Perceiving AI-Generated Music in the Classroom: The Influence of Contextual Framing and Musical Experience

Eva Maria Klimpel
Paderborn University
33098 Paderborn, Germany
e-mail: eva.klimpel@outlook.de

Axel Berndt
Paderborn University
33098 Paderborn, Germany
e-mail: axel.berndt@uni-paderborn.de

Abstract— **This paper investigates how contextual framing and musical experience influence the perception of Artificial Intelligence (AI) generated classical music. While previous research suggests that AI-generated art and music are often evaluated less favourably when labelled as artificial, it remains unclear whether such attribution effects also occur among younger listeners and whether different forms of contextual framing influence their perception of AI-generated classical music. Previous studies indicate that younger individuals tend to hold more positive attitudes towards artificial intelligence technologies, suggesting that familiarity with AI may reduce scepticism towards AI -generated cultural artefacts. A two-stage research design was employed, consisting of a concert-based pilot study and a controlled experiment with secondary school students. In the latter (N = 60), participants evaluated an AI-generated musical excerpt under three conditions: no context, AI attribution, and didactic framing. The results indicate no significant effect of contextual framing on aesthetic appreciation, perceived creativity, or emotional response. In contrast, musical experience significantly influenced aesthetic evaluation. Despite participants belonging to a younger age group generally considered more receptive towards AI technologies, the findings reveal no clear positive bias towards AI-generated classical music. These findings suggest that listener perception in educational contexts is more strongly associated with individual experience than with externally provided contextual information. The study contributes to research on AI in music by demonstrating that framing effects may be less influential in non-expert populations than commonly assumed, highlighting the importance of listener characteristics in the evaluation of AI-generated music.**

Keywords—AI-generated music; auditory perception; contextual framing; music education; attribution bias; listener evaluation.

I. INTRODUCTION

Recent advances in generative AI have enabled the production of stylistically convincing musical material. However, evaluating AI-generated music remains a fundamental challenge, as subjective human preferences often correlate only weakly with widely used automatic evaluation metrics [1]. Previous research has also shown that AI-generated music and art are often evaluated less favourably when audiences believe that they were created by artificial intelligence rather than by humans [2]-[4]. This so-called anti-AI bias has been observed across different artistic domains, including music, visual arts, poetry, and performance contexts [5], [6].

In music research specifically, contextual information and attribution knowledge appear to substantially shape aesthetic judgements [7]. Existing studies, however, have largely focused on adult populations and decontextualised listening settings. Less attention has been paid to younger audiences who have grown up with AI technologies as part of their everyday lives and interact with AI-based systems on a regular basis. Previous research suggests that younger individuals are often more familiar with AI technologies and may therefore be more receptive to their use in everyday life [8]. However, recent studies also indicate that attitudes toward AI among younger populations are complex and may include concerns about authenticity, creativity, and the growing role of AI in educational and artistic contexts [9], [10]. It therefore remains unclear whether adolescents show the same scepticism towards AI-generated classical music or whether increasing familiarity with AI may alter aesthetic evaluation patterns. At the same time, genre-specific expectations associated with classical music are likely to influence listener responses independently of general attitudes towards AI. This question is particularly relevant because young listeners represent a potential future audience for classical music concerts and digital concert formats.

This paper investigates whether contextual framing influences the perception of AI-generated classical music and how such effects relate to individual listener characteristics. To address this question, a two-stage research design was employed, consisting of a concert-based pilot study and a controlled listening experiment.

The main contribution of this paper lies in demonstrating that, in a non-expert educational context, the perception of AI-generated music is not significantly influenced by contextual framing but is more strongly associated with individual musical experience. Furthermore, the findings suggest that generally positive attitudes towards AI among younger audiences do not necessarily translate into greater acceptance of AI-generated classical music.

This paper is structured as follows: Section II reviews related work. Section III presents the methodology. Section IV reports the results. Section V discusses the findings. Section VI provides a conclusion and outlines future work.

II. RELATED WORKS

This section reviews relevant research on AI-generated music, listener perception and contextual framing.

Evaluation remains one of the central challenges in AI music research [11]-[13]. Unlike traditional machine learning tasks, music generation cannot be assessed solely through objective accuracy metrics because musical quality is strongly connected to human perception, creativity, emotion, and cultural context [14]-[16].

Many studies on AI-generated music rely on subjective listening tests to evaluate generated outputs. Common evaluation criteria include musicality, coherence, creativity, emotional expression, stylistic consistency, and perceived human-likeness. Researchers frequently employ Likert-scale ratings, pairwise comparisons, preference studies, and Turing-style tests in which listeners attempt to distinguish AI-generated music from human compositions. However, several studies criticize these methods for their limited reproducibility, small participant samples, and dependence on listeners' musical expertise [12], [13], [17].

In parallel, researchers have developed objective evaluation metrics to measure statistical and structural properties of generated music. Frequently used metrics include pitch distribution similarity, rhythmic consistency, tonal stability, entropy-based diversity measures, groove consistency, and style similarity. Nevertheless, prior work consistently reports that many objective metrics correlate only weakly with human musical preferences and perceived quality [1].

Because of these limitations, current literature increasingly advocates for hybrid evaluation approaches that combine automatic metrics with human evaluation and music-theoretical analysis. Recent benchmarking studies demonstrate that human preference ratings remain essential for assessing musical creativity and listener satisfaction. At the same time, the field continues to face major challenges, including the lack of standardized benchmarks, inconsistent evaluation protocols, genre-specific biases, and the difficulty of operationalizing abstract concepts, such as creativity and originality [1], [18].

While research on music perception emphasises the role of listener-dependent factors, such as prior knowledge, expectation, and cultural context, studies in auditory perception and empirical musicology suggest that musical meaning and evaluation are not solely determined by acoustic features but are shaped by cognitive and contextual processes [15], [16], [19].

Furthermore, research in media psychology and cognitive science has demonstrated that attribution and framing significantly influence evaluative judgements. Labelling content as machine-generated can activate biases that affect perceived quality, authenticity, and aesthetic value [20], [21].

However, the relationship between contextual framing, musical experience and the perception of AI-generated classical music among younger listeners remains largely unexplored. This study addresses this gap by integrating perspectives from AI research, musicology and music education.

III. METHODOLOGY

A two-stage research design was employed, consisting of a concert-based pilot study (Study 1) [22] and a controlled experimental study (Study 2). The pilot study served an exploratory function and informed the design of the controlled experiment. The present paper focuses primarily on Study 2, which was designed to isolate the effects of contextual framing under controlled conditions.

A. Study 1: Concert Study

A live concert study was conducted in an open-air setting with approximately N = 60, participants. The programme included multiple musical formats, such as AI-generated compositions, historically informed formats, and interactive elements.

Participants evaluated each format using rating scales assessing aesthetic appreciation and general acceptance. The results indicated that AI-generated music received lower ratings compared to other formats, whereas context-rich and interactive formats were evaluated more positively. No significant effects were observed for age or musical training.

These findings suggested that contextual and performative factors may influence perception in ecologically rich environments, motivating the controlled investigation conducted in Study 2.

B. Study 2: Controlled Experiment

1) Participants

The study was conducted with N = 60 secondary school students (34 female, 25 male, 1 diverse; mean age = 14.1 years, SD = 2,29). Participants were recruited from regular classroom settings and represent a non-expert listener group with limited prior exposure to classical music. The participant group included students from diverse socio-cultural and educational backgrounds, with varying levels of familial engagement in educational and cultural activities. Prior to the study, an internal pilot test with five participants was conducted to ensure the questionnaire's clarity for musically inexperienced users.

Musical experience was recorded as a binary variable (no experience vs. prior experience), with approximately 25% of participants reporting some form of musical background.

Participation was voluntary, and informed consent was obtained in accordance with ethical guidelines.

2) Experimental Design

A between-subjects design was employed with three experimental conditions:

1. No context
2. AI attribution without explicit positive or negative framing
3. Didactic AI framing

Participants were randomly assigned to one of the three conditions. All participants listened to the same musical excerpt, ensuring that differences between groups were solely attributable to contextual manipulation.

3) Materials

The stimulus consisted of an AI-generated piano excerpt in a Chopin style, with a duration of approximately 40 seconds. The musical material was generated using the AI music generation platform Udio. Minor corrections and production adjustments were subsequently made in Logic Pro, while the composition itself remained largely unchanged in order to preserve the original AI-generated output [23]. The piece was designed to approximate stylistic characteristics of Romantic repertoire, including tonal clarity, phrasing, and expressive dynamics.

The use of a single, stylistically coherent stimulus ensured that differences in evaluation could be attributed to contextual framing rather than variability in musical material.

4) Procedure

The experiment was conducted during regular classroom sessions. Participants received condition-specific instructions prior to listening:

- In the *no context* condition, no additional information was provided.
- In the *AI attribution* condition, participants were neutrally informed that the music was generated by artificial intelligence, without any evaluative or suggestive framing.
- In the *didactic framing* condition, participants received a brief explanatory introduction describing how AI systems generate music.

Following the listening task, participants completed a structured questionnaire assessing their evaluations of the musical excerpt.

5) Measures

Three dependent variables were assessed:

- Aesthetic appreciation (Liking)
- Perceived creativity
- Emotional response

All variables were measured using Likert-type scales ranging from 1 to 5.

Composite scores were computed by averaging multiple items per construct. Internal consistency was assessed using Cronbach's alpha, yielding moderate reliability values (α = .67 for liking, α = .58 for creativity, and α = .59 for emotional response), which were considered acceptable given the exploratory nature of the study and the non-expert sample.

6) Statistical Analysis

Statistical analyses were conducted using Jamovi.
To examine the effect of contextual framing, one-way Analyses of Variance (ANOVA) were conducted for each dependent variable. Independent samples t-tests were used to assess the influence of musical experience. In addition, an Analysis of Covariance (ANCOVA) was performed to evaluate whether contextual effects remained when controlling for musical experience.

In addition to the quantitative rating scales, participants were given the opportunity to provide open-ended comments after the listening task. These responses were included to capture qualitative aspects of music perception that may not be reflected in standardised rating measures.

The comments were analysed using a qualitative, exploratory approach. Responses were reviewed and grouped into recurring thematic categories, such as perceived expressiveness, stylistic familiarity, and overall musical coherence. The qualitative data were not subjected to formal coding procedures but were used to support the interpretation of the quantitative findings.

This mixed-methods approach was intended to complement the statistical analysis by providing additional insight into how participants experienced and evaluated the musical material.

IV. RESULTS

A. Descriptive Statistics

Mean ratings for aesthetic appreciation (Liking) were similar across contextual conditions: no context (M = 3.35, SD = 0.98), AI attribution (M = 3.24, SD = 0.89), and didactic framing (M = 3.08, SD = 0.97). Similar patterns were observed for perceived creativity and emotional response.

B. Effect of Contextual Framing

A one-way ANOVA revealed no significant effect of contextual framing on aesthetic appreciation, $F(2, 37.7) = 0.38$, $p = .684$, $\eta^2 = .02$. Similarly, no significant effects were found for perceived creativity, $F(2, 37.6) = 0.13$, $p = .879$, $\eta^2 < .01$, or emotional response, $F(2, 37.5) = 0.17$, $p = .845$, $\eta^2 < .01$. Figure 1 illustrates the similarity of mean liking ratings across contextual conditions as well as the higher liking scores among participants with musical experience.

Post-hoc comparisons confirmed that there were no significant differences between any of the three conditions (all $p > .05$).

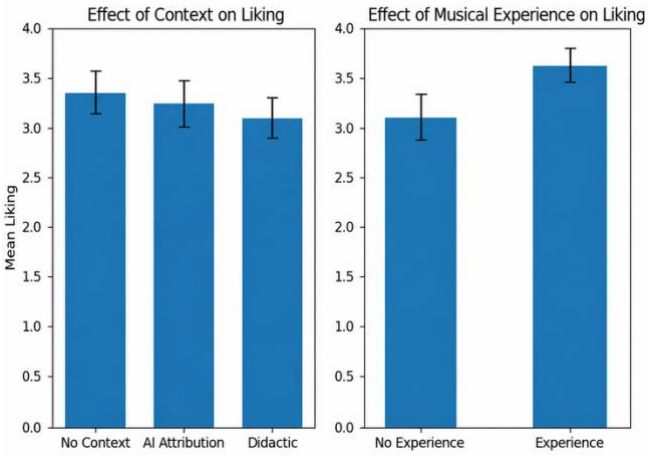

Figure 1. Mean ratings of aesthetic appreciation (Liking) across contextual framing conditions (left) and levels of musical experience (right). Error bars represent standard errors.

C. Effect of Musical Experience

An independent samples t-test revealed a significant difference in aesthetic appreciation between participants with and without musical experience, $t(58) = -2.03$, $p = .047$, $d = -0.61$. Participants with musical experience ($M = 3.64$, $SD = 0.68$) reported higher liking scores than those without ($M = 3.09$, $SD = 0.98$).

No significant differences were found for perceived creativity, $t(58) = -1.61$, $p = .113$, or emotional response, $t(58) = -1.48$, $p = .143$.

D. Combined Analysis

An ANCOVA with contextual framing as the fixed factor and musical experience as a covariate revealed no significant effect of contextual framing on aesthetic appreciation, $F(2, 56) = 0.37$, $p = .695$, $\eta^2 = .01$. The effect of musical experience approached significance, $F(1, 56) = 3.91$, $p = .053$, $\eta^2 = .06$.

E. Summary of Findings

The results indicate that contextual framing did not significantly influence listener evaluations, whereas musical experience showed a measurable effect on aesthetic appreciation.

F. Qualitative Findings

Participants' open-ended comments provided additional insight into the factors shaping their evaluation of the musical excerpt. Three recurring themes emerged from the responses:

First, several responses highlighted a perceived lack of coherence, describing the music as "not harmonious", "assembled", or "not forming a coherent whole". Participants also reported inconsistencies and perceived certain musical developments as "somewhat random", suggesting that structural consistency played an important role in aesthetic evaluation.

Second, a number of participants referred to stylistic characteristics of the piece. Some described the excerpt as "very romantic" and initially attributed it to a human composer. These comments indicate that listeners were able to recognise stylistic features associated with the Romantic piano repertoire, even without extensive musical training.

Third, several responses reflected assumptions about authorship and intentionality. Some participants stated that "one could tell that it was not human" or suggested that the music lacked purpose or compositional intention. These observations indicate that perceptions of agency may have influenced evaluations independently of the experimental framing conditions.

Across these themes, some participants explicitly expressed uncertainty regarding their own evaluative capacity, indicating limited confidence in their ability to analyse the music. This highlights the role of musical experience not only in perception, but also in the ability to articulate and justify aesthetic judgements.

Taken together, the qualitative findings suggest that participants primarily based their judgements on perceived musical coherence, stylistic familiarity, and assumptions about intentionality rather than on the contextual information provided prior to listening.

V. Discussion

The present study investigated the extent to which contextual framing influences the perception of AI-generated classical music in a non-expert listener group. Contrary to expectations derived from prior research on attribution and framing effects, the quantitative results did not reveal any significant influence of contextual information on listener evaluations. Across all three conditions—no context, AI attribution, and didactic framing—ratings of aesthetic appreciation, perceived creativity, and emotional response remained largely stable.

A. Absence of Context Effects

The absence of significant differences between contextual conditions suggests that, within this sample, externally provided information did not substantially shape perceptual judgements. This finding challenges the assumption that simple attribution cues, such as labelling music as AI-generated, are sufficient to influence aesthetic evaluation.

One possible explanation lies in the characteristics of the participant group. As non-expert listeners with limited exposure to classical music, participants may rely more strongly on immediate auditory impressions than on contextual or conceptual framing. In such cases, perception is likely driven by surface-level musical features rather than by reflective interpretation of authorship.

In addition, the controlled classroom setting may have limited the cognitive engagement required for framing effects to emerge. Unlike immersive performance contexts, the experimental design intentionally reduced environmental complexity, thereby isolating contextual information as a single variable. The results suggest that such isolated cues may be insufficient to alter perception in this context.

B. Role of Musical Experience

In contrast to the non-significant context effects, musical experience was associated with differences in aesthetic appreciation. Participants with prior musical experience reported higher liking scores, indicating that even limited familiarity with musical structures can influence evaluative judgement.

However, this effect should not be interpreted as evidence that musically experienced participants were more favourable towards AI-generated music as such. Rather, the qualitative data suggest that these participants were more receptive to the stylistic characteristics of the musical excerpt. Musical experience may therefore have functioned as a mediator of stylistic accessibility, enabling participants to recognise and engage with features associated with classical idioms.

C. Insights from Qualitative Responses

The qualitative findings provide additional insight into why contextual framing may have shown little influence on listener evaluations. The observations suggest that listener responses were shaped by two interacting processes: on the one hand, the recognition of stylistic familiarity, and on the other, sensitivity to perceived deviations from expected musical coherence. Importantly, these processes appear to operate largely independently of explicit contextual framing

and may be mediated by individual listening competence and musical experience.

D. Implications for AI Music Perception

The findings suggest that the perception of AI-generated music in non-expert populations may be less susceptible to simple contextual manipulation than assumed. Instead, listener responses appear to be shaped primarily by perceived musical properties and individual listening experience.

This challenges common assumptions in discussions of AI-generated art, which frequently emphasise attribution bias as a dominant factor. While such effects may be observable in expert or highly reflective contexts, they may be considerably attenuated in everyday listening situations.

E. Educational Implications

The results have direct implications for music education. The lack of strong framing effects suggests that simply providing contextual or explanatory information about AI-generated music may not be sufficient to alter students' evaluations. Instead, fostering musical experience and listening competence appears to play a more critical role.

From a pedagogical perspective, this highlights the importance of developing students' ability to recognise stylistic features, evaluate musical coherence, and reflect on their own listening processes. AI-generated music may therefore serve as a useful tool for promoting critical listening and discussion, rather than as an object of evaluation based solely on its origin.

F. Methodological Implications

A key methodological insight of this study concerns the limitations of purely quantitative approaches. While the rating scales successfully identified overall patterns—such as the absence of context effects and the influence of musical experience—they did not fully explain the underlying reasons for these evaluations.

The inclusion of open-ended comments proved essential for interpreting the results. The qualitative data revealed that participants' judgements were shaped by perceived coherence, stylistic familiarity, and attributed intentionality, dimensions that are not easily captured through standardised scales.

These findings support the use of mixed-methods approaches in the evaluation of AI-generated music, combining quantitative measures with qualitative data to capture both evaluative outcomes and underlying perceptual processes.

G. Relation to Pilot Study

The findings of the controlled experiment can be contrasted with the pilot study conducted in a concert setting, where contextual and performative factors appeared to play a more prominent role. This suggests that the impact of context may depend on the richness of the listening environment.

In immersive settings, context is embedded in multiple dimensions, including visual presentation, social interaction, and spatial acoustics. In contrast, the present study isolated contextual framing as a single variable, which may explain the absence of measurable effects.

H. Limitations

Several limitations should be considered when interpreting the findings of this study. First, the study relied primarily on a relatively homogeneous sample of students in a German secondary school within an educational setting, which may limit the generalisability of the results to broader or more diverse listener populations. In addition, participation required parental consent, which may have introduced a degree of self-selection bias related to parental educational engagement. Future research should therefore include participants with varying levels of musical expertise, cultural background, age, and technological affinity to better understand how listener characteristics shape the perception of AI-generated music. Second, the study employed a single AI-generated classical music excerpt as stimulus material. While this allowed for a controlled experimental design, it restricts the extent to which the findings can be generalised across different musical styles, compositional qualities, or levels of perceived sophistication. Furthermore, the stimulus was not a fully raw AI output, as minor corrections were applied during post-production. Future studies should investigate a broader range of musical stimuli, including different genres, varying aesthetic complexity, and human-composed comparison material. Third, the contextual framing manipulation was intentionally subtle and text based. It therefore remains possible that stronger or more immersive forms of contextualization — such as visual concert environments, performer interaction, multimedia presentation, or social listening contexts — may produce more pronounced framing effects. Future research should explore how multimodal and socially embedded listening situations influence the reception of AI-generated music. In addition, the reliability of some measurement scales was moderate, reflecting the exploratory character of the study and the conceptual difficulty of operationalising constructs such as creativity, authenticity, and emotional engagement in the context of AI-generated music. Future work may benefit from the development of more specialised and validated instruments for measuring listener responses to AI-generated artistic content.

Furthermore, the moderate reliability values observed for some of the questionnaire scales should also be considered when interpreting the findings. Lower internal consistency may have introduced additional measurement error, thereby reducing statistical power and potentially attenuating small effects. Although the absence of significant framing effects was consistent across all dependent variables, future studies would benefit from the use of more robust and extensively validated scales.

Finally, the study focused primarily on explicit self-reported evaluations. As a result, more implicit or unconscious forms of listener response may not have been captured adequately. Future research could therefore combine subjective ratings with physiological measurements, behavioural observation, qualitative interviews, or longitudinal designs to obtain a more comprehensive

understanding of how AI-generated music is perceived over time and across contexts.

VI. CONCLUSION AND FUTURE WORK

This study investigated the influence of contextual framing and musical experience on the perception of AI-generated classical music. The results show that contextual framing does not significantly affect listener evaluations in a non-expert educational setting, whereas musical experience plays a measurable role.

An important finding is that contextual framing effect can depend largely on the listener´s expertise and situational context. Simple informational cues may be insufficient to alter perception in controlled settings.

Future research should therefore investigate more immersive and ecologically valid forms of contextualization. Rather than relying solely on textual attribution, subsequent studies could examine multimodal framing strategies, including visual concert settings, performer embodiment, live performance simulations, or social framing effects generated through group listening situations. Such approaches may reveal whether attribution effects emerge more strongly when listeners experience AI-generated music in socially and emotionally enriched environments. Further studies should also compare different listener populations, including professional musicians, experienced concert audiences, and participants with varying levels of technological affinity. Although younger individuals are often described as highly familiar with AI technologies, familiarity does not necessarily imply uncritical acceptance. Recent research suggests that attitudes toward AI among younger populations may be characterised by both curiosity and scepticism, particularly in domains associated with creativity, authorship, and education [24]. Against this background, the absence of a clear positive bias toward AI-generated classical music observed in the present study appears less surprising and warrants further investigation. Methodologically, future work could employ longitudinal designs, cross-cultural comparisons, physiological measurements, or qualitative interview approaches to capture more nuanced forms of listener response. In addition, future experiments may compare different musical genres, levels of stylistic complexity, and varying degrees of transparency regarding the compositional process. Taken together, these directions suggest that the perception of AI-generated music cannot be reduced to a simple question of attribution alone, but instead emerges through a complex interaction of contextual, cultural, educational, and experiential factors.

REFERENCES

[1] F. Grötschla, A. Solak, L. A. Lanzendörfer, and R. Wattenhofer, ‘Benchmarking Music Generation Models and Metrics via Human Preference Studies’, in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025, pp. 1–5. doi: 10.1109/ICASSP49660.2025.10887745.

[2] A. Ansani *et al.*, ‘AI Performer Bias: Listeners Like Music Less When They Think it was Performed by an AI’, *Empirical Studies of the Arts*, 2025, doi: 10.1177/02762374241308807.

[3] L. Bellaiche *et al.*, ‘Humans versus AI: whether and why we prefer human-created compared to AI-created artwork’, *Cogn. Research*, vol. 8, no. 1, p. 42, Jul. 2023, doi: 10.1186/s41235-023-00499-6.

[4] R. L. Dunham, G. A. van Kleef, and E. Stamkou, ‘The threat of synthetic harmony: The effects of AI vs. human origin beliefs on listeners’ cognitive, emotional, and physiological responses to music’, *Computers in Human Behavior: Artificial Humans*, vol. 6, p. 100205, Dec. 2025, doi: 10.1016/j.chbah.2025.100205.

[5] C. B. Horton Jr, M. W. White, and S. S. Iyengar, ‘Bias against AI art can enhance perceptions of human creativity’, *Sci Rep*, vol. 13, no. 1, p. 19001, Nov. 2023, doi: 10.1038/s41598-023-45202-3.

[6] S. Altay and F. Gilardi, ‘People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation’, *PNAS Nexus*, vol. 3, no. 10, p. pgae403, Oct. 2024, doi: 10.1093/pnasnexus/pgae403.

[7] C. Kroger and E. H. Margulis, ‘“But they told me it was professional”: Extrinsic factors in the evaluation of musical performance’, *Psychology of Music*, vol. 45, no. 1, pp. 49–64, Jan. 2017, doi: 10.1177/0305735616642543.

[8] J.-P. Stein, T. Messingschlager, T. Gnambs, F. Hutmacher, and M. Appel, ‘Attitudes towards AI: measurement and associations with personality’, *Sci Rep*, vol. 14, no. 1, p. 2909, Feb. 2024, doi: 10.1038/s41598-024-53335-2.

[9] I. Alpizar-Chacon, H. Keuning, I. de Jong, I. Lykourentzou, and S. Rings, ‘Excited, Skeptical, or Worried? A Multi-Institutional Study of Student Views on Generative AI in Computing Education’, in *Proceedings of the 25th Koli Calling International Conference on Computing Education Research*, Nov. 2025, pp. 1–11. doi: 10.1145/3769994.3770008.

[10] A. Bewersdorff, M. Hornberger, C. Nerdel, and D. Schiff, *AI Advocates and Cautious Critics: How AI Attitudes, AI Interest, Use of AI, and AI Literacy Build University Students’ AI Self-Efficacy*. 2024. doi: 10.13140/RG.2.2.31045.87521.

[11] Z. Xiong, W. Wang, J. Yu, Y. Lin, and Z. Wang, ‘A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music’, Aug. 26, 2023, *arXiv*: arXiv:2308.13736. doi: 10.48550/arXiv.2308.13736.

[12] L.-C. Yang and A. Lerch, ‘On the evaluation of generative models in music’, *Neural Comput & Applic*, vol. 32, no. 9, pp. 4773–4784, May 2020, doi: 10.1007/s00521-018-3849-7.

[13] A. Lerch *et al.*, ‘Survey on the Evaluation of Generative Models in Music’, Sep. 20, 2025, *arXiv*: arXiv:2506.05104. doi: 10.48550/arXiv.2506.05104.

[14] M. Rohrmeier, ‘On Creativity, Music’s AI Completeness, and Four Challenges for Artificial Musical Creativity’, *Transactions of the International Society for Music Information Retrieval*, vol. 5, no. 1, Mar. 2022, doi: 10.5334/tismir.104.

[15] C. S. Pereira *et al.*, ‘Music and Emotions in the Brain: Familiarity Matters’, *PLoS ONE*, vol. 6, no. 11, p. e27241, Nov. 2011, doi: 10.1371/journal.pone.0027241.

[16] A. Tubadji, H. Huang, and D. J. Webber, ‘Cultural proximity bias in AI-acceptability: The importance of being human’, *Technological Forecasting and Social Change*, vol. 173, p. 121100, Dec. 2021, doi: 10.1016/j.techfore.2021.121100.

[17] S. JI, X. YANG, and J. LUO, ‘A Survey on Deep Learning for Symbolic Music Generation: Representations, Algorithms,

Evaluations, and Challenges.', *ACM Computing Surveys*, vol. 56, no. 1, pp. 1–39, Jan. 2024, doi: 10.1145/3597493.

[18] X. Zhou, 'Analysis of Evaluation in Artificial Intelligence Music', *Journal of Artificial Intelligence Practice*, 2023, doi: 10.23977/jaip.2023.060802.

[19] D. Stammer, H. Strauss, and P. Knees, 'Perception of AI-Generated Music -- The Role of Composer Identity, Personality Traits, Music Preferences, and Perceived Humanness', Dec. 02, 2025, *arXiv*: arXiv:2512.02785. doi: 10.48550/arXiv.2512.02785.

[20] W. Zhang *et al.*, 'Neural correlates of evaluative bias against artificial intelligence-labeled versus human-labeled artworks', *Social Cognitive And Affective Neuroscience*, vol. 20, no. 1, p. nsaf071, Jan. 2025, doi: 10.1093/scan/nsaf071.

[21] A. Schreiber, K. Sander, R. Kopiez, and R. Thöne, 'The Creative Performance of the AI Agents ChatGPT and Google Magenta Compared to Human-Based Solutions in a Standardized Melody Continuation Task', *Yearbook of Music Psychology*, pp. 1–14, 2024, doi: 10.5964/jbdgm.195.

[22] A. Berndt, D. A. Mauro, H.-P. Gasselseder, S. A. Etezazi, and E. M. Klimpel, 'AI and Audience Interaction in Classical Concerts', 2026, submitted.

[23] A. Deniz Gokcin, 'Chopin Regenerated: AI Prelude by AyseDeniz'. Jun. 04, 2024. Accessed: Jun. 04, 2026. [Online]. Available: https://www.youtube.com/watch?v=ZmYCTBwaOV4

[24] P. Her and P. Hamerski, *Investigating student perceptions of creativity and generative ai in computational physics*. 2026. doi: 10.48550/arXiv.2603.12154.