



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

AI in the public eye

Investigating public AI literacy through AI art

Citation for published version:

Hemment, D, Currie, M, Bennett, SJ, Elwes, J, Ridler, A, Sindere, C, Vidmar, M, Hill, RL & Warner, H 2023, AI in the public eye: Investigating public AI literacy through AI art. in *FACCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery (ACM), New York, pp. 931-942, The 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, Illinois, United States, 12/06/23. <https://doi.org/10.1145/3593013.3594052>

Digital Object Identifier (DOI):

[10.1145/3593013.3594052](https://doi.org/10.1145/3593013.3594052)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

FACCT '23

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



AI in the Public Eye: Investigating Public AI Literacy Through AI Art

DREW HEMMENT, University of Edinburgh, United Kingdom

MORGAN CURRIE, University of Edinburgh, United Kingdom

SJ BENNETT, University of Edinburgh, United Kingdom

JAKE ELWES, Artist, United Kingdom

ANNA RIDLER, Artist, United Kingdom

CAROLINE SINDERS, Researcher and Artist, United Kingdom

MATJAZ VIDMAR, University of Edinburgh, United Kingdom

ROBIN HILL, University of Edinburgh, United Kingdom

HOLLY WARNER, Unaffiliated Scholar, United Kingdom

Recent advances in diffusion models and large language models have underpinned a new generation of powerful and accessible tools, and some of the most publicly visible applications are for artistic endeavour. Such tools, however, provide little scope for deeper understanding of AI systems, while the growing public interest in them can eclipse notice of the vibrant community of artists who have long worked with other forms of AI. We explore the potential for AI Art – particularly work in which AI is both tool and topic – to facilitate public AI literacies and consider how tactics developed before the current generative AI boom have continued relevance today. We look at the strategies of critical AI artists to scaffold public understanding of AI and enhance legibility for non-experts. This paper also investigates how collaborations between artists and AI researchers and designers can illuminate key technical and social issues relevant to the development of AI. The study entailed workshops between three professional artists who work with AI and a cross-disciplinary set of academic participants. This paper reports on these workshops and presents the intentions and strategies expressed by the artists, as well as insights of relevance to the research community on public AI literacies. We find that critical AI art can link underlying technical systems to structural issues of power and facilitate experiential learning that is situated and embodied, valuing interpretation over explanation. The findings also demonstrate the importance of transdisciplinary conversations around art, ethics and the political economy of AI technologies and how these dialogues may feed into AI design processes.

CCS Concepts: • **Applied computing** → **Media arts**; • **Computing methodologies** → **Artificial intelligence**.

Additional Key Words and Phrases: Society, Culture, AI literacy, Art, Computational Art, AI Art, Creative AI, Artificial Intelligence, Transparency, Education, Dialogue, Interdisciplinary Research

ACM Reference Format:

Drew Hemment, Morgan Currie, SJ Bennett, Jake Elwes, Anna Ridler, Caroline Sinders, Matjaz Vidmar, Robin Hill, and Holly Warner. 2023. AI in the Public Eye: Investigating Public AI Literacy Through AI Art. In *2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*, June 12–15, 2023, Chicago, IL, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3593013.3594052>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

FAccT '23, Chicago, USA.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0192-4/23/06...\$15.00

<https://doi.org/10.1145/3593013.3594052>

1 INTRODUCTION

Art has long offered a space for fostering new, critical perspectives on technological development. As cultural theorists, social scientists and philosophers have argued, art provides a distinct setting to discuss new, controversial technologies, whether genetic engineering or drones – art can make this technology more accessible, relatable and friendly or, paradoxically, frightening and dangerous [36][45][52]. In the current generation of artists working with AI, we define critical AI arts as artistic practices that explore these technologies as media in their own right while adopting a stance of critique or advocating for possible futures beyond current technological limitations [28]. By testing the creative limits of new technologies, and by creating public fora for discussing and interacting with them, the arts can provide opportunities to foster greater literacies around their promises and risks.

In this paper, we are concerned with the potential for artists to scaffold public understanding of Artificial Intelligence (AI) in their works. Scholars, designers and civil society organisations alike have argued that societies must understand how to foster public AI literacy, since public opinion in part determines how various AI applications are accepted and ultimately implemented [58][2]. Better critical engagement also helps users understand a technology's dependence on the context of its development and whether it can be designed more in alignment with social goals. Current approaches to AI literacy range from mathematical forms of explicability and statistical fairness to shedding light on broader social dimensions also shaping the design and use of these systems [69][44]. A question this study asks of relevance to the AI community is how art might offer alternative, novel approaches to developing public AI literacies – and what AI developers and scientists might learn from this. We ask how creative practice and tangible interactions with artworks contribute to greater legibility and critical comprehension of AI processes to address current limitations of more technical approaches to explainability and accountability in AI systems and thereby to wider concerns of fairness, accountability and transparency (FAccT).

This study looks at strategies of artists working with AI in the decade before the release of tools such as Midjourney, Stable Diffusion, Dall-E 2 and ChatGPT. The study entailed workshops between three professional artists who work with AI and a cross-disciplinary set of academic participants working in a University setting. The three workshops, held in July and August 2020, form the first phase

of a multi-year co-operative research enquiry carried out in collaborative partnership with the artists, where the research team supports the creative development of the artists and the artists participate in the research. The overarching research objective was to understand how AI can fuel significant cultural works and, conversely, how artistic practice can enrich or inform new paradigms for legible, inclusive and responsible AI [37]. Thus, beyond public engagement, this paper is also interested in how collaborations between artists and AI researchers and engineers can facilitate novel perspectives on AI design, contribute new interpretations of terms and illuminate key technical and social issues relevant to AI.

The guiding questions of this study were

- RQ1: What are the current intentions and strategies of the artists towards making artificial intelligence tangible and explicit and building critical public literacies about those systems?
- RQ2: How might artistic experiences that foster public AI literacies differ from current best practices in public understanding of AI?
- RQ3: Can knowledge-exchange between artists and AI designers and researchers help to sharpen meanings and expressions for artists and/or researchers, potentially facilitating conceptual leaps or reinterpretations of certain terms?

Our key findings are:

- (1) The intentions of critical AI artists include linking AI systems to structural issues of power, presenting edifying experiences for audiences, defamiliarising AI to cause people to think of it in fresh ways, and raising normative questions of what people should use AI for. [Q1]
- (2) Strategies of artists to make AI tangible and explicit include making visible design choices in the AI pipeline, exposing gaps in training data, revealing human labour and showing the key role of training datasets as they interact with algorithms. [Q1]
- (3) The arts offer an experiential approach to AI public literacies that engages people tangibly, emotionally and cognitively by enabling audiences to interact with an AI system or outputs, using spectacle to draw people in before provoking discussion and combining the aesthetic experience with wrap-around activity to deepen engagement. [Q2]
- (4) Arts practice can be further enriched by providing more accessibility to AI tools and creating multiple opportunities to connect art to science. [Q3]
- (5) AI design can be further enriched by illuminating situated and embodied meaning, connecting work in the lab to real-world applications and consequences, fostering critical and poetic perspectives, valuing interpretation over explanation and accounting for a wider range of stakeholders. [Q3]

This paper begins by reviewing a body of literature on public AI literacy to examine how public understanding of AI has been of academic and civic interest. Next we look at the role of AI in art – both as subject and as a medium for artistic expression by artists wanting to probe AI processes, concepts and public issues around AI. We then explain the project background and methodology before offering an analysis of workshop results and final reflections

detailing the intentions and strategies expressed by the artists, as well as insights of relevance to the research community on public AI literacies arising through the interactions between artists, AI developers and social science and humanities researchers. The final section of the paper considers how strategies developed within this critical art community during this period have equipped practitioners to respond to the fast paced developments in the field today, how the findings of the workshop have been built on tangibly in AI design and policy and their relevance to future research.

2 LITERATURE REVIEW

2.1 AI Public Literacies

Human-computer interaction (HCI) and critical data studies literature have conceptualised AI literacy in a number of ways. Long and Magerko [39] think of AI literacy as “a set of competencies that enables individuals to critically evaluate AI technologies” as well as use it (p. 1). Kandhofer et al. [34] define AI literacy as the ability to understand the basic techniques and concepts behind AI in various applications beyond simply using them. But the notion of ‘literacy’ or ‘understanding’ by a non-expert public in these studies vary.

2.1.1 Conceptualisations of Public Understanding of AI. In some studies, understanding AI means how accurately people generally define it – what the term ‘Artificial Intelligence’ means to a person [3]. Researchers have also inquired about the public’s basic comprehension or awareness of a technology’s prevalence in daily life as part of popular platforms, for instance whether users are aware that their Google search or social media newsfeed is algorithmically controlled [72][58][1][12]. The notion can include more technical comprehension of different pipelines of AI tasks and processes, including the various roles humans play in designing AI systems. Research in this area also focuses on normative understanding – what the public thinks a technology should or should not do or whether it is for good or ill [56][72]. To this end researchers have asked about public opinions and responses to emerging AI technologies in specific contexts, such as autonomous vehicles [6][64][32], biometrics [41], or facial recognition[7][72]. Ng et al [46] reported that while there has been a dramatic increase in AI literacy publications from 2014 to 2021, AI ethics is largely neglected. Still others have asked about the various imaginaries and metaphors people associate with AI [26][14].

Racial justice and design justice approaches addressing the area of FAccT (fairness, accountability and transparency) in AI call for involving the expertise of affected and minority communities in efforts around algorithmic transparency and governance [43][51]. This literature also stresses the importance of understanding the broader political economic contexts of algorithmic systems and look at “contextual and historical antecedents” explaining algorithmic patterns, rather than focusing on accuracy of the outputs [47][4]. Here, literacy is about connecting technical systems to the broader historical and socioeconomic power dynamics that shape them.

Studies on public AI literacy are usually distinct from what many understand as explainable AI (XAI), a technical field that typically focuses on bringing transparency to the technical decision-making processes of AI systems more narrowly by making the mechanics of a machine learning (ML) model easier to understand [8]. Morley et

al [44] make the point that these purely mathematical approaches to fairness can be reductive – that a more comprehensive approach to literacy would move beyond understanding post hoc algorithmic outputs. Therefore, while research into public understanding can encompass comprehension of the technical workings of AI systems, its remit is broader, including questions of people’s general awareness of AI systems in various contexts as well as their opinions and normative takes on AI applications and impacts [58].

Public literacies, according to Rader et al. [50] facilitate a more critical approach to support “appropriate reliance, rather than blind faith” in a system’s results; literacy can help people be aware of different types of AI applications, especially those embedded in well-known technologies used daily, and of their possible biases. Others find that public understanding is core to AI’s development and oversight: democratic citizens influence policy makers – exemplified in city-wide bans on facial recognition and predictive policies technologies – or as shareholders and consumers impacting the governance of big corporate tech [2][58]. Mechanisms for public AI literacy can encourage diverse communities to come together, especially those who have been most vulnerable to experiencing AI’s negative impacts [20][2], and it can stimulate shared understandings between AI experts, policy makers and the general public around interpretations and values that should drive implementation of AI [3]. Several studies ask whether greater understanding leads to more or less support or shapes ideas about the governance of AI [59][38][22]. Studies argue that both trust and understanding can be improved when citizens are given a role in oversight and governance of AI technologies [2] and when public values guide their design and implementation from the start [3].

2.1.2 Current Approaches to Developing Public Understanding of AI. Another key contribution of these studies is to explore what mechanisms might promote public AI literacy. Scholars have shown that literacy can be improved through giving an algorithmic system visibility, which can be as simple as explaining that features of a platform, like a news feed or a set of user-generated reviews, are in fact driven by an algorithm [50][19]. The MIT Moral Machine experiment built an interactive website that asked participants to say how they would navigate the moral dilemmas confronted by driverless cars [6]. XAI experiments – those focused on exposing the underlying mechanics – include ‘example-based explanations,’ which reveal examples of data from an underlying training set to explain algorithmic behaviour [11]; interactive ‘explanation interfaces’ showing the variables comprising an algorithm and their weighted values [13]; and graphics showing how algorithms are influenced by user-generated content, such as the ratings of others with similar profiles to you [31]. These tools focused on explanation can be technically simple to carry out, yet XAI experiments often don’t link AI visibility back to broader discussions around accountability, justice and fairness.

Other studies have looked beyond these more individualised or consumer-based approaches to public literacy. Researchers and civil society groups have called on governments to encourage public understanding of machine learning used by public agencies by drawing on democratic processes of algorithm impact assessments [53] and

algorithm audits [42] that make design choices and results transparent to the public. Civil society organisations have also trialled efforts towards more communal forms of understanding of AI. The Royal Society’s Public Views on Machine Learning carried out public workshops on AI, reaching over 15,000 people in the UK, while the RSA’s Forum for Ethical AI designed a citizens’ jury to look at AI in decision-making by public and private sectors [3]. Several studies argue that public literacy around AI should be a key component of public institutions, such as libraries and schools, calling for its addition in secondary school and college curriculum [2][54][67].

2.1.3 Summary. As we can see, approaches to public AI literacy range from technical methods for exposing underlying mechanics (such as many XAI approaches) to more or less deliberative and communal (such as citizen juries); they also range from addressing the problem at more individual levels of experience to targeting a structural level through educational policy and algorithmic governance. These existing studies raise important questions that have informed our research: What are useful strategies to foster greater literacies? Should AI literacy focus simply on awareness that AI is part of a particular application, should it encompass how well a person understands its underlying mechanics, or should it be an ability to make normative judgements about a technology or to connect it to wider political-economic structures?

To date we find that most studies assume such interventions will be made by engineers, researchers, educators or public bodies, while very little studies have looked at the role that art can play in public AI literacy and how artists may contribute interventions in this area. To this end, we turn to literature on how the arts can provide interactive experiences for the public around AI technology.

2.2 AI + Art

Artists have worked with AI and emerging digital technologies since at least the 1960s [63] [66]. A body of literature in cultural AI [40] [57] looks at direct applications of AI in the creative industries (Davies, 2020) or within the wider landscape of computational and new media arts [68] [49]. Some scholars, such as Fry [24], point out that AI literacy can be an issue for people deploying creative tools that are opaque in how they operate; Bryan-Kinns et al. [9] note, regarding AI-generated music, “only a few creative AI systems explain what their models are doing, how they do it, and why” (p. 4). Beyond this literature about AI as artistic tool is research about art that makes AI a topic for critical engagement [27]. In this area, scholars ask about the role of creativity and artistic value of using AI to generate novel artworks [16], issues of agency and authorship [70] and the socio-political foundations of AI [73].

Researchers have also looked at strategies of AI artists to address complex and multi-dimensional issues alongside aesthetic and technical themes when working with creative applications of AI [30]. We identify a field of critical practice – critical AI arts – in which artists work with AI technologies and data both as medium and as theme, both as tool and as topic. Stark and Crawford (2019) similarly observe artists who used AI to craft critical literacies in their audience, writing, “Many artists sought to provoke audiences by defamiliarizing the opaque and proprietary nature of software tools

commonly covered by trade secret protections and functioning as 'black boxes' " [62] (p.5).

Artists who engage with AI as both tool and theme have addressed issues such as accessibility, by allowing non-technicians to build interactive systems (Fiebrink and Cook's Wekinator [23]); bias, through engagement with the underlying politics of a system (Crawford and Paglen's ImageNet Roulette [15]) or the absences in datasets (Mimi Onuoha's What is Missing Is Still There [48]) and the failures of AI systems in recognising iconic African American women (Buolawmini's AI, Ain't I A Woman? [10]); labour (AI Oracle, no:topia collective 2018); prejudice, via an interactive artwork examining the concept of 'normal' and algorithmic prejudice (Zer-Aviv's Normalising Machine [71]), and surveillance (Elahi's Tracking Transience [18]).

The approaches and strategies of artists working with AI as both tool and topic here fold into a longer lineage of artists that have engaged in political critique [25], continuing to more contemporary socially-engaged participatory art [5][55] and speculative design, which uses non-functional design artefacts to stimulate critical discussion [17]. As Simoniti [60] notes, such practices are dedicated pragmatically to impact, aligning art with social work and activism around technology development. Such work raises questions about how art may feed into scientific processes of design and development and how artists and practitioners can mutually inform each other's practices. Tromble [65] charts the entanglements between art practice and scientific research in AI, arguing that artists, often steeped in the humanities, are likely to raise questions about how AI can be beneficial, as opposed to commercially successful, and keep technological possibilities in play. In a similar vein, with regard to accountability, Jensen [33] observes that art can offer checks on practitioners' conscience and keep designers accountable.

As yet there remain few systematic reviews of current strategies in use to develop critical public literacies through audience reception of AI art and still very little work on how artists can gain from dialogue with researchers, engineers and ethicists, and vice versa. Srinivasan and Uchino [61] have stated that most academic studies on AI ethics have not investigated the role of arts in this area, highlighting this as a neglected area for enquiry, despite great potential. The role of art in public literacy in AI is therefore under-examined in the current landscape.

3 RESEARCH DESIGN

To explain the design of the workshops informing this study, it is necessary to contextualise it as part of an ongoing project called The New Real. The New Real is a research effort that facilitates exchanges between the arts and AI comprising transdisciplinary fields of applied science, engineering, design, art and social science. A primary aim of the group is developing the area of *experiential AI*, a concept first envisioned for the The New Real in 2018 and elaborated in a preliminary internal workshop at The University of Edinburgh in 2019. A hypothesis for the research is that art and tangible experiences can mediate between computer code and human comprehension to overcome the limitations of explainability and accountability in AI systems [30]. The group uses an open prototyping [29] approach to broker between disciplinary viewpoints

and to engage public audiences in the research. The ambition for the group is to support the creation of significant artistic works and inspire new concepts and paradigms for fair and inclusive AI. The research continues at time of publication and is currently in its third generation of projects, including artistic commissions and exhibitions, the development of an Experiential AI system and wide engagement with the AI and arts communities.

As an early stage in this effort, the research team designed the cooperative study reported in this paper to develop the research theme and, subsequently, new bodies of artistic work informed by the theme. The team recruited two artists, Anna Ridler (Figure 1) and Caroline Sindors (Figure 2), through an open call in collaboration with Ars Electronica in Linz. The principal investigator (PI) directly invited the third artist, Jake Elwes (Figure 3), following their participation at a workshop at ZKM in Karlsruhe. The outcome was an open-ended commitment to work together over a period of eighteen months to two years to develop a new body of work for the artists and a phase of the study for the research team. The artists already had bodies of work that used AI technologies that were discussed as case studies in the workshop. The research team offered insight and directions to explore that the artists might find enriching to their practice and the works they create. The engagement with the artists included commissioning new works to explore new dimensions of their creative practice and develop bodies of work.



Fig. 1. Myriad (Tulips). Anna Ridler, 2018. Detail of an installation of thousands of hand-labeled photographs of tulips. Image credit: Emily Grundon.

The specific objectives for the workshops were to work closely with AI artists to understand their intentions, strategies and methods, while also creating the conditions for reciprocal exchange with engineers and researchers. Alongside the artists, participants in the workshops included AI engineers, social scientists and philosophers; these had been participants in the preliminary internal workshop in 2019 or were researchers at University of Edinburgh recruited by direct invitation. The project PI and a research assistant designed the workshops to elicit and explore the artists' intentions and strategies, to invite semi-structured contributions from different disciplinary perspectives and to facilitate an open dialogue between the multidisciplinary group. The team developed a workshop information sheet,



Fig. 2. Sorting Feminist Data. Caroline Sindors, 2018. A social justice art workshop that interrogates every step of a machine learning process from a feminist perspective. SOHO20, July 2018.



Fig. 3. Closed Loop. Jake Elwes, 2019. A film of two neural networks interacting: a language captioning recurrent neural network writing what it sees in the images generated, and a generative neural network generating images responding to the words generated. Installation shot from Writing the History of the Future 2019 - ZKM, Karlsruhe.

script and questions based on their knowledge and understanding of art curation and AI research along with insights from the interactions with the artists through the early stages of the collaborative partnerships. Two workshops were held on Zoom on 21 July (one with Jake and another with Anna and Caroline), and a third held on Zoom a month later, on 26 August (with Anna and Jake). The workshops were supported by interviews of the artists conducted by the research assistant.

The team conducted the workshops online; these were recorded, transcribed, and anonymized. Members of the research team coded the workshop transcripts based off of themes raised by the literature review, though more codes were added as other themes arose.

The team analysed interview and workshop transcripts through a thematic analysis approach [21], a deductive method for qualitative data analysis that allows researchers to surface patterns from the data while remaining open to emergent themes. The research received approval through the ethical review process at the University of Edinburgh, and all participants consented to the study as well as to their direct quotes used in this paper. The artists consented to be named in this article and also took part in the article's writing, in keeping with the project's goal of including artists in all phases of the research process. After the workshops, the artists collaborated with The New Real on the development and presentation of new artworks at Edinburgh International Festival and Ars Electronica as part of a later research phase.

4 FINDINGS

The workshops shed light on the intentions and strategies of the artists and the key issues and themes of interest to the group. Our analysis discerned, in particular, four themes: the strategies that artists use to illuminate AI issues, the types of public literacy participants thought artists who work with AI might engage with, how art that illuminates AI issues offers distinct approaches from other efforts at public AI literacy, and how art can feed back into AI design processes in industry and academia.

4.1 Artists Strategies Towards Public AI Literacy

In this section we detail how the artists discussed, in their own words, ways to create experiences around AI that lead to more public awareness or critical understanding (addressing RQ1). The artists reported using a range of artistic techniques to create critical understandings of AI. Some of these strategies entailed making underlying systems more visible, but we also detail other approaches artists took to connect technologies to broader political issues.

4.1.1 Making Underlying Systems Visible. All three artists talked about strategies they use to make design choices more visible and so expose the pipeline of AI. Caroline commented that AI art can have “whispers of the data appearing in the outcome itself” in ways that made the underlying processes more transparent to audiences. By displaying the entire dataset underlying a ML system in gallery settings, for example, Anna said that her work aimed to “open these systems up and allow people to understand them and, once they understand them, hopefully change.” Caroline commented that such artwork challenges the “magic wand effect” of AI as “something magic we don’t understand.” Caroline actively engages people in building AI systems in a way that is visible and rewarding of their effort. This artist deployed co-creation – working with affected communities to build an entire ML pipeline, and so exposed these communities to features of how these systems work. Such artwork resonates with Caroline’s civil society efforts in consumer protection to bring more transparency around algorithms.

4.1.2 Revealing gaps in the data. A strategy used by Jake is to add new data to expose gaps in underlying datasets. The artist works with Generative Adversarial Networks (GANs), and they discussed a work of theirs that brought in additional data to an original training dataset to create a GAN. Their additional curated dataset generated

results that addressed a lack of diversity in the original training data. In this instance, the artwork started public conversations that made the technology more accessible to non-technical audiences. By exposing these gaps and omissions, the artist hoped the project pulled back the curtain on the human choices and biases in the datasets that construct many ML algorithms: "datasets are often gathered by engineers without necessarily as much of an interest in including diversity and marginalised groups."

4.1.3 Revealing human labour. Another of Anna's strategies is to expose the hidden labour and materiality behind ML. Belying the common assumption that AI is, as Jake put it, a "magic piece of mathematics that then digests the world on its own and generates outputs and insights on its own," Anna has created a ML artwork by producing the training dataset herself, a series of photographs of tulips, as well as painstakingly labelling each data point by hand, such that the data input becomes part of the public work along with the ML output. Such effort, Anna says,

shows the labour that goes into something. The handwriting also indicates that there is this very human aspect that people often mistake, I think, when thinking about the digital.

Anna also saw how her dataset decisions led to certain outcomes, as she was "controlling the GAN output with various different conditions and labels."

4.1.4 Defamiliarisation. Another approach that participants raised was presenting AI in a new environment that can cause people to think of it in fresh ways, exposing its taken-for-grantedness. This technique draws on the concept of defamiliarisation [62] – Jake talked about this tactic in a project of theirs called Cusp that projected ML-generated imagery of birds into a natural landscape, transplanting the digital to physical:

It led an artistic audience that were not used to these conversations around machine learning, it got them to discuss machine learning in a different way. And inevitably, the conversation would lead to all sorts of things around datasets, how training works.

4.1.5 Interactivity. AI art can be similar to public literacy projects that require active engagement from an audience, an approach that fits with one participant's (participant 7) thoughts on education. As an educator in robotics, this participant said she believes that "if you want people to learn things, then you have to have them do things, not just watch things." The pedagogical role of interactivity holds true for some of the artworks under discussion. Caroline has designed a chatbot that can be manipulated by audiences to create new instantiations with each interaction, and her work also engages different audiences as co-designers to explore the design choices throughout the pipeline of AI.

4.1.6 Linking AI systems to structural issues of power through aesthetics, satire, scale and speculation. Caroline compared her artwork to a 'trojan horse' that drew people in first through aesthetics before provoking political discussion on social change or alternate possibilities. Anna illustrated the tactic of scale to create this effect, stating that an artwork does not need to be large scale or technically

complex to engage politically; it can be done with "something very minute, very tiny to think about these wider, bigger, ethical, philosophical, political questions." Jake draws on satire to present the GANs they create to directly confront some of the "fears that society have" around AI capabilities. The artists also discussed the tactic of speculation, to envision alternative or contradictory futures, drawing on the speculative design field. Caroline described an example of how a playful, speculative approach can provoke discussion on very troubling topics, such as her Critical Design Chatbot created to speak to people who have experienced online harassment: "The bot also tells people it shouldn't exist but it does to fill this void created by platforms, which is providing victim help and care."

4.1.7 Wrap around engagement to enrich an experiential journey. The artists in our sample all viewed the piece itself – the artwork – as one dimension of the process of creation and presenting work, which also includes public lectures, discussion, writing and political activism. Their art practice also includes other outputs and forms of audience engagement, which provide opportunities and an environment for public literacies. In some cases, the artwork is an avenue towards these conversations, which become the primary loci of audiences' critical engagement, rather than the piece itself, according to Anna:

Thinking of how you visualise or how you talk about algorithms that, for me, at the moment is something that fits more into the writing and speaking part of my practice rather than the more artifact gallery-driven part of my practice.

The workshop in this sense gave rise to a view of AI Art as a process rather than a thing, and an understanding that wrap-around engagement is an important component of the artists' practice and the presentation of their work. Anna pointed out that her work is durational and should be treated as a process that changes from exhibit to exhibit rather than a finished product.

4.2 Distinctive Contributions of Artists' Strategies Towards Public AI Literacy

The workshops presented an opportunity to ask what the arts can offer that other approaches to public AI literacy, such as those discussed in the literature review, may not (RQ2). Participants also considered the types of AI literacy art should engage with, a question over which there was some disagreement among both the artists and other workshop participants.

The experiential aspect of art – already evident above – can distinguish it from other AI public literacy approaches by giving the audience a particular role that values interpretation over explanation and final answers. Anna discussed how her artworks do not explain so much as ask audiences to complete a piece through their own interpretations of it, as shaped by the context of the art work: "I never really like to explain 100 percent, I like to leave, like, 10, 15, 20 percent of something that someone has to work out themselves."

The artists pointed out that art can reach wider audiences and offer an alternate forum to other XAI or public literacy efforts to engage audiences with technical issues that some would ordinarily be unaware of, avoid or feel uncomfortable with. Jake described this function as the "poetic aspect" of art to engage people first, then,

with their attention captured, “open up the black box a bit.” Anna said a goal was not to alienate audiences who had no experience before with a particular technology, such as GANs.

Other participants also saw the opportunity to use experiential and creative methods from the arts to expose the underlying mechanics of opaque AI systems with the intent of giving people more agency in their interaction with them in other contexts. Jake cautioned that artistic works that – deliberately or not – repeat or amplify misrepresentations of AI techniques can generate “fear-mongering from the public.” Rather, artistic methods can be used to illuminate “the gaps and omissions in datasets in AI itself, but also the constructions and human choices that give rise to it.” Participant 5, an AI engineer, agreed, saying that working with “technology in this creative way... might open up people to what exactly the technology’s doing to them and the people around them. I think that’s the key point.” Another participant (participant 6) pointed out that, even if an artwork can’t precisely explain how algorithms work, it could show the key role of training datasets as they interact with algorithms. “There’s lots of things people misattribute to AI, and 90 percent of the problems come because of data baked in.”

In contrast, other participants saw the distinctive potential for art to foster poetic and critical perspectives on the political economy of AI and raise normative questions of what people should use AI for. Participant 8, also a social scientist, felt strongly that the artist need not reveal the inner technical operation – to explain how systems work in a narrow operational sense – but should connect AI systems to their broader social and political economic contexts. An artists’ role is not, she argued

So what is being created here are questions around ethics, politics, economics, the social repercussions and these more... well, less tangible but equally as important issues. So to sum that up, I would say demystification is the wrong word and it’s not your responsibility to demystify anything around AI.

Participant 4, a philosopher of AI ethics, making a similar point, highlighted the need to address a broad range of social and ethical considerations. Specifically, she argued there is a problem with AI literacies always focusing on the specific issue of bias in training datasets rather than asking questions about the social valences and applications of a technology:

It’s also important not to over-focus on issues of under-representation and misrepresentation in the datasets. Not that that’s not a major concern, but that focusing on that alone can obscure some of the broader issues around AI implementation and its interactions with other social systems.

The participant illustrated this with the problem with facial recognition systems that, when de-biased, simply make tools that over-police minority ethnic groups more effective. A greater technical understanding of how to make systems less biased, in other words, will not necessarily address what the AI actually does in the world.

These points resonated with the artists’ own interest in connecting technology and data to higher level questions of power and ethics. Caroline explores how the production of a community-produced ML system could embed feminist principles. She described

an artist workshop in which feminist literature is related to each step of an AI pipeline as that pipeline is built, “from data collection data training to generating a model then creating new forms of user interface and then an output of a chatbot.” Anna’s example was a project she created capturing iPhone autocorrects in a dictionary – she notes that words related to sexuality, homosexuality and drugs were most likely to be autocorrected, revealing “a lot about how corporations think.” While the work didn’t explain how autocorrect worked at a technical level, it documented the choices made by Apple around what words to autocorrect, and, as she put it, that the “kinds of decisions which ordinarily would have been taken by academic institutions who produce dictionaries are now being taken over by corporations.”

Participant 1, a researcher of human-algorithm interaction, pointed out the importance of this “contextual view” of AI and how art can show “the connections between the algorithms and the broader social context that they work in” – for instance, showing feedback loops of certain algorithms that reinforce social inequalities.

Anna also highlighted how her artwork connects to the natural sciences and other themes. She talked about pushing the limits of AI’s autonomy to better understand human consciousness and theories of mind. Anna said she explores how natural seasons and concepts from the natural sciences and taxonomy dictate her data collection practices, and therefore let the digital and analogue inform each other. Her project about tulips, she said,

...was about speculation, it was about collapse, it was about how GANs work, it was about datasets, it was the history of natural history, all of these different things could be brought together in one project.

4.3 The Responsibilities of Artists

Here we consider what workshop participants thought is and isn’t an artist’s responsibility in regard to AI literacies (RQ2).

One of the main themes discussed was the degree to which it is or isn’t for artists to be positioned as educators. The workshop participants agreed that art can play a pedagogic role, though there was some nuanced understanding of how this might play out. There was agreement among participants not to ‘instrumentalise’ art. Participant 7 talked about the role of artistic autonomy, separate from academic or commercial spheres: “an artist is not an educationalist.” An artist can contribute pedagogically, but in ways that don’t undermine their autonomy. As Anna put it, art is not “glorified graphic design for how ML works,” functioning at the service of one specific edifying purpose – the aesthetic dimensions of an artwork cannot be tied to a particular utility.

A difference, therefore, between art and other AI public literacy approaches is that an artwork can function at the same time as an aesthetic piece in its own right, even if it also aims to educate. The artists who took part, for instance, mentioned other motivations for their work, beyond edifying.

The artists in the workshop all saw their art as political, and a sense of responsibility is integral to their practice. Jake viewed themselves as part of a relatively small group of creators who used ML as a tool of their practice but also to comment on ML as a political and ethical concern. Caroline positioned her work explicitly within

traditions of social justice and human rights-driven art. Caroline linked the ability for art to engage in explainability and legibility of AI systems directly to a political project itself – she didn't see a distinction between giving greater visibility to the technical aspects of systems and linking them to a broader political economic context. To her, “art can totally provide solutions, be explanatory and also at times even ridiculously pedantic.” Making AI systems legible was a political project and part of her work as an activist.

4.4 Potential Contributions to AI Design

In this section we bring together points by the artists and other participants concerning the ways artistic strategies could complement or contribute to AI design (RQ3).

By its very nature, art provides a diverse range of opportunities for real world applications that optimise creative work flows or enhance audience experiences; these are valuable opportunities for user-centred research and evaluation that ML designers in industry and academia can often struggle to access. Participant 6 thought that artworks allow a uniquely exploratory approach, moving beyond explaining the technical “function of something” towards showing “the degrees of freedom in how it might operate” – pointing, in other words, to alternate possibilities within the realm of actual affordances and constraints. As such, participants argued the art can act as a sandpit for uncovering issues before engineers and designers find them. For instance, Anna pointed out that GANs, which she works with, have few other real world applications at the time of the workshop, but could be soon adopted by the medical community. She also reported that as an artist she is contributing findings that are novel to science. She found that her attempts to reduce bias in a model did not work as well as expected:

I spent all of the time being very conscious about bias and making sure that I had a super-balanced dataset. But what I found, even though when I did this, it wouldn't give me a totally kind of representative kind of output.

Anna pointed out that this is a problem that has not yet been well covered in the literature on GANs, or more widely among the community – “I'm finding things that aren't in the academic paper, that aren't being talked about on Twitter.” Caroline described how artworks – like her critical design chatbot – can suggest alternate solutions that do not yet exist, and so call into question the designs we casually use day to day.

The practices of the artists were seen to suggest ethical and inclusive approaches to designing AI that may not be found in typical R&D and lab settings more driven by commercial imperatives. For instance, Caroline's Feminist Dataset, which used participatory design and incorporated questions about the gendered dimensions of data at each step, slowed down the process of building with ML and forestalled adopting, uncritically, pre-made datasets – an approach that could help designers interrogate standard procedures.

Participant 4 pointed out that artists do not need to make technical possibilities or pitfalls more visible to practitioners, but rather the ethical and political stakes of such choices. She cautioned that she still encounters large portions of the engineering and design practitioner community who believe social, political and moral choices

around AI happen “after their work is done, in circles that they do not see themselves as being a part of, policy circles, regulatory circles.” Instead, for the AI research and practitioner community, art can contribute by

...making visible the ways in which the members of those communities, when they design, develop, train, build, tune AI systems, are making social, political and moral choices, consciously or unconsciously. And making the social, moral and political agency of AI practitioners themselves more visible.

Artists can therefore expose the political dimensions of the choices made in commercial and research contexts.

Participants felt that art does not need to have an explicitly commercial or civic function, and for this reason, it can leave room for serendipity and the unexpected. An artwork can offer, as one participant put it, a “kind of poetry” in ways “that weren't necessarily explicitly coded into these models, and weren't intended by the sciences.” Indeed, Jake stressed that art should not be purely educational or it will be unengaging; it is necessary to delight audiences at the same time:

If I was going purely on explainability and demystifying, then it would not be engaging for 99 per cent of people who want to engage with artwork, it's going to be a very cold, technical thing.

4.5 Limitations to Critical AI Art and Cautionary Notes

Finally, we detail cautionary points and limitations of the strategies discussed, both with regards to general audiences and interactions between artists and AI engineers. One participant (participant 8), a social scientist, cautioned against assuming all audiences will find such artworks accessible:

I think there needs to be a recognition of the plurality of different audiences [...] So that's, you know, hard to reach groups, marginalised subsections of the population, as well as the audiences that would include activists and those that are more interested, such as practitioners.

This comment acted as a reminder that artists should not assume a singular audience and that reaching diverse audiences may require different tactics and contexts to present a work.

Participant 6 pointed out that many black box systems do not need explaining in order to be experienced – “we are perfectly happy to not understand anything about a plane and get on it.” Participant 3, a social scientist, challenged the claim that audiences need to be demystified, as if they did not already understand.

Something we talk about in science and technology studies which is the deficit model, the assumption that people don't like science because they don't understand science or they don't know enough about science. And often that is an assumption, it's not based on any kind of evidence. And assumptions are problematic.

At the same time, some participants worried that the assumption that art could influence ML designers was overly optimistic – participant 7 said she found art “self-contained” and didn't understand

how it could interact with her work as an engineer. Participant 5 said that the themes raised by many AI artworks were too technically simple and, while they may address public literacies, they would not necessarily create an exchange with engineers and designers:

I worry that to some extent these kinds of demonstrations might just hit a superficial aspect of algorithms and not really go deeper because they are not asking the right questions. They're kind of saying, yes, there's bias in the data, yes, there's immense human effort in collecting the data and it's something we recognise, but at the point where we need to say, okay, what can we take back to science, I kind of worry that this sometimes falls short.

That same participant made the point, however, that creative and experiential exploration of the materiality of AI is one way to deepen discussions with ML designers. To him, problems arise when artistic metaphor departs from the science or invents or amplifies misrepresentations of AI. A mistake some artists have made is relying on metaphors of machine consciousness or dreaming, contributing to mystification specifically through using anthropomorphised language and metaphors:

We should be careful to not give people the impression that, you know, this is way ahead, this is where the machines are dreaming and whatnot. And then you are in the zone when we simply cannot even have a current conversation with science.

5 DISCUSSION

5.1 Linking systems visibility to sociopolitical structures

Art provides opportunities for rethinking concepts and approaches around AI literacy. The literature on AI literacies suggests four main types of public AI literacy: 1) awareness, 2) technical understanding 3) normative assessment and 4) making critical links to broader structures of power. From the tactics described by artists here and the discussion with scientists and social scientists, the focus should be mostly on 4 – the researchers participating in the workshop particularly pointed out that there is a huge risk in ML of focusing too much on technical explainability and missing more important overarching issues, such as questioning whether a technology should even exist. However, shedding light on the technical layer, through the tactic of rendering underlying design choices visible, literacy type 2, can offer a way to link AI systems to broader power structures. For example, a project that exposes how designers curate datasets can link to issues of hidden labour, bias and gendered data.

Implicit in this discussion is that by illuminating power dynamics and political economy, and by linking underlying the data and models to these dimensions, audiences might start to make the kind of normative assessments needed to design and regulate AI. This theme of linking system's visibility to structural issues of power – if not explicitly than through surrounding documentation of the artworks – cuts across the projects discussed and aligned with the artists' own sense of political responsibility. This linking can be seen in the artistic tactics described: relating underlying design choices to power; exposing labour practices; and using defamiliarisation, satire, scale and speculation to provoke political discussion.

5.2 An experiential approach to public AI literacies

The workshops highlight how art functions quite differently from other attempts at public AI literacy that make technical understanding and interpretability of AI an explicit or sole goal. Artists can enjoy an autonomy separate from the aims and values of educational institutions, business or government policy – a distancing that fosters the works' critical and poetic dimensions. Research on experiential learning [35] has shown that artistic practice can act as a powerful pedagogic mechanism. Creative and experiential methods, and the poetic and exploratory dimensions of artistic works, therefore offer an effective and distinctive way to illuminate key technical and social issues. For the audience, the tangible interactions with the critical AI artworks form an integral part of their learning journey and outcomes. From this perspective, critical AI art has the possibility to connect statistical reasoning underlying a piece to human intuition and creativity, and to situated and embodied meaning, while valuing interpretation over explanation. Art, in this way, is distinct from more technical approaches discussed in the literature review – whether public XAI interfaces that expose internal data processes or pedagogical methods aimed at teaching lay people how to program AI.

The artists in our workshops describe strategies to inform audiences and aesthetically engage them at the same time through compelling experiences. The arts have the potential, in this manner, to increase the reach of literacies practices. One way is through what the participants called the Trojan horse function – putting AI technologies into contexts and using aesthetic strategies to reach audiences who otherwise may not engage with AI literacy projects. Critical AI art can reach people who initially respond to aesthetic and poetic appeal, then find themselves more aware of AI systems and their political dimensions.

Workshop participants agreed this edifying role is not all that these artworks do. Artworks can be evaluated for their aesthetic function, which falls outside the remit of didactic messaging. The participants also agreed that it was not the artists' role to spoon feed audiences whose interpretation of a piece cannot be fixed or foreclosed, only guided.

5.3 Enriching art and AI design practices

We conclude that arts practice can be further enriched by creating multiple opportunities to connect art to science, and that AI design can be enriched by the arts. We discussed the strategies of artists to explore the poetic dimensions to AI systems, in ways that can point to more legible, and, ultimately, more fair and responsible AI.

Scientists and engineers at the workshop were able to speculate that art can play a role in design and development, since art can act as a sandbox to discover particular properties of AI systems. Artists may be less constrained by market or institutional factors in their technical experimentations, and so they can suggest alternative processes to designing AI that deviate from typical design steps and have more criticality and reflection. Our workshops also demonstrated the importance of transdisciplinary conversations around art, ethics and the political economy of AI technologies and how these dialogues may feed into AI design processes.

Beyond technical affordances, the artists link statistical reasoning to human intuition, imagination and serendipity. Artists may use their platform to make the political dimensions of AI more visible to scientists and designers who would otherwise see their choices as neutrally disconnected from broader politics and social goals. However, some AI engineers in the group still expressed feelings that the art world remained disconnected from their work and worried that most artworks only engage with technologies at a superficial level, such that they wouldn't transfer much knowledge back to engineers. Ongoing conversations between artists and engineers facilitated by The New Real offer one avenue to address these misgivings.

6 CONCLUSIONS

This paper presents strategies of critical AI artists as a novel contribution and proposes how AI art – particularly work that uses AI as both tool and topic – can scaffold public understanding of AI in unique ways. The paper reports on strategies to foster greater legibility and critical comprehension of AI processes in ways that go beyond current limitations of more technical approaches to explainability and accountability in AI systems.

These strategies can equip practitioners to navigate emerging opportunities and threats. At the time of the workshops, the use of vast training datasets in machine learning was commonplace and underpinning major advances, but the use of datasets developed by scraping content indiscriminately from the Internet was not so widespread. As a result, prominent concerns relevant in 2023, the time of publication, did not emerge as central to the artists in the workshop – issues including authorship, consent, accreditation, rights, fair pay and threats to jobs for artists and creative workers. We have also seen in recent years a greater maturity in the critical debate on issues such as bias, more sustained efforts to address such concerns through engineering and design, and deepened understanding of the wider structural issues that generate these issues.

Nonetheless, the creative strategies detailed here remain relevant to the current landscape and can be applied to newer, emerging issues. The artists in our sample all considered a public literacies dimension to be an integral and vital part of their artistic practice, not something external or secondary. Moreover, in these cases, the art not only communicates established knowledge, it also explores possible futures beyond current technological limitations and new, critical perspectives on technological development. The research raises distinctions between tactics for AI literacies that could be reflected upon by artists and other AI practitioners, social scientists and AI ethicists generally. The workshops also provide insights into questions about how artists and designers might engage through ongoing conversation or collaboration around AI and offer examples of how AI engineers and designers could learn from artists, as well as caveats about the limitations of some artworks to transfer knowledge to the public and into other design contexts.

The New Real team has built on these findings with AI design and policy communities. The strategies outlined above – such as making underlying systems visible, defamiliarisation and interactivity – informed the development of an Experiential AI system and the participating artists' works as well as collaborative work with later cohorts of artists. The artists, for instance, developed

new artworks (The Zizi Show by Jake Elwes, Figure 4, and Mechanized Cacophonies by Anna Ridler and Caroline Sindors, Figure 5) presented at The New Real exhibition in 2021 at the Edinburgh International Festival, the festival's first AI Art event. A second cycle of the research culminated in an exhibition at Ars Electronica 2022, and the third cycle of the research is under the title Uncanny Machines. The artists' strategies have been built on in a system architecture encompassing machine learning and other computational techniques to enable artists to collaborate with AI to make underlying systems visible and make connections to higher level dimensions of AI. National programmes developed through The Alan Turing Institute, the UK's national data science and AI institute, and the Scottish AI Alliance, an initiative of the Scottish Government, have also built on these findings. The Turing has incorporated The New Real to introduce an Arts and Creativity theme at the Institute; part of this partnership is to design creative methods for XAI, leading to wider academic uptake. The Scottish AI Alliance invested in an AI Art commission in collaboration with The New Real to build public literacies around responsible AI. The workshop format – gathering together interdisciplinary scholars with artists – highlights how conversations across and between disciplines can illuminate key technical and social issues, and offers a template for future efforts to explore art and AI literacies.



Fig. 4. The Zizi Show. Jake Elwes, 2021. A deepfake drag cabaret, a virtual online stage featuring acts that have been constructed using deepfake technology, learning how to do drag by watching a diverse group of human performers. Commissioned by The New Real. Premiered at Edinburgh International Festival. www.newreal.cc/artworks/the-zizi-project.

ACKNOWLEDGMENTS

Thanks to the participants in the research workshops: Dave Murray-Rust, Pablo Schyfter, Shannon Vallor, Vaishak Belle, Michael Rovatsos, Ruth Aylet, Frank Boz, and Gill Haddow. The New Real at University of Edinburgh is a partnership with The Alan Turing Institute and Edinburgh's Festivals, supported by funding from Arts & Humanities Research Council, Engineering & Physical Sciences Research Council, and Creative Scotland.

REFERENCES

- [1] Thomas Abbott, Despina T. Tomboulides, Lexyngton McIntyre, and James Dooney. 2019. Strategies to Inform the Swiss Public on Artificial Intelligence. Publisher: Worcester Polytechnic Institute.



Fig. 5. Mechanized Cacophonies. Anna Ridler and Caroline Sindors, 2021. An immersive experience for a remote audience of orchestral layered soundscapes generated by training an ML model using field and on-line recordings of both natural and industrial environments. Commissioned by The New Real. Premiered at Edinburgh International Festival. <https://www.newreal.cc/artworks/mechanized-cacophonies>.

- [2] Lizzie Adams and Simon Burrell. 2019. How to stimulate effective public engagement on the ethics of artificial intelligence. https://www.involve.org.uk/sites/default/files/field/attachemnt/How%20to%20stimulate%20effective%20public%20debate%20on%20the%20ethics%20of%20artificial%20intelligence%20_0.pdf
- [3] Brhmie Balaram, Tony Greenham, and Jasmine Leonard. 2018. Artificial intelligence: real public engagement. <https://www.thersa.org/reports/artificial-intelligence-real-public-engagement>
- [4] Abeba Birhane and Fred Cummins. 2019. *Algorithmic Injustices: Towards a Relational Ethics*. Technical Report arXiv:1912.07376. arXiv. <https://doi.org/10.48550/arXiv.1912.07376> [cs] type: article.
- [5] Claire Bishop. 2012. *Artificial Hells: Participatory Art and the Politics of Spectatorship*. Verso Books. Google-Books-ID: iX8nQrLrybUC.
- [6] Jean-François Bonnefon. 2021. *The Car That Knew Too Much: Can a Machine Be Moral?* MIT Press, Cambridge, MA, USA.
- [7] Ben Bradford, Julia A Yesberg, Jonathan Jackson, and Paul Dawson. 2020. Live Facial Recognition: Trust and Legitimacy as Predictors of Public Support For Police Use of New Technology. *The British Journal of Criminology* 60, 6 (Oct. 2020), 1502–1522. <https://doi.org/10.1093/bjc/azaa032>
- [8] Andrea Brennen. 2020. What Do People Really Want When They Say They Want "Explainable AI?" We Asked 60 Stakeholders. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–7. <https://doi.org/10.1145/3334480.3383047>
- [9] Nick Bryan-Kinns, Berker Banar, Corey Ford, C. Reed, Yixiao Zhang, Simon Colton, and Jack Armitage. 2022. Exploring XAI for the Arts: Explaining Latent Space in Generative Music. *1st Workshop on eXplainable AI approaches for debugging and diagnosis* (2022).
- [10] Joy Buolamwini. 2018. When AI Fails on Oprah, Serena Williams, and Michelle Obama, It's Time to Face Truth. <https://medium.com/@JoyBuolamwini/when-ai-fails-on-oprah-serena-williams-and-michelle-obama-its-time-to-face-truth-bf7c2c8a4119>
- [11] Carrie J. Cai, Jonas Jongejan, and Jess Holbrook. 2019. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*. Association for Computing Machinery, New York, NY, USA, 258–262. <https://doi.org/10.1145/3301275.3302289>
- [12] Stephen Cave, Kate Coughlan, and Kanta Dihal. 2019. "Scary Robots": Examining Public Responses to AI. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AI/ES '19)*. Association for Computing Machinery, New York, NY, USA, 331–337. <https://doi.org/10.1145/3306618.3314232>
- [13] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F. Maxwell Harper, and Haiyi Zhu. 2019. Explaining Decision-Making Algorithms through UI: Strategies to Help Non-Expert Stakeholders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300789>
- [14] Harry Collins and Trevor Pinch. 2002. *The Golem at Large: What You Should Know about Technology*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511541353>
- [15] Kate Crawford and Trevor Paglen. 2019. Excavating AI. <https://excavating.ai>
- [16] Antonio Daniele and Yi-Zhe Song. 2019. AI + Art = Human. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AI/ES '19)*. Association for Computing Machinery, New York, NY, USA, 155–161. <https://doi.org/10.1145/3306618.3314233>
- [17] Anthony Dunne and Fiona Raby. 2013. *Speculative Everything*. Cambridge, MA, USA. <https://mitpress.mit.edu/9780262019842/speculative-everything/>
- [18] Hasan Elahi. 2006. Tracking Transience | Creative Capital. <https://creative-capital.org/projects/tracking-transience/>
- [19] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300724>
- [20] Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press. Google-Books-ID: pn4pDwAAQBAJ.
- [21] Ceryn Evans. 2017. Analysing semi-structured interviews using thematic analysis: Exploring voluntary civic participation among adults. *SAGE Research Methods Datasets* (2017).
- [22] Andrea Ferrario and Michele Loi. 2022. How Explainability Contributes to Trust in AI. In *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1457–1466. <https://doi.org/10.1145/3531146.3533202>
- [23] Rebecca Fiebrink and Perry R. Cook. 2010. The Wekinator: a system for real-time, interactive machine learning in music. In *Proceedings of The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010)(Utrecht)*, Vol. 3.
- [24] Christopher Fry. 2018. Organic Complexity and AI in Generative Art. (2018).
- [25] Jason Gaiger. 2009. Dismantling the Frame: Site-Specific Art and Aesthetic Autonomy. *The British Journal of Aesthetics* 49, 1 (Jan. 2009), 43–58. <https://doi.org/10.1093/aesth/ajyn058>
- [26] Roberto Musa Giuliano. 2020. Echoes of myth and magic in the language of Artificial Intelligence. *AI & SOCIETY* 35, 4 (Dec. 2020), 1009–1024. <https://doi.org/10.1007/s00146-020-00966-4>
- [27] Dejan Grba. 2022. Information Particles: Tracing the Ambiguities of the Creative AI. In *Proceedings of Art Machines 2: International Symposium on Machine Learning and Art 2021*. 2021–05.
- [28] Dejan Grba. 2022. Deep Else: A Critical Framework for AI Art. *Digital* 2, 1 (March 2022), 1–32. <https://doi.org/10.3390/digital2010001> Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.
- [29] Drew Hemment. 2020. Reordering the Assemblages of the Digital through Art and Open Prototyping. *Leonardo* 53, 5 (Oct. 2020), 529–536. https://doi.org/10.1162/leon_a_01861
- [30] Drew Hemment, Ruth Aylett, Vaishak Belle, Dave Murray-Rust, Ewa Luger, Jane Hillston, Michael Rovatsos, and Frank Broz. 2019. Experiential AI. *AI Matters* 5, 1 (April 2019), 25–31. <https://doi.org/10.1145/3320254.3320264>
- [31] Jonathan Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. 241–250.
- [32] Lynn M. Hulse, Hui Xie, and Edwin R. Galea. 2018. Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age. *Safety Science* 102 (Feb. 2018), 1–13. <https://doi.org/10.1016/j.ssci.2017.10.001>
- [33] Beth Jensen. 2020. How AI and Art Hold Each Other Accountable. <https://hai.stanford.edu/news/how-ai-and-art-hold-each-other-accountable>
- [34] Martin Kandlhofer, Gerald Steinbauer, Sabine Hirschmugl-Gaisch, and Petra Huber. 2016. Artificial intelligence and computer science in education: From kindergarten to university. In *2016 IEEE Frontiers in Education Conference (FIE)*. 1–9. <https://doi.org/10.1109/FIE.2016.7757570>
- [35] David A. Kolb. 2014. *Experiential Learning: Experience as the Source of Learning and Development*. FT Press. Google-Books-ID: jpbBQAAQBAJ.
- [36] Nikolas Kompridis. 2014. *The Aesthetic Turn in Political Thought*. Bloomsbury Publishing USA. Google-Books-ID: piw_AwAAQBAJ.
- [37] Benjamin Laufer, Sameer Jain, A. Feder Cooper, Jon Kleinberg, and Hoda Heidari. 2022. Four Years of FAccT: A Reflexive, Mixed-Methods Analysis of Research Contributions, Shortcomings, and Future Prospects. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 401–426. <https://doi.org/10.1145/3531146.3533107>
- [38] Steve Lockey, Nicole Gillespie, and Caitlin Curtis. 2020. Trust in Artificial Intelligence: Australian Insights. *The University of Queensland and KPMG* (Oct. 2020). <https://doi.org/10.14264/b32f129> Publisher: The University of Queensland and KPMG.
- [39] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300789>

- Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3313831.3376727>
- [40] Lev Manovich. 2019. *AI Aesthetics*. Strelka Press Books, Moscow. <https://y.shop.do.strelka.com>
- [41] Aaron K. Martin and Kevin P. Donovan. 2015. New surveillance technologies and their publics: A case of biometrics. *Public Understanding of Science* 24, 7 (Oct. 2015), 842–857. <https://doi.org/10.1177/0963662513514173> Publisher: SAGE Publications Ltd.
- [42] Danaë Metaxa, Joon Sung Park, Ronald E. Robertson, Karrie Karahalios, Christo Wilson, Jeff Hancock, and Christian Sandvig. 2021. Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Foundations and Trends® in Human-Computer Interaction* 14, 4 (Nov. 2021), 272–344. <https://doi.org/10.1561/1100000083> Publisher: Now Publishers, Inc..
- [43] Jacob Metcalf, Emanuel Moss, Ranjit Singh, Emnet Tafese, and Elizabeth Anne Watkins. 2022. *A relationship and not a thing: A relational approach to algorithmic accountability and assessment documentation*. Technical Report arXiv:2203.01455. arXiv. <https://doi.org/10.48550/arXiv.2203.01455> arXiv:2203.01455 [cs] type: article.
- [44] Jessica Morley, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics* 26, 4 (Aug. 2020), 2141–2168. <https://doi.org/10.1007/s11948-019-00165-5>
- [45] Anna Munster. 2013. *An Aesthetics of Networks: Conjunctive Experience in Art and Technology*. MIT Press. Google-Books-ID: uLi5EF9LMCC.
- [46] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* 2 (Jan. 2021), 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- [47] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
- [48] Mimi Onuoha. [n. d.]. What is Missing Is Still There. <https://mimionuoha.com/what-is-missing>
- [49] Christiane Paul. 2008. *New Media in the White Cube and Beyond: Curatorial Models for Digital Art*. University of California Press. Google-Books-ID: kS09MwAACAAJ.
- [50] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3173677>
- [51] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 33–44. <https://doi.org/10.1145/3351095.3372873>
- [52] Ingeborg Reichle. 2009. Art in the Age of Technoscience: Transgressing the Boundaries. In *Art in the Age of Technoscience: Genetic Engineering, Robotics, and Artificial Life in Contemporary Art*. Springer, Vienna, 1–14. https://doi.org/10.1007/978-3-211-78161-6_1
- [53] Dillon Reisman, Jason Schultz, Kate Crawford, and Meredith Whittaker. 2018. *Algorithmic Impact Assessments: A practical framework for public agency accountability*. Technical Report. AI Now Institute.
- [54] Michael Ridley and Danica Pawlick-Potts. 2021. Algorithmic Literacy and the Role for Libraries. *Information Technology and Libraries* 40, 2 (June 2021). <https://doi.org/10.6017/ital.v40i2.12963> Number: 2.
- [55] Gabriel Rockhill. 2014. *Radical History and the Politics of Art*. Columbia University Press. <https://doi.org/10.7312/rock15200> Publication Title: Radical History and the Politics of Art.
- [56] RSA. 2018. *Artificial Intelligence: Real Public Engagement*. Technical Report. Royal Society for the encouragement of Arts, Manufacturers and Commerce. https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf
- [57] Marcus Du Sautoy. 2020. *The Creativity Code: Art and Innovation in the Age of AI*. Harvard University Press. Google-Books-ID: uoPWDwAAQBAJ.
- [58] Neil Selwyn and Beatriz Gallo Cordoba. 2021. Australian public understandings of artificial intelligence. *AI & SOCIETY* (Sept. 2021). <https://doi.org/10.1007/s00146-021-01268-z>
- [59] Donghee Shin. 2022. How do people judge the credibility of algorithmic sources? *AI & SOCIETY* 37, 1 (March 2022), 81–96. <https://doi.org/10.1007/s00146-021-01158-4>
- [60] VID SIMONITI. 2018. Assessing Socially Engaged Art. *The Journal of Aesthetics and Art Criticism* 76, 1 (Feb. 2018), 71–82. <https://doi.org/10.1111/jaac.12414>
- [61] Ramya Srinivasan and Kanji Uchino. 2021. The Role of Arts in Shaping AI Ethics. http://ceur-ws.org/Vol-2812/RDAI-2021_paper_3.pdf
- [62] Luke Stark and Kate Crawford. 2019. The work of art in the age of artificial intelligence: What artists can teach us about the ethics of data practice. *Surveillance & Society* 17, 3/4 (2019), 442–455.
- [63] Grant D. Taylor. 2014. *When the Machine Made Art: The Troubled History of Computer Art*. Bloomsbury Publishing USA. Google-Books-ID: hAMQAwAAQBAJ.
- [64] Chris Tennant, Sally Stares, and Susan Howard. 2019. Public discomfort at the prospect of autonomous vehicles: Building on previous surveys to measure attitudes in 11 countries. *Transportation Research Part F: Traffic Psychology and Behaviour* 64 (July 2019), 98–118. <https://doi.org/10.1016/j.trf.2019.04.017>
- [65] Meredith Tromble. 2020. Ask not what AI can do for art... but what art can do for AI. *Artnodes* 26 (Sept. 2020), 1–9. <https://doi.org/10.7238/a.v0i26.3368> Number: 26 Publisher: Universitat Oberta de Catalunya.
- [66] V&A. [n. d.]. Digital art · V&A. <https://www.vam.ac.uk/articles/digital-art>
- [67] Henriikka Vartiainen, Tapani Toivonen, Ilkka Jormanainen, Juho Kahila, Matti Tedre, and Teemu Valtanen. 2021. Machine learning for middle schoolers: Learning through data-driven design. *International Journal of Child-Computer Interaction* 29 (Sept. 2021), 100281. <https://doi.org/10.1016/j.ijcci.2021.100281>
- [68] Stephen Wilson. 2002. *Information Arts: Intersections of Art, Science, and Technology*. MIT Press. Google-Books-ID: yRW0QgAACAAJ.
- [69] Feiyu Xu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. 2019. Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. In *Natural Language Processing and Chinese Computing (Lecture Notes in Computer Science)*, Jie Tang, Min-Yen Kan, Dongyan Zhao, Sujian Li, and Hongying Zan (Eds.). Springer International Publishing, Cham, 563–574. https://doi.org/10.1007/978-3-030-32236-6_51
- [70] Martin Zeilinger. 2021. *Tactical Entanglements: AI Art, Creative Agency, and the Limits of Intellectual Property*. Meson Press. Google-Books-ID: MKSCzgEACAAJ.
- [71] Zer-Aviv. 2018. The Normalizing Machine | An experiment in machine learning & algorithmic prejudice. <https://mushon.com/tnm/>
- [72] Baobao Zhang and Allan Dafoe. 2019. Artificial Intelligence: American Attitudes and Trends. *SSRN Electronic Journal* (Jan. 2019). <https://doi.org/10.2139/ssrn.3312874>
- [73] Joanna Zylińska. 2020. *AI Art: Machine Visions and Warped Dreams*. Open Humanities Press. <https://library.oapen.org/handle/20.500.12657/40042>