Composers' Evaluations of an AI Music Tool: Insights for Human-Centred Design

Eleanor Row
Centre for Digital Music
Queen Mary University of London
London, UK
e.r.v.row@qmul.ac.uk

György Fazekas
Centre for Digital Music
Queen Mary University of London
London, UK
george.fazekas@qmul.ac.uk

Abstract

We present a study that explores the role of user-centred design in developing Generative AI (GenAI) tools for music composition. Through semi-structured interviews with professional composers, we gathered insights on a novel generative model for creating variations, highlighting concerns around trust, transparency, and ethical design. The findings helped form a feedback loop, guiding improvements to the model that emphasised traceability, transparency and explainability. They also revealed new areas for innovation, including novel features for controllability and research questions on the ethical and practical implementation of GenAI models.

1 Introduction

Generative AI (GenAI) in creative fields faces challenges due to a divide between human-centric and model-centric research approaches, reflecting a broader disconnect between AI researchers and creative professionals overall, which we highlight in the music domain [13, 17, 1]. Concerns about trust, ethics, and usability stress a need for interdisciplinary dialogue, which can be used to help guide the development of GenAI tools that meet real-world needs. To address this in the music domain, we conducted a qualitative study with professional composers, exploring how generative music models can serve as creative aids in music composition. Through semi-structured interviews, composers interacted with a baseline transformer model designed to create musical variations [15, 14], providing feedback on outputs, evaluation metrics, transparency, and ethical concerns. This approach established a feedback loop that helps to inform GenAI tool design and foundational research questions, aligning AI development with the practical and ethical needs of end-users. Although centred on music, our findings offer valuable insights for broader applications of GenAI in creative contexts.

2 Methodology

Our methodology employed a User-Centred and Participatory Design-based approach [2, 16, 10, 8, 11]. Four professional composers participated in semi-structured interviews, where they were introduced to a novel music task, Music Overpainting, using a transformer model designed to generate jazz piano variations from basic melodies and chords [14, 15]. Adapted from the music transformer architecture proposed by Huang et al. [9], this model serves as a task-specific baseline for Music Overpainting. Composers could listen to the model's inputs and outputs as much as needed to familiarise themselves with its capabilities. We asked targeted questions for their opinions on how to define a variation, the success of the model, trust and ethical concerns, user integration and controllability aspects as detailed in Appendix A.1. These questions helped to address critical aspects of the model's development that were not fully covered by existing literature, aiming to fill knowledge

Creativity & Generative AI Workshop, 38th Conference on Neural Information Processing Systems (NeurIPS 2024).

gaps and reduce researcher bias, particularly given the novelty of the generative task. Responses were analysed by one coder, using a reflexive thematic analysis-inspired approach [7, 4, 3]. To minimise biases, the coder engaged in regular reflexive practices, including keeping a reflective journal and self-reflecting on biases that they brought in as a researcher. Additionally, we asked the composers to complete a short questionnaire based on [12], collecting demographics and information on their musical education and experience as seen in Appendix A.5. While the small sample size and focus on a specific generative task lead to context-specific findings that may not generalise to broader creative communities or demographics, this study serves as an initial exploration that underscores the value of engaging directly with end-users to identify potential sources of researcher bias and enrich GenAI model development, with insights that may otherwise be overlooked. As a preliminary qualitative study, future research could build upon these findings to enhance generalisability by expanding the outputs to additional musical genres and involving a larger and more diverse participant pool.

3 Findings and Discussion

The key findings from engaging with professional composers, along with the considerations that influenced the model's development, are summarised in Table 1. This table highlights the main themes identified during the study, the specific feedback from composers (see example transcripts in Appendix A.2), and the resulting design decisions and adjustments made to the model.

Table 1: Summary of key findings and considerations

Theme	**Findings**	**Considerations**
Definition of Variation	Composers defined variations as maintaining a recognisable connection to the source, regardless of transformation.	Helped to further define the novel generative task. Reframed evaluation metrics to focus on traceability, emphasising practical relevance over objectivity.
Traceability and Transparency	Composers valued transparency in how variations evolve from the source material.	Changes made to the sampling technique, enabling adjustments to the model and user interface (UI) to visually represent transformation steps.
Ethical and Trust Concerns	Concerns over usable data i.e. trained with consent and sources were traceable.	Questions to consider on model training data, its source, and how useful it is.
Applicability of Outputs	Composers found traditional subjective evaluations insufficient, preferring assessments of applicability to their workflows.	Reshaped evaluation questions to focus on how outputs fit within the composers' creative processes, shifting from "Is it good?" to "Is it useful?"
Practical Model Design	Highlighted practical needs such as controllable, lightweight models that integrate well into existing workflows.	Influenced decisions on model architecture and parameter size, sampling techniques, and UI considerations for an initially simple design that later advances.

We used composer feedback to improve the same generative variation model introduced in the study. Key concerns around trust, transparency, and ethical design were raised. The findings helped to clarify the generative task which had been defined through researcher bias, and exposed gaps in existing evaluation metrics for new creative tasks, revealing a need for subjective and objective criteria that reflect real-world creative practices. Feedback also emphasised traceability around data collection and model outputs, and explainability as essential for building trust and aligning AI outputs with artistic intentions [5, 6].

While many of the themes identified in Table 1 align with domain-general principles from fields like Human-Centred AI (HCAI)— such as interpretability, explainability, and data transparency [18] — the direct feedback from composers added a complementary, application-specific perspective, and helped to reveal how the straightforward application of domain-general principles may overlook subtle requirements unique to certain creative tasks. For instance, composers expressed a strong desire for traceability, viewing it not only as a means to track how variations related to, and evolved from the source material but also as a way to support their creative choices. Some suggested that having the ability to visually and interactively follow how the model transforms the output, as a transparent traceable process, could help them better understand and select variations that align with their artistic intentions, providing them with greater control over their creative decisions. Likewise, the call for controllable and lightweight models reflected not only a general design preference but a specific need for adaptable tools that fit seamlessly into the composers' iterative workflows. Additionally, the study identified new areas to innovate, including the need for novel controllable features. Research questions on how features could be implemented emerged, including approaches to model sampling and tuning.

Our collaboration helped reduce researcher bias and revealed previously unconsidered user assumptions about creativity and AI's role as a collaborative partner. It highlighted the need for adaptable, user-responsive AI systems, and encouraged researchers to shift from model-centric to human-centric approaches that prioritise real-world applicability. By incorporating these insights, we can refine model features and interface elements to align AI capabilities with artistic workflows, emphasising a balanced approach where technical functionality is grounded in practical relevance for creative end-users. This in turn fostered interdisciplinary mutual learning by involving composers directly in the development process. Maintaining a continuous feedback loop with creative professionals is crucial, ensuring that AI evolves to meet user needs rather than requiring users to adapt to the AI.

4 Conclusion and Future Work

Looking ahead, extending this approach to other GenAI domains can lay the groundwork for interdisciplinary collaborations that prioritise user-centred AI design. We urge the research community to adopt continuous engagement with end-users, explore innovative features, and focus on designing AI systems that genuinely support and enhance artistic workflows, striking a balance between AI as a creative agent, the needs of the creative community, and the evolving technical landscape of AI development.

5 Acknowledgements

The first author is a research student at the UKRI Centre for Doctoral Training in Artificial Intelligence and Music, supported by UK Research and Innovation *[Grant number: EP/S022694/1]*.

References

[1] Oded Ben-Tal, Matthew Tobias Harris, and Bob L.T. Sturm. How Music AI Is Useful: *Engagements with Composers, Performers and Audiences*. *Leonardo*, 54(5):510–516, October 2021.

[2] Hugh Beyer and Karen Holtzblatt. *Contextual design: defining customer-centered systems*. Morgan Kaufmann, San Francisco, Calif, 1998.

[3] Virginia Braun and Victoria Clarke. (Mis)conceptualising themes, thematic analysis, and other problems with Fugard and Potts' (2015) sample-size tool for thematic analysis. *International Journal of Social Research Methodology*, 19(6), June 2016. Publisher: Taylor & Francis (Routledge).

[4] Virginia Braun, Victoria Clarke, and Debra Gray, editors. *Collecting Qualitative Data: A Practical Guide to Textual, Media and Virtual Techniques*. Cambridge University Press, Cambridge, 2017.

[5] N. Bryan-Kinns, B. Banar, C. Ford, C. Reed, Y. Zhang, S. Colton, J. Armitage, and NeurIPS 2021-eXplainable AI Approaches for Debugging and Diagnosis Workshop. Exploring XAI for the Arts: Explaining Latent Space in Generative Music, December 2022.

[6] Nick Bryan-Kinns, Ashley Noel-Hirst, and Corey Ford. Using Incongruous Genres to Explore Music Making with AI Generated Content. In *Creativity and Cognition*, pages 229–240, Chicago IL USA, June 2024. ACM.

[7] Victoria Clarke and Virginia Braun. *Successful Qualitative Research: A Practical Guide for Beginners*. Sage, 2013.

[8] Rebecca Fiebrink and Baptiste Caramiaux. The Machine Learning Algorithm as Creative Musical Tool. In *The Oxford Handbook of Algorithmic Music*. Oxford Handbooks, online edition, February 2018.

[9] Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Ian Simon, Curtis Hawthorne, Noam Shazeer, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, and Douglas Eck. Music transformer. In *International Conference on Learning Representations*, 2019.

[10] Trevor Hunter, Peter Worthy, Ben Matthews, and Stephen Viller. Using Participatory Design in the Development of a New Musical Interface: Understanding Musician's Needs beyond Usability. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, pages 268–271, Nottingham United Kingdom, September 2019. ACM.

[11] Michael J. Muller and Sarah Kuhn. Participatory Design. *Communications of the ACM*, 36(6):24–28, June 1993.

[12] D. Müllensiefen, B. Gingras, J. Musil, and L. Stewart. Measuring the facets of musicality: The Goldsmiths Musical Sophistication Index (Gold-MSI). *Personality and Individual Differences*, 60:S35, April 2014.

[13] Michele Newman, Lidia Morris, and Jin Ha Lee. HUMAN-AI Music Creation: Understanding the perceptions and experiences of music creators for ethical and productive collaboration. In *Proceedings of the 24th International Society for Music Information Retrieval Conference*, November 2023.

[14] Eleanor. Row and György. Fazekas. Advancing AI in Music Composition: Refining the Generative Music Overpainting Task. In *AES International Symposium on AI and the Musician*, April 2024.

[15] Eleanor Row, Jingjing Tang, and György Fazekas. JAZZVAR: A Dataset of Variations found within Solo Piano Performances of Jazz Standards for Music Overpainting. In *Proceedings of the 16th International Symposium on Computer Music Multidisciplinary Research*, pages 265–276, Tokyo, Japan, July 2023. Springer.

[16] Douglas Schuler and Aki Namioka. *Participatory Design: Principles and Practices*. CRC Press, March 1993.

[17] Soumya Sai Vanka, Maryam Safi, Jean-Baptiste Rolland, and G. Fazekas. Adoption of AI Technology in the Music Mixing Workflow: An Investigation. In *AES EUROPE 2023*, September 2023.

[18] Daniel S. Weld and Gagan Bansal. The challenge of crafting intelligible intelligence. *Commun. ACM*, 62(6):70–79, May 2019.

A Appendix / supplemental material

A.1 Questions

Participants were asked fifteen semi-structured questions designed to encourage open dialogue and facilitate in-depth discussions. The semi-structured format allowed participants to freely express their thoughts and opinions, fostering a conversational environment. The questions focused on the participants' understanding of variations in music, their perceptions of AI-generated outputs, and potential improvements for AI tools in creative workflows.

1. How would you define a "variation" in the context of music?
2. On a scale of 1 to 10, how much do you agree with the statement "The outputs generated were successful variations of the original piece?"
3. Follow-up: Can you explain your rating and what you found successful or unsuccessful about the variations?
4. Would you consider using a model like this in your music creation process?

5. Follow-up: If yes, at what stage(s) would you use it - such as improvisation, ideation, or redrafting? If not, why not?
6. Do you think using a model like this could enhance or hinder your creative process?
7. How would you envision this model being integrated into your workflow?
8. Follow-up:Would you prefer it as a separate system or integrated into an existing UI? Why?
9. How would you prefer to interact with this model — through a simple interface with basic controls, or a more advanced system with deep customization options?
10. Follow-up: What specific features or controls would be most important to you in a user interface for this model?
11. How much trust would you place in the outputs generated by this model?
12. Follow-up: What would make you more confident in using AI-generated variations in your work?)
13. Are there any ethical concerns you have about using AI to generate variations of existing music?
14. Follow-up: How do you think these concerns could be addressed?
15. What improvements or changes would you suggest for this model to make it more useful or effective for your needs?

A.2 Transcripts

This section provides a selection of transcripts of verbatim responses to questions from interviews conducted with participants. Each participant is labelled anonymously (P01, P02, etc.). The discussions focused on the definition of variations, their impressions of the AI model, and its potential use in music composition.

Q: How would you define a "variation" in the context of music?

P01: I don't really know. I don't really know how to describe what a variation of something would be. Like if something is a certain way... I don't know how you'd... How would you describe like what the variation is? Variation... something that exists and is recognisably the same from the same thing but... or has a recognisable idiom or stylistic similarity to something else. Because the thing is a variation can be anything from one note changing to something almost completely absent from the original idea.

Interviewer: So there are many different aspects that could be considered varied?

P01: Yes. So it's not just the notes not just the harmonies not just the rhythm. Yeah a variation... How close to the original idea it can be... It's a really tricky definition I think. What is a variation? Because it doesn't even necessarily have to be recognisably similar to the original. It can be something completely absent. And then you can say "Well it's a variation of this because there's one tiny thing about it which has a connection to the original."

Interviewer: So do you think as long as it has a connection to the original?

P01: But I mean there's definitely an argument that even something that doesn't have any connection to the original... You could say it's a variation because you can trace back ten different transformations you've gone through to get to this point. But you started at an original position but you've gone through ten different iterations to get to the variation. But then when you look at the variation it might be completely different to the original.

P02: I think I would say a variation in the context of music is something that has some sort of core elements of the thing that it is varying but changes it in a way that retains some reference to it but introduces some new ideas. I understand that some people might be like "Oh that's… it just needs to be a variation and is entirely derived from it." But I think there's nothing wrong with introducing new ideas in a variation so long as there is a degree of reference to the original one.

P04: Oh! Well, I mean… it's just... A repetition. The same thing, but with something that's different.

P03: That's a good question. It's quite… I feel... like a jazzy technical term I should really know and I don't. No no just in your own words as a composer. Like adapt some definition for it but I suppose… yeah I guess like an alternative way of playing something or like a different... A reinterpretation or like... Yeah I don't know. My definitions like... everything that's springing to mind is about like kind of interpretation or about having some sort of individual like import or transmutation of something. Does that make sense?

Q: Can you explain what you found successful or unsuccessful about the variations?

P01: I mean if we're using this as a kind of context the point is can we trace what's actually been done to create these examples? If we can't then I would say they're not successful variations of the original idea. If there is a path in which they have been forged... If you know what I mean from the original... then I would say...

P01: Like I couldn't necessarily say how the model infers to create those variations. But in terms of if you were to just listen to it... yeah maybe without the traceability aspect... I don't know. Yeah without the traceability you could sort of infer there's some kind of similarity. But I think possibly this is an area where... I don't know whether it's all AI models but I think there needs to be a kind of rigour put back into analysing where an idea has come from. And I don't know how possible that is to do with AI models. It comes down to the same thing of... There's a certain randomness to the output that you don't necessarily have control over. That worries me slightly because I could not tell you that those are good variations or bad variations because I can't say whether they are... Other then you know a fleeting resemblance that you might slightly hear. But they're completely rhythmically harmonically melodically... melodically completely removed from the original ideas as far as I can hear myself.

Interviewer: So there's no sort of relation at all?

P01: Not that I can hear myself but I would be interested if there was some kind of breakdown of how it got to those ideas. But I don't know how possible that would be to get.

P02: Even though the first variation was slower I think that harmonically it felt closer to the original. The chords were sort of closer and it had a similar tonality to it. The second one started to introduce a lot of leading notes and chromatic transitions between them which were very much not present in the first one. I can understand why some might say that fits a variation but maybe that was a bit too much of a variation in my mind.

P03: Like... I suppose... they felt... I suppose first what I found about them is they felt kind of... like a little bit less intentional or like a bit less... But maybe this is my human bias but they felt less intentional like with less sort of direction or something. But then the flip side of that was what I kind of enjoyed about them was that they were... they were kind of rogue. Like I don't think they would have been... you know I don't know much like the variations I would have come up with and that is kind of interesting and a good thing you know what I mean?

Q: Would you consider using a model like this in your music creation process?

P01: I think... It would need to be more predictable and more... You know you had the control over it to actually be related to the original idea. So at the very basic level if it was at the same tempo and time signature that would be one step towards getting closer. If it had fragments of the original melody that would be another step to being closer. I can see there's like a fleeting harmonic resemblance but even the harmony seemed very non-functional in the examples. If it had all those things if it was just honed enough that you could have control over those factors I think it could be potentially useful.

P02: Yeah definitely. I think that particularly because a lot of what we do is… the kind of model that my music tends to take is very much 'ABA'. So if I've written the 'A' section and then I'm like "Play me a 'B' section" or "Help me write a 'B' section" I think that would be really helpful.

P03: Yeah I mean I'm not sure what the application would be for me. But I'd be intrigued to try something that was kind of similar and work with it. Yeah because I suppose... I'm like... Again it's kind of like this wrongness thing or something that always seems to come back a lot in this chat. Like the... I've generally leaned towards... Cause this model is dealing with MIDI data rather than audio data. From the examples I've heard in terms of applicability to my own work I felt like models that deal with audio I've generally found more interesting because they produced weirder and kind of shittier results.

Q: What specific features or controls would be most important to you in a user interface for this model?

P01: Well I don't really know how it's functioning. Maybe a slider of how far away from the original idea it gets and then you could vary the metric variability the tempo variability the rhythmic variability... Yeah maybe those are sliders. Like very basic that you could mess around with a little bit and see how far away from it you could get.

P04: Well hmm. I think my favourite plugins kind of have a multi-tier thing that it's like... we're gonna show you just like macros first so you don't lose your brains and then once you come to grips with what's happening here then we'll show you what's under the hood.

P03: Yeah I mean I would definitely be into having lots of sliders to slide around and figure out what's... I like the idea of being able to tweak your weights inside the model and being able to turn... I think being able to mess with the weights and produce... I would really want that to have enough range that you could pull it into the state of wrongness or extremes. I love the idea of being able to pull it really dissonant or really consonant or really far or close from the original. They're quite nebulous ideas even...

A.3 Codebook for Thematic Analysis

The following table provides a codebook for the thematic analysis conducted in the study [4], which can be used to help reproduce similar research. Each theme represents key insights derived from participant responses, grouped into broader categories that reflect the participants' concerns and observations.

Table 2: Codebook for Thematic Analysis

Theme	Description	Codes
Transparency, Traceability, and Controllability	The need for clear, traceable, and controllable processes in AI-generated variations to enhance trust and usability.	6
Intentionality and Interpretation	The importance of intentionality behind variations and how it influences the perceived quality and relevance of AI outputs.	4
Creativity as a Collaborative Process	Framing the model as a creative partner, inspiring new ideas rather than replacing human creativity.	5
Customization and Adaptability	A balance between simplicity and depth, allowing users to customize the AI's outputs to better fit their creative needs.	3
Predictability and Familiarity	Maintaining a degree of predictability in outputs to ensure they align with the user's expectations and creative intent.	4

A.4 Consent and Risk Forms

Participants were informed about the study, its purpose, and their rights, including voluntary participation, the ability to withdraw at any time, and the secure, anonymous handling of their data. The consent form outlined the scope of data collection, including screen, audio, and video recordings, and highlighted the data protection measures in place. Participants were required to acknowledge each statement before agreeing to take part in the study. A section from the participant information form on risks is also included below.

Title of Research Study: Composing Culture: A Brief Ethnography of Composers and their Creative Process

Principal Investigator: Eleanor Row

Thank you for your interest in this research. Should you wish to participate in the study, please consider the following statements. Before signing the consent form, you should initial all or any of the statements that you agree with. Your signature confirms that you are willing to participate in this research, however, you are reminded that you are free to withdraw your participation at any time.

Ethics of Research Committee Ref: QMERC20.565.DSEECS23.058

Risk assessment: Both the observation and interview sessions will require your dedicated time. While we prioritise your privacy and anonymity, sharing your screen and photographs of workspace involves some level of exposure. Precautions will be taken to ensure your anonymity. Remember, you always have the choice to opt-out or restrict any part of the process if it makes you feel uneasy. Possible technical issues with Zoom or other tools used during the study might affect the study experience.

Statement of Consent

Statement	**Initials**
1. I confirm that I have read the Participant Information Sheet dated (retracted) version 1.01 for the above study; or it has been read to me. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2. I understand that my participation is voluntary and that I am free to stop taking part in the study at any time without giving any reason and without my rights being affected.	
3. I understand that my data will be accessed by the principal investigator Eleanor Row.	
4. I understand that collected data is completely anonymous and information that I have provided cannot be withdrawn after submission.	
5. I understand that my data will be securely stored in the United Kingdom, and in accordance with the data protection guidelines of Queen Mary University of London for 5 years in fully anonymous form.	
6. I understand that the information collected about me will be used to support other research in the future, and it may be shared in anonymous form with other researchers.	
7. I understand that my shared screen will be recorded and that my audio and video will also be recorded via Zoom.	
8. I agree to take part in the above study.	

Participants should read the QMUL privacy notice for research participants, which contains important information about your personal data and your rights in this respect. If you have any questions relating to data protection, please contact the Data Protection Officer.

A.5 Questionnaire Responses

This appendix provides a detailed summary of the responses collected from the participant questionnaire. The questionnaire was designed to gather demographic information, and information on composers' musical backgrounds and experiences with questions stemming from Goldsmith's Musical Sophistication Index (Gold MSI) [12].

Table 3: Demographic Information and Musical Education of Participants

Question	Responses	Details
Age	25-34	All participants were in the 25-34 age range.
Ethnicity	White, Mixed/Multiple Ethnic Groups	Majority identified as White, with one identifying as Mixed/Multiple Ethnic Groups.
Sexual Orientation	Heterosexual (straight), Prefer not to answer, Queer, Homosexual (lesbian or gay)	Participants expressed varied sexual orientations, including Queer and Heterosexual.
Gender Identity	Man, Genderqueer/gender fluid, Woman	Participants identified primarily as men, with one identifying as a woman and another as genderqueer/gender fluid.
Disability Status	No, Yes	One participant reported having a disability.
Occupation	Composer, AV Technician, Tutor, PhD Student	Majority were composers, with other roles including AV Technician and Tutor.
Highest Educational Qualification	Postgraduate degree, Undergraduate degree	Most participants had postgraduate qualifications, with one still pursuing education.
Highest Qualification Expected (if still in education)	Not applicable, Postgraduate degree, PhD	Relevant to those still studying; one pursuing a PhD.
Music-Related Qualifications	ABRSM/Trinity Grade 5+, Postgraduate (Masters), Undergraduate, GCSE	Participants had various music-related qualifications, from ABRSM/Trinity to postgraduate degrees.

Table 4: Summary of Participant Questionnaire Responses based on Gold MSI [12]

Question	Response	Details
Complimented for musical talents	Neither Agree nor Disagree, Disagree, Completely Disagree	Participants mostly did not feel recognised for their musical talents.
Consider as a musician	Strongly Disagree, Neither Agree nor Disagree, Disagree, Completely Disagree	Most participants did not consider themselves musicians, with varying degrees of disagreement.
Years of regular practice	4-5, 2, 0	Ranged from no regular practice to up to 5 years of daily practice.
Hours practised daily	5 or more, 1, 3-4, 1	Practice ranged from an hour a day to over 5 hours at peak interest.
Instruments played	3, 2, 3, 4	Participants played multiple instruments, reflecting diverse musical backgrounds.
Listening time	1-2 hrs, 15-30 min, 4 hrs or more, 15-30 min	Varied listening habits from minimal to extensive daily music listening.
Best instrument	Piano, Voice, Violin, Piano/Keys	Main instruments varied widely among participants, showing different primary musical skills.
Years of formal training (Music Theory)	0, 4-6, 2, 0	Experience in music theory ranged from none to several years.
Years of formal training (Instrument)	7 or more, 4-6, 1, 4-6	Participants had varying lengths of formal instrument training, some with extensive experience.
Years of formal training (Composition)	4-6, 4-6, 4-6, 0	Most had some composition training, with one participant having none.
Engagement frequency	Everyday	All participants reported daily engagement in musical activities.
Main musical genre	Classical, Other, Rock/Pop, Rock/Pop	Participants' musical preferences spanned all the genres listed.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction reflect the main contributions of the paper - a study on user-centred design that emphasises the importance of composer feedback and the broader implications for AI in the creative fields.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalise to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [No]

   Justification: Yes and No. The paper tries to address some of the limitations of the work within the methodology section (including the sample size and context-specific findings), but due to restrictions of the page limit, it is not able to include all of the limitations of the paper.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: The paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The paper provides open access to the questions, transcripts and also code book for Thematic Analysis employed within the study.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This research conforms to the NeurIPS Code of Ethics. The study underwent an institutional ethics review process, ensuring adherence to ethical standards for research involving human participants. All data collected was anonymised, securely stored on an encrypted disk, and will be deleted after five years. Informed consent was obtained from all participants, with clear communication regarding data usage and privacy protections. Further information on the ethics review process can be found in the Appendix section.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: Due to the page limit, there was not enough space to discuss the negative societal impacts of the work. Only the positive societal impacts were discussed.

    Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The baseline model that was used in the study was appropriately cited and credited.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

    Justification: Participant instructions are included in the supplemental material.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

    Justification: Ethics approval was received for the study, and potential risks were laid out to applicants in the call for participants (provided in the supplemental material). However, due to space constraints, this was not explicitly mentioned in the paper but in the appendix section.

    Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.