Power, Data and Control

AI in the Museum

Oonagh Murphy

Museums are increasingly engaging with emerging data-led technologies that utilize cutting edge artificial intelligence (AI) systems to help them engage with their collections and visitors in new ways. These technologies present opportunities to increase impact and understanding and develop new ways for museums to evidence their contemporary relevance and reassert their position as necessary public institutions. In a drive to be relevant, museums often rush to develop new digital products and services focussing on providing strategic priorities around collections, access, and visitor engagement. This chapter argues that museums should also engage in a broader conversation about the power and impact of these digital tools and products, and outlines three models or frameworks that can be used to help development teams think about the impact these technologies have on visitors, as well as on the data, archives, and collections that museums hold as public institutions. The focus of the discussion will thus be on museums as public institutions and their role in shaping how visitors experience digitally mediated worlds.

Narratives and Practice

We quite often think of digital innovation in connection with self-driving cars and robots, but when it comes to AI in particular, we find contradictory narratives of voice assistants and robot housekeepers helping improve our lives, alongside dystopian narratives of totalitarian robot states, where lives are controlled by faceless machines (Fry 2018). Both narratives serve as helpful outliers, a framing through which to examine how we want to use these technologies now and in the future. A utopian future in which all labour is carried out by machines, while we enjoy a constant life of leisure, may feel unlikely, and, at times, a dystopian fear around the challenges posed by these technologies can often feel more probable. The reality is and will continue to be more nuanced (Parry 2011; Wittkower 2017; Zuboff 2019). Museums serve as an important public forum to help develop what this imagined future may look like (Ippolito 2012; Peters 2012). But before we talk

about how museums might help to design our data future, it is perhaps useful at this point to pause and reflect on how these technologies already shape our everyday life experiences.

Many of us already use a range of AI technologies in our daily lives, from plotting journeys on Google Maps, receiving recommendations on Netflix, visually searching photos on our phones, and asking for help from Alexa, to adding a filter to our photos to look younger or fresher. These technologies feel far from the dystopian and utopian narratives of AI we often see discussed in the media. They are technologies that we are familiar with, that we are comfortable with. They do not feel challenging; if anything, these technologies feel friendly, are ubiquitous parts of our lives. 'This information infrastructure characterizes a particular mode of governance, one that is rooted in a political economy in which the prevailing logic is to predict and modify human behaviour as a means to produce revenue and market control' (Dencik/Hintz/Cable 2016, 1). User interfaces can, however, be deceiving, since behind these technologies are powerful companies powered by vast amounts of users' data, some of which is being used to improve our everyday lives, while other data is being used to limit, and control, our life opportunities and experiences (Amaro 2019a; 2019b).

On a microlevel, I increasingly find that these technologies seek to control the way I look and speak. If we take Gmail, for example: its predictive email writing encourages me to write in the Queen's English, to be more formal, to be more professional. It quashes my regional dialect, those colloquialisms that make me who I am. Every time I take a photo, I am prompted by my phone, and a variety of apps, to edit my image, add filters, make myself fit in, and conform to a mould of what beauty is. On the web, my search results have improved, meaning that I find what I am looking for more quickly, but this quite often also limits any serendipitous opportunities for discovery. It takes away the joy of discovering things in the digital archive, of stumbling across an unknown female scientist or a local hairdresser who does not pay to advertise (Noble 2018). As Lury and Day reflect on the limitations of algorithmic: 'personalization', 'the familiar recognition that personalization seems to provide—knowing you better than you yourself do—should not be considered merely a more precise form of individuation. To the contrary, personalization also constrains who and how we can be' (Lury/Day 2019, 19). What we see here is a form of efficiency: information is provided quickly, but this efficiency often has a negative impact on representation, creativity, and accuracy. Whilst the experience may be more efficient, the trade-off for an efficient life is less opportunity for discovery and a life viewed through the prism of those who program the machine.

AI in (Museum) Practice

A benchmarking study in 2019 and 2020 identified 120 projects using some form of AI technologies in museums, including computer vision, natural language processing, neural networks, robotics, predictive analytics, machine learning, and generative adversarial networks.[1] Whilst this study is not definitive, it does provide insights into the ways in which these technologies are being used in museums in the United Kingdom, the United States, and Europe (with some examples drawn from beyond this geographic scope). In addition to this benchmarking study, a series of associated toolkits published in English, German, and Spanish provide more detailed case studies of seven such projects in the UK, USA, Germany, and Spain from 2019 to 2021. The case studies feature projects from the National Gallery (UK), The Metropolitan Museum of Art (USA), the American Museum of Natural History (USA), the Badisches Landesmuseum (Germany), the Ludwig Forum Aachen (Germany), the ZKM (Germany), and the Museo Nacional Del Prado (Spain), and provide in depth insights into a cross-section of current practices in this area. These case studies can be used to highlight two bodies of practice that have emerged in how these technologies are being applied within a museum context (Murphy/Villaespesa 2020; 2022; Murphy et al. 2022).

The first body of practice might be defined as applications focussing on visitor experience and operations. For example, the American Museum of Natural History has piloted natural language processing (NPL) to analyse large amounts of visitor feedback data. While in the UK, the National Gallery has developed a predictive analytics dashboard to create more accurate visiting profiles so as to allow for more appropriate staffing levels and provide data that can create a business case for exhibition run times, gallery size, and even exhibition title. The ZKM (Germany) has created the 'intelligent.museum' in partnership with the Deutsches Museum Nuremberg. It seeks to use machine learning and natural language processing to support voice identification technologies in identifying a variety of spoken languages, with a view to creating responsive, language-specific exhibition content for visitors.

The second body of practice exists around digital collections, collections data, and collection management. For example, The Metropolitan Museum of Art (USA), has used computer vision tools to generate metadata for digital collections so that they become more searchable and discoverable. The Museo Nacional Del Prado (Spain) is using natural language processing as a mechanism for automatic reading and comprehension in order to enrich the scope of searchable contextual data

1 See the list of artificial intelligence initiatives in museums compiled by Elena Villaespesa, Oonagh Murphy, and Kate Nadel: https://docs.google.com/spreadsheets/d/1A7IVnucQZoICxYSOCjqQ1oV3xGgNzDKtIYGrk6smV7w/edit?usp=drive_web&ouid=102621042281518178063&usp=embed_facebook (all URLs here accessed in August 2023).

available to support discoveries in its collection. The Training the Archive project at the Ludwig Forum Aachen (Germany) seeks to use machine learning technologies to embed contextual collections data, with a focus on pattern recognition as a means to influence new approaches to curatorial thinking and practices. The Badisches Landesmuseum's (Germany) xCurator project is exploring how predictive analytics and natural language processing can be used to engage visitors with the entire collection, not only the objects on display, and looks at recommendation modelling and storytelling mechanisms to support discovery.

These case studies show the diversity of approaches currently being developed within a museum context. They also demonstrate the richness and diversity of the datasets being used to develop projects in this area, and it is the collection, processing, and generation of output data that is of critical significance from an ethical perspective. In a rush to remain relevant, digitize collections, optimize operations, meet strategic goals, and hit key performance indicators, it can be easy to forget the social purpose that underpins the museum as a social construct, the museum as a public institution. These early examples of AI practices in museums show that these technologies provide valuable new ways to develop collections and engage with audiences. As these technologies become more powerful and pervasive, it will, however, become increasingly important to understand the power they have in the museum context and beyond. It is my argument that museums should take a pause to think about the impact these technologies have on their visitors and users. Yes, AI might help us sell more tickets and to have more users engage with our digital collections, and might help us to run more efficient buildings. But taking a purely operational approach means missing out on a potentially greater calling, the strategic and curatorial vision needed to show the contemporary relevance of museums as a place where ideas, culture, and society are made, not simply displayed. Museums provide a unique space to have such civic conversations, since they are organizations driven by purpose not profit. They are underpinned by an international code of ethics set out by the International Council of Museums and, in addition, often by codes of ethics defined by national governing bodies (Sandahl 2019; Murphy/Villaespesa 2021).

Models of Ethical AI (Museum) Practices

The long-term success of AI applications across museum practice rests on the ability of museum professionals to adopt and adapt these technologies to be of service to the broader institutional and social mission of museums as public institutions. This requires behind-the-scenes work, the type of work that is not mentioned in press releases. Whilst it is often slow and complicated, it is this foundational work that creates inclusive, accessible, engaging, representative, and diverse digital products and

services. The most important time to have conversations about power, data, control, creativity, access, and representation is thus at the very start of the process. Facilitating these conversations can be challenging, but using existing development frameworks can provide the language, format, and skills necessary to embed these ethics-focussed considerations within the development of an AI project. Though many such models and frameworks are available, in this chapter I will introduce three that provide elements that align particularly well with the non-profit, purpose-driven remit of museums. The tools which I will discuss in turn are: 1. the Data Ethics Canvas, 2. Consequence Scanning, and 3. the Museums + AI Toolkit.

Figure 1: Data Ethics Canvas. Source: Open Data Institute.

The Data Ethics Canvas was developed by the Open Data Institute, a non-profit institute focussed on data and society established in London in 2012 by Tim Berners-Lee and Sir Nigel Shadbolt (fig. 1).[2] The Data Ethics Canvas is designed to provide a framework that can be applied to any context, whatever the project's size or scope. It is particularly useful for museum practitioners because it provides a framework and structured pathways for thinking about the broader ethical context of any data-led innovation, and therefore has benefits beyond AI-specific work, since it is

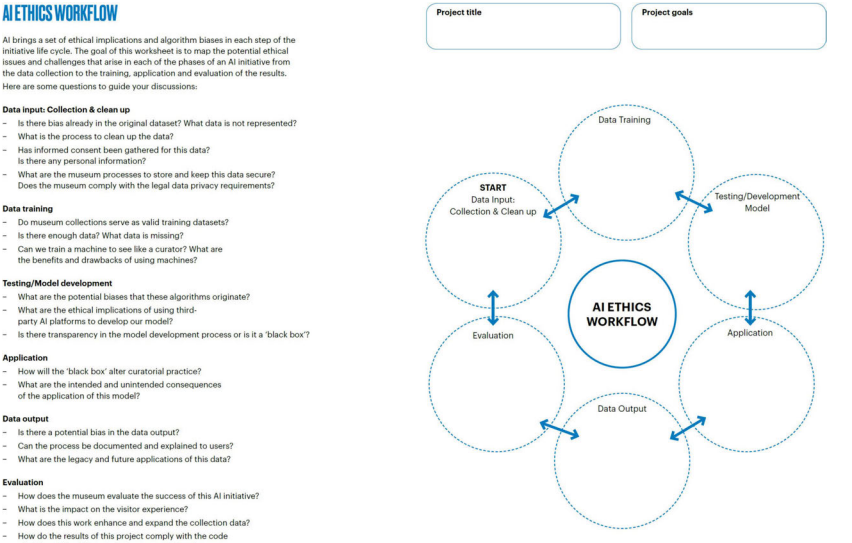
2 https://www.theodi.org/article/the-data-ethics-canvas-2021/.

applicable to all data-led projects. This framework is designed to be a multidepartmental strategic planning tool. Due to the scalability of scope within this framework it can support project teams in thinking beyond personal project goals, and situate project development within a wider operating context, whether of a department, museum, cultural policy, or broader sectoral strategy. The framework invites users to consider the implications of the data being collected, the data that will be processed, and any output data generated by the process. Using this canvas, we can start to develop more robust approaches to using digital technologies in museums. Whilst the micro-aspects of the framework are helpful in designing ethically robust data-led projects, the real strength comes from the supported pathway to the creation of a more macro-view of the impact such projects can have beyond a direct product or project delivery and encouraging users to position their work in the wider context of the institution, its values, and the broader impact the projects can have on society.

Figure 2: Introductory flow map outlining the Consequence Scanning method. Source: doteveryone.

Consequence Scanning is described by its creators as an 'agile practice that fits within an iterative development cycle' (fig. 2).[3] It is a way for organizations to consider the potential consequences of their product or service on people, communities, and the planet. This practice is an innovation tool that also provides an opportunity to mitigate or address potential harms or disasters before they happen. This framework was developed by the digital think tank doteveryone, and, like the Data Ethics Canvas, is an interdisciplinary tool that can be used in a commercial or non-profit context. Whilst such frameworks often seek to facilitate group thinking, a key feature of Consequence Scanning is dedicated space for 'quiet time', which invites participants to respond independently to design challenges and prompts. This model of ethics development provides opportunities for diverse voices and perspectives to be harnessed and embedded within the development of data-led projects. As the title suggests the focus of this framework is on mapping the consequences of a particular project being realized. Uniquely, however, it puts equal time and emphasis on mapping and anticipating the intended consequences of a project, product, or service as it does on unintended consequences. By identifying unintended consequences, we can problem solve before the problem becomes an issue.

Figure 3: Narrated AI Ethics Workflow guidance and worksheet. Source: The Museums + AI Network.



AI ETHICS WORKFLOW

AI brings a set of ethical implications and algorithm biases in each step of the initiative life cycle. The goal of this worksheet is to map the potential ethical issues and challenges that arise in each of the phases of an AI initiative from the data collection to the training, application and evaluation of the results.
Here are some questions to guide your discussions:

Data input: Collection & clean up
- Is there bias already in the original dataset? What data is not represented?
- What is the process to clean up the data?
- Has informed consent been gathered for this data? Is there any personal information?
- What are the museum processes to store and keep this data secure? Does the museum comply with the legal data privacy requirements?

Data training
- Do museum collections serve as valid training datasets?
- Is there enough data? What data is missing?
- Can we train a machine to see like a curator? What are the benefits and drawbacks of using machines?

Testing/Model development
- What are the potential biases that these algorithms originate?
- What are the ethical implications of using third-party AI platforms to develop our model?
- Is there transparency in the model development process or is it a 'black box'?

Application
- How will the 'black box' alter curatorial practice?
- What are the intended and unintended consequences of the application of this model?

Data output
- Is there a potential bias in the data output?
- Can the process be documented and explained to users?
- What are the legacy and future applications of this data?

Evaluation
- How does the museum evaluate the success of this AI initiative?
- What is the impact on the visitor experience?
- How does this work enhance and expand the collection data?
- How do the results of this project comply with the code of ethics of the different museum associations?

3 https://doteveryone.org.uk/project/consequence-scanning/.

AI: A Museum Planning Toolkit was developed by The Museums + AI Network and builds on the work of the Data Ethics Canvas and the Consequence Scanning framework to create a sector-specific model for the ethical development and deployment of AI technologies within the museum context (fig. 3).[4] Through workshops and public events, the network brought together a range of stakeholders including policymakers, funders, museum professionals, and visitors to examine current and future uses of AI within museums. Through these meetings, it was established that there were no ethics policies or practices accepted industry-wide when it comes to using artificial intelligence. Through iterative testing, the Data Ethics Canvas and the Consequence Scanning framework were deemed to be useful tools for beginning to think about ethics and data within a museum. When applied, they, however, lacked the nuance required for publicly funded, social purpose organizations such as museums, libraries, and archives. A series of three 'frameworks' with focuses on capabilities, ethics, and stakeholder management within the museum context were subsequently co-produced in cooperation with a range of stakeholders. To enable these worksheets to be used by a range of museum professionals in a wide variety of institutional contexts, they were published within a more extensive toolkit providing case studies, a glossary, and a strategic introduction. This enables the planning tool to be used by individuals from both technical and nontechnical backgrounds, as well as across museum departments. The toolkit has been published as open access (Murphy/Villaespesa 2020; 2022a; 2022b).

Conclusion

By being transparent about the technologies they are using and developing, museums can encourage members of the public, their visitors, and users to develop their digital literacy so that they can be more confident and, when necessary, critical of how they use digital technologies, as well as regarding the data they share in their everyday lives. While we think that the work we are now doing is helping to correct historical data biases that exist in museums, the likelihood is that, in 100 years people, will look back at the work being done today and be embarrassed or maybe even outraged at how racist, sexist—or perhaps a term of discrimination that has not even been defined yet—this work is. So, as much as I am advocating for a strategic engagement with issues regarding power, data, and control, with a particular focus on consequences and bias mitigation, I think it is perhaps a helpful reality check to remember that our goal is to do the best we can today, with all the information,

4 https://themuseumsainetwork.files.wordpress.com/2020/02/20190317_museums-and-ai-toolkit_rl_web.pdf.

knowledge, and experience that we as a sector currently have available to us, whilst knowing that tomorrow we may need to alter course.

References

Amaro, Ramon (2019a). AI and the Empirical Reality of a Radicalised Future. In: C. Wood (Ed.) AI—More than Human. London, Barbican Art Gallery, 125–27.

Amaro, Ramon (2019b). As if, e-flux. Available online at https://www.e-flux.com/architecture/becoming-digital/248073/as-if/ (all URLs here accessed in August 2023).

Dencik, Lina/Hintz, Arne/Cable, Jonathan (2016). Towards Data Justice? The Ambiguity of Anti-Surveillance Resistance in Political Activism. Big Data & Society 3 (2). https://doi.org/10.1177/2053951716679678.

Fry, Hannah (2018). Hello World: How to be Human in the Age of the Machine. London/New York/Toronto/Sidney/Auckland, Doubleday.

Ippolito, Jon (2012). Which Commons: Market, Zoo, or Tribe? In: Sarah Schultz (Ed.). Open Field: Conversations on the Commons. Minneapolis, Walker Art Center, 74–75. Available online at https://walkerart.org/magazine/which-commons-market-zoo-or-tribe.

Lury, Celia/Day, Sophie (2019). Algorithmic Personalization as a Mode of Individuation. Theory, Culture & Society 36 (2), 17–37. https://doi.org/10.1177/0263276418818888.

Murphy, Oonagh/Villaespesa, Elena (2020). AI: A Museum Planning Toolkit. London, Goldsmiths. Available online at https://themuseumsainetwork.files.wordpress.com/2020/02/20190317_museums-and-ai-toolkit_rl_web.pdf.

Murphy, Oonagh/Villaespesa, Elena (2021). Innovation, Data and Social Responsibility. In: H. Eid/M. Forstrom (Eds.). Museum Innovation: Building More Equitable, Relevant and Impactful Museums. London/New York, Routledge, 109–21. https://doi.org/10.4324/9781003038184-9.

Murphy, Oonagh/Villaespesa, Elena (2022a). Künstliche Intelligenz und Museen. Ein Toolkit. Translated and expanded by Johannes Bernhardt/Tabea Golgath/Sonja Thiel. London/Karlsruhe, Goldsmiths/Badisches Landesmuseum. Available online at https://www.landesmuseum.de/fileadmin/user_upload/Barrierefreie_PDFs/KI-und-Museen-Toolkit-web_UA.pdf.

Murphy, Oonagh/Villaespesa, Elena (2022b). La Red De Museos + Inteligencia Artificial. Translated by C. Moreu Lopez. London, Goldsmiths. Available online at file:///C:/Users/Johannes/Downloads/toolkit-la-red-de-museos-e-inteligencia-artificial.pdf.

Parry, Ross (2011). Transfer Protocols: Museum Codes and Ethics in the New Digital Environment. In: Janet Marstine (Ed.). Routledge Companion to Museum

Ethics: Redefining Ethics for the Twenty-First Century Museum. London, Routledge, 316–31. https://doi.org/10.4324/9780203815465

Peters, Sarah (2012). When Bad Things Don't Happen. In: Sarah Schultz (Ed.). Open Field: Conversations on the Commons. Minneapolis, Walker Art Center, 127–44. Available online at https://walkerart.org/magazine/open-field-when-bad-things-dont-happen.

Sandahl, Jette (2019). The Museum Definition as the Backbone of ICOM. Museum International 71 (1–2), vi–9. Available online at https://doi.org/10.1080/13500775.2019.1638019.

Wittkower, Dylan E. (2017). Technology and Discrimination. In: Joseph C. Pitt/Ashley Shew (Eds). Spaces for the Future: A Companion to Philosophy and Technology. London, Routledge, 37–64. https://doi.org/10.4324/9780203735657.

Zuboff, Shoshana (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. London, Profile Books.