Cambridge Forum on AI: Culture and Society (2026), **2**, e7, 1–9
doi:10.1017/cfc.2026.10014

Cambridge Forum

REFLECTION

Slicing the past to predict the future: Recasting data slicing as curatorial work in ML development and evaluation

Anna Schjøtt

Department of Media Studies, University of Amsterdam, Amsterdam, The Netherlands
Email: a.s.hansen@uva.nl

(Received 2 April 2025; revised 15 August 2025; accepted 1 December 2025)

Abstract

Data slicing is an inherent practice in machine learning (ML) evaluation, where different subsets of a dataset are used for training and evaluating ML systems. Drawing on ethnographic fieldwork conducted between September 2023 and February 2024 among the data scientists who develop ML-driven recommender systems for the British Broadcasting Corporation (BBC), this reflection piece highlights the important, yet often overlooked, ML practice of data slicing. Building on archival influences in Critical Dataset Studies and scholarship on critical curation, the paper proposes recasting data slicing as a curatorial practice. This shift makes visible the often highly tacit slicing practices undertaken by data scientists when working with ML datasets. Specifically, the paper traces three central considerations regarding data slicing: replicability, representativeness and generalisability. Using these as examples, the paper reflects on the implications of the selection, organisation, and choices of how to best represent the past to predict the future preferences of audiences. By engaging with data slices as curatorial constructs, we can better understand and intervene in their material politics by making visible how their orderings convey certain meanings more effectively than others. Through this approach, the paper expands on existing work within Critical Dataset Studies by identifying the political role of data slicing in the evaluation of ML systems.

Keywords: critical data studies; curatorial practices; data slicing; machine learning (ML); politics

1. Introduction

To initially ground this reflection piece, I want to return to a specific moment in my fieldwork with the British Broadcasting Corporation's (BBC) Personalisation Team.[1] This team develops and deploys

[1]The fieldwork was conducted as part of a research project focusing on responsible decision-making in AI development processes at the BBC, which was guided by questions of "*What forms of knowledge are prioritised at different stages of the development process of AI systems and who is invited to participate in the decision-making processes?*" and "*When are development decisions opened up for discussion with other BBC teams and when are they taken for granted and why is this the case at these exact moments?*". These questions resulted in a central focus on the knowledge devices used in the process of development, such as visualisation devices, evaluation metrics and the data slicing practices explored in this reflection piece. The fieldwork was conducted hybridly with online and physical participation in the team's activities from September 2023 to February 2024. The hybrid approach was chosen as the team is located across BBC offices, so their work is inherently hybrid. I would be physically present – mainly in the London offices – once a month, when the team would have joined activities and otherwise participate via MS Teams in their ongoing meetings. Beyond observations, I conducted 20 follow-up interviews with the editors, product

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/cfc.2026.10014 Published online by Cambridge University Press

machine learning ML-based recommender systems[2] to personalise content distribution across BBC News, Sounds, and iPlayer – the BBC platforms that provide online news and on-demand video and audio content. The moment I wish to highlight was not particularly significant, in that it did not change the trajectory of a project or spark controversy. In fact, it was a somewhat comical moment. However, it made me aware of the significance of a specific ML practice I had not encountered before, in my fieldwork or in the literature.

During the fieldwork at the BBC, a newly formed group of data scientists was working on what they called "pan-BBC recommendations" – a new type of recommendations aimed at recommending content across BBC News, Sounds and iPlayer. At a meeting in early December 2023, the team's principal data scientist, Maxine,[3] presented their initial results to editorial colleagues from BBC iPlayer. As a first experiment, the team had used users' listening history on BBC Sounds to generate recommendations for content on iPlayer, Maxine explained. These results were visualised in a mock interface that displayed the recommendations in rails (i.e., rows of content), mimicking the online interface. She noted: "Essentially, what we are looking at is a month's worth of user interaction on Sounds from October, and then we have recommendations based on the content available on iPlayer on 31st of October" (Quote Maxine, fieldnotes). While the purpose of the meeting was to narrow down a tentative list of what they called "business rules" (post- and pre-processing steps to be added as filters to the recommendations),[4] the example highlighted something unintended. The choice of training data and the day of predictions led to a slightly "horrifying" mock interface. Due to the period's co-occurrence with Halloween, the recommendations featured many titles related to the paranormal, as well as some horror films and TV shows, such as "Ghost" and "Interview with a Vampire." To account for this overrepresentation of Halloween-themed content, it is necessary to understand the role of "data slicing" in ML development, which involves creating specific subsets of an ML dataset for training and evaluation. In this reflection paper, I wish to use this observation to reflect on the significant, yet often overlooked, role of data slicing in ML evaluation practices.

The BBC is an intriguing case, as it functions both as a broadcasting archive and an active Public Service Broadcaster (PSB).[5]While the on-demand content platforms do not provide direct access to the archival collections, they continually make parts of the collections accessible to the public, as the available content on these platforms is regularly updated, also in consideration of shifts in audiences' sensitivities. The recommender systems deployed on these platforms, therefore, play a crucial role in providing access to the past by predicting what content users might enjoy watching in the future.[6] The development of new recommenders at the BBC was generally motivated by editorial goals of

managers, engineers, data scientists, and managers who participated in these meetings. Furthermore, I conducted two workshops – one physical and one online – aimed at qualifying my initial findings and creating a reflective space for those involved to gain insights into their own practices. An overview of interviews and workshops can be found in supplementary materials.

[2]Recommender systems are ML algorithms (they can also be rule-based) that filter information based on contextual data according to different technical logics (Bobadilla et al. 2013). The most common logics in the media domain are *content filtering*, which uses item features to recommend other items similar to what the user has previously clicked on, and *collaborative filtering*, which uses similarities between users and items simultaneously to provide recommendations (Bobadilla et al. 2013).

[3]The names of those involved are pseudonymised to protect their identity and in accordance with the consent for their participation. The pseudonyms reflect the genders of those observed and interviewed. While data science teams often remain male-dominated, (Seaver, 2022), this is not the case at the BBC, where more than 50% of data scientists in the team are female. Additionally, many described having specifically chosen to work for the BBC, as they felt they were more positively contributing to society here, as opposed to purely commercial companies.

[4]Business rules have commonly been used in recommender systems for news distribution to make the recommenders more aligned with editorial principles. For example, to ensure the timeliness of content (see e.g., Møller 2024).

[5]The BBC is the UK's largest public service broadcaster (PSB). As a public service institution, the BBC is grounded in its public service mission, as set out by the "Royal Charter" (BBC, n.d.). Public service media operate differently from commercial media organisations, as their remit lies in fulfilling their public service role and serving the public interest with no governmental or economic influences (see e.g., Born 2018; Ferraro et al. 2024; Jones 2022).

[6]Providing better and easier access to archival collections has been a key motivation for using AI in audiovisual archives (Noordegraaf and Schjøtt 2025)

https://doi.org/10.1017/cfc.2026.10014 Published online by Cambridge University Press

showcasing the depth and breadth of the collection in support of the BBC's public service mission. During the development of new recommender systems, editors, curators and commissioners would continually assess the quality of recommendations to ensure they aligned with the BBC's values and delivered on the desired objectives. To evaluate the recommendations, the data scientist would produce visualisations[7] of future recommendations, as was also evident in the meeting described above. Yet these visualisations would rely on specific "slices" of historical user data, including what specific users had watched or listened to on the BBC platforms at that time. While ML systems are always grounded in the past, as they use historical data to identify patterns and predict the future (Ziedler 2024), how that past is "sliced" when evaluating ML systems can have significant cultural implications by shaping how the deployed systems, for example, strengthen or weaken cultural citizenship in society (Ferraro et al., 2024). Building on archival influences in Critical Dataset Studies (Thylstrup 2022), I draw on scholarship on critical curation to recast data slicing as a curatorial practice to critically engage with these practices and foreground the politics and implications of data slicing – particularly in the context of broadcasting archives, such as the BBC.

1.1. Critical dataset studies: from data composition to data slicing

Research within Critical Dataset Studies has already drawn parallels between the production of ML datasets and archival practices (Crawford and Paglen 2021; Gebru et al. 2021; Jo and Gebru 2020; Pipkin 2020; Thylstrup 2022). By reframing ML dataset production as sociocultural data collection, Jo and Gebru (2020), for example, highlighted five lessons from archival practices that could inform a more responsible approach to this work, including consent, power, inclusivity, transparency, ethics and privacy. Later, Thylstrup (2022) expanded on this work, emphasising how archival considerations can inform an expanded approach to data deletion in ML datasets that better capture the ethical and political issues at stake in such practices. In this piece, I further extend this archival parallel to data slicing practices.

As an emerging field, Critical Dataset Studies has offered insights into how "datasets are collected, organised, distributed and deployed" (Thylstrup 2022, 659). These investigations have taken more artistic forms (Crawford and Paglen 2019; Pipkin 2020), used historiographic methods to study the content of datasets to discern their values and assumptions (Denton et al. 2021; Scheuerman et al. 2019) or been grounded in ethnography or interviews to explore the construction of datasets (see e.g., Engdahl 2024; Henriksen and Bechmann 2020; Jaton 2017; Orr and Crawford 2024). What unites these studies is their attention to how datasets are composed and annotated in practice, with the aim of uncovering the inherent politics of dataset production. In doing so, these studies foreground what is materialised (e.g., values, categorisations, etc.) in the production process but also what becomes invisible, such as the doubts and moral dilemmas of crowd workers or data scientists (Jaton 2021; Miceli et al. 2021). This form of politics is what Law and Mol (2008, 141) understand as material politics, which describes a "material ordering of the world in a way that contrasts this with other and equally possible alternative modes of ordering." Yet, the opening anecdote draws attention to how the material politics of datasets extend beyond their composition. The choice of how to slice the data equally orders the world in a specific way that significantly impacts what can be known and evaluated on the basis of that particular slice. Such orderings consequently shape whose or what interests are made negotiable, essentially, "who or what can count ethicopolitically" (Amoore 2020, 69), when deciding how to curate the past for the future.

Data slicing[8] is essential to ML evaluation, as different subsets of a dataset are always used to train and validate ML systems (often referred to as the test-train split). However, data slicing can also be

[7]In a separate paper, I explore the visual politics of these visualisation tools and how their epistemic qualities shape the development of recommender systems at the BBC and how core Public Service Media (PSM) and editorial values are represented.

[8]Data slicing is also often discussed in terms of "splitting" (see e.g., Nguyen et al. 2021). Here, I use slicing, as this was how it was discussed by the data scientists.

https://doi.org/10.1017/cfc.2026.10014 Published online by Cambridge University Press

employed to explore specific parts of the dataset or as a methodology for achieving higher accuracy by training the ML system on critical subsets of the dataset (see e.g., Ye 2021). In this context, slicing refers to splitting a dataset into two or more parts. These splits can be random based on percentages (e.g., 80/20) or targeted to evaluate critical or underrepresented parts of the dataset (Chung et al. 2019; Ovaisi et al. 2022; Zhang et al. 2023). However, how these splits are conducted can affect the overall performance of the ML system (Joseph and Vakayil 2022) and obscure its poorer performance on a subpopulation within the data (Chung et al. 2019; Ovaisi et al. 2022). Due to the increasing size of datasets, various tools now automate and scale data slicing (Liu et al. 2022). Such standardised techniques typically involve ensuring an equal and diverse distribution of classes within the dataset. However, these approaches have faced criticism for their arbitrariness and for overlooking outlier tendencies in the data (Chung et al. 2019). In response, statistical methods for more granular data slicing have been developed, enabling more interpretable ways to slice the data and address questions related to fairness in the data (Chung et al. 2019; Ovaisi et al. 2022; Zhang et al. 2023).

These developments illustrate a growing awareness of the politics and effects of data slicing, although the debates remain framed within a computer science, solution-oriented approach. In this paper, I advocate for social science and humanities scholars to join these efforts by providing critical analyses of data slicing that can better inform our understanding of the material politics and implications of these practices. In advancing this proposal, I remain within the archival analogy. In prior studies, data scientists were cast as *archivists* collecting, recording and archiving data (Jo and Gebru 2020). In this reflection, I propose recasting data scientists as *curators* when engaging in data slicing, as they decide how to slice and display the past when evaluating future predictions.[9] In doing so, I aim to highlight the agency of data scientists in selecting the slices and how these slices, by definition, become curatorial constructs that convey certain meanings more effectively than others. In the following, I situate this argument within the literature on critical curation before returning to my observations from the BBC.

2. Data slicing as curatorial work

Broadly speaking, a curator is someone who selects and organises materials in a display to tell a specific story (Hansen et al. 2019; Imaz-Sheinbaum 2024; Obrist 2015). As a practice, curation is often associated with galleries and museums. However, as archives are increasingly showcased in exhibits (Lester 2022; Williamson 2013), it is crucial to understand how the past is curated in such exhibitions and the effects of different curatorial strategies (Lester 2022). Hansen et al. (2019, 4) argue that scholars should seek to understand, "What sort of knowledge is produced in and through curatorial strategies?" The critical analysis of curatorial practices follows a broader trend in archival studies that challenges the understanding of archives and their exhibits as neutral sites of knowledge (Lester 2022; Michelle et al. 2022). Here, scholars increasingly examine how archivists and the archive as an institution impart "particular meaning to the past which, in turn, refutes or silences other perspectives" (Lester 2022, 28).

In the context of data slicing as part of ML practices at the BBC, there is no public exhibit. However, the choice of data slices establishes the basis for evaluating the ML system, which shapes how the system will later distribute the BBC's vast collection to its audience. These data slices, consequently, create a selective view of the past that influences the meaning that can be inferred from them. To critically engage with the data scientists' considerations around data slicing at the BBC, I draw on the work of Imaz-Sheinbaum (2024), who similarly recasts historians as curators to illustrate how historians, in their selection of evidentiary materials, actively curate the past to narrate a specific story of historical events. By reframing historiography as a curatorial practice, she emphasises three interrelated activities that shape this work: "selection, organisation, and choice" (Imaz-Sheinbaum 2024, 191). *Selection*

[9] Archivists also engage in curatorial practices as part of their work; however, here I make the distinction to make the point clearer and to signal a focus on the curation of displays meant for others.

https://doi.org/10.1017/cfc.2026.10014 Published online by Cambridge University Press

refers to the activity of choosing the materials to be exhibited, which is often guided by the specific concept or frame surrounding the exhibit. *Organisation*, in turn, describes the subsequent activities of how materials are positioned in relation to one another and presented to direct attention (Imaz-Sheinbaum 2024, 191–192). Finally, *choice* articulates the navigational space of the curator in these decisions. Can they freely decide how to frame the exhibition, or are these choices constrained by, for instance, the materials or the space (Imaz-Sheinbaum 2024, 197–198)? In what follows, I use these three interrelated activities to reflect on and discuss three interrelated considerations that the data scientist had regarding the data used to evaluate the recommenders at the BBC. These considerations are specific to recommender systems, where data slices are typically temporal constructs composed of historical audience data collected through continuous, live audience analytics (Ovaisi et al. 2022). Yet, such reflections could be generative in understanding data slicing practices across ML applications and domains.

2.1 Considering replicability, representativeness and generalisability

During my fieldwork at the BBC, I observed the use of various data slices for evaluation purposes and conducted interviews with data scientists to understand their considerations surrounding such slicing practices. Through this work, I found that the data scientists tended to focus on three main qualities when slicing the data: replicability, representativeness and generalisability. The first quality, *replicability*, relates closely to the classic scientific ambition of replicating experiments to validate their findings (Pinch 1993). The data scientists at the BBC explained that they often reused specific data slices across different experiments to directly compare results. As one data scientist, Emily, explained:

> Around maybe almost two years ago. (…) We generated one slice that is very close to what production data looks like. (…) So, every time we launch a new experiment, we use the static data. We want it to basically be comparable with the past experiments, but this data is now very stale, as it was a long time ago. So now we are working on a process where we can basically regenerate the same version of data, but in a different time frame every time we launch a new experiment (Interview, Emily).

Replicability was critical for understanding and comparing different ML models to assess their performance. However, since some ML models had been deployed years earlier, replicability could, at times, undermine another quality of a data slice, as the quote also reveals. Specifically, replicability could challenge the *representativeness* of the data slice. The BBC's collections are constantly changing and evolving, with new content being commissioned and published. Therefore, it was important that the data slices were temporally representative of the current BBC, as represented in the available collection across various on-demand platforms. However, this temporal representativeness was perceived as quite fleeting. Returning to the initial meeting in the opening anecdote, held in early December, here the data from October was already considered "slightly out of date," according to the principal data scientist (Quote from Maxine, fieldnotes). The issue with "old" data slices was that they hindered the editors' ability to relate the results to the current collection on offer, which affected how well the editors could interpret the results, evaluate the quality of the recommendations, and identify any potentially problematic recommendations that required mitigation, for example, by adding "business rules."

The "Halloween dataset," as the data slice from October had been informally dubbed, posed another challenge. This data slice was considered an outlier due to its overrepresentation of seasonal content, which meant it lacked *generalisability*. Another data scientist, Tom, explained that even certain days of the week could exhibit biased viewing patterns, such as Sundays, and that seasons like Christmas and Halloween could introduce bias, as evidenced by the increased number of horror movies during Halloween (Interview, Tom). These biases meant that the recommender's performance could not be generalised to other months of the year. Tom went on to argue that relying on a specific

https://doi.org/10.1017/cfc.2026.10014 Published online by Cambridge University Press

weekday when generating the predictions could even skew the results. Over time, they aimed to eliminate this weekday bias by aggregating the consumption of users from the 7 days following the training data period (Interview, Tom).

From these examples, it becomes possible to untangle how the *selections* of slices were generally grounded in data science concerns regarding the validity and comparability of the results. Hartley and Thylstrup (2024) have similarly demonstrated that, in the context of journalistic news production, the ethical considerations surrounding the construction of a recommender system were primarily grounded in data science epistemologies rather than journalistic ones. However, the data scientists at the BBC did consider how these qualities of the data slices affected their interpretability and, consequently, the editors' ability to assess the results. This aspect is important because the data slices were generally *organised* in a similar manner, specifically in a visualisation tool that provided a mock interface of how the recommendations would look in the "real" interface. This visualisation tool displays what a particular user has listened to or watched in the last 30 days (i.e., what was included in the training data) and subsequently shows the recommendations produced by one or more recommender systems based on the user's consumption. The selection of data slices, therefore, significantly influenced what could be inferred from the visualisation, which remained the same. As a curatorial construct, the data slices and their visual organisation offered a specific view of the past that would inform future predictions.

The *choice* of how to slice the data primarily belonged to the data scientists. However, a technical manager, John, mentioned in an interview that there was a desire to explore alternative ways of slicing and visualising the data – especially in dialogue with their editorial colleagues.

> (…) How about if we slice it by age or something (…), it feels like there should be a dialogue between the people who can develop the tools and interrogate the data and the people who ultimately make a decision about it (Interview, John).

The slicing itself would not be difficult, he explained. However, visualising it would require reallocating resources to develop new visualisation tools that could meaningfully capture these slices. Such developments, however, were not currently a priority due to other organisational demands for technical resources. Consequently, the existing organisation of the data slices had an implicit effect on the initial decisions regarding how to slice the data in the first place.

3. Concluding: caring for and beyond ML data

The term "curator" originates from the Latin word "curare," which means to care for or attend to something (Hansen et al. 2019; Imaz-Sheinbaum 2024).[10] This origin reflects the responsibility that curators have for the materials they care for – a sense of care that I have also observed among the data scientists I have encountered at the BBC and beyond. A sentiment that also aligns with ideas around ethics of care in both archival practices and ML (see e.g., Asaro 2019; Caswell and Cifor 2016; Gray and Witt 2021; Thylstrup 2022). Recasting data scientists as curators encapsulates this sense of care while also enabling us to critically engage with the material politics of data slicing, namely the ordering of who and what can be cared for in the data slices – and ultimately what is left out.

When slicing the ML data for evaluation, the BBC's data scientists predominantly cared about ensuring the quality and validity of the slices according to established data science practices. Yet they also considered the interpretive needs of the editors who depend on these slices to provide their input during the development process. However, as the examples demonstrate, the current constraints on how the data slices could be organised (i.e., visualised), along with the slicing itself, influenced which relationships could be cared for when evaluating the slices. Among the professionals working on recommender systems at the BBC, there was a strong shared sense of responsibility towards the audience,

[10]For a longer history of the origin and changing role of curatorship, see Obrist (2015).

https://doi.org/10.1017/cfc.2026.10014 Published online by Cambridge University Press

grounded in public service values. As part of this responsibility, there was an increased emphasis on ensuring that, for example, minority audiences (e.g., specific geographies in the UK or racial and gender minorities) felt seen and represented by the BBC.[11] Specific normative aims, therefore, guided the development of recommender systems, but these aims were not explicitly reflected in the selection and organisation of data slices. The editors, curators and commissioners would continuously discuss how well the visualised recommendations represented the interests of these groups. Yet, they would have to infer these considerations from the material orderings offered by the chosen slices and visualisations, which did not make such relations directly visible. Consequently, the selection, organisation, and choice of the curators (i.e., data scientists) when slicing the data materially shaped what normative relations could be more directly cared for, and which ones would be more loosely inferred or entirely left out of sight. These very mundane and often invisible decisions of how to slice the data could, therefore, have a significant impact on the cultural representation on the BBC's platforms.

To conclude, this short reflection piece aimed to illustrate how recasting data slicing as a curatorial work allows for critical engagement with the often highly tacit practices undertaken by data scientists when working with ML datasets. By actively considering the selection, organisation and choices that govern data slicing in ML evaluation, it becomes possible to understand the values and assumptions that implicitly shape these practices, as well as what considerations or relations are excluded and why. In doing so, I also expand on existing work in critical dataset studies by foregrounding the material politics of data slicing in the evaluation of ML systems. My hope is that this short reflection can serve as a starting point for critical engagement with the curatorial choices, challenges, and implications of data slicing across various ML applications and domains. Even more broadly, the reflections on data slicing as a curatorial practice may enable further considerations of how other ML practices can benefit from established debates in cultural heritage (see also Bunz, 2024), and, vice versa, how such practices can help us better understand curatorial practices, as these too become entangled in ML systems.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/cfc.2026.10014.

Acknowledgements. I would like to first and foremost thank the data scientists and the entirety of the Personalisation Team at the BBC, who made me curious about data slicing and generously shared their considerations and experiences with me. Without you, there would have been nothing to reflect on in the first place. Second, I wish to acknowledge the role Rhianne Jones and Natali Helberger played in facilitating this project and making it a fruitful experience. Third, thank you to Katie Mackinnon and Nanna Thylstrup for encouraging me to write this piece and to the rest of the editorial team on the AI & Archives issue, and to Nadja Schaetz and Tobias Blanke for thinking along with me on this reflection piece. Fourth and finally, thank you to the two anonymous reviewers for your constructive feedback, valuable suggestions and ideas for the future.

Funding statement. This research was supported by AI4Media, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911.

Open access funding provided by the University of Amsterdam.

Competing interests. No competing interests.

References

Amoore, Louise. 2020. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham and London: Duke University Press.

Asaro, Peter M. 2019. "AI Ethics in Predictive Policing: From Models of Threat to an Ethics of Care." *IEEE Technology and Society Magazine* 38 (2): 40–53. https://doi.org/10.1109/MTS.2019.2915154.

BBC. n.d. 'Charter and Agreement.' Accessed 22 March 2025. https://www.bbc.co.uk/aboutthebbc/governance/charter/.

Bobadilla, J., F. Ortega, A. Hernando, and A. Gutiérrez. 2013. "Recommender Systems Survey." *Knowledge-Based Systems* 46 (July): 109–132. https://doi.org/10.1016/j.knosys.2013.03.012.

[11]This focus on previously underrepresented groups is also a focus in archival research that focuses, for example, on community archives (see e.g., Caswell et al. 2017; Michelle et al. 2022).

Born, Georgina. 2018. "Taking the Principles of Public Service Media into the Digital Ecology." *A Future for Public Service Television*, edited by Des Freedman and Vana Goblot. Goldsmiths Press. London. https://doi.org/10.7551/mitpress/9781906897710.003.0025

Bunz, Mercedes. 2024. "The Role of Culture in the Intelligence of AI." In *AI in Museums: Reflections, Perspectives, and Applications*, edited by Sonja Thiel and Johannes C. Bernhardt. Transcript Verlag. Bielefeld.

Caswell, Michelle, and Marika Cifor. 2016. "From Human Rights to Feminist Ethics: Radical Empathy in the Archives." *Archivaria* 81, (May 6): 23–43.

Caswell, Michelle, Alda Allina Migoni, Noah Geraci, and Marika Cifor. 2017. ""To Be Able to Imagine Otherwise": Community Archives and the Importance of Representation." *Archives and Records* 38 (1): 5–26. https://doi.org/10.1080/23257962.2016.1260445.

Chung, Yeounoh, Tim Kraska, Neoklis Polyzotis, Ki Hyun Tae, and Steven Euijong Whang. 2019. "Slice Finder: Automated Data Slicing for Model Validation". *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, April, 1550–1553. https://doi.org/10.1109/ICDE.2019.00139.

Crawford, Kate, and Trevor Paglen. 2019. "Excavating AI." *Excavating AI The Politics of Images in Machine Learning Training Sets*. https://excavating.ai.

Crawford, Kate, and Trevor Paglen. 2021. "Excavating AI: The Politics of Images in Machine Learning Training Sets." *AI and Society* 36 (4): 1105–1116. https://doi.org/10.1007/s00146-021-01162-8.

Denton, Emily, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole. 2021. "On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet." *Big Data and Society* 8 (2): 20539517211035955. https://doi.org/10.1177/20539517211035955.

Engdahl, Isak. 2024. "Agreements "In the Wild": Standards and Alignment in Machine Learning Benchmark Dataset Construction." *Big Data and Society* 11 (2): 20539517241242457. https://doi.org/10.1177/20539517241242457.

Ferraro, Andres, Gustavo Ferreira, Fernando Diaz, and Georgina Born. 2024. "Measuring Commonality in Recommendation of Cultural Content to Strengthen Cultural Citizenship." *ACM Transactions on Recommender Systems* 2 (1): 1–32. https://doi.org/10.1145/3643138.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, J W. Vaughan, H. Wallach, H D. Iii, and K. Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92. https://doi.org/10.1145/3458723.

Gray, Joanne, and Alice Witt. 2021. 'A Feminist Data Ethics of Care for Machine Learning: The What, Why, Who and How.' *First Monday*. https://firstmonday.org/ojs/index.php/fm/article/view/11833.

Hansen, Malene Vest, Anne Folke Henningsen, and Anne Gregersen, eds. 2019. *Curatorial Challenges: Interdisciplinary Perspectives on Contemporary Curating*. London: Routledge. https://doi.org/10.4324/9781351174503.

Hartley, Jannie Møller, and Nanna Bonde Thylstrup. 2024. "The Algorithmic Gut Feeling – Articulating Journalistic Doxa and Emerging Epistemic Frictions in AI-Driven Data Work." *Digital Journalism* 0 (0): 1–20. https://doi.org/10.1080/21670811.2024.2319641.

Henriksen, Anne, and Anja Bechmann. 2020. "Building Truths in AI: Making Predictive Algorithms Doable in Healthcare." *Information, Communication & Society* 23 (6): 802–816. https://doi.org/10.1080/1369118X.2020.1751866.

Imaz-Sheinbaum, Mariana. 2024. "Curating the Past." *Journal of the Philosophy of History* 18 (2): 189–201. https://doi.org/10.1163/18722636-12341526.

Jaton, Florian. 2017. "We Get the Algorithms of Our Ground Truths: Designing Referential Databases in Digital Image Processing." *Social Studies of Science* 47 (6): 811–840. https://doi.org/10.1177/0306312717730428.

Jaton, Florian. 2021. "Assessing Biases, Relaxing Moralism: On Ground-Truthing Practices in Machine Learning Design and Application." *Big Data and Society* 8 (1): 20539517211013569. https://doi.org/10.1177/20539517211013569.

Jo, Eun Seo, and Timnit Gebru. 2020. 'Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning'. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (New York, NY, USA), FAT*' 20, January 27, 306–316. https://doi.org/10.1145/3351095.3372829.

Jones, Elliot. 2022. *Inform, Educate, Entertain… and Recommend? Exploring the Use and Ethics of Recommendation Systems in Public Service Media*. Ada Lovelace Institute. https://www.adalovelaceinstitute.org/report/inform-educate-entertain-recommend/.

Joseph, V. Roshan, and Akhil Vakayil. 2022. "SPlit: An Optimal Method for Data Splitting." *Technometrics* 64 (2): 166–176. https://doi.org/10.1080/00401706.2021.1921037.

Law, John, and Annemarie Mol. 2008. "Globalisation in Practice: On the Politics of Boiling Pigswill." *Geoforum, Environmental Economic Geography* 39 (1): 133–143. https://doi.org/10.1016/j.geoforum.2006.08.010.

Lester, Peter. 2022. *Exhibiting the Archive: Space, Encounter, and Experience*. London: Routledge. https://doi.org/10.4324/9781003159193.

Liu, Zifan, Evan Rosen, and Paul Suganthan. 2022. 'AutoSlicer: Scalable Automated Data Slicing for ML Model Analysis'. arXiv:2212.09032. Preprint, arXiv, December 18. https://doi.org/10.48550/arXiv.2212.09032.

Miceli, Milagros, Tianling Yang, Laurens Naudts, Martin Schuessler, Diana Serbanescu, and Alex Hanna. 2021. 'Documenting Computer Vision Datasets: An Invitation to Reflexive Data Practices'. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, March 3, 161–172. https://doi.org/10.1145/3442188.3445880.

https://doi.org/10.1017/cfc.2026.10014 Published online by Cambridge University Press

Michelle, Caswell, Ricardo Punzalan, and T.-Kay Sangwand. 2022. "Critical Archival Studies: An Introduction." *Journal of Critical Library and Information Studies* 1 (2). https://doi.org/10.24242/jclis.v1i2.50.

Møller, Lynge Asbjørn. 2024. "Designing Algorithmic Editors: How Newspapers Embed and Encode Journalistic Values into News Recommender Systems." *Digital Journalism* 12 (7): 926–44. https://doi.org/10.1080/21670811.2023.2215832.

Nguyen, Quang Hung, Ly Hai-Bang, Lanh Si Ho, N. Al-Ansari, H V. Le, V Q. Tran, I. Prakash, and B T. Pham. 2021. "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil." *Mathematical Problems in Engineering* 2021 (1): 4832864. https://doi.org/10.1155/2021/4832864.

Noordegraaf, Julia, and Anna Schjøtt. 2025. "From Preservation to Access and Beyond: The Role of AI in Audio-Visual Archives." In *Navigating Artificial Intelligence for Cultural Heritage Organisations*, edited by Lise Jaillant, Claire Warwick, Paul Gooding, Katherine Aske, Glen Layne-Worthey and J. Stephen Downie, 93-112. London: UCL Press.

Obrist, Hans Ulrich. 2015. *Ways of Curating*. London: Penguin Books.

Orr, Will, and Kate Crawford. 2024. "The Social Construction of Datasets: On the Practices, Processes, and Challenges of Dataset Creation for Machine Learning." *New Media & Society* 26 (9): 4955–4972. https://doi.org/10.1177/14614448241251797.

Ovaisi, Zohreh, Shelby Heinecke, Li Jia, Yongfeng Zhang, Elena Zheleva, and Caiming Xiong. 2022. 'RGRecSys: A Toolkit for Robustness Evaluation of Recommender Systems'. *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, February 11, 1597–600. https://doi.org/10.1145/3488560.3502192.

Pinch, Trevor. 1993. "Testing - One, Two, Three Testing!": Toward a Sociology of Testing." *Science, Technology, & Human Values* 18 (1): 25–41. https://doi.org/10.1177/016224399301800103.

Pipkin, Everest. 2020. 'On Lacework: Watching an Entire Machine-Learning Dataset'. *The Photographers' Gallery: Unthinking Photography*, July 20. https://unthinking.photography/articles/on-lacework.

Scheuerman, Morgan Klaus, Jacob M. Paul, and Jed R. Brubaker. 2019. "How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis Services." *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 144:1–144:33. https://doi.org/10.1145/3359246.

Seaver, Nick. 2022. *Computing Taste: Algorithms and the Makers of Music Recommendation*. University of Chicago Press. Chicago, IL.

Thylstrup, Nanna Bonde. 2022. "The Ethics and Politics of Data Sets in the Age of Machine Learning: Deleting Traces and Encountering Remains." *Media, Culture & Society* 44 (4): 655–671. https://doi.org/10.1177/01634437211060226.

Williamson, Ashley. 2013. "The Archive on Display: Issues of Curating Performance Remains." *Canadian Theatre Review* 156 (156): 24–29. https://doi.org/10.3138/ctr.156.005.

Ye, Ziyuan. 2021. 'A Data Slicing Method to Improve Machine Learning Model Accuracy in Bankruptcy Prediction'. *Proceedings of the 2021 5th International Conference on Deep Learning Technologies* (New York, NY, USA), ICDLT' 21, November 12, 32–39. https://doi.org/10.1145/3480001.3480008.

Zhang, Xiaoyu, Jorge Piazentin Ono, Huan Song, Liang Gou, Ma Kwan-Liu, and Liu Ren. 2023. "SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation." *IEEE Transactions on Visualization and Computer Graphics* 29 (1): 842–852. https://doi.org/10.1109/TVCG.2022.3209465.

Ziedler, Sarah. 2024. 'One Step Forward, Two Steps Back: Why Artificial Intelligence Is Currently Mainly Predicting the Past.' *HIIG*, October 15. https://www.hiig.de/en/why-ai-is-currently-mainly-predicting-the-past/.

Anna Schjøtt is a technological anthropologist and PhD Candidate in the Media Studies Department at the University of Amsterdam. In her PhD research, she ethnographically explores how knowledge around Artificial Intelligence (AI) is being produced and intervened upon across different media-related sites, including industry-wide initiatives supporting AI in journalism, in-house development of AI at the BBC and within the Media Museum in Amsterdam. In doing so, she aims to critically examine the politics of knowledge production and its implications for the media sector. She is a member of the AI, Media & Democracy Lab and the Cultural AI Lab. Furthermore, she co-organises the online Critical AI Studies Seminar and is currently co-editing the topical collection on the Politics of Machine Learning Evaluation in Digital Society.

Cite this article: Schjøtt, A. 2026. Slicing the past to predict the future: Recasting data slicing as curatorial work in ML development and evaluation. *Cambridge Forum on AI: Culture and Society*, **2**, e7, 1–9. https://doi.org/10.1017/cfc.2026.10014

https://doi.org/10.1017/cfc.2026.10014 Published online by Cambridge University Press