Experiential AI: Enhancing explainability in artificial intelligence through artistic practice

Experiential AI: Enhancing explainability in artificial intelligence through artistic practice

Drew Hemment
Dave Murray-Rust
Vaishak Belle
Ruth Aylett
Matjaz Vidmar
Frank Broz

This is a preprint version of an article submitted to a journal for peer review.

Abstract

In this paper, we present a methodology that bring experiential methods to bear on the challenge of developing understanding of AI systems – their operations, limitations, peculiarities and implications. We describe an approach that uses art and tangible experiences to communicate black-boxed decisions and nuanced social implications in engaging, experiential ways, with high fidelity to the concepts. In this approach, that we call Experiential AI, scientists, artists and other interdisciplinary actors come together to understand and communicate the functionality of AI and intelligent robots, their limitations, and consequences, through informative and compelling experiences.

We propose that experiential methods offer significant contributions to both intelligence and interaction in the design of interactive intelligent systems for explainable AI. We specifically look at strategies and methods in the AI arts that offer new modalities of explanation for human-centred explainable AI, and reframe explanation as a more holistic form of understanding. This leads us to the hypothesis that art and tangible experiences can mediate between impenetrable computer code and human understanding, making not just AI systems but also their values and implications more transparent and legible. Through three case studies, we develop insights on inclusivity, empowerment and responsibility in machine intelligence and user interaction. We go on to present new methodology for the design, development, and evaluation of human-centred explainable AI, and argue that legible intelligent systems need to be open to understanding and intervention at four levels: Aspect, Algorithm, Affect and Apprehension.

1.
Introduction: An experiential approach to the explainability of AI systems

We are seeing a step change in the number of people both currently and potentially impacted by artificial intelligence (AI) and automated or semi-automated decisions. The explosion in the use of data -driven approaches such as Machine Learning is a key driver of this change. Discovering patterns in troves of data in an automated manner is a core element of data sciences, and drives applications in diverse areas such as computational biology, security, law and finance. However, everything from the way data is collected, the labelling and cleaning processes, and both training and testing regimes profoundly determine the type of decision-making deployed, and its impacts on end-users.

At the same time, the move to data-driven systems has increased their opacity. Earlier goal-driven systems could at least articulate in some form what their goal was, where data-driven approaches require a detailed understanding of the often dynamic data context in which they operate. Moreover powerful and widely applied algorithms such as deep learning encode the system's knowledge implicitly and in a distributed fashion so that even experts may not easily be able to determine what the system 'knows'.

Explainable AI (XAI: Gunning, 2017) – as initially positioned – investigates how the decisions and actions of machines can be made explicable to human users. Such work is important, as for interactive intelligent systems to be inclusive, empowering and responsible, they need to be legible and contestable. While significant advances have been made in XAI (Arrieta et al 2019, Belle and Papantonis 2021), to date it has predominantly targeted a narrow range of expert users, and it struggles to generate the kind of explanations needed from a human point of view. A person will want to know why there was one decision and not another, the causal chain, not an opaque description of machine logic. There is also the more general question of the nature of explanation, and more work to be done to account for how explanation works as a social practice - in particular, the idea that explanation is a process, not an artefact, and arises from an ongoing engagement (O'Hara 2020).

In this paper we aim to complement and enrich work on explanation by advancing a novel approach to the challenge of making data-driven AI and machine learning tangible, interpretable, and accessible to the intervention of end users. Here, we are not concerned only with the internal operations of algorithms. We are also concerned with opening up algorithms, the science behind them, and their potential impacts in the world to user intervention, public scrutiny and policy debate. This can complement existing work in XAI that traces details of an algorithm's functioning, by making tangible and illuminating underlying assumptions of machine learning models, processes that generate their data, or the social context in which automated decision making is situated.

The primary challenge to usability and acceptability to which our research responds is the inscrutability of intelligent systems, and we are primarily interested in understanding better ways of interacting with existing intelligent technology through more holistic and experiential explanation. To meet this challenge we propose that experiential methods can generate more holistic and meaningful ways of explaining AI. Long-established theories of experiential knowledge describe it as is a function of type of acquired information and one's attitude towards said information (Borkman, 1976). Hence, experiential methods for acquiring knowledge (Kolb, 2014) entail observation of and interaction with information or events. We aim to show that experiential methods in the arts are particularly relevant for a human-centred approach to XAI, and to enhancing inclusion, empowerment and the responsiveness of interactive intelligent systems situated in social settings.

A precursor to this work is Experiential AI (Hemment et al, 2019) which gave the outline of the multi-disciplinary perspective that we articulate and develop in this paper. Turning to the arts for inspiration, we develop a novel approach for designers of interactive intelligent systems that addresses important gaps in intelligibility. We draw on the arts as a novel way to place people at the centre of research and development, in order to develop systems that better meet the needs of users. We specifically ask whether and how artistic narratives, interaction and performance can help to explain how automated decision making is situated, as well as what the model actually does. We look at how such methods can help interacting humans to viscerally understand the complex causal chains in environments with data-driven AI components, including questions about: what data is collected, its nature, accuracy and freedom from bias, as well as who collects it; how the algorithms are chosen, commissioned and configured; and how humans are conditioned by their participation in algorithmic processes.

In this paper we report on work towards an experiential and holistic approach to machine intelligence and user interaction that can underpin new paradigms for human-centred XAI. We describe early stage empirical research on how the arts can answer questions for the AI community, with a specific focus on inclusivity, empowerment and responsibility in explainable AI. We present our early stage explorations, in which artists and other critical practitioners worked with scientists and engineers to create the scenarios in which ML systems, social robots or other technologies can be deployed and tested as experiences, in the form of interfaces, installations, performances, situations, interactions. We present and discuss three case studies that each provide an 'explanation' of a kind on themes that go beyond a narrow account of model interpretability to address the operation and implications of XAI in real-world settings: usability of intelligent interactive XAI tools, bias and inclusion in training data, and hidden human labour in automated systems. Through our case studies, we develop insights on the design for the needs of user groups that include non-experts (inclusivity), the actionability of AI when it becomes one commodifed component in complete systems (empowerment), and the fairness, accountability, transparency, and ethics of AI when situated in unpredictable social settings (responsibility). The main contribution of this paper is novel methodology developed through reflection on the co-creation process in these pilots. This research is ongoing and our future studies will further evaluate and aim to better understand the potential of experiential explanations. The ambition is to make a distinctive contribution to human-centred XAI, and to open up the AI field to greater understanding and collaboration between human and machine.

2.
Current advances and future directions in XAI

There is not yet a shared definition of explainability within the AI community. For the purposes of this paper, we take explainability – abbreviated XAI for short – to denote interpretable features within an algorithm that enable decisions to be justified, tracked and verified by a human (Montavon et al, 2018). In XAI, models attempt to give a human understandable account of their operation, to make their reasoning more transparent, build trust and allow humans to hold them to account. By and large, the vast majority of advances in XAI are technical in nature (Arrieta et al 2019, Belle and Papantonis 2021), and do not illuminate societal, political, cognitive, or regulatory aspects. Notable exceptions (such as Rudin 2019; Raji 2020) expose the dangers of blindly trusting standard accuracy measures in critical applications, for example. Some strands of research focus on using simpler models (possibly at the cost of prediction accuracy), others attempt ”local” explanations that identify interpretable patterns in regions of interest (Weld & Bansal, 2018; Ribeiro et al, 2016), while still others attempt human-readable reconstructions of high-dimensional data (Penkov & Ramamoorthy, 2017; Lake et al. 2015; Belle, 2017). A handful of approaches are now attempting the explicit construction of a user model, and offering a reconciliation of the machine’s assumption and the user model (Chakraborti et al., 2019; Kulkarni et al., 2019).

Perspectives in social science assert that explanations are primarily knowledge-producing in attempting to respond to the inquiry, and are minimal (focusing on the relevant entity) and contrastive (why this and not that) (Miller, 2019). Formulations from fields such as causality, information theory and statistics are plentiful, but a single solution is substantially elusive. Nonetheless, progress in the field has been exciting, and it is clear that explainability can facilitate the understanding of various aspects of a model, leading to insights that can be utilized by various stakeholders. Moreover, beyond improving human comprehensibility, research has shown that explainable techniques – in a banking application (Belle & Papantonis, 2021) – help to debug the model and check for robustness, correctness and bias, but also more generally to check for strategic fit, find possibilities for model improvement and transferability to other domains.

However, while there have been impressive advancements in XAI on local and counterfactual explanations, and model simplification strategies to provide proxies or otherwise isolate the underlying function to some extent, there is an urgent need to understand the entirety of AI systems. This includes not just the models at their core, but the data collection and processing that gives rise to them, the way the system has been commissioned and designed, and the relations between the system and the subjects of its decisions. As is now widely acknowledged, most machine learning systems cannot be disentangled from their data sources (Raji, 2020), raising privacy and ethical concerns, leaving the user with the burden of understanding this. This is all the more true of social robots that are composites of multiple models and directly act in the physical world.

While there are significant advances, current work in XAI addresses explainability as primarily a lower-level technical problem, and does not account for the higher-level – system-level, cognitive, political, legal, regulatory or institutional – aspects of AI. A Royal Society briefing argues that this is only the first step in creating trustworthy systems, and there is a need "to consider how AI fits into the wider socio-technical context of its deployment", in addition to explanation (Royal Society, 2019). In Seeing Without Knowing, Ananny and Crawford argue research needs not to look within a technical system, but to look across systems and to address both human and non-human dimensions (Ananny & Crawford, 2018). They call for "a deeper engagement with the material and ideological realities of contemporary computation" (Ibid.). However, it is not clear that abstract models can incorporate cultural and sociopolitical norms in a straightforward manner. This calls for moving beyond mainstream understanding of explanation in AI, to create deeper understanding of the nuances of socio-technical systems.

The design of interactive intelligent systems for explainable AI needs to account for a wide range of stakeholders, and yet to date it has predominantly targeted a narrow range of expert users. We recognise that, while explainability itself is a useful ideal, there are limitations, including that seeing inside something is not the same as understanding it, and that it places substantial burdens on people to seek out information and make sense of systems (Ananny & Crawford, 2018), mutating social into personal responsibility. Moveover, such work often struggles to generate the kind of explanations needed from a human point of view: the causal chain, not an opaque description of machine logic. A stakeholder won't always need to know in detail how black-box AI works, but will instead want to understand their limitations, and when their outputs can be trusted. Even when explanation can be provided, this may not always be sufficient (Edwards & Veale, 2017) and more intuitive solutions are required to, for example, to understand the changing relations between data and the world, or integrate domain knowledge in ways that connect managers with those at the frontlines (Veale et al, 2018) and that supports engagement by a broad segment of society.

Many people whose lives are affected by AI are not aware of its functions or complexities, and often not aware of how they are being affected by AI in the first place. As with many technological innovations, there are strong imaginaries around AI (Bory & Bory, 2015) seen in the breathlessly optimistic utopias and grim meathook dystopias of mainstream media, that can get in the way of meaningful public engagement with the direction of technology development. AI and robots are the subject of widespread illusions, for example that a machine 'belief state' is comparable to a human mental state. Anthropomorphism and an intentional stance can lead to unrealistic fears and even aggressive behaviour towards a robot or interface (Brščić et al, 2015). It can also produce an equally unrealistic fascination and over-imputation of authority (Robinette et al, 2016), and overconfidence by policy-makers about the actual capabilities of the technology based on their own limited understanding (Forrest, 2021). Alongside the sometimes exaggerated claims of current or immediate-future capabilities (Marcus & Davis, 2019), a broader set of fears about negative social consequences arise from the fast-paced deployment of AI technologies and a misplaced sense of trust in automated recommendations (Wagner et al, 2018). While some of these fears may themselves be exaggerated, negative outcomes of ill-designed data-driven machine learning technologies are apparent. Engaging with the imaginaries and misrepresentations that surround AI can help to "contest unduly optimistic visions that gloss over the potential harmful effects of technological" (Chan, 2021) and better orient discussion that navigates the myths surrounding the technology (Natale & Ballatore, 2020).

As AI-enhanced applications are increasingly deployed and situated in social settings, actionable insight on both their operation and implications can be communicated through experiential methods. Social robots are essentially interactive and thus themselves inescapably experiential; they are instances of embodied AI systems that

operate physically on the world and result from complex combinations of AI algorithms with advanced engineering. The quotidian experiences of daily AI that is aware of humans (Kambhampati, 2020) provide a rich ground for the design, development, and evaluation of human-centred XAI, and for interrogating intelligence and interaction in deployed systems. There are steps towards reframing explanation in a more holistic way, for example the Living Room of the Future (Sailaja et al, 2019) that makes experientially stark issues of agency, legibility, privacy and trust. Examples from human-robot interaction demonstrate that models of how we interact with technology that don't take experience into account may be incomplete or incorrect (Smedegaard, 2019). Such transdisciplinary practice can also be applied to interrogate the distinctions between artificial and augmented intelligences (Carter & Nielsen, 2017), and can help to advance both the science and the art of human-centred machine learning (Fiebrink & Gillies, 2018).

The potential contribution of the arts to explainability remains largely untapped, although there are promising developments around the use of generative models to explore the ethics of AI (Srinivasan & Parikh, 2021). Research on experiential learning (Kolb, 2014) confirms that experiential methods can act as a powerful pedagogic mechanism. Experiential methods have been shown (Vannini et al., 2011) to work better than a purely knowledge-based or data-driven approach, and are known to be more effective than merely providing information or logical arguments – showing rather than telling, to create deeper understanding. Experiential learning uses sensory and affective engagement to dramatise concepts, promote the freedom to act and explore the consequences, and generate reflection, by embedding relevant experiences in a story-world through narrative, and especially role-play, as for example Boals Forum Theatre (Boal, 2013). A range of methods can be used to create experiential learning settings and these methods can lead to outcomes such as user adaptation (Lim et al, 2018).

3.
Current directions in AI arts and explanation

While a full survey is beyond the scope of this paper, much work in AI arts has an explanatory aspect. Over recent years, artists and cultural institutions have increasingly come to experiment in AI, and several high profile programmes are testament to the fertility of this field (Zentrum fur Kunst und Medien, 2018; Barbican, 2019; Onassis Foundation, 2021). Recent advances in machine learning have made these tools more accessible to artists. The artist Memo Akten explores the creative affordances of AI and ML (e.g. http://www.memo.tv/works/learning-to-see/), often with the specific aim of making the operational logic, functional limits, and socio-economic implications of these technologies graspable for wider audiences. In Learning To See, an audience is invited to move everyday objects placed before a camera, and observe artificial but natural-looking images that result, to gain a direct experience of the so-called 'latent space' of neural networks. This visualisation is quite obviously based on training biases derived from the neural networks' computational protocols and infrastructure; in other words, the rendered video output inevitably represents something that was contained in the AI systems perceptual register.

Going beyond the typical human+computer view, artists are questioning the construction of prejudice and normalcy (Zer-Aviv, 2018), or working with AI driven prosthetics, to open possibilities for more intimate entanglements (Donnarumma, 2018). Accountability is variously addressed. Joy Buolamwini works with verse and code to challenge harmful bias in AI (see https://www.poetofcode.com), while the artist Trevor Paglen constructs a set of rules for algorithmic systems in such a way as to uncover the character of that rule space (see http://www.paglen.com). In ImageNet Roulette (https://imagenet-roulette.paglen.com), Trevor Paglen, and AI researcher, Kate Crawford, collaborated to produce an experience for an online and gallery audience that exposed the problematic outcomes when AI models are trained on undesirably biased datasets, accompanied by an essay discussing themes raised in the work (Crawford & Paglen, 2019). Users were invited to upload images, an application detected and classified faces in the submitted images, and the results included offensive, bizarre and problematic categories assigned to people's faces. The presentation of the resulting artwork generated high levels of public participation and led ImageNet and similar public image repositories to purge images from their datasets after the art project revealed their problematic bias (Ibid.). The artists then took the artwork offline, considering its work to have been done. It can be questionable how representative imagery generated by ML is of deep network structures, or whether it is a happy accident in machine aesthetics. Critique of ImageNet Roulette has questioned the artists' representation of the machine learning dataset, the ethics of the artwork itself in reproducing those results, and its compliance with the terms of use for the images (Lyons, 2020).

Taken in the round, we also see instances of artistic practice that serve to obscure rather than articulate, or amplify a misconception rather than demystify. Some artistic projects work with blackbox conceptualisations, that can be a long way from the real technology. Nonetheless, what we can say is that such works enable the artifacts of machine reasoning and vision to be made tangible, and create a concrete artefact or representation that can be used to illustrate attributes and concepts, and as an object to spark further enquiry. The ImageNet Roulette artwork illuminated in an accessible way concepts of transparency and fairness in AI, and enabled people to directly perceive the personal consequences of problematic bias in AI systems. Such works articulate high level attributes such as bias in tangible ways, and offer an exploratory rather than didactic mode of 'explanation' on the operation and implications of AI.

Figure 1. Fondazione Prada, 2019. IMAGE-NET ROULETTE, “Training Humans” Osservatorio Fondazione Prada. Photo Marco Cappelletti.

4.
Developing Experiential AI case studies

Experiential AI seeks to make the opaque mechanisms of AI artefacts and algorithms transparent to those who interact with them, in order to restore the basis for accountability and increasing the range of people engaged in shaping the field. The methodology has been developed through co-operative research inquiry (Heron and Reason, 2001) involving a cross-disciplinary team of professional artists and XAI, applied ethics and design scholars. The Open Prototyping design framework (Hemment, 2020) provided a six-stage process (Scope, Connect, Play, Produce, Display, Interpret) represented in Figure 2 to configure data, algorithms, models, people and situations to make explanations tangible as experiences. Between Summer 2020 and Winter 2022, three data-driven art projects were developed as case studies (Yin, 2009; Flyvbjerg, 2013) of AI Art from which strategies and recommendations for an experiential and holistic approach to explainability, and for greater inclusivity, empowerment and responsibility in XAI, could be derived. In each of the case studies, we examine the intelligence that extracts patterns from observed data to make predictions or judgments, and the interactivity by which users act directly on the model, or are supported in their action by the system, in ways that makes that system more intelligible. Oftentimes, the interactivity in the case studies is of the latter type, but the insights derived through the studies are shown to be of general relevance to the interactive intelligent systems field. We then reflect on these aspects of the work as well as the co-creation process through which they were developed to develop a novel conceptual framework and process model. Methods included semi-structured workshops and interviews, conceptual and creative design of cultural experiences, image processing with Generative Adversarial Networks (GANs; Goodfellow et al., 2014), word to vector transfers, linear discriminant analysis (Bishop, 2006), thematic analysis of qualitative data to build insights, and quantitative experiments to validate algorithms.

Experiential AI seeks to make the opaque mechanisms of AI artefacts and algorithms transparent to those who interact with them, in order to restore the basis for accountability and increasing the range of people engaged in shaping the field. The methodology has been developed through co-operative research inquiry (Heron and Reason, 2001) involving a cross-disciplinary team of professional artists and XAI, applied ethics and design scholars.

The Open Prototyping design framework (Hemment, 2020) provided a six-stage process (Scope, Connect, Play, Produce, Display, Interpret) represented in Figure 2 to configure data, algorithms, models, people and situations to make explanations tangible as experiences. Between Summer 2020 and Winter 2022, three data-driven art projects were developed as case studies (Yin, 2009; Flyvbjerg, 2013) of AI Art from which strategies and recommendations for an experiential and holistic approach to explainability, and for greater inclusivity, empowerment and responsibility in XAI, could be derived.

In each of the case studies, we examine the intelligence that extracts patterns from observed data to make predictions or judgments, and the interactivity by which users act directly on the model, or are supported in their action by the system, in ways that makes that system more intelligible. Oftentimes, the interactivity in the case studies is of the latter

type, but the insights derived through the studies are shown to be of general relevance to the interactive intelligent systems field. We then reflect on these aspects of the work as well as the co-creation process through which they were developed to develop a novel conceptual framework and process model.

Methods included semi-structured workshops and interviews, conceptual and creative design of cultural experiences, image processing with Generative Adversarial Networks (GANs; Goodfellow et al., 2014), word to vector transfers, linear discriminant analysis (Bishop, 2006), thematic analysis of qualitative data to build insights, and quantitative experiments to validate algorithms.

(i) Viewpoints and definitions workshop (Open Prototyping: Scope Stage)
Two 2.5 hour workshops in Summer 2020 brought together practitioners and researchers from XAI, AI arts, applied ethics, and design research were audio-recorded and transcribed. Presentations were developed to review and establish thematic areas, frame the research questions, and present strategies from the arts to make computational intelligence tangible and explicit. A semi-structured conversation then invited responses from different disciplinary viewpoints, to explore how these methods and concepts operate as modes of explanation, and the alignment or divergence of definitions in XAI and the arts.

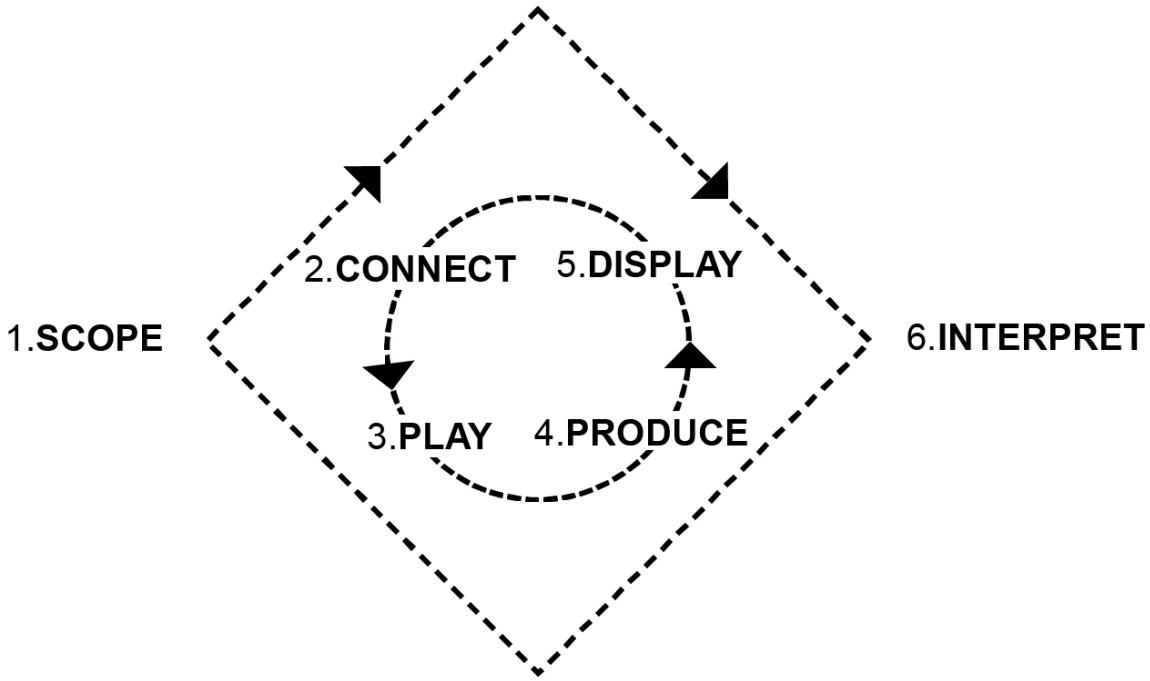


Figure 2: Open Prototyping Process Model: A graphical representation of a six-stage process, with an iterative process of knowledge creation represented in an external diamond, and a holistic process of artistic and technical development represented on an inner circle.

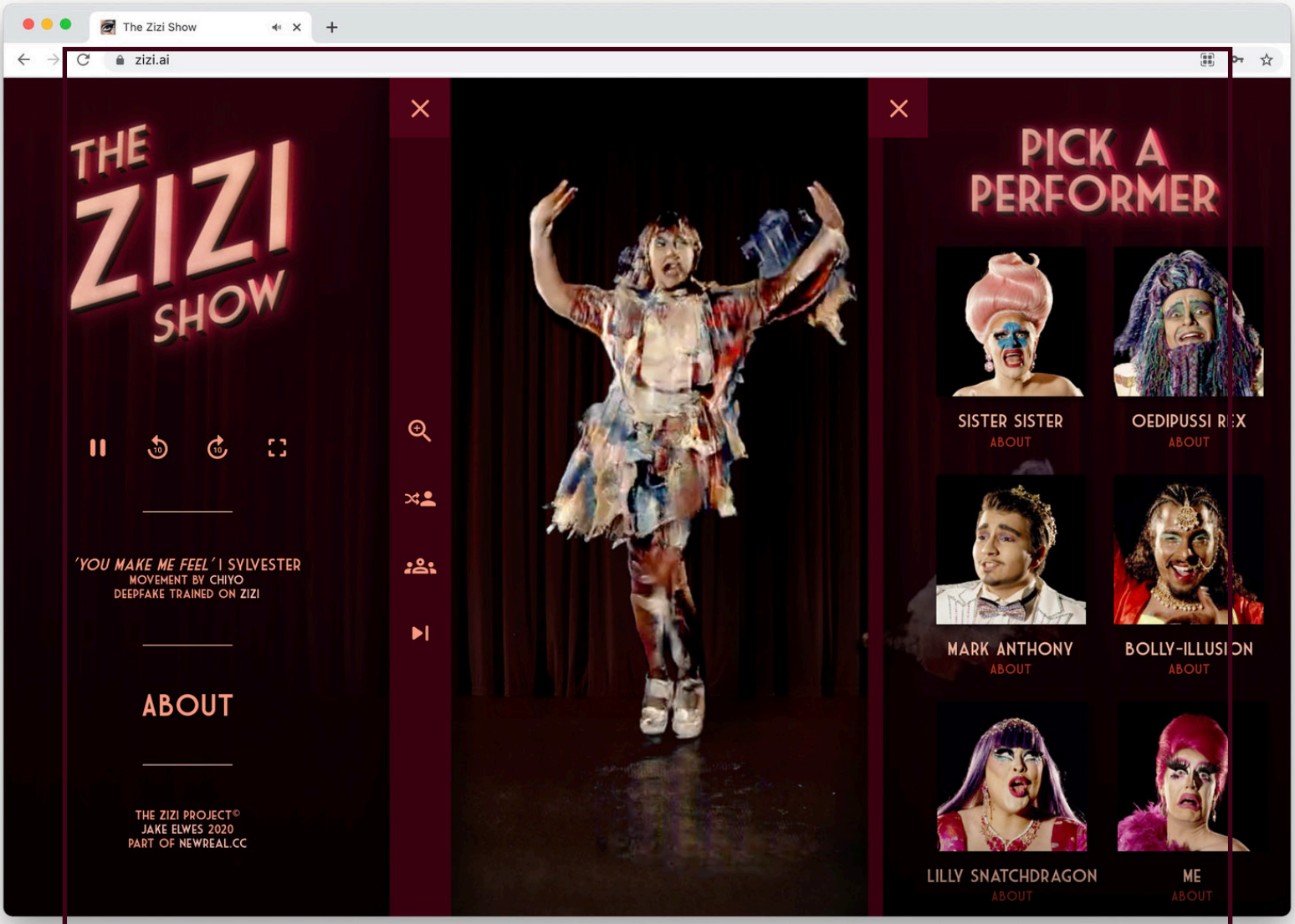
(ii) Data, model and algorithm (Open Prototyping: Connect & Play Stages)
In three iterations of the methodology, during Spring, Summer and Autumn 2021, a complete Experiential AI system was developed for each case study, reflecting these shared definitions, and combining off the shelf, commodified technologies with bespoke development. The artists were the primary users of the technology, and in some cases also developed their own solutions. In each case, the emphasis was on practical, creative real-world applications over the novelty of the algorithm design. All of the case studies involved GANs, and the curation and development of data by the artists. In one case, a 'science and technology team' identified XAI techniques and interpretable features in the algorithm to which the artists could respond, and co-designed with and for the artists data pipelines and AI processing engines. In two cases, the artists themselves developed explanatory concepts and methodologies, and led development teams to integrate data, model and algorithm. A platform approach was established through the case studies, where pre-trained algorithms were offered fro fine-tuning, examination and experimentation, coupled with accessible data streams, relevant to the theme of the project.

(iii) Design, development and presentation of digital experiences and data capture (Open Prototyping: Produce & Display Stages)
The artists conceived and developed digital experiences responding to a distinct theme for each project, and the common problem of making computational intelligence explicit and legible. The use of glitches, processing limitations and interrogation of publicly available datasets allowed for design of experiences that address the materiality of data and AI, as well as enable a core interrogation of machine learning from data, whilst not requiring advanced knowledge of AI for artists involved. The experiences were presented to online audiences in April 2021 (with Edinburgh International Festival), June-July 2021 (Edinburgh Science Festival), September 2021 (Ars Electronica) and October 2021 (at COP26 UN Climate Change Conference). Data was collected across these pilots on user/audience engagement, their direct feedback and follow up interviews, as well as expert analysis of the design process, AI and XAI integration involved, cognitive shifts for users of the experiences, and the implications of the works for reflection on critical emerging themes of accessibility and literacy.

(iv) Analysis and dissemination (Open Prototyping: Interpret Stage)
Data was collected, analysed and synthesised to generate insights. Early stage results are reported in this paper. In addition, wrap-around activities were delivered to generate engagement in the research themes, including conference talks, festival events, and an online magazine of accessible 'think pieces' for both AI and cultural audiences.

This process was iterative, and there were loops through different stages of the Open Prototyping process in each case study. The separation between stages was not always neat, and it was observed that creative and technical development often proceeded hand-in-hand.

Based on this process, three case studies were created that engaged with the development of intelligence, interactivity and explanation in AI systems. In each case, we look at the theme and manifestation of the work, the potential for interactivity and give a short summary of learnings and findings that relate to public understanding of AI.



4.1. Case Study
The Zizi Show by Jake Elwes

The Zizi Show (Figure 3), by London-based visual artist Jake Elwes, is an online interactive artwork in which a GAN has been trained on video footage of thirteen diverse 'drag' performers, filmed at a London cabaret venue during the COVID19 lockdown. This work, builds on previous work (simply titled Zizi) that exposes the latent space of the ML model, and highlights the way the model outputs are shaped by the training data. Where many generative works have been trained on opportunistically collected data, the purposeful curation of Zizi's dataset explores the question of how human identity is represented within complex models. The Zizi Show develops this through digital avatars, that have been learned from real performers to create an interactive work that allows user control. Significantly, it connects low level technology to high level, social, cultural and political aspects of AI, such as ideas of cultural appropriation and machine bodies. It exposes the limits to machine intelligence, and inverts what is otherwise a deficiency in the technology, through a positive use of deep fake technology, in which a marginal identity is celebrated and embellished, rather than obscured or misrepresented.

Figure 3: The Zizi Show (Jake Elwes, https://zizi.ai/). An algorithmically generated compere asks the audience to select performers and songs. Each performer has a body blended from video capture of drag performers that morphs and changes as they perform each work. All Rights Reserved © Jake Elwes 2021.

4.1.1 Theme:
Bias in ML data, and misrepresentations of AI.

4.1.2 Intelligence:
The project engages with the current wave of machine learning techniques, using a StyleGAN network architecture re-trained on a modified version of Flickr-Faces-HQ (FFHQ) dataset, to which an additional 1,000 portraits were added, alongside a custom video, sound and interactive web interface. Machine learning here allows the creation of a generative space that includes bodies and faces.

4.1.3 Interactivity:
Zizi rethinks what interactivity is, at scale, and enables us to ask what forms of interactivity are important. In Zizi, the artist interacts with the model by manipulating data and weightings. The audience do not interact with the model directly, but with its artefacts and outputs. Moreover, they are able to do so at scale, as it has been designed for large numbers of simultaneous users. The audience view the output in different settings, and are able to select from a menu to switch between AI generated personas of drag artists for different music tracks. This, however, is representative of the interactivity experienced by a majority of users of intelligent systems. Many end users access the outputs of AI systems through interfaces that are not themselves intelligent.

4.1.4 Explanation:
Zizi is an explanation of bias in ML and the power of the dataset through experiential means. Zizi highlights the way data and design choices shape what ML does. It shows how the model learns a representation of people, that is embedded social life. Zizi engaged a marginalised group, developing their literacy surrounding bias in ML, thereby supporting their agency in contesting its fairness and accountability. Zizi shows end users there is something to contest, even if that do not interact directly with the model themselves. Zizi specifically targets anthropomorphised misrepresentation of AI, by constructing an AI persona, and then deconstructing it, and exposing its construction in software by the human artist.

Figure 4: The Zizi Show. All Rights Reserved © Jake Elwes 2021.

Figure 5: Cypress Trees (Anna Ridler and Caroline Sinders, https://ars.electronica.art/newdigitaldeal/en/cypress-trees/), showing data collection (top), dataset display (bottom). All Rights Reserved © Anna Ridler & Caroline Sinders 2021.

4.2. Case Study
Cypress Trees by Anna Ridler and Caroline Sinders

How can AI help us to face the climate crisis and other entwined challenges? This machine-learning generated moving image piece (Figure 5) gives insights into the complexity of data sets and raises questions about deforestation and the politics of climate change, memory and loss. Anna Ridler and Caroline Sinders created a special dataset of the Bald Cypress on the gulf coast of the USA, where both have family ties. These trees, which can live thousands of years, are currently considered to be “threatened” by climate change. Rather than for problem-solving, this human-machine intelligence is applied to produce imagery and gallery installations that represent the ordering of knowledge by AI. The artists in this sense ‘perform’ a part of the machine learning algorithm. They turn a foundational definition in AI on its head, by the human artist doing a task usually done by the computer and associated with machine intelligence.

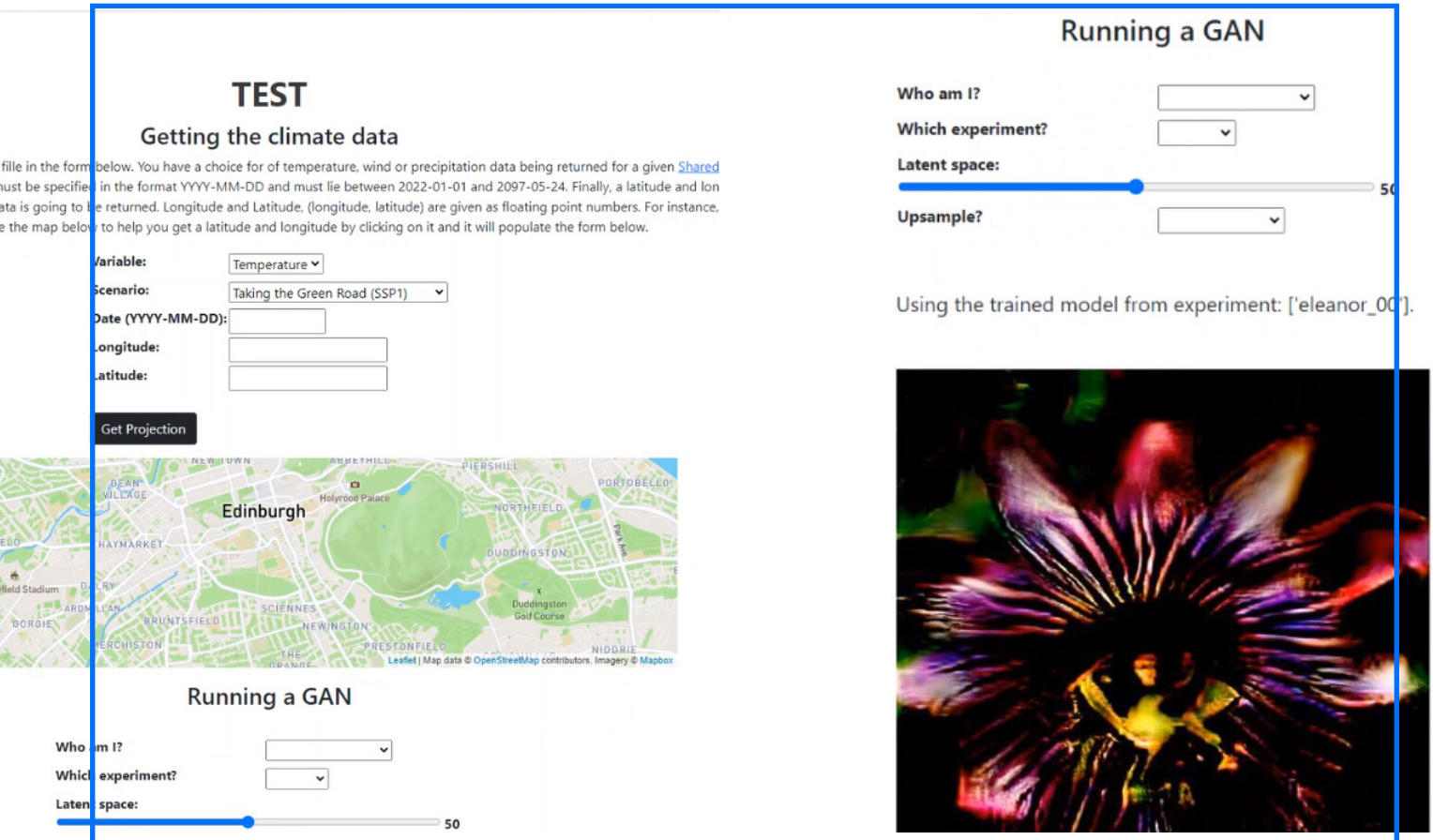
4.2.1 Theme:
Hidden human labour in ML

4.2.2 Intelligence:
The artists painstakingly and meticulously extract patterns from observed data using manual methods in order to make judgements. The piece involves a generative network that produces images of cypress trees. However, the heart of the piece is the collection and curation of the datasets that this model requires, drawing on the situational, embodied nature of machine learning systems.

4.2.3 Interactivity:
It foregrounds the painstaking work by the artists to develop a bespoke data, through photography of hard to access trees, to labelling and cataloguing. The 'interaction' for an audience is to observe and move between artifacts the artist has curated and positioned in a gallery.

4.2.4 Explanation:
The project highlights that what we think of as ML intelligence is actually human intelligence at many points in the system. It explains the labour in ML systems, often performed by so-called click workers, or the call centre workers having to listen to user responses. The way the artists build their own data in a very meticulous and painstaking way points to wider issues of hidden human labour in ML data. It debunks the neat representations of 'autonomous' systems, and by forcing an attention to hidden labour, raises wider questions around human bias and worker exploitation.

Figure 6. Initial experiements with GAN output (bottom right). All Rights Reserved © Anna Ridler & Caroline Sinders 2021.



4.3. Case Study
The New Real Observatory

The New Real Observatory (Figure 7) is an interactive intelligent system that integrates climate data and models and AI algorithms, developed with artists and for artists. Its primarily objective is to expose and explore through artistic outputs the the link between global climate information and 'ground truth', and to offer a world-wide audience personal encounters with environmental phenomena beyond human scale. Locally situated interactive art, sound, movement and play are generated in dialogue with global climate data and processed using pre-configured AI engines. A pilot experience, AWEN, was presented at Edinburgh Science Festival and at COP26, combining game design and visual art to test ideas to integrate in the platform. Now, the New Real Observatory sources global information on a selected number of climate features (temperature, precipitation, wind) using satellite data and Copernicus Climate Data Service, including forecasting scenario modelling. This is combined at the platform with processing engines including GANs and word to vector transfers, to manipulate images, words, sounds and numbers using the climate data and forecasts as the exploratory parameters altering present into future realities. Five artists were commissioned to test different ways to make AI processing and climate information explicit, combining XAI with artistic strategies, each one developing a prototype experience testing the platform.

Figure 7. The New Real Observatory (by The New Real, www.newreal.cc/thenewrealobservatory), artists interact with the system by providing a geographical coordinate and a date to obtain a value for a future climate parameter, and then curating and uploading an annotated image library to train a GAN with a weighting provided by that future parameter.The latent space is then explored through AI generative images (GAN test sample, bottom right). Images: University of Edinburgh.

4.3.1 Theme:
Make AI accessible to artists, and make data processing in climate information legible to audiences and bridge the global-local data divide.

4.3.2 Intelligence:
The New Real Obseravtory highlights machine intelligence outside the current trends of generative networks and deep learning, specifically looking at planetary scale AI considerations (Bratton, 2016). Rather than centering on a single model, here predictive modelling is combined with an ecology of data flows, and then brought together around small interactive AI processes that produce images, words, numbers and sounds as responses to future possibilities and user input.

4.3.3 Interactivity:
The artists are able to interact with the model by, firstly, changing a future climate parameter that controls one of the weights for the model by specifying location and date, and, secondly, by curating and annotating data on which the model is trained. The platform has been used in five pilot projects with artists, and in future will be developed as a plug-and-play tool. In future iterations, audiences will be able to interact with the model by matching real-time (or past) GPS tracking with climate data-points, and by submitting new images, which are brought into the functioning of the piece. The platform generates visual, syntactic, audio or numerical outputs that the artists can use as material for art pieces, and, by extension, experiential explanations.

4.3.4 Explanation:
The prototypes built on The New Real Observator platform enable audiences to tangibly experience the operation of a machine learning algorithm as a means to alter possible futures and thus expose the learning principles behind the future casting used in modelling with data. This particularly highlights the possible discrepancies and local variations in interpretation

Figure 8. The platform was developed through co-creation with artists on the AWEN pilot experience (https://awen.earth/), showing user interaction.

5.
Discussion: Surfacing strategies to improve inclusivity, empowerment, responsibility.

Our early stage results indicate these works are highly effective in making AI and machine learning tangible and interpretable for the artists and their audiences. Our future research will seek to better understand the cognitive shifts for users, but here we discuss the implications of these strategies for inclusivity, empowerment, responsibility, and then go on to develop novel methodology through reflection on the co-creation process.

We see that the AI arts offer a set of methods and resources to address explanation in an experiential way. These explorations have surfaced ways in which art and creativity can make AI systems transparent and intelligible to users. The case studies reveal ways in which narrative, visual art, interaction, performance and game design can be used to probe, explore and communicate significant aspects of technology in highly imaginative and engaging modalities of explanation.

The three case studies each entail a complete AI system that generates an experience for an audience. In each case, there is a high level aspect of the ML system that is being made explicit and communicated: the bias in the dataset, the hidden human labour, and the latent space of the model. The explanation is baked into the design of the experience, either through the curation of data, the design of the algorithm, or the way the components are connected.

In each case, the 'explanation' is not a technical account of the model or algorithm, so much as a creative representation of higher level aspects of an AI system situated in the world, and its attributes or implications. The strategies in these studies do not present 'explanations' as such. Indeed, the arts are not instruction, and it would be wrong to instrumentalise the arts for system design. Nonetheless, we can derive from the case studies strategies and methods which can add to the toolbox for human-centred XAI. We further envision future projects might combine the two by using the low-level approaches along with higher level aspects to further clarify or otherwise illustrate what is going on.

The arts make a merit of leaps of imagination. Here we look for explanatory methods that are scientifically rigorous, with high fidelity to the concepts. At the next stage in our research we will evalate the experience and evaluation to understand cognitive shifts in users, and test the validity of the algorithms through quantitative experiments.

Below we reflect on the design and development journey to build insights on inclusivity, empowerment and responsibility that may be of relevance to designers of interactive intelligent systems for explainable AI:

5.1 Inclusivity

The case studies suggest strategies that can help to make AI more accessible and help to reach those currently excluded from the creation and deployment of systems. These projects engage and illuminate a vital class of users who are often taken for granted and who can be exploited by common business models and practices, namely click workers, system moderators, and other data contributors. They also demonstrate that it is possible to engage demographic groups who are underrepresented in training data in the design and evaluation of these systems. More broadly, in creating artistic experiences for the general public, these projects develop explanatory experiences for the majority of people who are impacted by intelligent systems, namely, non-experts.

Interactive explainable AI technologies and systems are often designed to meet the needs of system designers and of end users in domains such as healthcare and finance who use AI systems for decision making. These can be, literally, life and death decisions, and so this is vital work. However, as AI technologies become increasingly commodified, and used as components in complete systems, the range of users and contexts of user only increased. This can range from the driver of an autonomous vehicle, to a healthcare patient, a consumer of financial information, or an operator of a commercial wind farm who depends on seasonal-to-decadal forecasts derived from an ensemble of climate models.

The case studies bring into view ways to engage and account for a broader sweep of end users and people who may encounter and depend on automated decisions. Future work will connect these artistic practices with mainstream AI developments (see for example Bhatt et al, 2020).

5.2 Empowerment

Through these real-world applications, the case studies illuminate the question of actionability when AI is one such component in complete systems. As well as reaching more people, the artworks suggest strategies to improve actionability for specific groups. They illustrate how to improve user actionability for non-experts who do not directly interact with the model and so are not supported by mainstream XAI. Likewise, these projects engage and empower marginalised groups. The case studies demonstrate innovative strategies to engage those excluded groups and to incorporate their voices in design. These strategies can be added to the toolbox of system designers, and help them to optimise around a broader range of user needs.

The design and evaluation of interactive intelligent systems can learn much from the imaginative strategies employed by artists to develop critical literacies in their audiences. Here, the interaction is designed and the information is consumed in ways that may be more similar to the ways people engage in entertainment or social media. Hence the design of interfaces to algorithmic systems by artists can have wide relevance.

Artists are themselves a user group whose voice needs to be heard in designing explainable intelligent systems for improved actionability. In some cases, the artists themselves integrated XAI technology and techniques in bespoke systems they had designed. In other cases, the artists used narrative and storytelling to communicate general insights on AI.

5.3 Responsibility

In working with marginalised groups to design representative systems, and making explicit human labour that otherwise goes uncounted, the case studies remind us of the need for systems to be legible in ways that makes them meaningful and contestable for groups who may be obscured or misrepresented. Indeed, all of the projects have entailed thematic enquiry on themes related to fairness, accountability, transparency, and ethics

(FATE), and suggest novel and imaginative ways to evaluate and design for FATE in AI.

A common feature of the case studies is that they entail artistic exploration of the social entanglements and implications of AI. Where current advances in XAI tend to focus on the operation of the technology, such artistic perspectives and methods could widen the frame to address it's embedding in society, the politics and sociology that surround it, and the effects that it has in the world. The artistic experiments bring to life and question not only what an algorithm does, but also what a system could be used for, and who is in control.

These art projects reveal to users the way that the technology is connected to social understanding of the world. We see specific attention to the cultural differences of marginalised groups, and strategies to tackle bias in ML data. These experiences do not only relate to low-level technological deficiencies in the deployed systems, but also to system-level, cognitive, political, legal, regulatory or institutional factors in those deployments.This can advance cross-disciplinary understanding in AI and help to build literacy in those systems.

Reflecting on the case studies we can envision an enhanced field of human-centred XAI design. We contend that sensory and affective engagement can dramatise critical issues of intelligibility, ethics and trust, making connections beyond the models, algorithms and datasets out to their social contexts and implications. Here, the arts and tangible experiences enable complex philosophical, political, moral and technical questions to be explored experientially and when embedded in real situations. This leads us to the hypothesis that art and tangible experiences can mediate between impenetrable computer code and human understanding, making not just AI systems but also their values and implications more transparent and legible.

5.4 Speculation and Futuring

All of the pieces here invoke a sense of speculation about futures. The Zizi Show engages most directly with speculation about the future capabilities of ML models – what happens when they start to reproduce bodies, movements and other parts of cultural ways of being. This is in line with a general debate about appropriation within models and interactive technologies, whether the emerging use of deepfakes in video production (e.g. the combination of Mark Hamils early image with his current body to produce a 'de-aged' version, Giardana, 2020), the creation of 'digital twins' of an artist's voice (e.g. Holly Herndon's Holly Plus: https://auction.holly.plus) or the capture and use of dancer's movements in interactive works (Masu, 2019).

Artistic works allow an investigation of these possibilities before they become mainstream, supporting engagement with potential problems at an early stage. Zizi connects directly to this speculation, thinking into what it means for bodies to be captured and re-used, through imagining a joyous amalgamation of movements and characters. Cypress Trees uses speculation as an impetus, thinking into how we can use machine learning technology to remember lost species, and asking what tools we might create. ARWEN uses speculation as a material, drawing on scenarios thinking but using interactive intelligence as a compelling way in. There is a connection here to speculative design practices (e.g Auger, 2013), particularly those that look into creating experiences around emerging technical possibilities, using interaction to explore legibility, participation and embedding of human values (Murray-Rust, 2022). The pieces here engage with this speculation, with a similar but different remit, drawing on artistic reactions to current trends, rather than designerly forms of crafting to engage audiences.

6.
Towards a methodological framework for experiential XAI: Aspect, Algorithm, Affect, Apprehension

The collaborative experiments reported in the case studies were developed using the Open Prototyping process, and we extend this to propose a methodological framework for the use of experiential and artistic methods with human-centred XAI. We have reported on creative experiments in which AI researchers and artists are jointly engaged to make AI and machine learning tangible, interpretable, and accessible to the intervention of a user or audience. Elsewhere (Hemment, 2020) we have reflected on how replicable methodology can be built on the individual practice and methods of artists, and in how doing so it is vital to respect and not instrumentalise the individual voices of artists. Here we build methodology by incorporating insights from these collaborative experiments between artists within a multidisciplinary research team in modification and further development of the Open Prototyping process model. We arrive at a vision for human-centred explanation in interactive intelligent systems where an explanation is not a thing a model produces, but a learning journey, a process and an interaction between the AI and the user (O'Hara, 2020).

If we can think of XAI as a gear wheel that connects to the lower level mechanism, then the experience or art connects that technical explanation to higher level understanding (expert and non-expert). In our discussion there are four knowledge domains that have emerged as significant (FATE, XAI, AI arts, experiential learning). These loosely correspond to the four attributes of our case studies (theme, intelligence, interaction and explanation). This leads us to a formulation of Experiential AI characterised by these four dimensions: FATE provides the higher level aspect, the socio-technical theme or insight; XAI gives us the interpretable features in the algorithm; AI art generates meaning or affect; experiential learning is the apprehension of a user or audience, through which the explanation is taken up. We name these dimensions Aspect, Algorithm, Affect and Apprehension, and propose that, for there to be Experiential AI, the AI system needs to be open to understanding and intervention at these four levels. Aspect is the higher-level theme or feature, which here is FATE. Algorithm is the technology concept or capability. Affect is the passage from one experiential state to another. Apprehension is the contribution to knowledge for a specific user or audience. In Experiential AI, we bring these building blocks together in the creation of a space, represented in Table 1, where hypotheses can be actively tested, leading to concrete experiences that can be reflected on in order to generate new hypotheses (Zull, 2002). In the table we characterise these dimensions through reflection on the case studies in terms of type, material, theory of affect, input and output. We offer this as a heuristic ('good enough') conceptual framework which we will test and validate in future empirical experiments.

Level	Type	Material	Theory of Affect	Input	Output
Aspect	Theme	Regulation, Ethics	Milieu	Identified system-level, cognitive, political, legal, regulatory or institutional problems	Socio-technical transition
Algorithm	Technology	Tool, Technique	Functions	Technical understanding of what can be explained / interpretable features within an algorithm	Technical requirements
Affect	Art	Art, Practice	Precepts	Creativity and situated, embodied, intuitive meaning	Situated Experiences
Apprehension	Learning	Knowledge, Theory	Concepts	A framework for how people learn through those experiences	Pedagogy, debate, engagement

Table 1: The four dimensions for experiential human-centred XAI

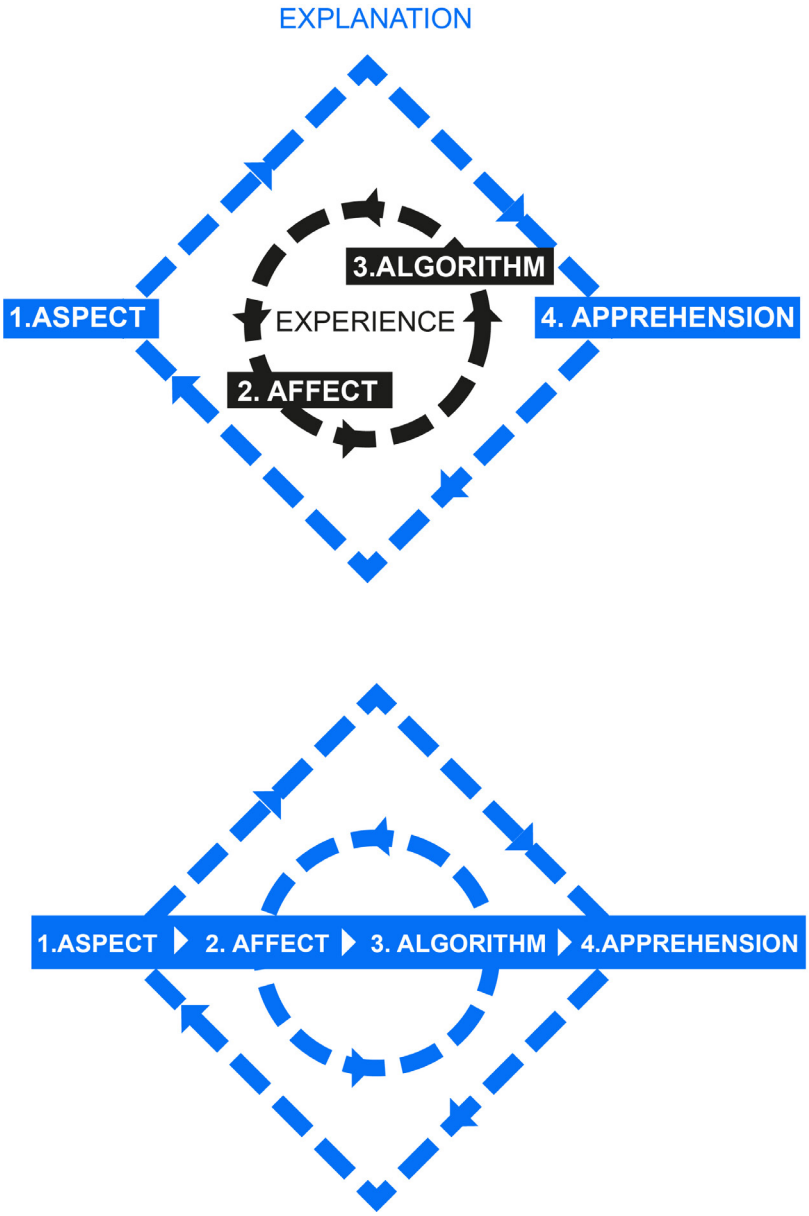


Figure 9: Experiential AI Process Model: A modified version of the Open Prototyping Process Model in which the outer diamond represents explanation, the inner circle represents experience, and the four dimensions of the proposed conceptual space have been mapped onto the diagram in place of the 6 Open Prototyping stages.

Looking back at the methods in our case studies, we see that we can loosely map the steps in the Open Prototyping process (Hemment et al, 2020) onto these dimensions or levels: Scope to Aspect; Connect and Play to Algorithm; Produce and Display to Affect; and Interpret to Apprehension. We also note that the two graphical components in the Open Prototyping Process Model (Ibid.), a concentric diamond and circle, can be mapped conceptually onto the two foundational components of Experiential AI: explanation and experience. We arrive at the Experiential AI Process Model, represented in Figure 9, as a modified version of the Open Prototyping Process Model.

The Experiential AI Process Model is a graphical representation of these four dimensions as steps in a process, mapped onto a concentric diamond and circle. In each of the case studies we see an enquiry on a particular theme conducted through the technical and creative development of an artistic experience, generating knowledge and understanding both for the artist and the participants. We can say there are two components, the creation of knowledge or understanding (explanation) through participation in the artwork (experience), that work together and reinforce one another. The concentric diamond and circle of the original Open Prototyping Process Model are here used to represent the concurrence of these two components, where the outer diamond represents the explanation, and the inner circle represents the experience. Onto this we have mapped the four dimensions of the framework from Table 1 as stages of the process. This figure therefore represents both the conceptual framework and process in diagrammatic form. Aspect and Apprehension on the outer diamond represent the interpretable features and the learning outcome, and Algorithm and Affect on the inner circle to represent the algorithm's functioning and the artistic expression. The figure represents the progression through the four stages of Experiential AI, visualising the iterative process of divergence and convergence through which explanation leads to understanding (diamond), and the holistic practice through which algorithm design and artistic inquiry generates experiences for users (circle).

This leads us to the conceptualisation of Experiential AI as a model for human-machine learning and development where knowledge is created through the transformation of experience, represented in Figure 9. This offers a distinct and complementary approach to XAI. Here we are not concerned only with the internal operations of algorithms. We are also concerned with opening up algorithms, the science behind them, and their potential impacts in the world to public scrutiny and policy debate. This approach bridges between applied science, engineering, design, art and social science, to understand the challenge of connecting the core concepts to the socio-political implications and contextualisation. The outcome is a methodology for human-centred XAI where explanations are created by reconfiguring data, algorithms, models, interfaces and situations as experiences. This practice of _experiential human-centred XAI_ deploys AI systems in order to empower users through through practical contact with and observation of tangible and legible representations of their attributes and implications.

7.
Developing Human-centered Explainable AI

We propose this methodology can be used to empower users through through practical contact with and observation of tangible and legible representations of its attributes and implications. To use this methodology to enhance explainability, designers of interactive intelligent systems create scenarios in which a higher-level aspects or assumption can be communicated to a user, then develop the architecture and algorithm, and integrate data and model, in a tangible output that makes the aspect explicit, and that users can have contact with and observe to derive a demonstrable learning outcome. The methodology provides a scaffold for the experiences to generate cognitive engagement and learning. This in turn, can feed back into the design of technologies, shaping XAI development.

The conceptual framework (Table 1) and process model (Figure 9) can be used by XAI practitioners to design, develop and evaluate a structured process of open and inclusive co-creation and co-investigation. This methodology can be applied to develop experiences with explanatory skill for various aspects of the life cycle of AI systems, from data collection – as seen in the case studies – systems design, algorithm selection and deployment, through to the interests and ideologies vested in their decisions and the social implications that follow. This includes not just the models at their core, but the data collection and processing that gives rise to them, the way the system has been commissioned and designed, and the relations between the system and the subjects of its decisions. This is a form of experiential learning, through which that affect is translated into knowledge and practice, represented by Apprehension.

This requires a collaboration between artists or experience designers and the XAI team, in the same way that an ethnographer might be included in a user-centred design team. In our case studies, digital artists and other interdisciplinary actors create the Affect to translate between human meaning and lower-level aspects, represented here by Algorithm. Our future research envisions multi-disciplinary teams, following a process to articulate definitions of interpretable aspects and features, and then giving access to data and algorithms to which AI artists and other critical practitioners can respond. These provide materials for creative experiments with AI and emerging technology that generate sensory experiences and affect. Technical systems are built that enact or implement the artists' strategy for exploring transparent and responsible AI, to create experiences that dramatise the AI concept.

Our early stage research has identified a number of promising themes for the future development of this methodology, such as: the strategies of the artists towards making AI explicit; the building of transparency and critical literacies into those systems; divergent meanings and definitions in the arts and sciences; and cognitive shifts or conceptual reinterpretations of certain terms or features. Design research has a specific contribution to the further development of this framework and to understanding of the challenges in

developing grounded yet generative work that allows publics to form opinions, bringing a sense of viscerality to explanations and understanding of AI. To make this research usable for the community, we plan to develop guidelines and heuristics to support XAI designers to form briefs and themes around datasets, algorithms and questions so that a team with the right knowledge and skills can use them as a springboard for critical artistic work.

Adoption of this approach requires engagement between communities with different logics and rigours, different vocabularies and goals. Future research will evaluate whether and how XAI practitioners and end users can derive meaningful and actionable insights, through experiences created with artists. Evaluation frameworks from human-computer and human-data interaction as well as experiential learning will help to investigate how cognitive-affective shifts come about through experiences, and the learning outcomes of such experiential explanations. Research can also help tp understand reinterpretations that can be made with the input of artists, while staying faithful to the AI models. Such research can evaluate how experiences may produce shifts in understanding for the public, but also, potentially, actionable insights for AI researchers and practitioners.

8.
Conclusion: The implications for explainability in AI systems

Our reflection on the potential contribution of the arts to XAI has led to our conceptualisation of Experiential AI, as a field in which scientists, artists and other interdisciplinary actors come together to understand and communicate the functionality of AI and intelligent robots, their limitations, and consequences, through informative and compelling experiences.

We contend that the arts can help to dispel the mystery of algorithms and make their mechanisms vividly apparent. Our early explorations have surfaced artistic strategies to make AI systems transparent and legible to users. We see in these practices a rich, multifaceted engagement in AI that has much to offer to science and society. This has led us to our central theme in Experiential AI, namely, how we can explore new modalities of explanation, to augment and enrich the field of XAI.

In Experiential AI, we advance methodology to use methods from the arts alongside and/or as a mode of explanation to engage people emotionally, cognitively and tangibly with the large scale effects of pervasive AI deployments. This involves using experiential methods to explore AI technology in concert with publics, and drawing on XAI philosophies, but going beyond functional explanation to develop contexts and implications for AI systems.

These experiences can help interacting humans to viscerally understand the complex causal chains in environments with AI components, including questions about: what data is collected, its nature, accuracy and freedom from bias, as well as who collects it; how the algorithms are chosen, commissioned and configured; and how humans are conditioned by their participation in algorithmic processes.

For the XAI community, we hope the contribution of this paper will be to contribute to research and practice on explainability, with a particular focus on supporting non-expect users and broadening the scope of current work to address the higher levels of explanation, understanding and context. We identify a potential for the arts to complement existing work in explanation that traces details of an algorithm's functioning, by illuminating underlying assumptions of machine learning models, processes that generate their data, or the social context in which automated decision making is situated. Immediate future work can try to combine artistic practices with existing XAI solutions to better situate and present such "solutions" and their limitations. We envision further work on creative ways to explain what the model actually does, as well as how automated decision making is situated. More holistic questions can then be asked of the entanglements of humans and machines, going beyond model interpretability, such as how does AI challenge our world view, or how we can avoid anthropomorphism and misplaced trust in AI.

The benefits can be reciprocal: to enhance XAI with the arts, and enhance the arts with XAI. Through Experiential AI, the arts are connected to significant science and

technology, and, in turn, produce situated, embodied and intuitive meaning around algorithms and the effects of their deployments. Artists who are already well versed in the current science can benefit from access to XAI concepts and tools, and the depth and rigour of AI practice in the wider arts and creative sectors can be enhanced by enabling structured access to the significant science.

We conclude that the arts, and the cross-disciplinary practice we call Experiential AI, can make a significant contribution. Reflecting on our early stage explorations, we identify promising opportunities for future research between the arts and XAI. Such experiential and holistic interventions can work alongside XAI to reach new audiences, and to create spaces for debate and engagement with populations outside the technical centre. By making explicit how AI systems and automated decisions are embedded in multidimensional social situations, we can demonstrate to users and the general public the implications of their power.

References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20: 3, 973–989.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2019). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. arXiv preprint arXiv:1910.10045.

Auger, J. Speculative design: crafting the speculation, Digital Creativity, vol. 24, no. 1, pp. 11–35, Mar. 2013, doi: 10.1080/14626268.2013.767276.

Barbican. (2019). AI: More than Human. https://www.barbican.org.uk/whats-on/2019/event/ai-more-than-human.

Belle, V. (2017). Logic Meets Probability: Towards explainable ai systems for uncertain worlds. In IJCAI, 5116–5120.

Belle, V. & Papantonis, I. (2021). Principles and Practice of Explainable Machine Learning. Front. Big Data 4:688969. doi: 10.3389/fdata.2021.688969.
Machine Learning Explainability for External Stakeholders. ICML Workshop XXAI: Extending Explainable AI Beyond Deep Models and Classifiers, 2020.

Bhatt, U., McKane, A., Weller, A., Xiang, A. (2020). Machine Learning Explainability for External Stakeholders. ICML Workshop on Extending Explainable AI: Beyond Deep Models and Classifiers

Bishop, C. M., (2006). Pattern Recognition and Machine Learning. Springer
Boal, A. (2013). The Rainbow of Desire: The boal method of theatre and therapy. Routledge.

Borkman, T. (1976). Experiential knowledge: A new concept for the analysis of self-help groups. Social service review, 50(3), 445-456.

Bory, S., & Bory, P. (2015). New Imaginaries of the Artificial Intelligence. Im@go. A Journal of the Social Imaginary. 6, 66-85.

Bratton, B. H., (2016). The Stack: On Software and Sovereignty. Cambridge, MA, USA: MIT Press.

Brščić, D., Kidokoro, H., Suehiro, Y. and Kanda, T. (2015). Escaping from children's abuse of social robots. In Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI), 59-66.
Carter, S., & Nielsen, M. (2017). Using artificial intelligence to augment human intelligence. Distill. 2: 12, e9.

Chakraborti, T., Sreedharan, S., Grover, S., and Kambhampati, S. (2019). Plan Explanations as Model Reconciliation. 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea, March 2019. IEEE, 258–266.

Chan, J. (2021). The future of AI in policing: Exploring the sociotechnical imaginaries. Predictive Policing and Artificial Intelligence. Routledge, 41-57.

Crawford, K. and Paglen, T. (2019). Excavating AI: The Politics of Training Sets for Machine Learning. https://excavating.ai.

Donnarumma, M. (2018). Is artificial intelligence set to become arts next medium? https://marcodonnarumma.com/ works/ai-ethics-prosthetics/.

Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. Duke L. & Tech. Rev., 16, 18.

Fiebrink, R., & Gillies, M. (2018). Introduction to the special issue on human-centered machine learning. ACM Trans. Interact. Intell. Syst., 8(2), 7:1–7:7. doi: 10.1145/3205942.

Flyvbjerg, B., 2013. Case Study, in: Denzin, N.K., Lincoln, Y.S. (Eds.), Strategies of Qualitative Inquiry. SAGE, Thousand Oaks, Calif.; London; New Delhi, pp. 169–204.

Forrest, K, B. (2021). When Machines can be Judge, Jury and Executioner: Justice in the Age of Artificial Intelligence. World Scientific.

C. Giardina, 'ILM Head Talks AI, Deepfakes and "Mandalorian" Visual Effects', The Hollywood Reporter, May 08, 2020. https://www.hollywoodreporter.com/movies/movie-news/ilm-head-talks-ai-deepfakes-mandalorian-visual-effects-1293243/ (accessed Feb. 21, 2022).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative Adversarial Nets. Advances in Neural Information Processing Systems 27 (NIPS 2014)
Gunning, D. (2017). Explainable artificial intelligence (XAI). https://tinyurl.com/yccmn477.

Hemment, D. (2020). Reordering the assemblages of the digital through art and open prototyping. Leonardo. 53: 5. Cambridge, MA: MIT Press, 529-536. DOI: 10.1162/leon_a_01861.

Hemment, D., Bletcher, J., & Coulson, S. (2020). Open Prototyping: A framework for Combining Art and Innovation in the IoT and Smart Cities. In Eds. Hjorth, L., de Souza e Silva, A., Lanson, K. The Routledge Companion to Mobile Media Art. London: Routledge, 270-283. ISBN 9780367197162.

Hemment, D., Aylett, R., Belle,. V., Murray-Rust, D., Luger, E., Hillston, J., Rovatsos, M., Broz, F. (2019) Experiential AI. AI Matters. 5: 1. ACM New York.

Heron, J., and Reason, P. (2001). The practice of co-operative inquiry: Research with rather than on people. In Reason, P., and Bradbury, H. (eds.), Handbook of Action Research: Participative Inquiry and Practice, London: Sage, London, pp. 179–188.

Kambhampati, S. (2020). Challenges of Human-Aware AI Systems. AI Magazine. Vol. 41 No. 3: Fall 2020

Kolb, D. A. (2014). Experiential learning: Experience as the source of learning and development. FT press.

Kulkarni, A., Zha, Y., Chakraborti, T., Vadlamudi, S. G., Zhang, Y., and Kambhampati, S. (2019). Explicable Planning as Minimizing Distance from Expected BehaviorProceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, Montreal, QC, May 2019. (International Foundation for Autonomous Agents and Multiagent Systems), 2075–2077.

Lake, B. M., Salakhutdinov, R., Tenenbaum, J. B., (2015) Human-level concept learning through probabilistic program induction. Science 350.6266 (2015): 1332-1338.

Lim, S. M., Tan, B. L., Lim, H. B., Goh, Z. A. G. (2018) Engaging persons with disabilities as community teachers for experiential learning in occupational therapy education. Hong Kong Journal of Occupational Therapy. https://doi.org/10.1177/1569186118783877

Lyons, M. (2020). Excavating 'Excavating AI': The Elephant in the Gallery. IEEE Transactions on Pattern Analysis and Machine Intelligence. 21, 1357-1362.

Marcus, G. & Davis, E. (2019). Rebooting AI: Building Artificial Intelligence We Can Trust. Pantheon Books.

Masu, R., Correia, N. N., Jurgens, S., Druzetic, I., Primett, W. (2019). How do Dancers Want to Use Interactive Technology? Appropriation and Layers of Meaning Beyond Traditional Movement Mapping. ACM Proceedings of the 9th International Conference on Digital and Interactive Arts: 1–9. https://doi.org/10.1145/3359852.3359869

Miller, T. (2019) Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence. Vol 267, Feb 2019: 1-38.

Montavon, G., Samek, W., Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. Digital Signal Processing. 73, 1–15. doi:10.1016/j.dsp.2017.10.011. ISSN 1051-2004.

Murray-Rust, D, Elsden, C., Nissen, B., Tallyn, E., Pschetz, L. and Speed, C. 'Blockchain and Beyond: Understanding Blockchains through Prototypes and Public Engagement', Transactions on Computer-Human Interaction, 2022, Available: http://arxiv.org/abs/2112.11891

Natale, S., & Ballatore, A. (2020). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. Convergence. 26.1, 3-18.
Onassis Foundation. (2021). You and AI: Through the Algorithmic Lens. Curated by FutureEverything. https://www.onassis.org/whats-on/festival-you-and-ai-through-the-algorithmic-lens

O'Hara, K. (2020). Explainable AI and the philosophy and practice of explanation. Computer Law & Security Review. Vol. 39, Nov. 2020. doi: 10.1016/j.clsr.2020.105474.
Penkov, S., & Ramamoorthy, S. (2017). Using program induction to interpret transition system dynamics. arXiv preprint arXiv:1708.00376.

Raji, I. D., et al. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing." Proceedings of the 2020 conference on fairness, accountability, and transparency.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should i trust you?: Explaining the predictions of any classifier. In Proceedings of the 22nd ACM Sigkdd International conference on knowledge discovery and data mining, 1135– 1144.

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 101-108. doi: 10.1109/HRI.2016.7451740.

Royal Society (2019). Explainable AI: The Basics Policy Briefing. https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf
Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1.5: 206-215.

Sailaja, N., Crabtree, A., Colley, J., Gradinar, A., Coulton, P., Forrester, I., Kerlin, L., and Stenton, P. (2019). The living room of the future. Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video.

Smedegaard, C.V. (2019). Reframing the role of novelty within social HRI: from noise to information. In Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI '19). IEEE Press, 411–420.

Srinivasan, R. & Parikh, D. (2021). Building Bridges: Generative Artworks to Explore AI Ethics, arXiv:2106.13901 [cs], Jun. 2021, Accessed: Aug. 24, 2021. [Online]. Available: http://arxiv.org/abs/2106.13901

Vannini, N., Enz, S., Sapouna, M., Wolke, D., Watson, S., Woods, S., Dautenhahn, K., Hall, L., Paiva, A., André, E., Aylett, R., & Schneider, W. (2011). "FearNot!": A computer-based anti-bullying-programme designed to foster peer intervention. European Journal of Psychology of Education. 26, 21–44.

Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 440.

Wagner, A. R., Borenstein, J., Howard, A. (2018). Overtrust in the Robotic Age. Communications of the ACM. 61: 9, 22-24. doi 10.1145/3241365.

Weld, D. S., & Bansal, G. (2018). Intelligible artificial intelligence. arXiv preprint arXiv:1803.04263.

Yin, R. K. (2009). Case study research: Design and methods (4th Ed.). Thousand Oaks, CA: Sage

Zentrum fur Kunst und Medien. (2018). Encoding cultures: Living amongst intelligent machines. https://zkm.de/en/event/2018/04/encoding-cultures-living-amongst-intelligent-machines.

Zer-Aviv, M. (2018). The Normalizing Machine. http://mushon.com/tnm/

Zull, J. E. (2002). The Art of Changing The Brain: Enriching the Practice of Teaching by Exploring the Biology of Learning. SCHOLE: A Journal of Leisure Studies and Recreation Education. 24:1, 181.

Resilience in the New Real is funded by the Arts and Humanities Research Council, and The New Real Observatory is funded by Turing 2.0/Engineering and Physical Sciences Research Council.

Acknowledgements:

Thanks to Jake Elwes, Anna Ridler, Caroline Sinders, Alan Butler, Lex Fefegha, Adam Harvey, Inés Cámara Leret, Keziah MacNeill, Mario Antonioletti, Julie Ann Fooshee, Keili Koppel, Sarah MacKinnon, Evan Morgan, Daga Panas, David Sarmiento Pérez, Sohan Seth, Miriam Walsh and Holly Warner.

How to cite this article:
Hemment, D., Murray-Rust, D., Belle, V., Aylett, R., Vidmar, M., Broz, F. (2022) Experiential AI: Enhancing explainability in artificial intelligence through artistic practice. Preprint: Pure URL