Identification of Computational and Cultural Explainable AI (CCXAI)

COURTNEY FORD and SUN PARK, School of Information & Communication Studies, University College Dublin, Ireland

Explainable AI (XAI) has made significant strides in technical transparency, yet it maintains a cultural blind spot regarding arts and heritage contexts. While traditional XAI focuses on model behaviour and post-hoc logic, it overlooks the contextual and representational choices embedded in creative processes. This paper introduces the Computational and Cultural Explainable AI (CCXAI) framework to bridge this gap. We distinguish between computational explainability clarifying what a model does and cultural explainability addressing who has made specific choices and under what socio-historical conditions. By synthesising these dimensions, CCXAI provides a shared framework for interdisciplinary practitioners. Our future project on CCXAI aims to integrate algorithmic logic with human-led meaning-making, fostering collaboration across computer science, humanities and social sciences while legitimising AI-enabled cultural production as a sophisticated and transparent practice.

Additional Key Words and Phrases: transparency, computational explainability, cultural explainability, AI-enabled cultural production

ACM Reference Format:
Courtney Ford and Sun Park. 2026. Identification of Computational and Cultural Explainable AI (CCXAI). In *Proceedings of Explainable AI for the Arts Workshop 2026 (XAIxArts 2026).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

AI transparency has a cultural blind spot. The Explainable AI (XAI) field has made significant progress in addressing the opacity of AI systems, including post-hoc explanations, example-based techniques, and human-centred approaches that move explanation closer to the needs of non-technical users. But the growing use of AI in arts and heritage contexts, such as digitisation of heritage sites and artistic practices in the creative industries, surfaces a problem that existing technical and human-centred XAI approaches have not fully addressed. Computational and cultural explainability are fundamentally distinct: the former explains model behaviour and decision making, while the latter concerns the contextual and representational choices embedded in a creative process as it unfolds.

For XAI researchers, transparency and interpretability are distinct concepts [13], and full model transparency is often constrained by model complexity and opacity. For arts and cultural practitioners, transparency extends beyond model behaviour to the representational choices involved in ensuring that datasets and model training adequately and "culturally" reflect a human group and their traditional cultural practices. Human-centred approaches to XAI have made meaningful progress in accounting for user context and decision making needs, including in domains such as a clinical diagnosis [19], and in communicating AI decisions to non-expert audiences [18], but neither technical nor human-centred approaches have bridged the gap between these two understandings of transparency.

Authors' Contact Information: Courtney Ford, courtney.ford@ucd.ie; Sun Park, sun.park@ucd.ie, School of Information & Communication Studies, University College Dublin, Dublin, Ireland.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

This paper therefore proposes Computational and Cultural Explainable AI (CCXAI) as a conceptual framework addressing this gap. CCXAI provides a shared framework for transparency across computational and cultural dimensions of AI-assisted creative practice, grounded in a planned empirical programme. In this paper, we first conceptualise each component in turn, before arguing for why and how they must be considered together. We conclude by outlining the framework's expected contributions to both academic and professional settings at the intersection of computer science, the arts, humanities, and social sciences.

2 Conceptualisation of Computational and Cultural XAI

2.1 Computational Explainability

In this paper, we use the term computational explainability to refer to the methods and techniques developed to make AI model behaviour interpretable to human observers. By situating XAI within a computational frame, we create space for a parallel concept of cultural explainability, arguing that explainability is not a singular concept but one that takes different forms depending on the context in which AI is deployed. Computational explainability encompasses a range of approaches, including post-hoc explanation methods [1], example-based techniques including factual and counterfactual explanations [3, 9], and concept-based approaches [10]. While XAI research distinguishes between transparency and interpretability as distinct concepts, we use transparency here as a shared umbrella term to enable meaningful comparison with cultural explainability, noting that the absence of common vocabulary across these domains is itself a symptom of the gap the CCXAI framework seeks to address.

Empirical work has shown that users' understanding of post-hoc explanations changes significantly depending on their familiarity with the domain being explained [8], suggesting that explanation methods effective at describing model behaviour do not necessarily serve the needs of the people who must act on them. Human-centred XAI (HCXAI) emerged from growing recognition that explainability must account for the needs, contexts, and varying expertise of the people who interact with AI systems [6, 12]. By prioritising the question of who explanations are for and drawing on participatory and co-design methods from HCI, HCXAI has made meaningful progress in moving explainability research toward user-facing explanation design. What constitutes a useful explanation remains contested, with usefulness contingent on the user, their expertise, and the context in which the explanation is received [14].

Computational explainability, even in its most human-centred forms, remains oriented toward explaining what a model does and why [11]. It does not address the question of what a model culturally represents nor does it account for the contexts, values, and representational choices embedded in AI-assisted creative practice. Addressing these questions requires a different conception of explainability which is one concerned not with what a model produces but with the cultural and creative conditions under which it is developed and deployed.

2.2 Cultural Explainability

Computational explainability is grounded in the identification of clear cause-and-effect relations. In cultural contexts, however, this replicability assumption does not always hold. The same AI model and datasets can produce divergent outputs depending on the subjective creative choices of the practitioner, for instance, two artists working with Irish basketmaking data will generate different results if their curation reflects distinct stylistic preferences, such as a traditional Creel style versus a more contemporary one [15]. Explainability in AI-human creation in the cultural field must account for the creator's subjective creative process, which generates (a) layer(s) of non-replicable artistic intentionality that transcends the underlying model behaviour.

Given that the notion of explainability is grounded in addressing the opacity inherent in AI systems [2], the black-box problem within the cultural sector needs to be addressed prior to conceptualising cultural explainability. Park [16] defines cultural black boxes as "the opaque decisions that researchers need to uncover about their cultural subjects as far as possible to legitimise their ways of designing and deploying AI systems to create cultural content," referring to cultural explainability as "a procedural clarity stemming from revealing the cultural black boxes". While computational explanations are typically generated after a models output is complete [17], cultural explainability is implemented during, or simultaneously with, a creative process. Suppose that a heritage digitisation project is required to satisfy only computational XAI criteria: the emphasis would likely fall on rendering model outputs legible by identifying which datasets and fine-tuning protocols contributed to a particular classification. By contrast, cultural explainability raises a different set of questions from the outset: who selected the datasets under what institutional or historical assumptions, and with what understanding of the relevant communities represented in the heritage dataset.

A similar distinction of cultural explainability applies in contemporary AI art. For an AI-assisted visual art project, the artist(s) should be able to justify the situated conditions underlying the work: the artistic histories mobilised through the generated output, how aesthetic preferences informed the curation process, and what institutional or market conditions was considered. Cultural explainability thus moves beyond the technical intelligibility of the system's operations to reveal the broader cultural references and human-led meaning-making that define the human-AI creative process. It does not replace computational explainability, but rather complements it by foregrounding the social and institutional conditions that shape the development of AI systems and the decision-making practices of their users. By shifting the focus from technical outputs to the artist's intent, this approach seeks to empower human creativity rather than replace it, while simultaneously addressing the political biases that are embedded within machine learning models [4].

3 The CCXAI Framework and a Future Direction

The preceding sections have established two distinct conceptions of explainability operating within shared domains. Computational explainability addresses what a model does and why, producing explanations after a model has generated its output. Cultural explainability addresses who made what choices and under what conditions, operating concurrently with a creative process rather than retrospectively. Read together, these two conceptions diverge in important ways: they operate at different points in time, serve different transparency functions, and use vocabularies that do not readily translate across disciplines.

Despite growing interest in applying AI to arts and heritage contexts, the application of XAI in these domains remains limited [5, 7]. Computational and cultural practitioners working on the same project may each have an understanding of transparency within their own domain, while having no shared language to articulate what the other's obligations even are. Current computational XAI frameworks tend to assume a universality of cognition that does not account for the socio-cultural context practitioners in creative and cultural fields bring to their work [11]. This gap reflects a division between technical and cultural knowledge that existing XAI frameworks have not yet bridged.

Addressing this gap requires more than applying existing XAI methods to cultural contexts. It raises questions about whether current explanation methods are adequate for creative and cultural practice, what evaluation approaches could capture both algorithmic and cultural transparency, and how interdisciplinary teams might be structured so that cultural practitioners function as co-designers rather than end users of explanation systems. CCXAI is proposed as a first step toward that integration, centring interdisciplinary practitioners as active architects of AI legibility and ensuring that explanations are both technically logical and culturally persuasive.

To empirically test our CCXAI framework, we will design a participatory workshop bringing together AI practitioners and cultural practitioners. We seek to discover how the academic formulation of CCXAI may be translated into more user-oriented language and practical guidance.

To this end, we will select three distinct types of heritage and artistic practices utilising AI technologies: simulation video creation on falconry (preservation of past culture), modern-style rearrangement of traditional music (adjustment of past culture) and generative short film production (creation of contemporary culture). Workshops will be organised, bringing together five AI practitioners and five heritage and arts practitioners to discuss the implementation of each case study. The workshops will adopt a semi-structured and participatory format: Four sessions will be structured around the general AI project life cycles (from ideation and data collection to model training and output selection), and participant discussion will be designed employing HCI methods such as personas and collaborative storyboarding. The workshop data will be qualitatively analysed to produce four core outcomes: a conceptual taxonomy to assist interdisciplinary collaborators in unifying their key terminology; metadata-related checklists for data explainability; paradata checklists for model training and output selection explainability; and insights into new modes of co-creation for both AI developers and cultural practitioners.

This project contributes across methodological social dimensions of the implementation of interdisciplinary XAI, offering guidance on documenting both meta-data and para-data of human-AI creation, thereby capturing not only the characteristics of datasets and models, but also the curatorial choices that inform AI-based cultural production. Our project aims to promote the recognition of AI-enabled creation as a legitimate mode of cultural production.

References

[1] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115. doi:10.1016/j.inffus.2019.12.012

[2] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3, 1 (2016), 2053951715622512. arXiv:https://doi.org/10.1177/2053951715622512 doi:10.1177/2053951715622512

[3] Ruth M. J. Byrne. 2019. Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 6276–6282. doi:10.24963/ijcai.2019/876

[4] Kate Crawford. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press. http://www.jstor.org/stable/j.ctv1ghv45t

[5] Natalia Díaz-Rodríguez and Galena Pisoni. 2020. Accessible Cultural Heritage through Explainable Artificial Intelligence. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 317–324. doi:10.1145/3386392.3399276

[6] Upol Ehsan and Mark O. Riedl. 2020. Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach. In *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*, Constantine Stephanidis, Masaaki Kurosu, Helmut Degen, and Lauren Reinerman-Jones (Eds.). Springer International Publishing, Cham, 449–466.

[7] Anna Foka and Gabriele Griffin. 2024. AI, Cultural Heritage, and Bias: Some Key Queries That Arise from the Use of GenAI. *Heritage* 7, 11 (2024), 6125–6136. doi:10.3390/heritage7110287

[8] Courtney Ford and Mark T. Keane. 2023. Explaining Classifications to Non-experts: An XAI User Study of Post-Hoc Explanations for a Classifier When People Lack Expertise. In *Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges*, Jean-Jacques Rousseau and Bill Kapralos (Eds.). Springer Nature Switzerland, Cham, 246–260.

[9] Eoin M. Kenny and Mark T. Keane. 2020. On Generating Plausible Counterfactual and Semi-Factual Explanations for Deep Learning. arXiv:2009.06399 [cs.LG] https://arxiv.org/abs/2009.06399

[10] Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, and Rory Sayres. 2018. Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV). arXiv:1711.11279 [stat.ML] https://arxiv.org/abs/1711.11279

[11] Hana Kopecka and Jose Such. 2020. Explainable AI for Cultural Minds. https://sites.google.com/view/dexahai-at-ecai2020/home Workshop on Dialogue, Explanation and Argumentation for Human-Agent Interaction , DEXAHAI ; Conference date: 07-09-2020.

[12] Q. Vera Liao and Kush R. Varshney. 2022. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences. arXiv:2110.10790 [cs.AI] https://arxiv.org/abs/2110.10790

[13] Zachary C. Lipton. 2018. The mythos of model interpretability. *Commun. ACM* 61, 10 (Sept. 2018), 36–43. doi:10.1145/3233231

[14] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38. doi:10.1016/j.artint.2018.07.007

[15] National Inventory of Intangible Cultural Heritage. [n. d.]. Basketmaking. https://nationalinventoryich.ccs.gov.ie/basketmaking/

[16] Sun Park. 2025. Revisiting Transparency in AI-based Creation: Unpacking 'Cultural Black Boxes' towards 'Cultural Explainability'. In *2025 IEEE International Symposium on Technology and Society (ISTAS)*. 1–7. doi:10.1109/ISTAS65609.2025.11269653

[17] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215. doi:10.1038/s42256-019-0048-x

[18] Beatriz Severes, Carolina Carreira, Ana Beatriz Vieira, Eduardo Gomes, João Tiago Aparício, and Inês Pereira. 2023. The Human Side of XAI: Bridging the Gap between AI and Non-expert Audiences. In *Proceedings of the 41st ACM International Conference on Design of Communication* (Orlando, FL, USA) *(SIGDOC '23)*. Association for Computing Machinery, New York, NY, USA, 126–132. doi:10.1145/3615335.3623062

[19] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3290605.3300831