From Bias Paralysis to Bias as a Category of Analysis

Introducing the Bias-Aware Framework

Mrinalini Luthra[1], Amber Zijlma[2]
[1] Huygens Institute, Royal Dutch Academy of Arts and Sciences, Amsterdam, mrinalini.luthra@huygens.knaw.nl
[2] Huygens Institute, Royal Dutch Academy of Arts and Sciences, Amsterdam, amber.zijlma@huygens.knaw.nl

ABSTRACT (ENGLISH)

As digital humanities (DH) increasingly rely on computational methods—from handwritten text recognition to automated metadata extraction—addressing bias becomes both more urgent and more complex. Datasets inherit and perpetuate biases through multiple channels: discriminatory language in archives, unequal representation in collection practices, and algorithmic biases in AI-assisted processing. These biases are compounded throughout the research process, yet the term "bias" itself lacks a clear definition, often causing "bias paralysis." This paper introduces the Bias-Aware Framework, a research framework that transforms bias from a paralyzing problem into an analytical category with clear action points.

The framework consists of three key components: (1) a *Bias Thesaurus,* which establishes a shared vocabulary of 'bias' across disciplines to address its conceptual instability; (2) a *Data Lifecycle Model* showing where biases enter the research process; and (3) *Guidelines* for documenting, describing, and mitigating bias. Together, these components can transform abstract concerns about bias into concrete opportunities for methodological improvement. We therefore approach bias not simply as an error, but as a revealing analytical lens that shapes knowledge production. By explicitly describing these conditions of production, researchers can improve transparency, improve dataset documentation, and enable more informed reuse of their data.
Keywords: bias, data ethics, taxonomy, computational methods, digital humanities

ABSTRACT (ITALIANO)[1]

Titolo del paper: Dalla paralisi da pregiudizio al pregiudizio come categoria di analisi: Introduzione al Framework Bias-Aware

Poiché le scienze umane digitali (DH) si affidano sempre più a metodi computazionali—dal riconoscimento del testo manoscritto all'estrazione automatica dei metadati—affrontare i pregiudizi diventa sia più urgente che più complesso. I dataset ereditano e perpetuano pregiudizi attraverso molteplici canali: linguaggio discriminatorio negli archivi, rappresentazione diseguale nelle pratiche di raccolta e pregiudizi algoritmici nell'elaborazione assistita dall'intelligenza artificiale. Questi pregiudizi si amplificano durante tutto il processo di ricerca, mentre il termine "pregiudizio" stesso manca di una definizione chiara, causando spesso una "paralisi da pregiudizio." Questo articolo introduce il Framework Bias-Aware, un quadro di ricerca che trasforma il pregiudizio da problema paralizzante a categoria analitica con chiari punti d'azione.

Il framework consiste di tre componenti chiave: (1) un *Thesaurus dei Pregiudizi*, che stabilisce un vocabolario condiviso del 'pregiudizio' tra discipline per affrontare la sua instabilità concettuale; (2) un *Modello del Ciclo di Vita dei Dati* che mostra dove i pregiudizi entrano nel processo di ricerca; e (3) *Linee guida* per documentare, descrivere e mitigare i pregiudizi. Insieme, questi componenti possono trasformare preoccupazioni astratte sui pregiudizi in opportunità concrete per il miglioramento metodologico. Pertanto, affrontiamo il pregiudizio non semplicemente come un errore, ma come una lente analitica rivelatrice che modella la produzione di conoscenza. Descrivendo esplicitamente queste condizioni di produzione, i ricercatori possono migliorare la trasparenza, perfezionare la documentazione dei dataset e consentire un riutilizzo più consapevole dei loro dati.
Parole chiave: pregiudizio, etica dei dati, tassonomia, metodi computazionali, scienze umane digitali

[1] Please note that the translation of the Italian title and abstract at the start of this document were created using generative AI (DeepL), as neither of the authors is a speaker of Italian.

1. INTRODUCTION

The digital humanities (DH) face an increasingly critical challenge: initial enthusiasm for the seemingly infinite possibilities of computational methods has made way for a more critical stance towards their implementation in current DH research (Prescott, 2023). Within this shift, the question of bias has taken on a new urgency - not just as a theoretical concern, but as a practical challenge that shapes every aspect of research. This urgency stems from the recognition that unexamined biases in institutions, (training) data, metadata, and computational methods risk amplifying historical exclusions and further marginalising already underrepresented narratives. These concerns are compounded by the challenge of applying computational methods to historical materials that contain inherently biased perspectives. To illustrate, when using digital infrastructure based on historical sources, descendants of enslaved peoples are being forced to search for their ancestors in lists of 'commodities' or with the use of painful search terms; alternatively, individuals mentioned in sources are anonymised or erased through opaque indexing practices (Luthra et al., 2023). While bias has long been a concept of inquiry in the humanities (McCullagh, 2000; Trouillot, 1995), DH researchers are now having to carefully perform a balancing act, in which both historicity and social justice are valued and attended to in their research practices, while also addressing technical concerns about algorithmic fairness and data representation.

But, the term "bias" itself lacks a clear definition, causing 'bias paralysis' among researchers: after all, if everything is biased, what should we do? This paper proposes that instead of intending to fully 'remove' bias from research - impossible and, for humanities and social sciences research, unwanted - bias should be treated as a productive category of analysis for digital humanities research, a productive lens through which to further enhance research. For this, we have developed a Bias-Aware Framework, consisting of a Bias Thesaurus, a Data Lifecycle Model and Guidelines for bias identification, articulation and reduction. This contribution therefore aims to:

- **Deconstruct and Articulate Bias** in DH research through the explanation and implementation of the Bias-Aware Frameworks.
- **Encourage Ethical Intervention** for mitigating bias through the data life cycle.
- **Promote** bias analysis and mitigation as urgent and necessary inclusion to the (digital humanities') research agenda.

2. BACKGROUND

As computational methods are increasingly entangled with humanities research, the issues of bias take on renewed urgency. While AI and digital technologies enable increased access to historical sources, they also risk perpetuating historical and contemporary biases embedded in archival sources, metadata, tools and technology (Navigli et al., 2023; Thylstrup, 2019). DH researchers must grapple with multiple forms of bias that intersect across different disciplines and stages of research. A single DH project might simultaneously confront archival biases in source selection (Trouillot, 1995), historical power structures in interpretation (van Rossum, 2019), representational biases in digitisation (Kizhner et al., 2021), and algorithmic biases in computational analysis (Buolamwini & Gebru, 2018; Mehrabi et al., 2021; Søgaard et al., 2014; Suresh & Guttag, 2021).

What makes this challenge particularly concerning is that these diverse forms of bias don't simply exist in parallel—they compound and amplify each other throughout the research lifecycle. Historical biases in archival selection get encoded into metadata structures, which then inform algorithmic design, creating feedback loops that can dramatically magnify originally subtle imbalances. This cascading effect of bias—from historical sources through metadata design to algorithmic processing—creates a distinctive challenge that requires a systematic approach rather than isolated interventions. This compounding effect means that seemingly minor biases at early stages can result in significantly skewed outcomes during analysis, interpretation, and beyond.

Additionally, due to the ubiquitous use of the term 'bias', there is no consistent vocabulary to rely on when wishing to address the issue. Even within specific academic fields, the concept of bias proves elusive: Blodgett et al.'s (2020) analysis of 146 papers in the field of natural language processing reveals significant confusion in defining 'bias', while in digital cultural heritage, the characterization of offensive terminology as bias remains unclear.[2] Without a coherent transdisciplinary framework for understanding

[2] For instance the Words Matter (Modest & Lelijveld, 2018), a publication on sensitive words in the museum sector, doesn't use the term 'bias', but projects such as DE-BIAS (Masschelein et al., 2023), based at the Dutch Institute of Sound and Vision, use the term in context of developing a tool to identify harmful language in archives.

and addressing bias, there is a risk of either oversimplifying it or becoming stunned by it - what we term “bias paralysis”.
Drawing inspiration from Joan Scott's (1986, 2010) concept of gender as an analytical category and Sherman et al.'s (2024) treatment of algorithmic absences, we argue that bias, like Foucault's concept of power (2008), is relational and dynamic, actively shaping and being shaped by social, cultural, and historical contexts. This perspective shifts our focus from attempting to "solve" bias - an impossible task - to using bias as a critical tool for reflection and analysis.
Unlike existing bias taxonomies in computer science and AI ethics that focus primarily on statistical and algorithmic biases (Blodgett et al., 2020; Navigli et al., 2023), our Bias-Aware Framework specifically addresses the unique challenges faced by digital humanities researchers who must simultaneously consider historical context, cultural sensitivity, and computational aspects. The Framework's distinctiveness lies in its recognition that for humanities scholars, bias is not merely a technical problem to be eliminated but also a historically significant phenomenon that requires careful documentation and contextualization. Where AI ethics frameworks typically aim to remove bias entirely, our approach acknowledges that humanistic inquiry necessitates preserving certain original biases as historical evidence while preventing their harmful amplification through computational methods.
There have been several valuable contributions towards practical mitigation of bias, including templates for transparent documentation and alternatives to harmful terminology (Bender & Friedman, 2018; Chilcott, 2022, 2022; Luthra & Eskevich, 2024; Masschelein et al., 2023; Modest & Lelijveld, 2018; Scheuerman et al., 2020). However, these approaches typically address specific manifestations of bias. DH researchers need tools that can help them address the full spectrum of biases they encounter, from historical biases to technical biases to representation biases. Moreover, they need to understand how these different forms of bias interact and transform across the stages of their work. Furthermore, there is the need to make this approach specific to DH: because unlike the gross of the strategies and taxonomies proposed within the data and computer science fields, humanists necessarily need to ‘respect’ the original form of their (historical) data, as its biases provide valuable insights. Reflexively supplementing the data or changing the source data is therefore out of the question - to humanists, it’s precisely these imbalances that are significant to research, but not to be perpetuated through machine learning models. We believe that bias mitigation in DH currently faces the following fundamental questions as obstacles to identifying, articulating, and mitigating bias:

1. What exactly do we mean by “bias” in digital humanities research?
2. Where does bias occur in the dataset creation process?
3. How can researchers practically address bias with their available resources?

In order to address these challenges, we are developing a “Bias Aware Framework” which has the the three corresponding components:

1. **A Bias Thesaurus**: A comprehensive list of the concepts connected to bias (such as representation, offensive language, FAIR, CARE, silences, etc.) that creates a shared vocabulary for discussing bias across disciplines.
2. **A Bias-Aware Data Lifecycle Model**: Showing where and how bias manifests at different research stages, allowing for targeted interventions at critical points.
3. **Guidelines and Toolkit**: Reflective questions at each stage of the dataset lifecycle, illustrative examples, and “good-better-best” recommendations for bias analysis, description, and mitigation.

Our development of the Bias-Aware Framework is a response to the gap in the DH field, that “a set of guidelines is missing, a serious lack when one might want to think through ethical concerns” (O’Sullivan, 2024). It is an actionable framework that demystifies ‘bias’ and transforms it into a productive tool for improving current and future research and knowledge production. Our focus on the dataset specifically

3. METHODOLOGIES

Our development of Bias-Aware Framework has followed a three-phase iterative approach combining theoretical analysis with practical validation:

1. **Literature Review:** To gain a better overview of current theories about and strategies against bias, we systematically reviewed literature in the fields of archival studies (Modest & Lelijveld, 2018; Stoler, 2010; Trouillot, 1995), epistemology (Foucault, 2002) and computer sciences (Bender & Friedman, 2018; Gebru et al., 2021; Orr & Crawford, 2023; Scheuerman et al., 2020). These are fields in which (origins of) bias has received much attention. From here, we expanded out to include sources that were not academic and/or text-based, such as video’s, art installations,

and fiction. The importance of including these resources was to critically confront our own biases for the written and academic. For each resource, we focused on what forms of bias or strategies to mitigate bias were present, compiling these in an open-to-access list of resources.[3]

2. **Insights from Partner Projects:** The Framework's development is guided by semi-structured interviews with stakeholders from four major digital infrastructure projects focusing on colonial and slavery archives: Slave Voyages[4], GLOBALISE[5], Exploring Slave Trade in Asia[6], and the Historical Database of Suriname and Caribbean[7]. These partnerships provide crucial insights into practical challenges and implementation requirements. We also conduct conversations with (external) advisors, based in different parts of the world and with expertise from a range of disciplines: cultural heritage, critical archival studies, community (memory) work, natural language processing, and the FAIR principles. As a result, the Framework is rooted in plurivocality (Sitzia, 2023). The experiences of our partners and advisors, both in terms of challenges and strategies to navigate these, form the foundation of the Framework.
3. **Framework Validation and Refinement:** We are validating and refining the Bias-Aware Framework through two parallel tracks: expert consultations and interactive workshops with digital humanities projects. Workshops have been organised that serve as practical testing grounds where participants apply the Framework's toolkit to analyse bias in their own datasets. This implementation phase aims to reveal the framework's strengths and limitations in practice and identify potential blind spots. Participant feedback and documented use cases will drive iterative improvements to the Framework's components, ensuring its broader applicability and effectiveness.

4. BIAS-AWARE FRAMEWORK

The Bias-Aware Framework consists of three elements:

1. **The Bias Thesaurus** maps the various *expressions of bias*—concrete forms bias takes in research practices, such as harmful language, uneven descriptive depth, or limiting categorization schemes. The Thesaurus creates a shared vocabulary across disciplines, visualises interconnections between different expressions of bias through network graphs, and provides researchers with a conceptual map for navigating bias-related concerns. For each expression of bias, the Thesaurus includes their definition, where they occur in the life cycle (therefore linked to the Lifecycle Model), what questions to consider regarding the specific expression throughout their research and good-better-best practices (therefore linked to the Guidelines). In this way, the Thesaurus is the theoretical grounding of our Framework (see Figure 1).
2. **Bias-Aware Dataset Lifecycle Model** (Figure 2) forms the structural backbone of our framework, grounding abstract bias considerations in familiar research workflows while addressing a gap in digital humanities methodology. Our model[8] identifies five key stages (Set Up, Collect, Process, Analyse, Preserve & Share) and maps how different expressions of bias defined in the Thesaurus manifest at each stage. A key insight from our research is the "stickiness" of certain bias expressions across multiple stages, though they manifest differently depending on the stage's focus. For example, representation concerns appear throughout the lifecycle: in Set Up, they relate to whose scholarship informs the project; in Collect, they concern whose perspectives are captured in the data; in Process, they involve how categories represent complex realities; and in Analyse, they address whose stories are highlighted in the subsequent story the data tells.

---

[3] Combatting Bias Resources List: https://docs.google.com/spreadsheets/d/17mAClY06JuPQm9qf3Z7ZmdYi08VrZboCUTOv27J6J-E/edit?usp=sharing

[4] Slave Voyages: https://www.slavevoyages.org/

[5] GLOBALISE: https://globalise.huygens.knaw.nl/

[6] Exploring Slave Trade in Asia (ESTA): https://esta.iisg.nl/

[7] Historische Database Suriname and Caribbean (HDSC): https://www.ru.nl/onderzoek/onderzoeksprojecten/historische-database-van-suriname-en-de-cariben

[8] See https://www.rd-alliance.org/wp-content/uploads/2024/09/D1_The-creation-of-a-harmonised-research-data-lifecycle-RDL-model-and-crosswalkpdf

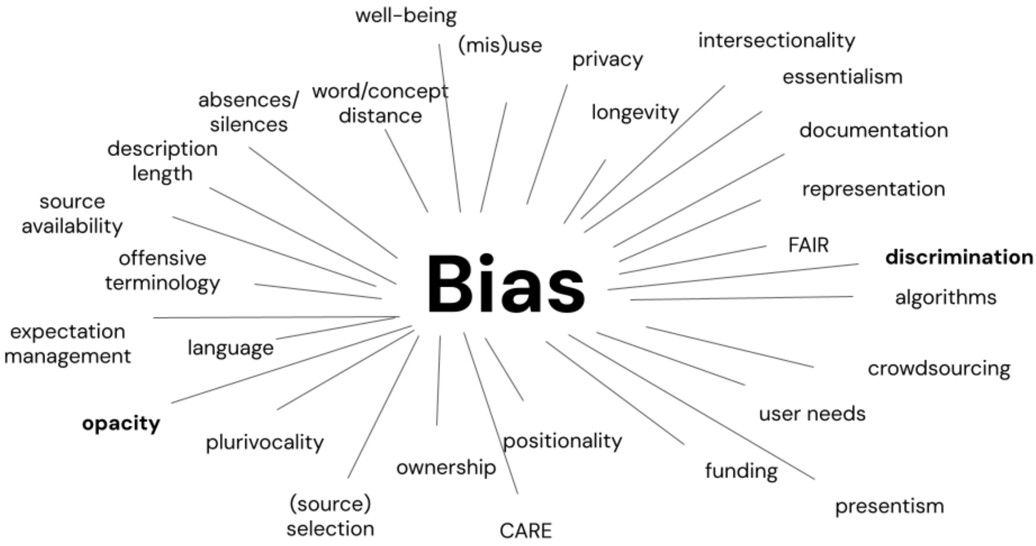

Figure 1: Bias as heuristic for distinct yet interrelated issues. These expressions of bias are explored in the Bias Thesaurus.

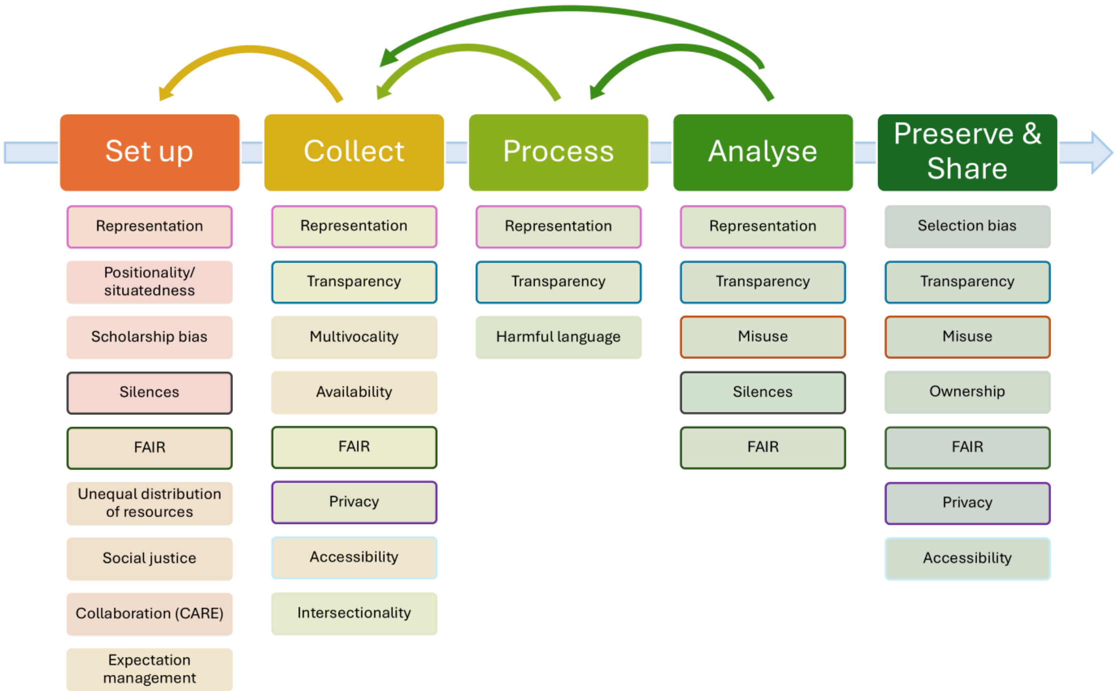

Figure 2: The Bias-Aware Lifecycle Model. The expressions of bias associated with each stage are contextualised in the Thesaurus and Guidelines.

3. **Guidelines and Toolkit:** The final component transforms theoretical understanding into practical action through structured guidelines for each stage of the dataset lifecycle. These guidelines provide reflective questions, curated resources, documentation templates, "good-better-best" recommendations[9] that accommodate varying resource constraints, and example strategies drawn from successful DH projects. Figure 4 illustrates our guideline approach for addressing archival silences, offering tiered intervention strategies from basic documentation to participatory community engagement. The guidelines emphasize that addressing bias is not an all-or-nothing proposition—even resource-constrained projects can implement basic bias-aware practices. This

[9] The good-better-best-framework has been humbly taken over from Alicia Chilcott (2022).

scaffolded approach helps prevent “bias paralysis” by making intervention accessible regardless of project scale or resources.

Essentially, the Framework compiles existing resources in the field - both published (literature) and unpublished (internal project strategies, collected through conversation) - to create a visual and practical toolkit on bias identification, articulation and reduction.

Dealing with Discriminatory language in databases[a]

Scenario: A dataset includes outdated or harmful language, such as colonial-era terms or descriptors that marginalize specific groups based on ethnicity, gender, sexuality, or ability.

Reflective questions

- What does your research understand by ‘offensive terminology’? What is the source of this understanding?
- What processes exist in your project for identifying potentially offensive terms?
- Is it relevant for your project to document offensive terminology?
- Along what lines is terminology offensive (e.g. ethnoreligious, gender and sexuality, race, (dis)ability)?[b]

Example dataset: Thesaurus Creation in the GLOBALISE project[c]

Good practice	Better practice	Best practice
• Present offensive terms taken directly from the archive in quotation marks. • Do not substitute offensive terms with a modern equivalent. • Include a content warning explaining usage of offensive terms • Reference community-approved guidelines (such as Words Matter, DE-BIAS, Cultural Heritage Terminology Network). • Integrate this information into the data-envelope[d] and documentation[e] accompanying the dataset.	*In addition to ‘good practices’:* • Create research guides with relevant search terms and context about offensive language in historical records. • Acknowledge the evolving nature of terminology guidelines.	*In addition to ‘better practices’:* • Implement participatory methods for tagging and description. • Ground practices in ethical representation and empathy. • Integrate a discussion of problematic language into the project plan and workflow early-on in the project. • Consult communities and experts on terminology. • Select tools that enable separate processing of sensitive terms (e.g. DE-BIAS). • Document all decisions in documentation and data-envelopes. • Maintain versioned documentation with clear update plans.

[a] Chilcott’s (2019) framework addresses racially offensive language specifically - this example therefore incorporates her proposed protocols as well.

[b] For more, see the ‘wheel of bias’ developed by the DE-BIAS project, in Beirigo & Groot (2024).

[c] This dataset which is currently under construction defines and creates a taxonomy of concepts that feature in the archives of the Dutch East India Company. As many of these concepts are referred to by labels that are problematic and derogatory, conceptualizing a method of addressing problematic language in this dataset is an important component of dataset creation.

[d] Developed by Luthra & Eskevich (2024), data-envelopes are machine-readable forms of dataset documentation (also known as datasheets in machine learning), tackling the specific needs and challenges of cultural heritage datasets.

[e] Documentation refers to the descriptive account of dataset creation which accompanies the publication of a dataset. Documentation provides the contextualizing information for a dataset from a disciplinary point of view, that a data-envelope does not, or only briefly provides.

Figure 3: Example *Guideline* for dealing with discriminatory language in datasets.

5. ABOUT COMBATTING BIAS

Combatting Bias[10], is a one-year project focusing on the ethical creation of datasets for the social sciences and humanities. It particularly focuses on the dataset, because creators, users, and digital infrastructures intersect at this point, making it a good unit of analysis. CB’s work is enriched by partnerships with leading digitisation initiatives and advisors from different geographies and disciplines such as museum studies, critical archival studies, ethnomusicology, history, computer science, and economics.

ACKNOWLEDGEMENTS

The authors of this paper are part of the *Combatting Bias* project, funded by the NWO via the Thematic Digital Competence Centre Social Sciences & Humanities (TDCC-SSH). We thank the anonymous reviewers for their critical feedback.

REFERENCES

Bender, E. M., & Friedman, B. (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, *6*, 587–604. https://doi.org/10.1162/tacl_a_00041

[10] https://combattingbias.huygens.knaw.nl/

Beirigo, Isabel & Monique Groot (2024). *D2.2 Community Interactions: Scenarios and Results* (No. D2.2; DE-BIAS). Netherlands Institute of Sound and Vision.

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of "Bias" in NLP. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454–5476). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.485

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Chilcott, A. (2022). Towards protocols for describing racially offensive language in UK public archives. In V. Frings-Hessami & F. Foscarini (Eds.), *Archives in a Changing Climate—Part I & Part II* (pp. 151–168). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19289-0_10

Foucault, M. (2002). *Archaeology of Knowledge* (2nd ed.). Routledge. https://doi.org/10.4324/9780203604168

Foucault, M., & Hurley, R. (2008). The history of sexuality. Volume 1, The will to knowledge. Penguin.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., III, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86–92. https://doi.org/10.1145/3458723

Kizhner, I., Terras, M., Rumyantsev, M., Khokhlova, V., Demeshkova, E., Rudov, I., & Afanasieva, J. (2021). Digital cultural colonialism: Measuring bias in aggregated digitized content held in Google Arts and Culture. *Digital Scholarship in the Humanities*, *36*(3), 607–640. https://doi.org/10.1093/llc/fqaa055

Luthra, M., & Eskevich, M. (2024). Data-Envelopes for Cultural Heritage: Going beyond Datasheets. In I. Siegert & K. Choukri (Eds.), *Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies @ LREC-COLING 2024* (pp. 52–65). ELRA and ICCL. https://aclanthology.org/2024.legal-1.9

Masschelein, A., Truyen, F., Taes, S., van Mulder, J., Stynen, A., & Pireddu, R. (2023). *Report on research into bias types and patterns, including a typology applied to Europeana use cases and a vocabulary co-created with communities* (D2.1; DE-BIAS). KU Leuven.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, *54*(6), 115:1-115:35. https://doi.org/10.1145/3457607

Modest, W., & Lelijveld, R. (2018). *Words Matter: An unfinished guide to word choices in the cultural sector*. The National Museum for World Cultures (Tropenmuseum, Afrikamuseum, Museum Volkenkunde, Wereldmuseum). https://amsterdam.wereldmuseum.nl/en/about-wereldmuseum-amsterdam/research/words-matter-publication

Navigli, R., Conia, S., & Ross, B. (2023). Biases in Large Language Models: Origins, Inventory, and Discussion. *Journal of Data and Information Quality*, *15*(2), 10:1-10:21. https://doi.org/10.1145/3597307

Orr, W., & Crawford, K. (2023). *The social construction of datasets: On the practices, processes and challenges of dataset creation for machine learning*. OSF. https://doi.org/10.31235/osf.io/8c9uh

O'Sullivan, J. (2024). *The Bloomsbury Handbook to the Digital Humanities*. Bloomsbury Publishing. http://www.bloomsburycollections.com/collections/monograph

Prescott, A. (2023). Bias in Big Data, Machine Learning and AI: What Lessons for the Digital Humanities? *Digital Humanities Quarterly*, *17*(2). https://www.digitalhumanities.org/dhq/vol/17/2/000689/000689.html

Scheuerman, M. K., Spiel, K., Haimson, O. L., Hamidi, F., & Branham, S. M. (2020). *HCI Guidelines for Gender Equity and Inclusivity*. https://doi.org/10.13016/M2NW1F-P0JX

Scott, J. W. (1986). Gender: A Useful Category of Historical Analysis. *The American Historical Review*, *91*(5), 1053–1075. https://doi.org/10.2307/1864376

Scott, J. W. (2010). Gender: Still a Useful Category of Analysis? *Diogenes*, *57*(1), 7–14. https://doi.org/10.1177/0392192110369316

Sherman, J., Morrison, R., Klein, L., & Rosner, D. (2024). The Power of Absence: Thinking with Archival Theory in Algorithmic Design. *Designing Interactive Systems Conference*, 214–223.

https://doi.org/10.1145/3643834.3660690

Sitzia, E. (2023). Multiple Narratives and Polyvocality as Strategies of Inclusive Public Participation: Challenges and Disruption in the History Museum. *Muséologies. Les Cahiers d'études Supérieures*, *10*(2), 51–63. https://doi.org/10.7202/1108037ar

Søgaard, A., Plank, B., & Hovy, D. (2014). Selection bias, label bias, and bias in ground truth. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Tutorial Abstracts*, 11–13. https://aclanthology.org/C14-3005.pdf

Stoler, A. L. (2010). *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*. Princeton University Press.

Suresh, H., & Guttag, J. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–9. https://doi.org/10.1145/3465416.3483305

Thylstrup, N. B. (2019). *The politics of mass digitization*. MIT Press.

Trouillot, M.-R. (1995). *Silencing the Past: Power and the Production of History*. Beacon Press.

van Rossum, M. (2019). Labouring Transformations of Amphibious Monsters: Exploring Early Modern Globalization, Diversity, and Shifting Clusters of Labour Relations in the Context of the Dutch East India Company (1600–1800). *International Review of Social History*, *64*(S27), 19–42. Cambridge Core. https://doi.org/10.1017/S0020859019000014