understanding our own identity, how others perceive us, and where we fit into the world.Internal self-awareness refers to our ability to clearly perceive our own values, passions, aspirations, reactions, and how they align with our environment and impact others. On the other hand, external self-awareness pertains to how others perceive us in relation to our environment, values, passions, aspirations, and reactions. We can be aware of these perceptions too (Eurich 2018).

Positionality and bias

Bias refers to the tendency to view certain ideas, things, beliefs, individuals, groups, or cultures in an unfair (i.e., overly positive or negative) and unproven way. Biases can be conscious (explicit) or sub/unconscious (implicit). Our biases influence our thoughts, attitudes, and actions, whether we are aware of them or not.

Unconscious or implicit bias describes situations where our background, personal experiences, societal stereotypes and cultural context can impact our decisions and actions without us realising. Implicit biases are the preferences and prejudices that we're unaware of.

The problem with (the language and practice of) ‘bias’

I. ‘Bias’ is misleading

The problem with the term “bias” is that it is an umbrella term, often quite suggestive, used in very specific ways in different fields, and as such can lead to distraction.

II. Is de-biasing data or AI systems the solution?

The cultural heritage community has been actively trying to address bias in the collections they develop or work with.

- A recent DE-BIAS project is dedicated to detecting and contextualising the use of harmful language in cultural heritage collections aggregated in Europeana.eu, while supporting cultural heritage professionals to address bias in the collections they work with. Over the course of two years, the project will develop an AI-powered tool to automatically detect problematic terms in cultural heritage metadata and provide information about their problematic background. It will use vocabularies that combine offensive language with contextual information and suggestions for appropriate terms. These vocabularies will focus on three themes: migration and colonial history; gender and sexual identity; and ethnicity and ethnoreligious identity.
- The cultural heritage and the ML/AI community are dedicating significant effort towards the mitigation and removal of bias, by focusing on the creation of **representative or better datasets** by adding more data/rows, but that alone will not de-bias data and models neither eliminate undesirable and unacceptable outcomes of ML/AI systems.
- Current efforts to de-bias data and models suggest AI can be objective, impartial in principle. Eliminating AI bias is of course unrealistic ; however what is really useful is to identify all potential sources of bias by drilling down into training datasets, machine learning algorithms and other elements of AI systems.

III. A bias-free future is an illusion

The language of “bias” suggests the attainability of a bias-free future, successful and complete bias removal, suggesting AI can be impartial in principle. But this is not possible because all ML/AI systems are shaped by human choices that we make in data collection, the design of classification systems and the models that bring in positionality.

Why a “positionality-aware” framework?

By proposing and designing a positionality-aware framework, I seek to explore the following areas:

1. Challenge the traditional model of assessing bias found in individual areas of the pipeline, that is in the data or the algorithm. This approach reproduces an outdated and rather limiting way of doing digital scholarship, presupposing that the data is independent from the algorithm or the process /systems or, more importantly, the human agent(s), and the other way around. Especially with AI pipelines, we are able to understand how deeply interrelated all the parameters/factors are. Thus a positionality aware framework offers a modular, comprehensive way to investigate and assess such interrelations.
2. Place the human and social element of the social machine (human-in-the-loop) at the forefront of our data-driven processes, by adopting a three-fold approach on $<\mathrm{people}, \mathrm{data}, \mathrm{systems}>$. I seek to introduce a more humane perspective within the pipeline concept and foster a more inclusive, responsible and equitable landscape for data/AI-driven digital cultural heritage.
3. Suggest that it **is way more important to acknowledge and reveal positionality than mitigate bias.** While current bias mitigation strategies are focusing on locating

and correcting obvious or visible gaps or errors, a positionality aware framework helps us unveil and analyze the subtle ways in which positionality infiltrates the pipeline, often going unnoticed.
4. Explore the modular, complex & interconnected nature of the organisational concept of the ‘pipeline’ in digital cultural heritage scholarship.

By embracing positionality-aware AI, we aim to transcend the limitations of traditional bias mitigation strategies and foster a more transparent and equitable landscape for AI in digital cultural heritage. Such a framework, focused on continuous assessment and evaluation of the fit between the positionality embedded throughout the AI pipelines and the scenarios within which it is deployed, promotes the values of responsibility, transparency and accountability in the AI ecosystem.

In the end, positionality is not (always) something bad we need to eliminate; however, we must acknowledge its existence. Through this positionality-aware framework we can make sure the knowledge and perspectives that are part of and baked into our AI systems are aligned with what these systems are intended and used for.

Literature Review

A . Lucy Havens developed a bias-aware NLP methodology to engage with power relations as revealed in archival metadata description (Havens et al. 2020). Although this approach has a different focus and methodology, the original concept of providing an assessment framework for uncovering bias throughout a Machine Learning project, has been a great inspiration for this positionality-aware framework.

Havens’ comprehensive methodology includes the following stages (although they are not all quite distinct and clear in terms of process and documentation):

1) Examining power relations

Stakeholder identification
Stakeholder collaboration
Unavailable stakeholders

2) Explaining the bias of focus– type of bias

Data statement curation rationale
Data biography

3) **Applying NLP methods** , bias in algorithms
< train a discriminative classifier on the dataset using supervised learning ; evaluate classifier>

B. A collaborative project by Google AI , MIT media LAB named **Kaleidoscope** seeks to introduce the concept of Positionality-Aware Machine Learning, through a specific focus on **classification** and **data annotation.** This project has been eye-opening for the concept of positionality broadly speaking, the discourse around positionality and how it relates with AI systems, as well as the concept of awareness and management instead of elimination. However, this project is not proposing a reproducible model for assessing positionality-aware ML systems.

The team developed a white paper titled Towards Better Classification providing an overview on positionality and its effects on ML and AI. They have also been organising workshops on specific application domains, where they work with domain experts to uncover positionality in their work.

They proposed a three level approach to deal with positionality in AI systems :
Step 1: Uncovering Positionality
Step 2: Working with Positionality
Step 3: Embedding Positionality in Workflows

Introducing a Positionality-Aware approach for AI-driven cultural heritage pipelines

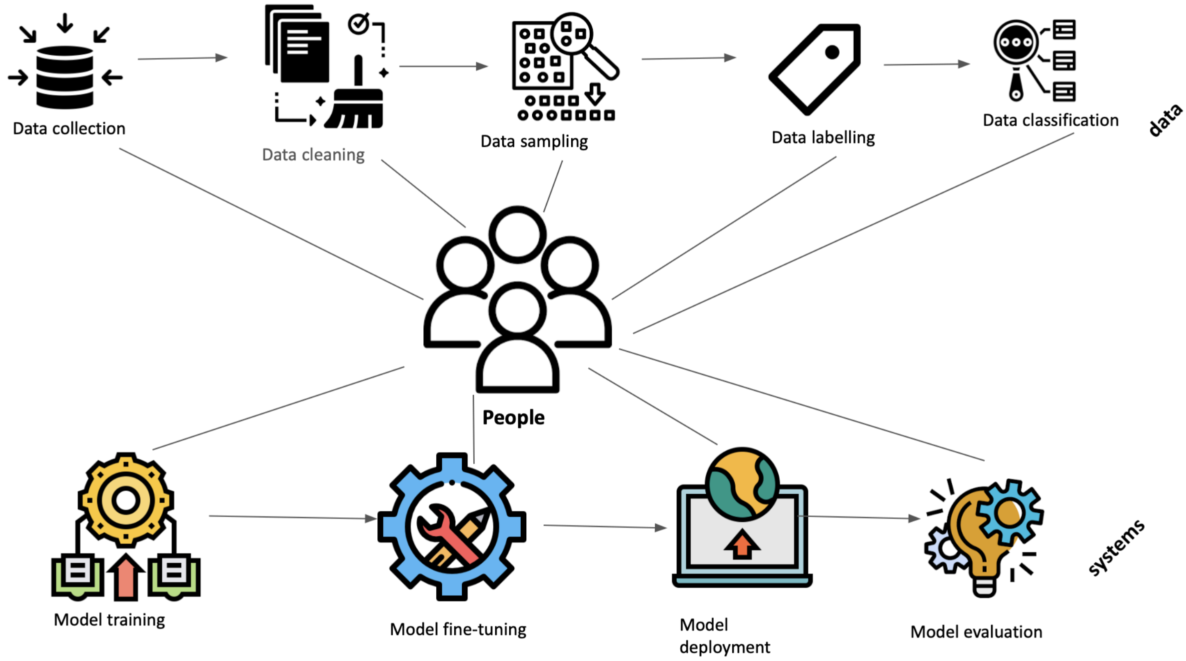
Positionality is unavoidable in every human-related activity, including information infrastructures. AI-driven pipelines and AI systems are designed by humans, who inevitably bring in their positionality. AI pipelines, thus, exhibit the positionality of their human agents.

A positionality-aware approach places the human (and social) element at the forefront when assessing data/AI-driven processes. By recognizing and incorporating the various positions, perspectives, and experiences of individuals involved throughout the development and application of machine learning models, this approach brings a crucial - while misrepresented - dimension to the AI discussion, that of the **human agent.**

AI-driven pipelines

The proposed AI-driven pipeline is an abstract, simplified diagram to help us uncover positionality throughout its various components/lifecycle. As such, it does not encompass distinctions between supervised and unsupervised learning and does not include specific processes/ stages implied by different technologies (Machine Learning, Computer Vision , neural networks, etc). In the end, AI systems are very varied and way more complex than the one represented here.

AI-driven pipelines are designed by humans, who inevitably bring in their positionality. At the same time, AI pipelines are composed of a series of practices and processes related to **data** and **systems**. By placing the human element (**people**) at the forefront of our data-driven processes, we seek to uncover how positionality enters **data** and **systems** in AI-pipelines. By adopting a three-fold approach on < people, data, systems>, we aim to introduce a more human-centered perspective within the pipeline concept and foster a more responsible and equitable landscape for data/AI-driven digital cultural heritage.



Making positionality-aware cards - **a walkthrough**

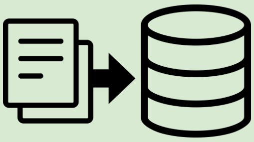
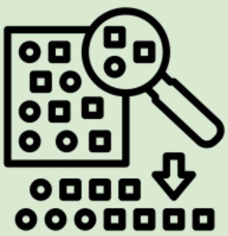
This approach is based on the - well-known within the Responsible AI community - concept and practice of card-based documentation and assessment through data cards (Pushkarna et al 2022) and model cards (Mitchell et al 2019). Data and model cards offer a way to document AI components and to ensure transparency, accuracy and reproducibility within the AI ecosystem.

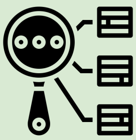
The framework is described through the template below representing three different cards <people, data, systems>, corresponding to three components in the pipeline structure/concept. Each card includes a number of processes related to the corresponding component in question; feel free to include as many as you think it's useful for the pipeline in question. For each component here are a number of indicative questions to help you uncover positionality – of course each pipeline and case study is unique so please feel free to adjust them as you see fit.

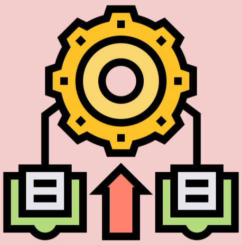
In order to perform a valid and comprehensive positionality-aware approach, for each pipeline we need to have (at least) **three different cards** < people, data, systems> in place. In case of multiple datasets or models, it is recommended to use different cards for the corresponding parts of the pipeline. We propose to start with the <People> card, continue with the data one and then move to the systems one. In each card feel free to include as many processes as you see necessary.

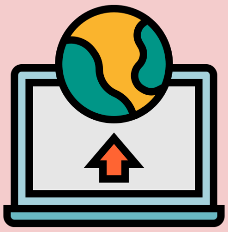
One (ethical) caveat: as this approach is profoundly human-centered, we have to ensure that the type, level and quality of personal information concerning individuals that can be included in the positionality-aware cards respect their right to privacy.

 **People**	**Identification of stakeholders** <Who is involved and in which capacity with this pipeline?> **Brief description of their background & relationship to data and systems** < Social positions related to Eg educational, professional background cultural background technical expertise>

Data	
 **Data Collection**	**How does positionality impact the dataset's composition and provenance?** How has this dataset been gathered or chosen for use? Who and how is involved in the data collection? ! User-generated data may lead to a positionality feedback loop.
 **Data cleaning**	**How does positionality impact the data preprocessing process?** How has this dataset been preprocessed and curated? Who and how was involved in the data cleaning process?
 **Data Sampling**	**How does positionality impact the data sampling process?** {Data sampling ranks dataset samples via relevance scores to select the most representative subset of data to train ML models} Who and how is involved in the data sampling process? What are the criteria for selecting representative samples?

 **Data annotation/labelling**	■ **How does positionality impact the data labelling process?** { Annotation is the act of assigning a category or label to an individual data point. } <Who assigns labels to data samples? What perspectives and experiences do they bear with them during this process? How do people decide on and allocate labels? How much does the selection of a label depend on an individual judgement, and therefore inevitably informed by perspectives, experiences, and knowledge of the labeller? Is there any quality control involved in the data annotation process? >
 **Data classification**	■ **How does positionality enter the classification systems?** { Classification is the act of assigning a label to an object, situation, idea, or person. Classification systems are contextual, thus informed by the knowledge, experiences, perspectives, and value commitments of their creators at the moment of creation and designed to organise items and ideas into discrete categories.} <Who performed and decided on the data classification? How much does the assignment of a label depend on an individual judgement, and therefore inevitably informed by perspectives, experiences, and knowledge of the person who performed the classification?

Systems	
 **Model training**	■ **How does positionality get embedded in the models and applications?** <To what extent does the choice of the model have been influenced by background knowledge and expertise of the people involved? How much do the tests or simulations and the quality of the results and the accuracy of predictions are influenced by background knowledge and expertise of people involved?

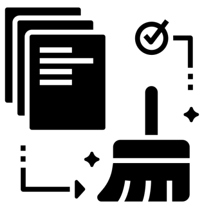
 **Model fine tuning**	■ **How does positionality impact the model fine tuning process?** <How much does the fine-tuning of a model depend on an individual judgement, and therefore inevitably informed by perspectives, experiences, and knowledge of the person who performed the fine-tuning? How much research questions influence the fine tuning process?
 **Model deployment**	■ **How does positionality impact the contexts within which these models will be deployed?** {Model deployment means integrating a trained machine-learning model into a real-world system or application to automatically generate predictions or perform specific tasks. } <How much does the deployment of a model depend on an individual judgement, and therefore inevitably informed by perspectives, experiences, and knowledge of the person who performed the deployment?
 **Model evaluation**	■ **How does positionality impact the model evaluation process?** <How much does the evaluation of a model depend on an individual judgement, and therefore inevitably informed by perspectives, experiences, and knowledge of the person who performed the evaluation?

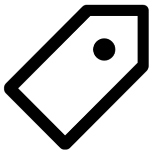
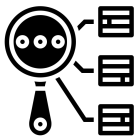
A case study: Gender in BT archives

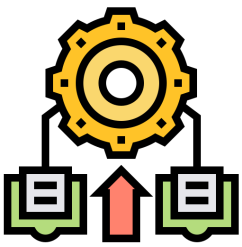
This investigation will investigate BT Archives full catalogue to identify how gender is represented in the catalogue. Thinking towards proposing a design specification for a national collection, this would create a valuable tool that allowed those looking at combining collections to understand how gender is situated in the catalogues. Our goal is to raise awareness, and even contextualise the results we find. This investigation would highlight the way archives are not neutral, and provide tools to use prior to linkage.

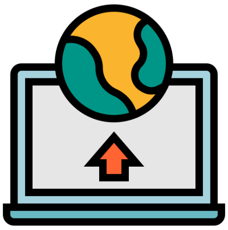
How is gender represented in BT Archives catalogue descriptions?

People	**Identification of stakeholders** **Brief description of their background & relationship to data and systems** ***Natasha Kitsher NK :*** a trained historian, with expertise in UK communications history and experience working with various archival collections including BT archives /she/her/ native British/ educated to PhD level/ ***Anna-Maria Sichani AMS***: a digital humanist, cultural /media historian without prior knowledge of UK communications history / expertise & experience in digital cultural heritage, digital research infrastructures, data management and curation, ethics / she/her/ Greek / educated to PhD level/ ***Lucy Havens LH:*** *a data scientist, developed a bias-aware NLP methodology and a model to trace bias in archival descriptions/* originally from the USA, educated in USA/ UK ***Kunika Kono: DHRH*** technical lead, software engineer, with experience in data wrangling, systems engineering , educated ***Kaspar Beelen :*** technical lead, digital historian with data science expertise, worked in various data-intensive projects

 **Data**	**1. BT Archives full catalogue data (provided by JE to DW in 2022).** **2. An export of BT Archives full catalogue data linked with authority records. Authority records are how BT's catalogue, CALM, saves all Person, Organization, and Institution records that are linked to catalogue entries.** → NK provided BT Archive context to AMS through notes and useful resources, as well as an explanation of the datasets' fields.
 **Data Collection**	**How does positionality impact the dataset's composition and provenance?** The initial BT Archives full catalogue data has been provided by BT archives as an export from their catalogue without any editing. NK created an export of BT Archives full catalogue data linked with authority records.
 **Data cleaning**	**How does positionality impact the data preprocessing process?** NK performed an initial exploration of the Person Export from Catalogue and a reconciliation via OpenRefine using the ' Code' field, the 'Name', 'Record', 'Surname', and 'Forename' fields. NK's previous familiarity with the BT archive and the datasets allowed her to quickly assess and decide on the 'important' fields and proceed with the data cleaning process. AMS struggled to understand the chaotic dataset and to familiarise herself with the 'important' fields.
 **Data Sampling**	**How does positionality impact the data sampling process?**

 **Data annotation/labelling**	■ **How does positionality impact the data labeling process?**
 **Data classification**	■ **How does positionality enter the classification systems?**

Systems	- a. Genderize.io has successfully been used against a sample of names to show what proportion of men and women are linked to the catalogue: this would be used to look at how frequently items in the catalogue linked to each gender of name . - b.Gender Bias Decoder: not applicable as we don't have access to the model/ datasets atm - c. Lucy Havens' gender bias NLP: this could be used to look at catalogue descriptions as well as a sample OCR'd dataset to see how gender is spoken about by archivists and by the actual data held by BT (TBC)
 **Model training**	■ **How does positionality get embedded in the models and applications?**

 **Model fine tuning**	■ **How does positionality impact the model fine tuning process?**
 **Model deployment**	■ **How does positionality impact the contexts within which these models will be deployed?**

BIBLIOGRAPHY

Christine Kaeser-Chen, Elizabeth Dubois, Friederike Schüür, and Emanuel Moss. 2020. Positionality-aware machine learning: translation tutorial. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 704. https://doi.org/10.1145/3351095.3375666

Takacs, David. How Does Your Positionality Bias Your Epistemology?, Thought & Action Thought & Action 27 (2003). Available at: https://repository.uchastings.edu/faculty_scholarship/1264

Bayeck, R. Y. (2022). Positionality: The Interplay of Space, Context and Identity. International Journal of Qualitative Methods, 21. https://doi.org/10.1177/16094069221114745

Merriam S. B., Johnson-Bailey J., Lee M.-Y., Kee Y., Ntseane G., Muhamad M. (2001). Power and positionality: Negotiating insider/outsider status within and across cultures. International Journal of Lifelong Education, 20(5), 405–416. https//doi.org/10.1080/02601370120490

Mahima Pushkarna, Andrew Zaldivar, and Oddur Kjartansson. 2022. Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA 51 Pages. https://doi.org/10.1145/3531146.3533231

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru, Model Cards for Model

Reporting", FAT* '19, January 29–31, 2019, Atlanta, GA, USA, https://doi.org/10.1145/3287560.3287596

Lucy Havens, Melissa Terras, Benjamin Bach, and Beatrice Alex. 2020. Situated Data, Situated Systems: A Methodology to Engage with Power Relations in Natural Language Processing Research. In Proceedings of the Second Workshop on Gender Bias in Natural Language Processing, pages 107–124, Barcelona, Spain (Online). Association for Computational Linguistics.

Eurich, T. (2018). What self-awareness really is (and how to cultivate it). *Harvard Business Review*.

https://www.queensu.ca/ctl/resources/equity-diversity-inclusivity/positionality-statement

https://pressbooks.bccampus.ca/undergradresearch/chapter/1-2-a-note-on-reflexivity-and-positionality/