Reflecting Human Values in XAI: Emotional and Reflective Benefits in Creativity Support Tools

Samuel Rhys Cox
srcox@cs.aau.dk
Aalborg University
Aalborg, Denmark

Helena Bøjer Djernæs
hbd@cs.aau.dk
Aalborg University
Aalborg, Denmark

Niels van Berkel
nielsvanberkel@cs.aau.dk
Aalborg University
Aalborg, Denmark

Abstract

In this workshop paper, we discuss the potential for measures of *user-centric benefits* (such as emotional well-being) that could be explored when evaluating explainable AI (XAI) systems within the arts. As a background to this, we draw from our recent review of creativity support tool (CST) evaluations, that found a paucity of studies evaluating CSTs for user-centric measures that benefit the user themselves. Specifically, we discuss measures of: (1) developing intrinsic abilities, (2) emotional well-being, (3) self-reflection, and (4) self-perception. By discussing these user-centric measures within the context of XAI and the arts, we wish to provoke discussion regarding the potential of such measures.

ACM Reference Format:
Samuel Rhys Cox, Helena Bøjer Djernæs, and Niels van Berkel. 2025. Reflecting Human Values in XAI: Emotional and Reflective Benefits in Creativity Support Tools. In *Proceedings of Explainable AI for the Arts Workshop 2025 (XAIxArts 2025)*. ACM, New York, NY, USA, 3 pages.

1 A Brief Background and Motivation

Explainable AI (or XAI) is a field focused on making the outputs and inner workings of AI models more interpretable and understandable to human users [20]. This could be via visual [1, 8] or text-based [12, 17] explanations of AI models. For example, an XAI system could explain a decision in relation to a multitude of complex impacting variables (such as visualising physiological measures affecting a health diagnosis [20], or factors affecting the price of a home [1]). This can be applied to domains such as stress coping strategies, where, for example, text-based counterfactual explanations [12] can be provided to help users select a preferred stress coping strategy.

However, the provision of AI explanations has historically been more productivity-centred [5, 19], which may conflict with the purpose of the arts where one may want to focus on improving aspects to the *user themselves* such as increased feelings of empowerment, enjoyment, and ownership of creative artefacts. The feeling of AI impeding these creative ideals has been found in multiple domains (such as creative writing [4, 6]) where people receiving AI support wish to retain feelings of control, autonomy, and ownership during the creative process. Additionally, the arts have been shown to improve both people's physical health and well-being (see a scoping review of over 3000 papers by Fancourt and Finn for a very clear and broad discussion of this [9]).

Given this background, in this workshop paper we discuss how explanations from XAI systems could be evaluated to focus on outcomes that benefit the *users themselves*, such as emotional well-being or self-reflection. This is contrast to perhaps "typical" evaluations of XAI systems, that may deploy more productivity-focused measures. For example, Rong et al. reviewed measures used to (empirically) evaluate XAI systems, and found that prior work measured explanations in terms of trust, understanding, usability, and human-AI collaboration performance [19]. While these measures certainly make sense (as one may expect understandable and trustworthy explanations) we wish to discuss the potential for measures of benefits to the *user themselves* such as emotional well-being.

This motivation comes from our review of measures used in the empirical evaluation of creativity support tools (CST) [6]. Here, our review of ACM DL publications (2015–2024) found that, although most studies evaluated user experience or creative artefact quality, only 15% assessed CSTs using user-centric benefits (i.e., developing intrinsic abilities, emotional well-being, self-reflection, and self-perception). We highlight these user-centric outcomes, contextualise them within XAI and the arts, and discuss future directions for XAI in CSTs beyond productivity [5, 6]).

2 What Measures of User-Centric Benefits Could We Use?

We frame our discussion of user-centric benefits around measures that emerged in our review of CST evaluations [6] (see [6, §4.3] for further prior-work). Further, this workshop paper is differentiated against our review, as our review did not contain or discuss applications to XAI. Additionally, a limitation of this paper is that we cannot claim to have reviewed all existing measures using XAI in the arts. However, while this will in no way be exhaustive of *all* potential measures of user-centric benefits, we hope to spur discussion and potentially motivate focus of future XAI studies.

We highlight four user-centric benefits: (1) Developing Intrinsic Abilities, (2) Emotional Well-being, (3) Self-Reflection, and (4) Self-Perception. These outcomes have seen limited empirical evaluation in CST, and we frame them in XAI and the arts to spur future work.

Developing Intrinsic Abilities: When evaluating XAI for CSTs, the development of intrinsic abilities, such as learning concepts and terminology, or improving creative skills could be measured. Connecting learning and XAI may seem most apparent where one may expect a detailed explanation (that is transparent about an AI's workings or recommendations) to help users form better mental models and learn more effectively. Contrasting this however, may be a perspective that XAI should should scaffold (e.g., provide hints, ask probing questions, or offer appropriate resources) to help users *discover* solutions. Findings that generative-AI may reduce critical thinking could enforce this [14].

XAIxArts 2025, Online
2025.

Within prior work, providing different visual explanations has been found to aid medical students when training to administer ultrasound [8]. In our review of CST evaluations [6], studies measuring developments of intrinsic abilities typically use pre- and post-tests to measure change [6], and such measures and experiment procedures would likely be necessary in XAI evaluations. For example, Alves et al. used validated measures and procedures to evaluate improvements in creative skills [2]. Drawing these together, measuring the impact of XAI on development of intrinsic abilities would provide seemingly fertile ground for investigation, with possibility to measure improvements in learning and creative skills specific to the arts.

Emotional Well-being: Creative activities are often pursued for their emotional and psychological rewards (think the therapeutic effects and joy of creating [9]), and AI tools should ideally amplify, not dampen, these benefits. Therefore a potential avenue of XAI and the arts, would be to develop explanations that are designed to improve people's emotional well-being. Here, XAI could: be emotionally-positive in *style* and *content* [15], focus on autonomy and empowerment (moderating well-being), and avoid negative comparisons and framing (e.g., explanations could deter perfectionism bias: "*Please remember: reference images are guides, not rules!*").

Within CSTs, Gonçalves et al. [11] found that creative writing support tools helped improve the well-being of marginalised youth in a two-week longitudinal study. Within XAI, a recent study found that counterfactual explanations can be given to help people choose coping strategies that lead to reductions in stress (using a validated measure of stress) [12].

Self-Reflection: XAI could be designed to promote people's self-reflection. Within the arts and HCI, validated measures have been developed to measure reflection as a result of creative activities. For example, prior CST evaluations have used the creativity-specific "*Reflection in Creative Experience*" (RiCE) questionnaire [10], and the more general "*Technology-Supported Reflection Inventory*" (TSRI) [3]. Further, prior work has investigated whether interactions with CSTs led to users learning new things about themselves, or ideas to overcome challenges as a result of interactions. From this, a natural progression could be the design of explanations to spur self-reflection beyond explaining the workings of AI models.

For example, Yan et al. developed NaCanva [22] to help children develop mood boards to aid in personal reflection. Here, they measured reflection from a variety of novel perspectives such as children's acting and caring on nature, awareness and emotional ties to nature, and feelings and identity within nature. Further, recent work investigated the impact of context and modality (text, image, audio, and video) on self-reflection [23] demonstrating a potential for XAI modality to be explored for self-reflection.

Self-Perception: Finally, XAI systems could be designed to enhance people's self-perceptions, such as self-efficacy and sense of achievement. On from this, AI systems that are more interactive and explainable could *invite the user* into the loop (so the process still feels like their creative journey), rather than a one-click solution that makes the user feel superfluous. Additionally, if users understand what the AI is doing and can guide it, they may retain a stronger sense of ownership and achievement.

Motivation behind these self-perceptions has received more attention within creativity [16, 21]. For one example, AI writing support has been found to improve performance but lower intrinsic motivation [21]. Additionally, the *style* of explanations can impact self-perception. For example, when completing learning tasks alongside conversational AI feedback, the tone of AI feedback can improve people's self-efficacy [13].

3 The Interplay of User-Centric Benefits

To some extent, the user-centric benefits discussed above are inherently connected. For instance, when providing an AI suggestion, XAI could centre the user to help *empower* them, such as the explanation: "*Your last five painting sessions have used warm hues — this suggested palette* [...]". Such an explanation can directly improve the user's **self-perception** and, in turn, enhance their **emotional well-being**. Similarly, without gamifying creativity sessions, XAI could focus on past user achievements to foster **self-reflection** and **emotional well-being**. For example, after offering a suggestion, an explanation could include: "*Your last three storyboards used high-contrast lighting to powerful effect. Take a moment to reflect on that success to boost your emotional well-being*". As a final example, XAI could help people **learn** and gain **self-confidence** by structuring explanations to be more conversational or omitting certain details (cf. studies showing that people are more comfortable expressing emotions to chatbots when they receive less feedback [7, 18]).

In conclusion, we have provided an overview of several user-centric benefits that could be the focus of future evaluations of XAI systems within the arts. These benefits, could both shape the evaluation of XAI tools, as well as the design of explanations used themselves. From this, we hope to spur discussion and motivate fellow researchers to focus on evaluations of XAI tools that focus on user-centric benefits, such as emotional well-being and self-reflection.

Acknowledgments

We would like to thank the workshop co-chairs Corey Ford and Elizabeth Wilson for organising the workshop, and committee members for detailed reviews that helped improve this paper. This work was supported by the Carlsberg Foundation, grant CF21-0159.

References

[1] Ashraf Abdul, Christian von der Weth, Mohan Kankanhalli, and Brian Y. Lim. 2020. COGAM: Measuring and Moderating Cognitive Load in Machine Learning Model Explanations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. doi:10.1145/3313831.3376615

[2] Patrícia Alves-Oliveira, Patrícia Arriaga, Matthew A Cronin, and Ana Paiva. 2020. Creativity Encounters Between Children and Robots. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 379–388. doi:10.1145/3319502.3374817

[3] Marit Bentvelzen, Jasmin Niess, Mikołaj P Woźniak, and Paweł W Woźniak. 2021. The Development and Validation of the Technology-Supported Reflection Inventory. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–8. doi:10.1145/3411764.3445673

[4] Oloff C. Biermann, Ning F. Ma, and Dongwook Yoon. 2022. From Tool to Companion: Storywriters Want AI Writers to Respect Their Personal Values and Writing Strategies. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference (DIS '22)*. doi:10.1145/3532106.3533506

[5] Nick Bryan-Kinns, Shuoyang Zheng, Francisco Castro, Makayla Lewis, Jia-Rey Chang, Gabriel Viglensoni, Terence Broad, Michael Paul Clemens, and Elizabeth Wilson. 2025. XAIxArts Manifesto: Explainable AI for the Arts. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. doi:10.1145/3706599.3716227

[6] Samuel Rhys Cox, Helena Bøjer Djernæs, and Niels van Berkel. 2025. Beyond Productivity: Rethinking the Impact of Creativity Support Tools. In *Proceedings of the 17th Conference on Creativity & Cognition*. doi:10.1145/3698061.3726924

[7] Samuel Rhys Cox, Rune Møberg Jacobsen, and Niels van Berkel. 2025. The Impact of a Chatbot's Ephemerality-Framing on Self-Disclosure Perceptions. In *Proceedings of the 7th Conference on Conversational User Interfaces (CUI '25)*. doi:10.48550/arXiv.2505.20464

[8] Helena Bøjer Djernæs, Rune Møberg Jacobsen, Simo Hosio, and Niels van Berkel. 2025. Visual Augmentations for Ultrasound Assessment Training of Medical Students. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. doi:10.1145/3706598.3714004

[9] Daisy Fancourt and Saoirse Finn. 2019. What is the evidence on the role of the arts in improving health and well-being? A scoping review. *Health Evidence Network synthesis report* 67 (November 2019). https://www.who.int/europe/publications/i/item/9789289054553

[10] Corey Ford and Nick Bryan-Kinns. 2023. Towards a Reflection in Creative Experience Questionnaire. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. doi:10.1145/3544548.3581077 (*Note*: RiCE questionnaire also found via: https://ricequestionnaire.github.io/).

[11] Frederica Gonçalves, Pedro Campos, Julian Hanna, and Simone Ashby. 2015. You're the Voice: Evaluating User Interfaces for Encouraging Underserved Youths to express themselves through Creative Writing. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. doi:10.1145/2757226.2757236

[12] Gyuwon Jung and Uichin Lee. 2025. CounterStress: Enhancing Stress Coping Planning through Counterfactual Explanations in Personal Informatics. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. doi:10.1145/3706598.3713730

[13] Naja Kathrine Kollerup, Maria-Theresa Bahodi, Samuel Rhys Cox, and Niels van Berkel. 2025. Enhancing Self-Efficacy in Health Self-Examination through Conversational Agent's Encouragement. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. doi:10.1145/3706598.3713142

[14] Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. doi:10.1145/3706598.3713778

[15] Han Li, Renwen Zhang, Yi-Chieh Lee, Robert E Kraut, and David C Mohr. 2023. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digital Medicine* 6, 1 (2023), 236. https://doi.org/10.1038/s41746-023-00979-5

[16] Michele Newman, Kaiwen Sun, Ilena B Dalla Gasperina, Grace Y. Shin, Matthew Kyle Pedraja, Ritesh Kanchi, Maia B. Song, Rannie Li, Jin Ha Lee, and Jason Yip. 2024. "I want it to talk like Darth Vader": Helping Children Construct Creative Self-Efficacy with Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. doi:10.1145/3613904.3642492

[17] Saumya Pareek, Niels van Berkel, Eduardo Velloso, and Jorge Goncalves. 2024. Effect of Explanation Conceptualisations on Trust in AI-assisted Credibility Assessment. *Proc. ACM Hum.-Comput. Interact.* CSCW2. doi:10.1145/3686922

[18] SoHyun Park, Anja Thieme, Jeongyun Han, Sungwoo Lee, Wonjong Rhee, and Bongwon Suh. 2021. "I wrote as if I were telling a story to someone I knew.": Designing Chatbot Interactions for Expressive Writing in Mental Health. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21)*. doi:10.1145/3461778.3462143

[19] Yao Rong, Tobias Leemann, Thai-Trang Nguyen, Lisa Fiedler, Peizhu Qian, Vaibhav Unhelkar, Tina Seidel, Gjergji Kasneci, and Enkelejda Kasneci. 2023. Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023). doi:10.1109/TPAMI.2023.3331846

[20] 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*. doi:10.1145/3290605.3300831

[21] Suqing Wu, Yukun Liu, Mengqi Ruan, Siyu Chen, and Xiao-Yun Xie. 2025. Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. *Scientific Reports*. doi:10.1038/s41598-025-98385-2

[22] Zihan Yan, Yanhong Wu, Danli Luo, Chao Zhang, Qihang Jin, Wei Chen, Yingcai Wu, Xiang'Anthony' Chen, Guanyun Wang, and Haipeng Mi. 2023. NaCanva: Exploring and Enabling the Nature-Inspired Creativity for Children. *Proceedings of the ACM on Human-Computer Interaction* 7, MHCI (2023). doi:10.1145/3604262

[23] ShunYi Yeo, Zhuoqun Jiang, Anthony Tang, and Simon Tangi Perrault. 2025. Enhancing Deliberativeness: Evaluating the Impact of Multimodal Reflection Nudges. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 515, 26 pages. doi:10.1145/3706598.3714189