Research Ideas and Outcomes 9: e113334
doi: 10.3897/rio.9.e113334

OPEN ACCESS

Research Article

Developing responsible AI practices at the Smithsonian Institution

Rebecca B Dikow[‡], Corey DiPietro[§], Michael G Trizna[‡], Hanna BredenbeckCorp[§], Madeline G Bursell[|], Jenna T B Ekwealor[¶], Richard G J Hodel[#], Nilda Lopez[¤], William J B Mattingly[‡], Jeremy Munro[«], Richard M Naples[¤], Candace Oubre[»], Drew Robarge[§], Sara Snyder[ˆ], Jennifer L Spillane[‡], Melinda Jane Tomerlin[ˇ], Luis J Villanueva[¦], Alexander E White[‡]

‡ Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC, United States of America
§ National Museum of American History, Smithsonian Institution, Washington, DC, United States of America
| Bioinformatics Research Center, North Carolina State University, Raleigh, NC, United States of America
¶ Department of Biology, San Francisco State University, San Francicso, CA, United States of America
# National Museum of Natural History, Smithsonian Institution, Washington, DC, United States of America
¤ Smithsonian Libraries and Archives, Smithsonian Institution, Washington, DC, United States of America
« National Air and Space Museum, Smithsonian Institution, Washington, DC, United States of America
» National Museum of African American History and Culture, Smithsonian Institution, Washington, DC, United States of America
ˆ Office of Digital Transformation, Smithsonian Institution, Washington, DC, United States of America
ˇ National Museum of Asian Art, Smithsonian Institution, Washington, DC, United States of America
¦ Digitization Program Office, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC, United States of America

Corresponding author: Corey DiPietro (dipietroc@si.edu)

Reviewed v 1

Academic editor: Editorial Secretary

Received: 26 Sep 2023 | Accepted: 11 Oct 2023 | Published: 25 Oct 2023

Citation: Dikow RB, DiPietro C, Trizna MG, BredenbeckCorp H, Bursell MG, Ekwealor JTB, Hodel RGJ, Lopez N, Mattingly WJB, Munro J, Naples RM, Oubre C, Robarge D, Snyder S, Spillane JL, Tomerlin MJ, Villanueva LJ, White AE (2023) Developing responsible AI practices at the Smithsonian Institution. Research Ideas and Outcomes 9: e113334. https://doi.org/10.3897/rio.9.e113334

Abstract

Applications of artificial intelligence (AI) and machine learning (ML) have become pervasive in our everyday lives. These applications range from the mundane (asking ChatGPT to write a thank you note) to high-end science (predicting future weather patterns in the face of climate change), but, because they rely on human-generated or mediated data, they also have the potential to perpetuate systemic oppression and racism. For museums and other cultural heritage institutions, there is great interest in automating the kinds of applications at which AI and ML can excel, for example, tasks in computer vision including image segmentation, object recognition (labelling or identifying objects in an image) and natural language processing (e.g. named-entity recognition, topic modelling, generation of word and sentence embeddings) in order to make digital collections and archives discoverable, searchable and appropriately tagged.

A coalition of staff, Fellows and interns working in digital spaces at the Smithsonian Institution, who are either engaged with research using AI or ML tools or working closely with digital data in other ways, came together to discuss the promise and potential perils of applying AI and ML at scale and this work results from those conversations. Here, we present the process that has led to the development of an AI Values Statement and an implementation plan, including the release of datasets with accompanying documentation to enable these data to be used with improved context and reproducibility (dataset cards). We plan to continue releasing dataset cards and for AI and ML applications, model cards, in order to enable informed usage of Smithsonian data and research products.

Keywords

artificial intelligence, machine learning, GLAM, galleries, libraries, archives, museums, collections

Introduction

The Smithsonian Institution is the world's largest museum, education and research complex. It includes 21 museums, eight research centres, 15 archival repositories, 21 specialised library branches and a zoo. The collection holdings contain approximately 157.2 million objects and specimens, 148.2 thousand archival cubic feet (4.2 thousand cubic metres) and 2.3 million library volumes (Fiscal Year 2022; https://www.si.edu/dashboard/national-collections). In terms of digitised collections, 37 million objects and specimens have a digital record, 7.5 million objects and specimens have a digital image, 113 thousand archival cubic feet (3.2 thousand cubic metres) with a digital record and 27.2 thousand archival cubic feet (770 cubic metres) with a digital image. A total of 1.5 million library volumes have a digital record and 59.7 thousand library volumes have a digital image. The Digitization Program Office, part of the Smithsonian Office of the Chief Information Officer (OCIO), has been instrumental in collaborating with museums across the Smithsonian to investigate how to most efficiently digitise data and to do so at scale for representative projects. The Smithsonian Open Access Initiative (Smithsonian Institution 2019) was launched in February 2019 and, as of September 2023, there are more than 4.5 million 2D and 3D digital items online licensed as Creative Commons Zero (CC0) for public use. This number will increase as more collections are digitised and as collections that have already been digitised are designated CC0.

Museum collections, libraries and archives contain many kinds of information resources. These include:

- Printed material (including books, correspondence, diaries, journals, posters, manuscripts, pamphlets, journals, newspapers, maps) and their digital surrogates, as images or text transcriptions;
- Physical objects (including specimens, artifacts, photographs, artworks) and their digital surrogates;
- Sound, video and film recordings, which sometimes have associated transcripts;
- Electronic databases;
- 3D scans;
- Microforms;
- eBooks and eJournals.

All of these resources can also have associated metadata, some of which are generated automatically during digitisation (e.g. EXIF metadata for digital photographs), while others are added manually by content experts in many different roles, including but not limited to archivists, cataloguers, librarians, collections information specialists, data managers and curators.

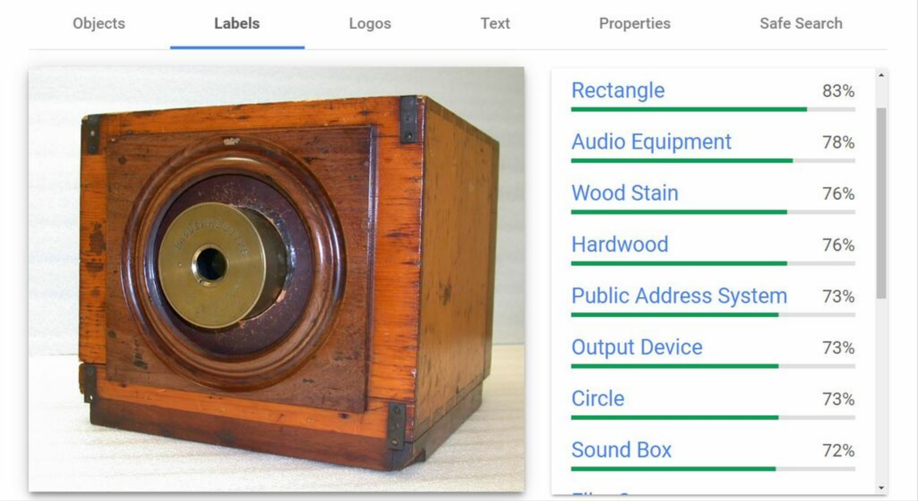
The opportunity to introduce AI and ML tools into parts of these workflows is appealing for many reasons. What rises to the top during conversations with knowledge workers at Galleries, Libraries, Archives and Museums (GLAM institutions) both large and small is the desire to make more of the collections available to the public, in a way that enables education, discovery and research. This desire is tempered with the recognition that there will never be enough staff or funding to enable this digital transformation because existing workflows do not allow for the massive scale required for the size of the collections. While there have been a number of reviews and experiments on the use of AI in GLAM institutions (e.g. Padilla 2019Cordell 2020, Murphy and Villaespesa 2020Lee 2022, Borowiec et al. 2022), the surge of interest in AI following the release of ChatGPT (OpenAI 2022) highlights the need for more detailed considerations of AI guardrails, best practices and lessons learned at GLAM institutions.

One of the most promising uses for AI in GLAM institutions is to improve accessibility. Ensuring that digitised content and data are accessible to all users, whether by adding alt-text to images (alternative text that describes the function or appearance of an image), text transcriptions to audio and video or translations into multiple languages, is crucial and, in many cases, required (e.g. WCAG Web Content Accessibility Guidelines; W3C Web Accessibility Initiative (2018)), but difficult for most institutions to achieve using existing workflows given current staffing levels. When GLAM institutions like the Smithsonian were closed during the COVID-19 pandemic, dramatic increases in virtual visitation and viewing of online collections further emphasised the need for institutions to move their data online and make it accessible. The Smithsonian was closed to the public from March 2020 to July 2020, then again from November 2020 to January 2021 due to the Omicron variant surge.

The rapid increase in the pace of digitisation has been met with a dramatic increase in the availability and usability of pre-trained models for diverse machine-learning tasks. While there is an eagerness to test these models on Smithsonian collections, there are many reasons why these pre-trained models might not work well on data associated with GLAM institutions or, if applied broadly across collections, could produce outputs that are misleading or even harmful. Off-the-shelf computer vision models, for example, have been trained with image datasets that often include images with labels, which may be outdated, offensive or inaccurate (e.g. Lipton et al. 2018Huang et al. 2019Northcutt et al. 2021). The ImageNet dataset (Deng et al. 2009) is still used to benchmark image classification models and has been shown to include biased, racist and offensive labels (Crawford and Paglen 2021, Denton et al. 2021). While the researchers who created ImageNet have acknowledged this and are working to provide corrective action (Yang et al. 2019), it is an ongoing challenge.

These benchmark computer vision training datasets also are not representative of the kinds of collections held by GLAM institutions since the types of objects present in these training data are those for which many image examples can be found online and labelled by non-experts. Many of the collections held by GLAM institutions are historical objects, rare, unique or would require other nuanced labelling. In early experiments done at the Smithsonian with applying commercial computer vision models to digitised collections objects from the National Museum of American History, we found examples where the model was simply inaccurate (e.g. a photo of a Morse Daguerreotype camera processed by Google Vision was classified with high probability as a sound box; Fig. 1) and more severe examples where inaccurate labels have the potential for harming people (e.g. a photo of prop shackles worn by LeVar Burton as Kunta Kinte in *Roots* processed by Google Vision was classified with high probability as jewellery; Fig. 2). The harm to visitors or communities that could result from making that label public re-emphasises the absolute necessity for a manual review and release process when applying any ML labelling model. For the daguerreotype example, a subject matter expert is still needed to assess outcomes. There is also a risk that, without a domain expert, instances of harm could potentially close the door to an organisation using AI or ML altogether. In addition to human review of labels to be made public, AI generated labels or other metadata should be identified as such in metadata fields that are separate from the fields that contain metadata generated by humans.

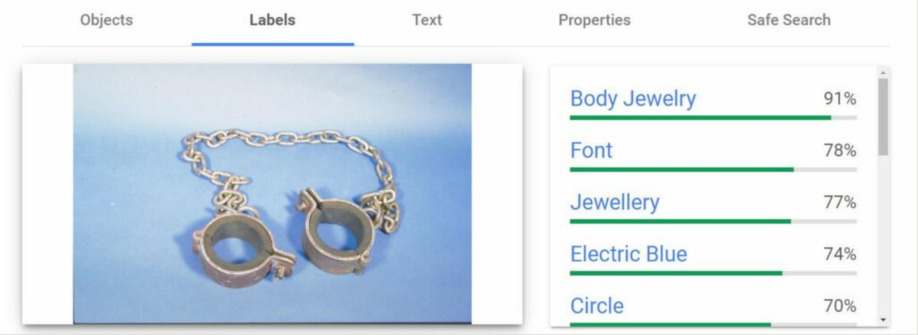
There is a robust body of scholarship around the topics of bias in AI and the disproportionate harm it can cause to people of colour (e.g. Caliskan et al. 2017, Buolamwini and Gebru 2018). “Big Tech” has had a tenuous relationship with these scholars, which was particularly noticeable when Timnit Gebru, co-leader of the Google AI Ethics team, was fired over a paper she co-authored, detailing the biases, risks and costs of large language models (Bender et al. 2021). While much of this work has highlighted the harm that specific AI algorithms can cause (Bolukbasi et al. 2016, Buolamwini and Gebru 2018), there has been an increasing focus on the datasets used to train these models; how they are gathered, curated, labelled (including impacts on human labour) and how errors and racist and biased labels on data are replicated and perpetuated (Denton et al. 2020).



Rectangle 83%
Audio Equipment 78%
Wood Stain 76%
Hardwood 76%
Public Address System 73%
Output Device 73%
Circle 73%
Sound Box 72%

Figure 1. doi

A photo of a Morse Daguerreotype camera. When processed by Google Vision, it was classified with high probability as a sound box. Source: https://n2t.net/ark:/65665/ng49ca746a6-6a45-704b-e053-15f76fa0b4fa. Date accessed: 22-06-2023.



Body Jewelry 91%
Font 78%
Jewellery 77%
Electric Blue 74%
Circle 70%

Figure 2. doi

A photo of prop shackles worn by LeVar Burton as Kunta Kinte in *Roots*. When processed by Google Vision, it was classified with high probability as jewellery. Source: https://n2t.net/ark:/65665/ng49ca746a9-d072-704b-e053-15f76fa0b4fa. Date accessed: 22-06-2023.

Methods to remediate or at least document these risks and biases have also been proposed. Data statements (Bender and Friedman 2018, McMillan-Major et al. 2023), model cards (Mitchell et al. 2019) and datasheets (Gebru et al. 2021) are all focused on better documentation of data and methods, making any work resulting from their use more transparent and reproducible. Algorithmic audits (Raji and Buolamwini 2019, Raji and Buolamwini 2022) can provide an additional layer of analysis after an AI system is developed to review system outputs and inspect documentation. Raji and colleagues also identify an additional layer beyond a technical audit and wrote that, "an AI system can be found technically reliable and functional through a traditional engineering quality assurance pipeline without meeting declared ethical expectations. A separate governance structure is necessary for the evaluation of these systems for ethical compliance. This evaluation can be embedded in the established quality assurance workflow, but serves a different purpose, evaluating and optimising for a different goal centred on social benefits and values rather than typical performance metrics such as accuracy or profit" (Raji et al. 2020). The environmental costs of training AI models is another area of active research (e.g. Schwartz et al. (2019)). As much of science increasingly relies on high-performance computing, consideration of the environmental and climate costs of implementing new technologies should be part of the project planning process.

Attempts at regulation and policy by government entities have lagged significantly behind the academic literature. Just in the past few years, however, the U.S. Executive Branch has convened AI experts from academia and industry and released multiple policy recommendations and proposed actions. Developed under Alondra Nelson's leadership at the Office of Science and Technology Policy, a *Blueprint for an AI Bill of Rights* was released in October 2022, which states, "The Blueprint for an AI Bill of Rights is an exercise in envisioning a future where the American public is protected from the potential harms and can fully enjoy the benefits, of automated systems" (Office of Science and Technology Policy 2022). In 2023, an additional Fact Sheet announcing new actions to promote responsible AI innovation was released (The White House 2023). The National Institute of Standards and Technology (NIST) also released their *AI Risk Management Framework* (RMF; National Institute of Standards and Technology 2023) and more recently launched the Trustworthy and Responsible AI Resource Center, which will facilitate implementation of, and international alignment with, the AI RMF. The Center for Security and Emerging Technology released a "Matrix for Selecting Responsible AI Frameworks" (Narayanan and Schoeberl 2023) and the Department of Defense (Defense Innovation Unit, Department of Defense 2022) released Responsible AI Guidelines. These federal guidelines are a useful reference point for the Smithsonian, but do not capture the fullness of our scope. The Smithsonian's educational and research activities as well as collections and data are generally much more diverse than those at other federal organisations.

Material and methods

In 2021, during COVID pandemic restrictions on in-person gatherings and meetings, a virtual AI and ML focused reading group was formed at the Smithsonian. In order to promote the broadest participation possible, we chose readings that were of interest not only to Smithsonian staff already building or implementing AI in their work, but also to staff from across the Institution who interact with data in all ways. We all came to quick agreement that focusing on data was important because the data themselves play such a central role in the downstream application of any computational tools. Due to the expansive footprint of the Smithsonian, the subject-matter expertise of staff is extremely broad. One constant challenge is finding ways to break down silos that form as staff work in their organisational "unit" (museum, department, research centre etc.) – sharing knowledge across units and even departments can be challenging. That was one impetus for the formation of the reading group – many are interested in similar topics, but it can be difficult to find time to stop what we are doing to talk to each other. The books that were chosen in this initial phase were: *Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence* by Kate Crawford (Crawford 2021), *Race After Technology: Abolitionist Tools for the New Jim Code* by Ruha Benjamin (Benjamin 2019) and *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* by Michael Kearns and Aaron Roth (Kearns and Roth 2019).

Almost immediately during the reading group conversations, we realised that we needed a document that could serve as best practices or guardrails, as AI and ML applications become more prevalent. Not having this in place was hindering our ability to work together on these topics across our distributed organisation. Our first goal was to draft an “AI Values Statement” with feedback from Smithsonian staff and affiliates with diverse expertise. We saw this as purposefully distinct from an official Smithsonian policy around the use of AI, which we felt and still feel, would be difficult to draft and implement while technologies are changing so rapidly and there may not be a one-size-fits-all solution to go across all Smithsonian data types and units. Indeed, during the months between drafting our Values Statement and compiling the community feedback, ChatGPT was released and added a new dimension to this work that could not be easily fit into our existing language. We found inspiration from the Stanford Special Collections and University Archives Statement on Potentially Harmful Language in Cataloguing and Archival Description (Stanford Special Collections and University Archives 2020).

We also felt the need to walk a bit of a tightrope; guidance should not stifle creativity and innovation in piloting experimental tools, but instead should empower potential users and consumers of AI tools and outputs to understand the potential risks and how to navigate the process when deciding whether and how to implement AI. We hope it can be a guide for users to ask appropriate questions before entering into a new project or partnership using AI tools. Particularly as AI tools have become part of applications we are already using (e.g. photo editing software, search engines, machine-generated transcription for video and audio, autocomplete in document and email programmes, computer vision weapons detection in security systems), there is really no avoiding this technology becoming a part of existing workflows, but that does not mean we cannot make choices about which tools we use and how we use them.

In order to begin to implement the recommendations from the Values Statement, we chose an initial handful of Smithsonian datasets on which pilot Dataset Cards and used the Dataset Card template from HuggingFace (Hugging Face 2023b) as a starting point. The AI in GLAM community seems to be coalescing around HuggingFace (Hugging Face 2023a) as the place to host models and datasets. The datasets we chose for this initial release of dataset cards are not meant to be representative of all Smithsonian data or content types, but they do span natural science, history and culture. From the National Museum of Natural History, we wrote dataset cards for both the digitised bumblebees and the digitised herbarium. We also created dataset cards for the National Museum of American History's Phyllis Diller Gag file, as well as presently-transcribed documents from the National Museum of African American History and Culture's Freedman's Bureau

Archive (digital surrogates stewarded by the National Museum of African American History and Culture and previously available on the 1918 rolls of microfilm held by the National Archives and Records Administration).

Data resources

All dataset cards have been posted to GitHub (https://github.com/smithsonian/dataset-cards) and archived at Zenodo (https://zenodo.org/doi/10.5281/zenodo.8381116).

Results and Discussion

During community discussions, we thought it was important to distinguish between two main tracks of work at the Smithsonian which use or may use AI tools, which we refer to as "research" and "strategic" tracks. The research track has been vibrant at the Smithsonian since 2017, particularly for digitised natural history datasets. For these projects, Smithsonian researchers have generally built custom convolutional neural networks (most recently using transfer learning on open-source models, for example, ResNet) or natural language processing pipelines, which were trained or fine-tuned using Smithsonian content. It is in our best interests to make sure that methods, techniques and lessons learned during the course of these research projects are discussed and shared. Some examples of these research projects include:

- A segmentation model to identify plant pixels on digitised herbarium sheets (White et al. 2020).
- A classification model to identify mercury staining on digitised herbarium specimens and to distinguish morphologically-similar families of fern allies (Schuettpelz et al. 2017).
- A combined segmentation and classification model to identify genera of Amazonian fish from images to provide conservation-related taxonomic baselines (Robillard et al. 2023).
- A model to classify and measure morphological variation in digitised herbarium specimens from the plum genus (a collaboration between OCIO and NMNH).
- A Natural Language Processing pipeline to extract named entities and the pronouns surrounding them in context (led by the OCIO Data Science Lab).
- A model to classify DNA sequence reads as host or contaminant for marine organisms using NLP techniques (led by the OCIO Data Science Lab).
- A classification model to identify species of ferns and fern allies and subsequently assess morphospace compared to species richness (a collaboration between OCIO and NMNH).
- AstroAI, new centre dedicated to the development of artificial intelligence to enable next generation astrophysics at the Center for Astrophysics (Harvard and Smithsonian).

At a strategic level, we are still trying to determine where AI can have the most impact and improve the efficiency of current collection workflows and practices to the greatest extent. While AI applications to research do indeed provide efficiencies (e.g. a person would not be able to measure leaves on 4.5 million herbarium sheets), the goal is often not focused solely on efficiency, but on new ways of capturing data or features to generate scientific insights (e.g. analysing total plant shape as opposed to restricting to a handful of traditional measurements). The strategic applications may rely more heavily on commercial models and, thus, may require more scrutiny after implementation to identify inaccuracies or harmful outputs.

The Smithsonian AI Values Statement is below and is also posted at https://datascience.si.edu/ai-values-statement. Fig. 3 summarises the actions proposed in our AI Values Statement, which are also detailed below.

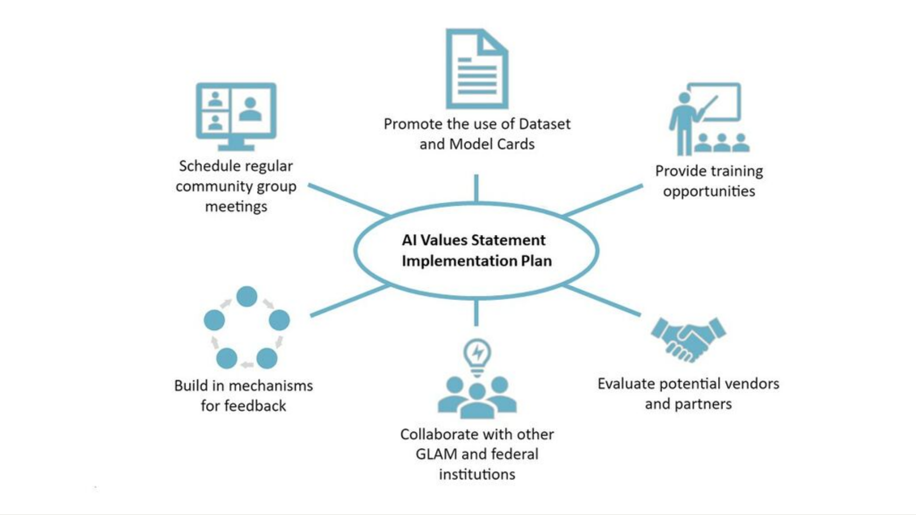

Promote the use of Dataset and Model Cards
Schedule regular community group meetings
Provide training opportunities
AI Values Statement Implementation Plan
Build in mechanisms for feedback
Evaluate potential vendors and partners
Collaborate with other GLAM and federal institutions

Figure 3. doi

Summary of the major actions proposed for implementing the AI Values Statement.

Engage internal community: In order to maintain open lines of communication across AI and ML and data practitioners across all Smithsonian units, we plan to continue our reading group as well as institute regular AI community meetings. At these meetings, community members can present projects using or building AI tools either in planning or implementation phases to receive feedback from other community members. We also see the opportunity to connect practitioners from different units that may be using the same tools or working on developing methods with shared challenges. We see this as an informal way of keeping track of which technologies, vendors and methods community members are using.

In addition to these gatherings, which may organically draw more technically-focused staff and AI practitioners, we plan to share learning opportunities and resources in non-technical venues, both synchronously and asynchronously. We think it is important to ensure that communications include as broad a group of Smithsonian staff and affiliates as possible as AI touches all of us. By plugging into existing committees and standing working groups including higher-level Director's meetings, strategic teams focusing on data governance, digital practitioners, webmasters and web developers, to a series of presentations at unit all-staff meetings, we hope to bring topics like AI literacy and policy advocacy to all Smithsonian staff. Plugging into and leveraging existing networks can help compliment and grow the existing community.

Promote the use of Dataset and Model Cards: The Smithsonian Open Access Initiative has made millions of digitised objects and records available for public use. Datasets made available by GLAM institutions can be difficult to use due to institution-specific metadata or cataloguing practices that are unclear to end-users. We also cannot easily anticipate all the ways these data will be used. Dataset Cards, which are human-readable README pages that contain general information about the data and how it should be used, can provide a way for institutional expertise, context and bias to be conveyed to end-users (and even our future selves). Our Dataset Cards completed to date are posted on GitHub (Smithsonian Institution 2023). Some items detailed on each Dataset Card include the original intent for gathering the dataset, its context, assumptions, changes to the data, normalisations, transformations that have occurred and explanation of known biases and social impact. Nevertheless, while there are clear advantages to creating and using Dataset Cards, it is also important to note the limitations of their applicable use. Table 1 details the advantages and disadvantages of Dataset Cards.

Table 1.

Dataset Cards advantages and disadvantages.

Dataset Card Advantages	Dataset Card Disadvantages
For discrete or "complete" datasets, these can provide comprehensive information for users of the data.	For broad datasets that span departments, answers to prompts may be too non-specific to be useful.
Provide context, awareness, cautionary information and potential risks for both the data content and data format.	For Smithsonian data, there is no real way to describe all Open Access data as a single set even though users may be interested in these data as a whole.
Gaps in collections scope and dataset biases (when known) can be identified and described up-front.	If the content of the card changes, it may be challenging to ensure users use the newer version.
Cards and their associated datasets and model cards can be integrated with HuggingFace and GitHub.	Incomplete or growing datasets may be more difficult to describe comprehensively and will require more extensive versioning.
Can be used for both "internal" as well as public-facing datasets, assisting both future staff and external users.	Datasets may be modified and manipulated into new versions depending on AI task or goal. Currently, there is no straightforward way to link to related datasets from a dataset card.

Provide training opportunities: The Smithsonian has a robust data science skills training programme coordinated by the OCIO Data Science Lab with more than 20 Smithsonian staff and fellows who have completed Carpentries (The Carpentries 2023) instructor training. The instructors volunteer their time to teach workshops on fundamental data science skills to Smithsonian staff and affiliates. More than 500 people have attended these workshops over the past five years. Recently, staff from the UK National Archives, the British Library and the Smithsonian OCIO Data Science Lab collaborated to develop an Intro to AI for GLAM Lesson within the Library Carpentry curriculum (Bell et al. 2021). The Intro to AI for GLAM Lesson is currently in the beta stage, but it was delivered by Data Science Lab members to a cross-discipline Smithsonian audience in autumn 2021. The Smithsonian OCIO Digitization Program Office also coordinates a Digital Foundations webinar series for Smithsonian staff and affiliates and sponsored a panel in May 2023 centred around the opportunities and risks of ChatGPT that almost 300 staff attended. Future training plans include both hands-on coding and technical deep-dives for AI practitioners, as well as AI and data literacy sessions, aimed at staff who may not build or implement AI models or systems, but may need to evaluate their outputs.

Build in mechanisms for feedback: The opportunity for AI to provide text descriptions, labels, captions and other enhanced metadata further enables more Smithsonian content to be shared online with the public. The increase in content also means that mechanisms must be in place to allow user feedback and suggestions for improving these machine-generated metadata. Both staff and the public should have opportunities to correct inaccuracies and flag terms that are outdated or harmful, particularly as vocabularies and language usage change over time or as objects develop new or different cultural relevance.

We also think it is crucial for any AI-generated content to be clearly labelled as such so that it is not confused with metadata that has been created by humans. While both machine-generated and human-generated metadata can be inaccurate and may need to change over time, it is important for users of the data to know how they were produced. Documentation of methods and model versions would be required of any scholarly research using these tools and we think it is as important when any AI methods are applied to data put online for public audiences.

Collaborate with other GLAM and federal institutions: For federal organisations, federal procurement regulations can sometimes limit our ability to be agile in the adoption of new technologies. In order to address such limitations, we have been developing collaborations with other federal organisations including the Library of Congress and the National Archives and Records Administration, as well as colleagues at Virginia Polytechnic Institute and State University, to discuss shared challenges and opportunities around the topic of AI. While policies will likely be institution-specific as our data as well as processes and organisational structure are all unique, we see great value in building on each other's progress. We hosted two workshops for staff from our institutions in 2022 and presented a panel discussion at the Joint Conference on Digital Libraries (Ingram et al. 2023). The Smithsonian is also represented on the Secretariat of AI4LAM (AI for Libraries, Archives and Museums), which sponsors monthly community calls and the annual Fantastic Futures conference. Participating in these collaborations and communities can help us keep up with emerging technologies, as well as providing valuable information about success stories or difficulties with implementation.

Evaluating vendors and partners: The use of contracts with outside vendors is common at many GLAM institutions and federal agencies, in particular for experimental work for which it would take a long time to develop the case for and dedicate funding to new staff positions. These institutions also often cannot match salaries in the for-profit sector, so contracting can be a mechanism to bring in expertise on a project-by-project basis. It can sometimes be difficult for institutions to evaluate vendor promises and for vendors to fully understand challenges before embarking on a project because GLAM data are often historical, messy and not uniform. The terms of the agreements signed by vendors and institutions may not detail the exact technology used when building or implementing an AI system and, in many cases, the training data, the model or the pipeline may be closed-source. How vendors will evaluate success may also vary from the metrics valued by the institution. This is particularly important given the status of our institutions as trusted sources. Inaccurate outputs can begin to erode public trust.

We expect the Values Statement to evolve over time, in response to changing technology, feedback from users and broadening of the types of data available for AI applications. We also foresee a time when a more structured governance framework is needed, as the number of users and use cases of AI applications grow. When the specifics of such a framework are considered, we hope that the steps put into place here lay the groundwork for a growing, vibrant, community of digital practitioners engaged in experimenting, evaluating and integrating new technologies into traditional collections practices.

Smithsonian AI Values Statement

Technology is not neutral.

The use of Artificial Intelligence (AI) tools[1] to describe, analyse, visualise or aid discovery of information from Smithsonian collections, libraries, archives and research data reflects the biases and positionality of the people and systems who built each tool, as well as those that collected, catalogued and described any data used for their training. These tools might hold extensive value in their use at the Smithsonian, but there are issues that will limit the applicability and reliability of their use due to the way they were planned and created.

We seek to only begin AI projects[2] that implement tools and algorithms that are respectful to the individuals and communities that are represented by the information in our museum, library and archival collections. We aim to be proactive in identifying and documenting biases and methodologies when building and implementing such tools and making the documentation available to audiences that will interact with the resulting products. We recognise that technology evolves over time and that our efforts must also evolve to ensure our ethical framework stays relevant and robust. We encourage any person, community or stakeholder involved with or affected by said tools and algorithms to provide feedback and point out any concerns.

We acknowledge the opportunities that AI tools present for cultural heritage organisations:

- As digitisation of museum, library and archival collections has become more prevalent, there is a need for tools to make digitised data available to our audiences.
- AI tools can be used to make museum, library and archival collections more discoverable to the public by efficiently extracting, summarising and visualising vast amounts of data.
- AI tools can help us become more representative of our audiences, through surfacing the histories of marginalised people and groups.

We urge anyone contemplating an AI project to consider:

- Is it the appropriate technology to solve the problem?
- The development of AI tools often requires the use of specialised computational hardware, the production of which relies on mining of rare earth metals and the operation of which can have a large carbon footprint. What is the environmental impact of choosing this technology or tool?
- There are no unbiased methodologies, datasets, collections, algorithms or tools. Therefore, what are the biases in the methodologies, datasets, collections, algorithms or tools you wish to use?

We strive to promote the following actions when implementing AI tools:

- Documentation of the biases in any methodologies, datasets, collections, algorithms or tools.
- Documentation of transparent data statements and that outline the intent of methodologies, datasets, collections, algorithms or tools.
- Creation of positionality statements of the creators of datasets or algorithms behind AI tools.
- Documentation of potential risks and regular updating of these risks as technology changes.
- Solicitation and inclusion of feedback from relevant members of the community.
- Documentation of how AI content was produced.
- Clear labelling of AI-generated content, so it is not confused with human-generated content.

We strive to recognise the following when implementing AI tools:

- Everyone at the Smithsonian is involved in data collection, creation, dissemination and analysis as a stakeholder.
- If any community or individual is harmed by the use of a technology, then that is one too many.

We strive to promote the following when partnering with outside organisations on AI tools or projects:

- We should seek projects and partnerships that adhere to our institutional values.
- We should not enter into contracts of collaborations with industry or other partners for the use of tools with unspecified or undisclosed methods and biases.
- We should require potential partners who create AI and machine-learning tools to explicitly evaluate and state if the datasets or data descriptions used in these tools were collected without consent or contain offensive or racist descriptions before we agree to use these tools.

[1] The term “AI tools” includes a variety of technologies that seek to create decision-making software. Some examples include facial and speech recognition, machine-learning based optical character recognition, language translation, natural language processing, image recognition, object detection and segmentation and data clustering. Common commercial examples include virtual assistants such as Siri or Alexa, website search and recommendation algorithms and tagging and identification of people in images on social media platforms.

[2] The term “AI project” refers to an intentional effort to utilise or create an AI tool in research or in an existing workflow.

Acknowledgements

We thank Deron Burba, Carmen Iannacone, Beth Stern and Diane Zorich for providing feedback on earlier versions of the AI Values Statement and this manuscript and Benjamin Charles Germain Lee and Lukas Hughes-Noehrer for their reviews of the preprint. We thank everyone who attended an AI reading group meeting and contributed to conversations about this topic. In addition to the authors, we also thank Smithsonian staff who contributed to dataset cards: Jessica Bird, Torsten Dikow, Sylvia Orli, Douglas Remley, Eric Schuettpelz and Kamilah Stinnett. RBD thanks Thomas Padilla, Liz Lorang, Harish Maringanti, Jefferson Bailey, Ryan Cordell, Abbey Potter, Meghan Ferriter, Jill Reilly, Bill Ingram, Sylvester Johnson, Laura Coyle and Courtney Bellizzi for discussions over the past few years that helped refine the ideas presented here. We dedicate this work to the memory of our colleague Effie Kapsalis, who led the Smithsonian Open Access Initiative.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Bell M, McGregor N, van Strien D, Trizna M (2021) Intro to AI for GLAM Library Carpentries Lesson. https://carpentries-incubator.github.io/machine-learning-librarians-archivists/
- Bender E, 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
- Bender E, Gebru T, McMillan-Major A, Shmitchell S (2021) On the Dangers of Stochastic Parrots. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency https://doi.org/10.1145/3442188.3445922
- Benjamin R (2019) *Race after technology: Abolitionist tools for the new Jim code*. Polity
- Bolukbasi T, Chang K, Zou J, Saligrama V, Kalai A (2016) Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. arXiv https://doi.org/10.48550/arxiv.1607.06520
- Borowiec M, Dikow R, Frandsen P, McKeeken A, Valentini G, White A (2022) Deep learning as a tool for ecology and evolution. Methods in Ecology and Evolution 13 (8): 1640-1660. https://doi.org/10.1111/2041-210x.13901
- Buolamwini J, Gebru T (2018) Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. Proceedings of Machine Learning Research 81: 77-91.
- Caliskan A, Bryson J, Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. Science 356 (6334): 183-186. https://doi.org/10.1126/science.aal4230
- Cordell R (2020) Machine learning and libraries: a report on the state of the field. Library of Congress. https://blogs.loc.gov/thesignal/2020/07/machine-learning-libraries-a-report-on-the-state-of-the-field/
- Crawford K (2021) Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press https://doi.org/10.12987/9780300252392-006
- Crawford K, Paglen T (2021) Excavating AI: the politics of images in machine learning training sets. AI & SOCIETY https://doi.org/10.1007/s00146-021-01162-8
- Defense Innovation Unit, Department of Defense (2022) Responsible AI Guidelines. https://www.diu.mil/responsible-ai-guidelines#:~:text=Responsible%20AI%20Guidelines%20in%20Practice&text=effectively%20examine%20test,%20and%20validate,across%20a%20variety%20of%20programs.
- Deng J, Dong W, Socher R, Li L, Kai Li, Li Fei-Fei (2009) ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition https://doi.org/10.1109/cvpr.2009.5206848
- Denton E, Hanna A, Amironesei R, Smart A, Nicole H, Scheuerman MK (2020) Bringing the People Back In: Contesting Benchmark Machine Learning Datasets. arXiv https://doi.org/10.48550/arxiv.2007.07399
- Denton E, Hanna A, Amironesei R, Smart A, Nicole H (2021) On the genealogy of machine learning datasets: A critical history of ImageNet. Big Data & Society 8 (2). https://doi.org/10.1177/20539517211035955

- Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, III HD, Crawford K (2021) Datasheets for datasets. Communications of the ACM 64 (12): 86-92. https://doi.org/10.1145/3458723
- Huang J, Qu L, Jia R, Zhao B (2019) O2U-Net: A Simple Noisy Label Detection Approach for Deep Neural Networks. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) https://doi.org/10.1109/iccv.2019.00342
- Hugging Face (2023a) Hugging Face:The AI Community Building the Future. https://huggingface.co
- Hugging Face (2023b) Dataset Card Creation Guide. https://github.com/huggingface/datasets/blob/main/templates/README_guide.md
- Ingram W, Dikow R, Potter A, Ferriter M (2023) AI and Public Archives: Collaborative Leadership for Responsible Adoption. ACM/IEEE Joint Conference on Digital Libraries. https://doi.org/10.1109/JCDL57899.2023.00079
- Kearns M, Roth A (2019) The Ethical Algorithm: The Science of Socially Aware Algorithm Design. Oxford University Press
- Lee BCG (2022) The "Collections as ML Data" Checklist for Machine Learning & Cultural Heritage. arXiv https://doi.org/10.48550/arxiv.2207.02960
- Lipton Z, Wang Y, Smola A (2018) Detecting and Correcting for Label Shift with Black Box Predictors. arXiv https://doi.org/10.48550/arxiv.1802.03916
- McMillan-Major A, Bender E, Friedman B (2023) Data Statements: From Technical Concept to Community Practice. ACM Journal on Responsible Computing https://doi.org/10.1145/3594737
- Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji ID, Gebru T (2019) Model Cards for Model Reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency https://doi.org/10.1145/3287560.3287596
- Murphy O, Villaespesa E (2020) AI: A Museum Planning Toolkit. https://themuseumsainetwork.files.wordpress.com/2020/02/20190317_museums-and-ai-toolkit_rl_web.pdf
- Narayanan M, Schoeberl C (2023) A Matrix for Selecting Responsible AI Frameworks. Center for Security and Emerging Technology. https://cset.georgetown.edu/wp-content/uploads/CSET-A-Matrix-for-Selecting-Responsible-AI-Frameworks.pdf
- National Institute of Standards and Technology (2023) AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework
- Northcutt C, Athalye A, Mueller J (2021) Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks. arXiv https://doi.org/10.48550/arxiv.2103.14749
- Office of Science and Technology Policy (2022) Blueprint for an AI Bill of Rights. https://www.whitehouse.gov/ostp/ai-bill-of-rights
- OpenAI (2022) ChatGPT. URL: https://openai.com/blog/chatgpt
- Padilla T (2019) Responsible Operations: Data Science, Machine Learning, and AI in Libraries. OCLC Research https://doi.org/10.25333/xk7z-9g97
- Raji ID, Buolamwini J (2019) Actionable Auditing. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society https://doi.org/10.1145/3306618.3314244
- Raji ID, Smart A, White R, Mitchell M, Gebru T, Hutchinson B, Smith-Loud J, Theron D, Barnes P (2020) Closing the AI accountability gap. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency https://doi.org/10.1145/3351095.3372873
- Raji ID, Buolamwini J (2022) Actionable Auditing Revisited. Communications of the ACM 66 (1): 101-108. https://doi.org/10.1145/3571151

- Robillard A, Trizna M, Ruiz-Tafur M, Dávila Panduro EL, de Santana CD, White A, Dikow R, Deichmann J (2023) Application of a deep learning image classifier for identification of Amazonian fishes. Ecology and Evolution 13 (5). https://doi.org/10.1002/ece3.9987
- Schuettpelz E, Frandsen P, Dikow R, Brown A, Orli S, Peters M, Metallo A, Funk V, Dorr L (2017) Applications of deep convolutional neural networks to digitized natural history collections. Biodiversity Data Journal 5 https://doi.org/10.3897/bdj.5.e21139
- Schwartz R, Dodge J, Smith N, Etzioni O (2019) Green AI. arXiv https://doi.org/10.48550/arxiv.1907.10597
- Smithsonian Institution (2019) Smithsonian Open Access Initiative. https://www.si.edu/openaccess
- Smithsonian Institution (2023) Smithsonian Dataset Cards. www.github.com/Smithsonian/dataset-cards
- Stanford Special Collections and University Archives (2020) Statement on Potentially Harmful Language in Cataloging and Archival Description. https://drive.google.com/file/d/1-2U14_QKT3N8FnAgmPK5OYhEIoClZvDG/view
- The Carpentries (2023) https://www.carpentries.org
- The White House (2023) FACT SHEET: Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans' Rights and Safety. https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/
- W3C Web Accessibility Initiative (2018) Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/
- White A, Dikow R, Baugh M, Jenkins A, Frandsen P (2020) Generating segmentation masks of herbarium specimens and a data set for training segmentation models using deep learning. Applications in Plant Sciences 8 (6). https://doi.org/10.1002/aps3.11352
- Yang K, Qinami K, Fei-Fei L, Deng J, Russakovsky O (2019) Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy. arXiv https://doi.org/10.48550/arxiv.1912.07726