

General Article

Experiential AI: Between Arts and Explainable AI

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

Drew Hemment (curator, researcher), University of Edinburgh, Edinburgh College of Arts, Edinburgh EH39DF, U.K. Email: drew.hemment@ed.ac.uk. ORCID: 0000-0002-0068-5500.

Dave Murray-Rust (artist, researcher), TU Delft, Faculty of Industrial Design Engineering, 2628 Delft, Netherlands. Email: D.S.Murray-Rust@tudelft.nl. ORCID: 0000-0001-6098-7861.

Vaishak Belle (researcher), University of Edinburgh, School of Informatics, Edinburgh EH89AB, U.K. Email: vbelle@ed.ac.uk. ORCID: 0000-0001-5573-8465.

Ruth Aylett (poet, researcher), Heriot-Watt University, School of Mathematics and Computer Science, Edinburgh EH144AS, U.K. Email: R.S.Aylett@hw.ac.uk. ORCID: 0000-0002-8606-5571.

Matjaz Vidmar (researcher), University of Edinburgh, School of Engineering, Edinburgh, EH93FB, U.K. Email: Matjaz.Vidmar@ed.ac.uk. ORCID: 0000-0002-1368-9762.

Frank Broz (researcher), TU Delft, Department of Intelligent Systems, 2628 Delft, Netherlands. Email: f.broz@tudelft.nl. ORCID: 0000-0002-9624-0599.

See <https://direct.mit.edu/leon/issue/57/3> for supplemental files associated with this issue.

Abstract

Experiential artificial intelligence (AI) is an approach to the design, use, and evaluation of AI in cultural or other real-world settings that foregrounds human experience and context. It combines arts and engineering to support rich and intuitive modes of model interpretation and interaction, making AI tangible and explicit. The ambition is to enable significant cultural works and make AI systems more understandable to nonexperts, thereby strengthening the basis for responsible deployment. This paper discusses limitations and promising directions in explainable AI, contributions the arts offer to enhance and go beyond explainability, and methodology to support, deepen, and extend those contributions.

The current explosive development of artificial intelligence (AI) has profound implications for the arts and society. However, due to the move to data-driven systems and increasingly complex machine learning (ML) algorithms, many current AI tools are inscrutable and opaque to human understanding, which limits the scope for informed human intervention. Generative AI tools powered by recent developments in diffusion models and large language models are powerful yet work in very particular ways and are black-boxed. There is a contradiction inherent in generative AI, in so far as the pervasive availability of prompt driven systems lowers the barrier to entry to AI, yet at the same time the user remains oblivious as to the mechanics of creation and retains little control of it. Users' agency over a tool's precise function and modes of use is further curbed by the prevalence of centralized, extractive business models, the trend towards Internet-scale training datasets, and the computing resources these systems require. More broadly, in current AI everything from the way data is collected, labelled, and cleaned to the way algorithms are trained and tested profoundly determines the types of outputs and decisions generated and their impacts on end users. Furthermore, powerful and widely applied deep learning algorithms encode the system's knowledge implicitly and in a distributed fashion, so that even experts may not be able to easily determine what the system "knows." These barriers limit the creative agency of artists and their ability to work with ML as material or tool; diminish the ability of individuals to make good decisions about using a system, contest or debug its outputs, and help shape its design and operation; and can more widely inhibit transparency, fairness, trust, and safety in AI deployments.

Explainable AI (XAI) is the current prevailing approach used in the AI community to enable human users to understand the results and outputs of ML algorithms [1]. XAI is predicated on the notion that informed human decisions and actions require not only that the operation of the system be visible, but also that the system can provide comprehensible explanations. We argue that despite significant achievements in XAI, the field is limited by a focus on technical and functional explanation. Turning to the arts for inspiration, we propose that experiential methods can enhance model interpretability and make the opaque mechanisms and meanings of AI artifacts and algorithms more legible to those who interact with them. This is a moment many artists are exploring the implications for their own practice, and voicing their perspectives on the profound upheavals

that the aforementioned developments bring. We see among AI artists ---particularly those concerned with AI as both tool and topic, material and theme ---strategies to engage people cognitively, emotionally, and in a directly graspable way with the implications of AI deployments. These include fostering critical and poetic perspectives, making visible design choices in the AI pipeline, and showing the key role of training datasets as they interact with algorithms [2].

We herein report on work in the arts and technology communities that has helped develop an experiential approach to machine intelligence, improve user interactions with AI, and underpin new paradigms for creative and responsible AI. We present experiential AI [3] as a transdisciplinary practice that opens up the AI field to greater understanding and collaboration between human and machine. This theme emerged from a 2019 workshop at the University of Edinburgh that identified the intersection of AI arts and XAI as a promising area for cross-disciplinary inquiry and was further developed to focus on methods and tools for exploring AI data and models. The ambition is twofold: to enable stimulating and intuitive co-creation with AI that gives rise to significant cultural works [4] and to make possible new modalities of explanation that open up algorithms, the science behind them, and their potential impact in the world to scrutiny, intervention, and debate. We begin with a review of current limitations and promising directions on explainability within the work of the AI community. We then look at three AI art case studies to understand how art and tangible experiences can address those limitations. Building on that understanding, we propose a novel methodology for designing and working with AI systems that can support artists, engineers, and other disciplinary actors to collaborate with each other and AI in order to better understand and communicate a system's functionality, limitations, and consequences through informative and compelling experiences. Aside from improving the explainability of AI tools, we ask how this methodology can both contribute to significant art and strengthen the basis for more inclusive and responsible technologies.

Recent Advances and Current Limitations in Explainability and Explainable AI

There is not yet a shared definition of XAI within the AI community. Hence, for the purposes of this paper, we take it to denote interpretable features within an algorithm that enable decisions to be justified, tracked, and verified by a human [5]. It can entail using simpler models (possibly at the cost of prediction accuracy), “local” explanations that identify interpretable patterns in regions of interest [6], or human-readable reconstructions of high-dimensional data [7]. Promising developments include the use of transparency-by-design and interactive explanations [8], the development of generative artworks as educational tools to explore AI ethics [9], and the growth of a community around XAI for the arts of which the authors of this paper are part [10]. Yet, while AI explainability can be enhanced by drawing on the social sciences [11], the potential contribution of the arts to explainability remains largely untapped. Our research has identified six gaps or areas of development to which work at the intersection of the arts and engineering may contribute:

Limitation 1. Providing explanations from human points of view

Solely technical and functional forms of XAI often struggle to generate the kind of explanations needed from a human point of view. A person won't always need to know in detail how black box AI works, nor will they always be helped by an opaque description of machine logic. They will instead want to understand the limitations in using a given AI tool, why there was one outcome and not another, or when their outputs can be trusted. Recent XAI work has acknowledged a need to develop alternative, human-centered routes to understanding through lenses like user experience [12] and sensemaking [13].

Limitation 2. Looking beyond model explanations to address the entirety of the AI system

While there have been impressive advancements in the socio technical aspects of XAI [14], there remains an urgent need to understand the entirety of AI systems, including the relationships between the system and the subjects of its operations. As it is now widely acknowledged, most ML systems cannot be disentangled from their data sources, raising ethical and privacy concerns and leaving the user with the burden of understanding those entanglements.

Limitation 3. Connecting technical aspects to higher-level dimensions of AI

The vast majority of advances in XAI focus on the technical and functional operation of AI [15]. Current work does not adequately account for the politics and sociology that surround it, its embedding in society, the effects that it has in the world, or its higher-level aspects ---whether systemic, cognitive, political, legal, regulatory or institutional [16].

Limitation 4. Accounting for a wider range of stakeholders in systems deployed in social situations
XAI has predominantly targeted a narrow range of expert users. As AI-enhanced applications become increasingly widespread, insight into both their operation and implications needs to be communicated to users and the people they impact in ways that enable them to regain agency in their interactions with those applications and support engagement by a broader segment of society [17].

Limitation 5. Engaging with the imaginaries surrounding AI
AI is the subject of widespread illusions: e.g. a machine "belief state" is comparable to a human mental state. There are strong imaginaries around AI [18] that can lead to exaggerated fears, heightened fascination, and unrealistic imputation of authority. Engaging with such imaginaries and misrepresentations can better orient policy, development, and public debate [19].

Limitation 6. Deeper engagement in material and ideological realities
It is unclear whether abstract models can incorporate cultural and sociopolitical norms in a straightforward manner. We echo Ananny and Crawford's call for "a deeper engagement with the material and ideological realities of contemporary computation" [20]. This would require creating a deeper understanding of AI in a manner that addresses the nuances of socio-technical systems.

Limitation 7. Connecting machine reasoning to embodied and situated human experience
The daily experiences of AI that is aware of humans [21] provide a rich ground for experiential knowledge, understood as a function of the type of acquired information and one's attitude towards said information [22]. 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 [23].

Case Studies of AI Art and Experiential Learning

In our empirical experiential AI research (see online supplemental materials), we conducted a co-operative research inquiry [24] (prolonged due to COVID-19) between summer 2019 and winter 2022 involving a multi-disciplinary team of professional artists and design, XAI, and applied ethics scholars. Collectively, the research team shaped the questions and approach, and the artists each brought their own methods to the co-operative inquiry. Open prototyping [25] provided a design framework to broker between disciplines while preserving autonomy for the practice of individual contributors and to configure data, algorithms, models, interfaces, and situations as experiences. Three projects in which professional artists were commissioned to participate were developed as case studies [26]; the unit of analysis for these case studies was operationally defined as the period of cross-disciplinary collaboration between artist, researchers, and stakeholders. One was an artist commission, one an artist residency, and the third a co-creative process involving artists and researchers. Insights into the case studies were developed through the thematic analysis of ethnographic material. Primary data included interviews with the artists, science team, audience members, and the participants' observation notes taken during development meetings and events. Secondary data included project documentation, artists' project websites, and commentary by audiences and critics published online and in press. Supported by online supplements (1-3) containing a summary of the work, a production note, and a description of key dimensions, a description of what each project can be understood to "explain" is presented below.

The Zizi Show



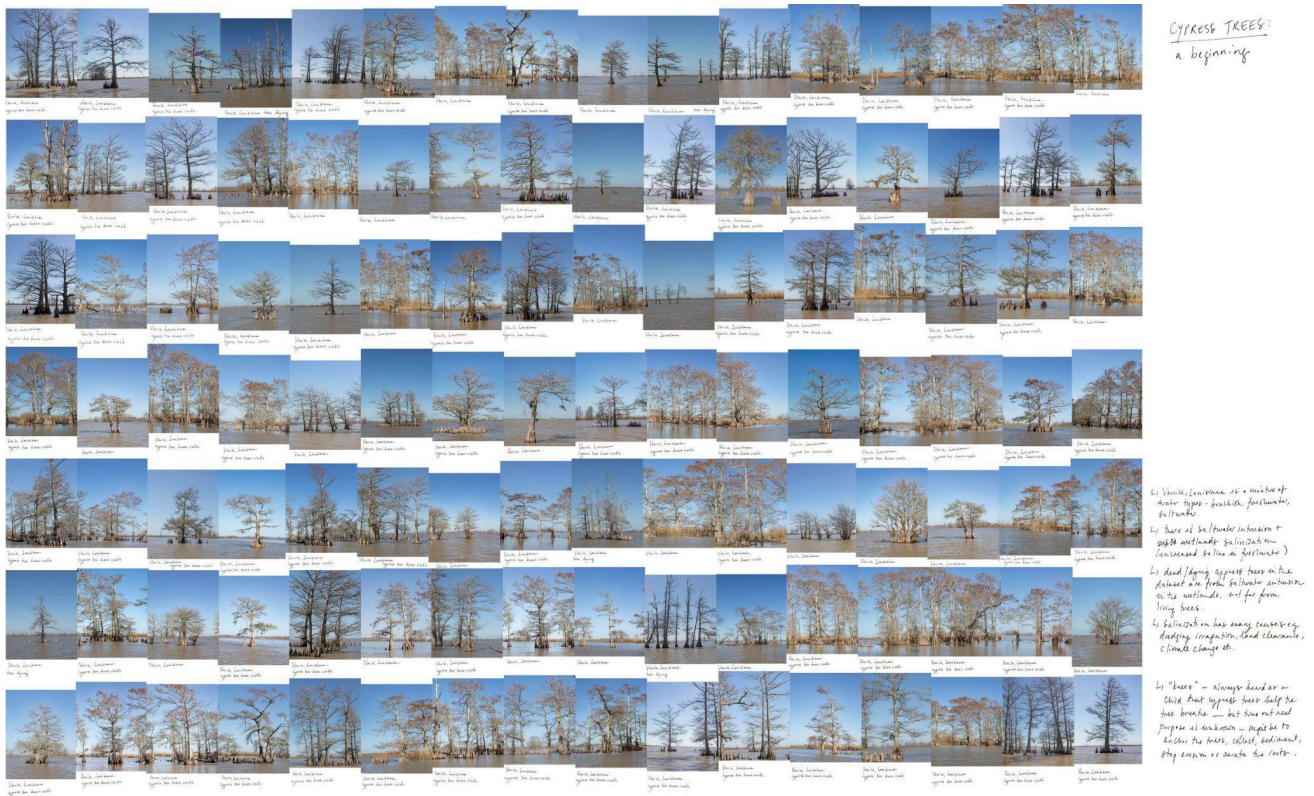
Theme

Bias in ML data; discriminatory design; misrepresentations of AI.

Explanation

The Zizi Show (Jake Elwes, 2020) (Fig.1 and supplemental materials 1) generates imagery of non binary bodies to bring attention to the underrepresentation of LGBTQ+ communities in ML training data. It provides an explanation through the experiential means of a dense clustering of issues: discriminatory design [27], bias in ML data, non binary identities, and anthropomorphism in AI. Zizi highlights the way data and design choices shape what ML does. It shows how the model learns a representation of people embedded in social life. Elwes directly engages and empowers an underrepresented community by ensuring there is fair pay for participating drag artists, crediting them as cast members, and providing agency to them over the way their data is stored on a secure server so it is not harvested for other models.. Zizi shows audiences there is something to contest even for those who do not interact directly with the model themselves. Zizi specifically targets anthropomorphized misrepresentations of AI by constructing an AI persona and then deconstructing it to expose its construction in software by the human artist. Highlighting correspondences between AI and drag at a surface level, the work asks deeper questions about how the character of statistical knowledge applies to shifting human identities.

AI is Human After All



Theme

Hidden human labor in ML; AI-mediated experience of the climate crisis.

Explanation

The artist residency project AI is Human After All (Anna Riddler and Caroline Sindere, 2019--2021) (Fig. 2 and supplemental material 2) highlighted that what we think of as ML intelligence is actually human intelligence at many points in the system. The way Riddler and Sindere meticulously develop, label, and catalogue bespoke data points to human labor that is often hidden or obscured. Their art practice debunks the neat representations of "autonomous" systems and exposes the situational, embodied nature of ML systems. It presses us to envision methodologies and pipelines for AI development in which human labor is acknowledged and honored while raising wider questions around human bias and worker exploitation. The artists extract patterns from observed data using manual methods and, in this sense, turn a foundational definition of AI on its head: the human artist performs a task usually done by the computer and associated with machine intelligence. Rather than solve problems, the human-machine intelligence produces imagery and gallery installations that represent the ordering of knowledge by AI and explores key concerns such as the impact of climate change.

The New Real Observatory



Theme

Interpretability of AI and climate data; co-creation and agency.

Explanation

The New Real Observatory (The New Real, Inés Cámara Leret, Adam Harvey, Keziah MacNeill, and Alex Fefegha, 2022) (Fig. 3 and supplemental materials 3) is an experiential AI system made up of a bespoke generative AI platform and accompanying artworks, each conceived as instances of the system. The platform provides a toolset that allows users to probe, manipulate and fine tune a model as an integral component of co-creation with AI and to provide a basis for substantive artistic works. Interaction with the platform makes explicit both the role of training datasets as they interact with algorithms and the fact that no machine can

create ad novo in isolation. The artists gain knowledge and skill through practical contact with and observation of the generative process and, by navigating through the latent space, are able to explore and generate new associations and instances. Three artworks offer specific insight, a perspective, a dimension, or a lens through which AI can be interpreted: Cámara Leret gives tangible expression to the entanglements that arise when translating global data to local environments and exposes the impact of AI in both enabling and hindering our understanding of, and relationship with, the environment; Adam Harvey questions the way generative AI automates artistic production, auto-completes visual concepts, and transforms low-dimensional ideas into high-resolution imagery while highlighting concerns ranging from computational plagiarism to excessive energy usage; Keziah MacNeill creates a dialogue between artificial intelligence, climate data, and photography, tuning into the operation of the neural network to stimulate novel thinking about the character of the algorithmically generated image. Wrap around activities including talks, workshops, and designed interactions supported further sense making with festival attendees.

Discussion: Enhancing and Moving Beyond Explainability through AI Art and Experiential Methods

In the case studies we see artists pushing at the boundaries of human-machine creativity to generate works that combine machine learning methods with human intuition and embodied experience. In Leonardo journal, Mateas (2001) argued that “AI-based art is not a subfield of AI... Rather it is a stance or viewpoint from which all of AI is reconstructed... The AI system becomes an artifact built by authors in order to communicate a constellation of ideas and experiences to an audience... [It questions] the deep technical and philosophical assumptions that underlie AI practice” [28]. More recently, in a comprehensive review of contemporary AI art, Dejan Grba notes that “Digital technologies offer a generous space for conceptual, as well as formal, methodological, and aesthetic experimentation that can transcend the technologically imposed limits of expression” [29]. Research on experiential learning [30] confirms that artistic practice can act as a powerful pedagogic mechanism. Experiential methods have been shown [31] to create deeper understanding than purely knowledge-based or data-driven approaches and to support situated and contextual understanding. Looking beyond explanation, human intuition and inventiveness can go beyond merely providing information or logical arguments to support other kinds of engagement, such as contestation around models and their design [32]. This in turn increases scope for human agency, intervention, and discovery.

Understanding the potential of art to improve XAI in this manner, our research identified the following contributions made by the reviewed projects to addressing the aforementioned limitations and pursuing promising directions in XAI.

Contribution 1. Providing explanations from human points of view

The artists involved in the case studies are adept at surfacing issues or ambiguities that are significant and relevant to human audiences and will benefit from explanation. They develop explanations as interactions with narratives and statistical information that explore and communicate concepts in relatable, imaginative and absorbing ways..

Contribution 2. Looking beyond model explanations to address the entirety of the AI system

The artworks consider AI systems, their operation, and their implications in a holistic way. They expose features of ML pipelines that may be obscured or not taken into consideration, such as human labor.

Contribution 3. Connecting technical aspects to higher-level dimensions of AI

These art projects look beyond the models, algorithms, and datasets and connect to higher-level aspects such as bias and inclusion. The artistic experiments bring to life and question not only what an algorithm does but also why the AI did one thing and not something else, what a system could be used for, and who controls it.

Contribution 4. Accounting for a wider range of stakeholders in systems deployed in social situations

The AI art projects improve understanding among non experts who are not supported by mainstream XAI. They engage demographic groups that are underrepresented in training data and in AI design and evaluation. The artworks suggest strategies that can help to make AI more inclusive and reach those currently excluded from the creation and deployment of systems.

Contribution 5. Engaging with the imaginaries surrounding AI

The art projects reviewed in our case studies speculate about the future capabilities of ML models and what happens when they start to reproduce bodies, movements, and other cultural ways of being. The artists seek to demystify AI by addressing anthropomorphism and exaggerated accounts of the autonomy of AI systems. In other cases, artists have deliberately obfuscated or subverted current knowledge and striven to reveal a deeper truth or challenge a knowledge regime.

Contribution 6. Deeper engagement in material and ideological realities

The investigated works enable a wide range of questions and facets to be explored alongside aesthetic and formal properties. They are not didactic explanations, but concrete artifacts that can be used to spark further inquiry and explore deeper meanings of the physical, moral, metaphysical and aesthetic qualities of AI.

Contribution 7. Connecting machine reasoning to embodied and situated human experience

The artworks enable AI to be explored experientially and in a context where they are embedded in real-world settings designed around human experience and relationships. The arts incorporate the embodied, intuitive, and affective character of human reasoning into XAI.

Building Methodology for Experiential AI

In the projects presented above, we see ways in which artworks and tangible experiences can enhance and go beyond explainability in AI. Similarly, for artists concerned with AI as both material and theme, enhanced model interpretability could provide a rich source of inspiration. However, while artworks may help to illustrate the operation and implications of AI, their purpose is not to generate functional or task-based descriptions. XAI is primarily concerned with task competence, and artists more so with shifting perspectives and the poetry that can be found [33]. It is not our intent to instrumentalize the arts for either science communication or system design. Therefore, to facilitate and extend these contributions and translations, we develop experiential AI as a transdisciplinary methodology that draws on and complements advances in both the arts and XAI. Building on components of the open prototyping design framework, we present: (1) a conceptual model named the 4As, (2) a schematic diagram, and (3) six principles.

Whereas case studies 1 and 2 build on the current practice of the artists, case study 3 was co-creative and cross-disciplinary and thus constitutes the most complete implementation of our proposed methodology. The novel, socio-technical experiential AI system in The New Real Observatory enables greater user control beyond prompt engineering when working with generative technologies and more synergetic interaction between human and machine. At the same time, the ability to probe and manipulate the model enables intuitive understanding of how the system works and how the interaction of variables may influence the outcome. The learning and intelligence that arise through this practical contact with and observation of AI -- by an artist-in-the-loop or audience-in-the-loop -- opens up ways to understand, respond to, and communicate model behavior and implications beyond functional explanations, which remains a focus in our ongoing research [34].

Conceptual model

A common feature of the case studies is that they were motivated by the participants' framing of an urgent problem and their critical understanding of the context surrounding deployed AI. Further, each case study involved creative and technical experimentation, as well as interaction between the artists, engineers, data, and system. They all generate novel interpretations and enhanced interactions. In our case studies (see supplemental material for more detail), we identify four key dimensions supported by prior research on the strategies and intentions of critical AI artists: (a) thematic, (b) technical, (c) experiential, and (d) learning dimensions. [35]. This informed our formulation of an experiential AI characterized by four dimensions: a spect, an algorithm, an affect, and an apprehension (Table 1). For there to be experiential AI, we propose that the AI system needs to be open to understanding and intervention at these four levels.

Process diagram

In the development of the case studies, we see the dimensions interact in two ways. First, each case study involves cross-disciplinary scoping and knowledge creation, characterized by an issues-based or challenge-led approach, providing impetus for the artistic works and types of higher-level insights that emerge. This is the interaction between aspect and apprehension, which we characterize as the iterative divergence and

convergence in a challenge-led research journey. Second, each case study also involves technical and artistic experimentation and outcomes that are intimately intertwined. This is the interaction between algorithm and affect, which we characterise as the ongoing communication or cooperation between AI and a human-in-the-loop. These two pairings can be represented by the outer diamond and inner circle of the process model [36] in the open prototyping design framework. Without implying a linear sequence of functional steps, we project the dimensions of the 4As onto the diagram to represent the layered and iterative character of the challenge-led, experiential learning journey in experiential AI (Fig. 4). Foregrounding the interactions between the dimensions like this helps us account for the way the learning of the artist feeds back into the fine tuning of the model in case study three and creates a feedback loop between human and machine learning. It also underlines a vision for experiential learning in intelligent systems where an explanation is not a static representation produced by a system but a learning journey, a process, and an interaction between the AI and the artist or audience [37].

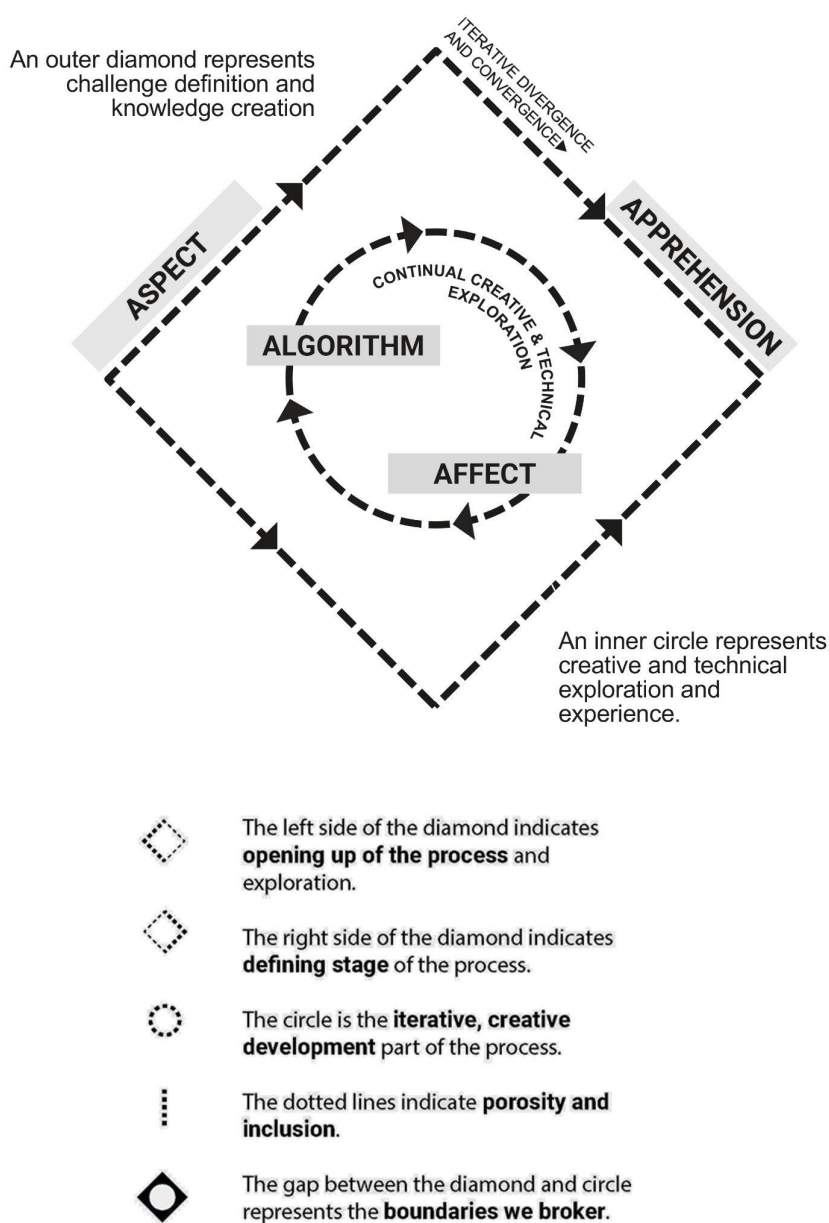


Figure 4: Experiential AI Process Model

Principles

Finally, we return to the design stages described in the open prototyping framework) and propose six corresponding principles for experiential AI:

1. connect technology to social goals;

2. combine the efforts of artists and XAI practitioners, while preserving autonomy for the practice of individual contributors;
3. encourage diverse interpretations, improvisation, and provocation, while remaining true to the technical concepts ;
4. develop experiences and wraparound engagement to enrich the learning journey ;
5. provide a platform for meaningful and multilayered interaction with users and stakeholders (industry, policy, academic, public) ;
6. enable rigorous evaluation.

Together, the conceptual model, process diagram and recommendations give us the three components of the experiential AI methodology, which together can be used to design and evaluate creative experiments in which technical and creative disciplines jointly make data-driven AI and ML more tangible, interpretable, and usable for a user or audience.

We see the three components in play when the case studies harness AI and human intuition and creativity to fuel cultural experiences for audiences and enhance artistic sensemaking and agency. The approach can, in turn, feed back into the design of technologies and shape XAI development. The methodology provides a versatile scaffold for multidisciplinary teams to scope concerns and future scenarios (aspect), articulate interpretable aspects within algorithms (algorithm) , and conduct creative experiments that generate sensory experiences (affect). Using the methodology, designers of intelligent systems can create scenarios in which a higher-level aspect or assumption can be communicated to a user. Design teams then develop the architecture and algorithm and integrate data and models into technical systems and creative works with which an audience can interact to derive learning outcomes that can inform the design of future systems (apprehension). Artistic narrative, imagery, interaction, and abstraction could augment the XAI toolbox or be used as a proxy for the challenging engineering task of developing bespoke explainability algorithms for many situations and users [38].

Combining strategies and methods from XAI and the arts offers the opportunity to develop richer understanding of both the inner workings and the wider implications of AI systems in engaging, experiential ways. We envision this methodology helping humans who interact with AI to viscerally understand various aspects of the life cycle of AI systems. This includes not just the models at their core but also the data collection and processing that inform their design, the way the system has been commissioned and designed, how automated decision- making is situated, what the model actually does, and the relationships between the system and the subjects of its decisions. The “ explanation” is not a technical account of the model or algorithm so much as a representation or instance of an AI system situated in the world and its attributes or implications. It is baked into the design of the experience through either the curation of data, the design of the algorithm, or the way the components are connected. Beyond communicating current knowledge, we propose that an experiential approach can generate new understanding of AI systems --- their operations, limitations, peculiarities, and implications.

The experiential AI methodology bridges humanities and engineering insight and practices. It provides a mechanism for how the arts can interact with science and technology and produce situated, embodied, and intuitive meaning around algorithms and the effects of their deployments. We can think of the process model as a cogwheel that connects machine intelligence to human learning and technical features to higher- level aspects. It can be used to reflect on the interactivity between, on the one hand, the machine intelligence that extracts patterns from observed data to generate decisions and outputs, and on the other, human judgment, creativity, and comprehension. The methodology furthermore involves drawing on XAI philosophies without stopping at functional explanation by embracing human traits such as intuition, risk, and play. Adoption of this transdisciplinary approach requires engagement between communities with different logics, vocabularies, cultures, and goals. The tool is not intended to take the place of a practitioner’s own methodology in developing XAI or practicing the arts but rather to facilitate and broker contributions from across different disciplines and help an individual practitioner guide a team through a transdisciplinary process.

This work is ongoing. Looking ahead, we foresee both weak experiential AI, which makes current XAI more relevant, interpretable, or interactive, and strong experiential AI, which is concerned with fundamentally different ways of inspecting, designing, and understanding AI. Our future research aims to deepen our understanding of the potential contribution of experiential explanations to AI and the arts by evaluating

cognitive shifts and learning outcomes among users through audience studies, the significance of the works through art criticism, and the validity of the algorithms through quantitative experiments.

Conclusion --- Beyond Explanation

We have presented research at the intersection of the arts and XAI. The result is an approach and methodology intended to provide practitioners with increased creative control and agency in co-creative experiments with AI, generate compelling cultural experiences for audiences, and expand human understanding of AI. Experiential AI supports a holistic engagement in the entanglements of people and machines particularly relevant to understanding the nuanced social implications of deployed AI technology. Here, tangible experiences mediate between impenetrable computer code and human comprehension, making not just AI systems but also their underlying values and implications more transparent, strengthening the basis for responsible deployment. This can have reciprocal benefits beyond disciplinary perspectives. For the arts, the methodology can open up algorithms and data as a material, provide structured access to emerging XAI paradigms, and enable artistic works based on rich modes of model understanding and interaction. For AI design, it can illuminate situated and embodied meaning and enrich design ideas with the reinterpretations that can be made with the input of artists while staying faithful to the AI models. For society, this could support decision-making in other fields and enable more inclusive and holistic understanding, helping to underpin fairer, safer, and more accountable technologies. Such experiential interventions can work to reach new audiences, increase the agency of people impacted by these systems, and create spaces for debate and engagement with populations outside the technical center. This transdisciplinary approach helps us to think differently about how algorithms should be designed and invites possibilities for radically new concepts and paradigms.

Acknowledgements

Thanks to Jake Elwes, Anna Ridler, Caroline Sindere, Inés Cámara Leret, Keziah MacNeill, Adam Harvey, Alex Fefegha, Roy Luxford, Veronika Liebl, Amanda Tyndall, Ricarda Bross, Daga Panas, Evan Morgan, Mario Antonioletti, Sohan Seth, Andy McGregor, Miriam Walsh, Janet Archer, and Suzy Glass. Funded by EPSRC/Turing 2.0, AHRC, Creative Scotland, Data-Driven Innovation Programme.

References and Notes

1. D. Gunning, “Explainable artificial intelligence (XAI)”: <https://tinyurl.com/yccmn477> (2017); V. Belle and I. Papantonis, “Principles and Practice of Explainable Machine Learning,” *Frontiers in Big Data* 4: doi:10.3389/fdata.2021.688969 (2021).
2. D. Hemment et al., “AI in the Public Eye: Investigating Public AI Literacy through Critical AI Art,” *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*: doi:10.1145/3593013.3594052 (2023).
3. D. Hemment et al., (2019). *Experiential AI*. *ACM AI Matters*. 5:1. doi:10.1145/3320254.3320264.
4. In this article, we use “co-creation with AI” when speaking of humans and AI, and “co-creative processes” for co-design between humans.
5. G. Montavon, W. Samek, and K.R. Müller, “Methods for interpreting and understanding deep neural networks,” *Digital Signal Processing* 73 (2018): doi:10.1016/j.dsp.2017.10.011.
6. D.S. Weld and G. Bansal, “The Challenge of Crafting Intelligible Intelligence,” arXiv:1803.04263 (2018); M.T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you?: Explaining the predictions of any classifier,” *ACM Sigkdd International conference on knowledge discovery and data mining* (2016).
7. S. Penkov and S. Ramamoorthy, “Using program induction to interpret transition system dynamics”: arXiv:1708.00376 (2017); V. Belle, “Logic Meets Probability: Towards explainable AI systems for uncertain worlds,” *International Joint Conference on Artificial Intelligence* (2017) pp. 5116--5120.
8. K. Sokol and P. Flach, “One Explanation Does Not Fit All: The Promise of Interactive Explanations for Machine Learning Transparency,” *KI--- Künstliche Intelligenz* 34 (2020): doi:10.1007/s13218-020-00637-y.
9. R. Srinivasan and D. Parikh, “Building Bridges: Generative Artworks to Explore AI Ethics,” arXiv:2106.13901.

10. N. Bryan-Kinns et al., "Exploring XAI for the Arts: Explaining Latent Space in Generative Music," arXiv.2308.05496. (2023).
11. T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, (2019) 267.
12. Q.V. Liao and K.R. Varshney, "Human-Centered Explainable AI (XAI): From Algorithms to User Experiences" (2022), arXiv.2110.10790.
13. H. Kaur et al., "Sensible AI: Re-imagining Interpretability and Explainability using Sensemaking Theory," *ACM FAccT* (2022), doi:10.1145/3531146.3533135.
14. For example, U. Ehsan et al., "Expanding Explainability: Towards Social Transparency in AI systems," *Conference on Human Factors in Computing Systems (CHI)* (2021), doi:10.1145/3411764.3445188.
15. A.B. Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI" (2019), arXiv:1910.10045.
16. Royal Society, "Explainable AI: The Basics Policy Briefing" (2019): <https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf>.
17. M. Veale, M. Van Kleek, and R. Binns, "Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making," *Conference on Human Factors in Computing Systems (CHI)* No. 440 (2018)..
18. S. Bory and P. Bory, "New Imaginaries of the Artificial Intelligence," *Im@go. Journal of the Social Imaginary* (2015) p. 6.
19. S. Natale and A. Ballatore, "Imagining the thinking machine: Technological myths and the rise of artificial intelligence," *Convergence* 26, No. 1 (2020).
20. M. Ananny and K. Crawford, "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability," *New Media & Society* 20, No. 3 (2018).
21. S. Kambhampati, "Challenges of Human-Aware AI Systems," *AI Magazine* 41, No. 3 (2020).
22. T. Borkman, "Experiential knowledge: A new concept for the analysis of self-help groups," *Social Service Review* 50, No. 3 (1976).
23. C.V. Smedegaard, "Reframing the role of novelty within social HRI: from noise to information," *ACM/IEEE International Conference on Human-Robot Interaction (IEEE Press, 2019)*.
24. J. Heron and P. Reason, "The practice of co-operative inquiry," *Handbook of Action Research: Participative Inquiry and Practice*, P. Reason and H. Bradbury, eds. (Sage, 2001).
25. D. Hemment, "Reordering the assemblages of the digital through art and open prototyping," *Leonardo* 53, No. 5, 529--536 (2020), doi:10.1162/leon_a_01861.
26. B. Flyvbjerg, "Case Study," in *Strategies of Qualitative Inquiry*, N.K. Denzin and Y.S. Lincoln, eds. (Sage, 2013).
27. See also O.G. Parry, "Review: Jake Elwes, *The Zizi Show---A Deepfake Drag Cabaret* (Edinburgh: The New Real, Edinburgh Futures Institute, 2020)," *Volupté* 4, No. 2 (2021), doi:10.25602/GOLD.v.v4i2.1595.g1709.
28. M. Mateas, (2001). "Expressive AI: A Hybrid Art and Science Practice," *Leonardo* 34, No. 2, 147--153 (2001), doi:10.1162/002409401750184717.
29. D. Grba, "Deep Else: A Critical Framework for AI Art," *Digital 2* (2022), doi:10.3390/digital2010001.
30. D.A. Kolb, "Experiential learning: Experience as the source of learning and development" (FTPress, 2014).
31. N. Vannini et al., "'FearNot!': A computer-based anti-bullying-programme designed to foster peer intervention," *European Journal of Psychology of Education*. 26 (2011).
32. K. Alfrink et al., "Contestable AI by Design: Towards a Framework," *Minds and Machines* (2022), doi:10.1007/s11023-022-09611-z.
33. Compare Mateas [28].
34. D. Hemment et al., "Agency and legibility for artists through Experiential AI," in *The 1st International Workshop on Explainable AI for the Arts (XAIxArts), ACM Creativity and Cognition (C&C)* (2023): <https://arxiv.org/abs/2306.02327>.
35. Hemment et al. [2].
36. Compare Design Council UK, (2005). "A study of the design process--- The Double Diamond" (2005): <https://www.designcouncil.org.uk/our-work/skills-learning/the-double-diamond>.
37. Compare K. O'Hara, "Explainable AI and the philosophy and practice of explanation," *Computer Law & Security Review* 39 (2020), doi:10.1016/j.clsr.2020.105474.
38. Sokol and Flach [8].

Figure 1: Jake Elwes, The Zizi Show (2020). Image © Jake Elwes.

Figure 2 - Anna Ridler and Caroline Sindere, Cypress Trees (2021). Image © Anna Ridler & Caroline Sindere.

Figure 3: From top left, clockwise: 1. The New Real Observatory generative AI platform (2022). 2. The New Real Pavilion and Research Hub (2022). 3. Circular Diffusion N°3, Adam Harvey (2022). 4. The Overlay, Inés Cámara Leret (2022). 5. Photographic Cues, Keziah MacNeill (2022). Photo 2,4,5 © DieFotoFrau. Image3 © Adam Harvey.

Fig. 4. Experiential AI process model.

Table 1: The 4As Conceptual Model

Dimension	Artists' strategies (Hemment et al. 2023)	Zizi (Supplement 2)	Cypress Hybrids (Supplement 3)	The New Real Observatory (Supplement 4)	Input	Output	Theory of Affect	4As
Theme or higher-level feature	Link AI systems to structural issues of power	ML bias; discriminatory design; mis-representations of AI	Hidden human labour in ML; AI-mediated experience of climate crisis	Interpretability of AI and climate data; co-creation and agency.	Current challenges and future scenarios (system-level, cognitive, political, legal, regulatory or institutional)	Socio-technical transition	Milieu	Aspect
Technology process or capability	Make AI tangible and explicit	StyleGAN re-trained on a modified dataset; interactive web interface	Photography; manual sorting, annotation; GAN; video projection	SLIDER; GAN; Word2vec; dimensionality reduction; various interfaces	AI and ML methods and tools, and technical understanding of interpretable features within an algorithm	Technical requirements	Function	Algorithm
Human percept or experience	Engage people emotionally, cognitively and tangibly	Online interaction; captivating imagery; fluid, unclassifiable identities; virtual cabaret	Fine art installation of artist video; curated artifacts from ML pipeline	Data curation; prompt tuning; latent space exploration; fine art installation.	Art making, curation, staging, interface design, performance, audience engagement	Situated experiences	Percept	Affect
Learning or knowledge contribution	Connect art to science, and technology to situated and embodied meaning	The way data and design choices shape what ML does	ML intelligence is actually human intelligence at many points.	The role of training datasets interacting with algorithms	A framework for how people learn through those experiences	Pedagogy, debate, action	Concept	Apprehension

