Do Generative AI Systems Hinder or Encourage Creativity in Music Education? Learnings from a National Project on Continuous Teacher Professional Development

Sönke Erdmann *
Chair for Music Education
University of Potsdam
Karl-Liebknecht-Str. 24-25, 14476 Potsdam
soenke.erdmann@uni-potsdam.de

Julia Benita Senfft
Faculty for Psychology and Sports
University of Bielefeld
Universitätsstr. 25, 33615 Bielefeld
julia.senfft@uni-bielefeld.de

Gergö Degen
Chair for Technology and Technical Education
University of Potsdam
Karl-Liebknecht-Str. 24-25, 14476 Potsdam
gdegen@uni-potsdam.de

Abstract

In recent years, the use of generative artificial intelligence (GenAI) in education has become widespread, even extending to aesthetic subjects. Recent studies have suggested the existence of a creative hivemind through homogeneous AI outputs in Large Language Models (LLMs). This article analyses AI systems used in continuous professional development courses for music teachers. The aim of this paper is to "unblack the box" of GenAI for music education. To this end, we classify these systems based on their architecture. We then evaluate their output using the Consensual Assessment Technique (CAT). Finally we also present a research design that focuses more on the songwriting process. This allows us to gain a deeper understanding of the affordances of these systems for instructional design.

1 ARTIFICIAL INTELLIGENCE IN MUSIC EDUCATION

Digital transformation is one of the most pressing topics in recent educational developments. With the rise of end-to-end generative Artificial Intelligence (GenAI) models (e.g. text-to-music), like 'Suno', it has moved from niche to mass markets, affecting the teaching of aesthetic subjects. An indication of this can be observed in the large-scale project *Lernen:Digital* in Germany that aims to provide courses and materials for the continuous professional development of teachers of all subjects in the use of digital technologies in the classroom (Lazarides et al., 2024). With its kick-off in July 2023, there was not that strong a notion on the topic of AI for music education, with just three modules out of roughly 100 taking this into account (one of them as part of a "Future Innovation Hub"). Almost three years later, however, this has changed. With the open release of "Suno" (Suhailudheen and Sheena, 2025; Schumacher et al., 2023) in March 2024, it became a hot topic in music education as well. The aim of this article is to demystify and classify the AI systems used for songwriting modules in teacher professionalisation courses, and to establish recommendations for teachers on how these systems might affect their teaching and their students' creative agency. To this end, we — a STEM educator, a psychologist, and a music educator — have joined forces to rate these systems based on their architecture and affordances, combining this with a creative evaluation of Audio GenAI system outputs using the Consensual Assessment Technique (Amabile, 1982). Although this approach only covers static system and output evaluations, we also propose a method for further evaluating creative processes with AI in music education.

The remainder of this article is structured as follows. Section 2 provides a brief introduction to AI professionalisation courses in music education, as well as a discourse about songwriting as a subject in (post) digital music education, offering guidelines for evaluating creative processes in songwriting. Section 3 provides an overview of creativity as a research topic, explaining how it is understood in this article and how it can be measured in humans and socio-technical systems. Section 4 presents the AI systems used in music lesson pilots and classifies them according to their architectures, inputs, and outputs. In section 5 the creative quality of the outputs of prompt-based generative audio systems is measured using the Consensual Assessment Technique (CAT) in two different genres, with a human-made reference track for each. Section 6 provides a research proposal for evaluating other aspects of creative songwriting, with a particular focus on processes rather than products. Finally, we summarise the findings and provide an outlook on future research and development directions.

2 COURSES FOR TEACHERS ON AI IN MUSIC EDUCATION

AI is still a very new topic in (school) music education. Most teachers are not even comfortable using Digital Audio Workstations (DAWs) in German classrooms, where the "liveness" norm still overshadows life-world oriented

*currently University of Amsterdam, Department of Media Studies

teaching (Godau, 2026). Thus, AI seems like a "future technology" to most music teachers, which could lead to technostress and non-evidence-based rejection or exaggeration of such applications. These findings are also supported by Luan et al. (2025), who conducted research on joint creativity using general-purpose AI and found that Human-AI co-creation as an evolutionary process is not a plug & play scenario. Onboarding and interventions are needed to prevent stalling. Although there are success stories of applied deep learning with audio courses for graduate arts students (Tampu and Lin, 2023), they seem rather challenging for an already packed school music study programme. Recent national publications focus solely on experiences with mostly web-based and proprietary AI applications in music education, rather than providing empirical insights. Developing and identifying attitudes towards GenAI in music education is crucial for the critical engagement with such tools, alongside gaining "operating skills" (Phung and Rolle, 2026; Rotsch and Werner, 2025; Rotsch, 2024).

However, studies on the use of GenAI music tools in classrooms have been conducted in other countries. The main finding is that it increases student participation among those without theoretical knowledge of songwriting, but also disempowers students by reducing their influence over the artefacts they produce, thus impacting their creative agency (Choi and Chang, 2025). Such studies can shed light on the educational affordances (i.e. capabilities for learning scenarios) of GenAI in music classrooms, an area in which little research has been conducted thus far.

We extracted the GenAI tools from courses developed within the *Lernen:Digital* project. These were piloted with teachers and pre-service teachers. Formats range from online to hybrid to in-person. Course materials are made available on the OpenMusicAcademy website.[1] Further information can be found in Krämer et al. (2026).

The following AI tools are derived from the courses:

- "AI Cube" (Hecht and Krämer, 2025; Krämer, Oliver and Hecht, Benjamin, 2024): *SunoAI, elevenlabs and Udio*
- "SongWrAIting" (Reich, 2026): *AIVA*
- "Songwriting with AI. Digital partners in a creative process?!" (Endres and Buchborn, 2026): *Melodea, StemRoller, GoGulilangu DeepDrum which is based on Google Magenta*(Roberts et al., 2019)

The main use of AI tools in music education that emerged from our project relates to songwriting. Recent studies in Germany suggest that there is a discrepancy between the songwriting techniques used by up-and-coming artists and those taught in classrooms (Gall et al., 2025), where songwriting still relies heavily on the foundations of music theory. What actually happens in smaller bands or "bedroom producing" is mostly independent from scholarly songwriting techniques. Therefore, Godau et al. (2025) proposed the 4-Room Model which incorporates four non-linear, intertwined rooms: an exploration room, a performance room, an inspiration room, and a recording room.

In this publication, we focus on the *Inspiration Room*, where generative AI can serve as an ideation tool and an expanded sample library that can be used without extensive domain knowledge (e.g. genre-specific stylistic elements). In Section 6 research designs to evaluate the usage of AI in other rooms are proposed. To determine whether GenAI outputs can be considered 'creative', we employed the Consensual Assessment Technique (Amabile, 1982) to evaluate the creativity of AI-generated products without a human in or on the loop. This is further described in Section 5.

3 HOW TO MEASURE CREATIVITY OF HUMANS AND SOCIO-TECHNICAL SYSTEMS

3.1 Conceptualising Creativity: Definitions, Dimensions, and Measurement

Creativity has become increasingly relevant not only for individual cognitive development but also as a core competency in formal education and broader societal adaptation. Reflecting this significance, the OECD Learning Compass 2030 identifies creativity—alongside critical thinking, collaboration, and communication—as one of the key transformative competencies required for learners to navigate an uncertain and rapidly changing world (Schleicher, 2019). However, precisely because of this growing contemporary importance, one of the central challenges within psychology and related disciplines remains the definition and measurement of creativity (Palmer, 2015). This challenge arises from the inherently complex and multidimensional nature of creativity, which encompasses individual cognitive processes as well as broader social, cultural, and technological contexts (Runco, 2021).

Among the many competing definitions, the combination of originality (novelty) and utility has established itself as the gold standard in creativity research (Runco and Jaeger, 2012). While this definition can be applied effectively to achievements in science and technology, its application to artistic works proves significantly more complex. Artistic products can nonetheless be understood as useful when they stimulate critical discourse, foster empathy, or evoke emotionally meaningful responses, with originality manifesting through stylistic ruptures, genre hybridity, or novel techniques (Uecker, 2008). To capture different levels of creative achievement, Merrotsy (2013) developed the influential model of *little c* (everyday creativity) and *big C* (eminent creativity), acknowledging that creative expression operates across a broad spectrum of social recognition and cultural impact.

A further important differentiation concerns the nature of novelty itself. According to Rausch, E (1952), the psychological differentiation of originality in problem-solving can be divided into three basic forms: absolute novelty (unprecedented in human history), subjective novelty (new to the individual), and relative novelty (new within a specific community or context). In educational settings, it is precisely subjective novelty that serves as the central starting point for the targeted cultivation of creative potential (Krampen, 2019). This outcome-oriented perspective has increasingly been criticised for neglecting the underlying mechanisms through which creative ideas emerge and develop over time, making it necessary to view creativity both as a trait and as a process (Green et al., 2024). The everyday assumption that something is creative only when it is entirely unprecedented leads to a considerable and unhelpful narrowing of the concept - one that may actively discourage creative engagement among learners who do not yet see themselves as capable of producing genuinely original work. In response to these limitations, more integrative frameworks have been proposed that conceptualise creativity as a multidimensional construct encompassing

[1] https://openmusic.academy/

not only products but also processes, individual characteristics, and environmental influences.

The most enduring integrative framework is the 4P model of creativity proposed by Rhodes (1961), which distinguishes between Person, Process, Product, and Press (environment), implying a genuinely multi-perspectival understanding of creativity (see Figure 1). For empirical inquiry, the specific dimension under investigation must be precisely defined, as this directly determines which methodological approaches and measurement procedures are appropriate (Simonton, 2018). At the cognitive level, divergent thinking plays a particularly central role in creative idea generation (Thakral et al., 2021), while convergent thinking is decisive for selecting, shaping, and implementing creative outputs (Puccio et al., 2020). The Torrance Tests of Creative Thinking (Torrance, 1974) remain one of the most widely used assessment instruments in this domain, with a reported reliability coefficient of McDonald's $\omega = .85$ (Acar et al., 2024). More recently, the Alternative Uses Task (AUT) and the Compound Remote Associates (CRA) test have been employed to assess the balance between divergent and convergent creative cognition, particularly in studies examining the neural correlates of creative insight (Beaty et al., 2016; Jung et al., 2013).

3.2 Measuring Creativity in AI Systems: Methods and Limitations

Assessing the creativity of AI systems introduces a distinct set of methodological challenges that go beyond those encountered in human creativity research. Several approaches have been proposed in the literature, each with substantial limitations when applied to contemporary generative systems (Carnovalini and Rodà, 2020).

The *Turing Test* (Turing, 1950) provides a behavioural baseline: if a system's output is indistinguishable from that of a human, it may be considered functionally equivalent. Applied to music generation, this amounts to asking whether listeners can identify AI-generated audio as non-human. While intuitively appealing, this criterion conflates perceptual quality with creative agency, and conflates output characteristics with the underlying generative process - precisely the distinction that matters most for educational contexts (Carnovalini and Rodà, 2020). The *Lovelace Test* (Bringsjord et al., 2001) sets a higher bar, requiring that a system produce outputs that its designers cannot explain - a criterion that current generative systems do not meet, as their outputs remain, at least in principle, traceable to training data distributions and architectural constraints. Both tests evaluate outputs in isolation and offer no insight into the process through which they are generated.

Self-assessments provided by AI developers - including benchmark reports, internal evaluations, and capability disclosures - represent a further class of evidence. However, their independence must be questioned in light of commercial interests: assessments produced by the companies developing and marketing these systems are structurally prone to confirmation bias and lack the external validity required for scientific use (Marcus and Davis, 2019). This concern is not merely theoretical; it reflects a broader pattern in AI evaluation in which capability claims have repeatedly outpaced independently replicated evidence (Liao and Vaughan, 2023).

Psychometric tests developed for human creativity, such as divergent thinking tasks or the Alternative Uses Task (Beaty et al., 2016), face a fundamental validity problem when applied to AI systems: the correct response patterns, and indeed the specific test items, are highly likely to have been present in the training data of large-scale models. What appears as creative performance may therefore reflect learned statistical associations rather than genuine generative flexibility (Stevenson et al., 2022; Organisciak et al., 2023). This data contamination problem renders standard psychometric instruments unreliable as measures of AI creativity, and calls for purpose-built evaluation protocols that are robust to training data exposure.

The *Consensual Assessment Technique* (CAT; Amabile 1982) offers a more defensible alternative. Rather than attempting to define creativity in advance, CAT operationalises it through the convergent judgements of independent domain experts, who rate creative products relative to one another without reference to fixed criteria. This approach has strong empirical support in human creativity research (Baer and McKool, 2012) and is recommended in the literature as a viable method for evaluating AI-generated creative outputs (Carnovalini and Rodà, 2020). Crucially, CAT assesses the *product* - not the process - and its ratings reflect the perceived creative value of an output within a specific domain and community of practice.

This points to a fundamental limitation that must be made explicit. All of the methods discussed above operate at the level of creative *outputs*; none provides access to the creative *process* that produced them. For AI systems, this distinction is especially consequential: as argued in Section 1 and elaborated in Sections 4 and 6, it is precisely the process dimension - divergent exploration, metacognitive reflection, iterative revision, and the development of creative self-concept - that carries the greatest educational significance. Output-level assessment, however methodologically rigorous, cannot capture whether and how an AI tool supports or suppresses these processes in the learner.

Given this constraint, the present paper focuses on the output level, and does so within a specific pedagogical framing. Drawing on the 4-Room Model of songwriting (Godau et al., 2025; Godau, 2026), we concentrate on the *Inspiration Room* - the space in which generative AI can serve as an ideation tool, providing stimuli that learners can respond to, modify, and build upon. Our CAT-based ratings therefore assess the creative utility of GenAI outputs specifically as inspirational material within this room: how stimulating, expressive, and generative of further creative activity might these outputs be for learners engaged in songwriting? This framing deliberately brackets the process dimension - not because it is unimportant, but because the methodological tools to assess it in ecologically valid educational settings do not yet exist at scale. How such process-oriented assessment might be designed and validated is addressed in Section 6.

3.3 Artificial Intelligence and the Challenge to Creativity Theory

While the theoretical and empirical frameworks outlined above were primarily developed to describe human creativity, recent advances in artificial intelligence and especially the widespread deployment of Large Language Models (LLMs) challenge the extent to which these assumptions can be transferred to non-human systems, and raise new questions about what it means to be creative in an AI-mediated world.

Empirical studies consistently show that the use of LLMs in creative production leads to a reduction in perceived individual creativity, a homogenisation of content, and a loss of personal expressive voice (Abdulhai et al., 2026; Jiang et al., 2025). This trend operates at both the micro and macro levels. At the micro level, structured prompt formats and pre-built templates constrain the generative output space and produce what has been termed a "diversity collapse" (Yun et al., 2025). In practice, this causes a narrowing of the range of ideas that users and systems actually produce. At the macro level, as large populations of users interact with the same underlying model architectures, similar patterns of thought, structure, and expression emerge across individuals and across models; a phenomenon Jiang et al. (2025) describe as an "Artificial Hivemind", which systematically reduces the overall diversity of creative ideas in circulation. In the context of creativity in music education, this concern becomes particularly salient: voice, style, and idiosyncrasy are not incidental features but constitutive of the creative act itself – the fine-grained interaction of these features is quintessential of what constitutes both the creative process and the creative output as an expression of the individual.

Beyond these empirical observations, the increasing capabilities of AI systems raise fundamental theoretical questions regarding the very definition of creativity. From a traditional perspective, it has been argued that AI generates only pseudo-creativity, as it lacks central features of human creative agency: intentionality, intrinsic motivation, and authentic idea generation embedded in lived experience (Runco, 2023). Building on the widely accepted novelty-utility definition (Runco and Jaeger, 2012), Da Pelo (2025) argues that AI systems can, at the level of observable outcomes, already fulfil both criteria. However, this apparent equivalence reveals a critical limitation of traditional definitions: they fail to account for essential qualitative differences in the underlying process by which an output is produced.

To address this gap, Da Pelo (2025) proposes two additional criteria: intentionality and authenticity. These dimensions refer to the embodied, experiential, and motivational foundations of human creativity - dimensions that are currently absent in AI systems. This gives rise to two fundamental distinctions. First, AI-generated outputs lack embodiment, resulting in an absence of intentionality and authenticity despite formally meeting the criteria of novelty and usefulness. Second, the creative process itself differs substantially: from a mechanistic perspective, it is constrained by computational boundaries such as model architecture and training data; from a psychological perspective, essential affective and volitional components - including motivation, curiosity, and the experience of effort - are entirely absent. This distinction between process-level and product-level creativity maps productively onto the 4P model: AI systems can simulate creative products while the person and process dimensions remain fundamentally human.

3.4 AI as a Tool in Creative Education: Affordances, Risks, and Conditions

However, although AI models do not produce creative outputs to the same degree as humans do, they still can be used

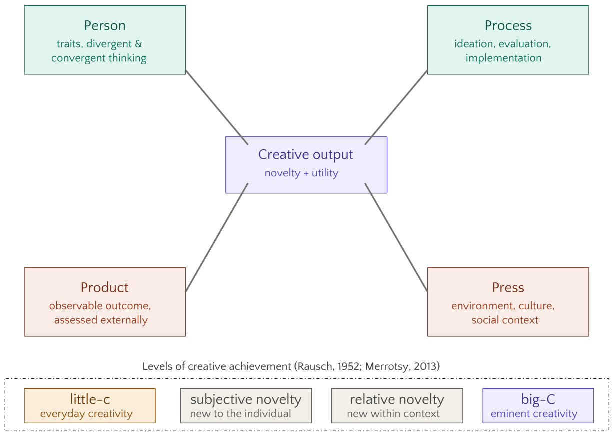

Figure 1: Rhodes' 4P model of creativity - including Merrotsy's (2013) little-c / big-C distinction.

as valuable tools to widen the space of possible teaching concepts in the context of music education. If the use of AI tools is well orchestrated (instructional design) in an educational sequence that allows a proactive, dynamical and iterative process of human-AI interaction, these tools can be productively reframed not as autonomous creators, but as tools that shape, scaffold, and extend human creative processes. In music education specifically, AI composition tools can lower technical barriers to creative participation by offering real-time harmony suggestions, chord generators, or stylistic transformations (Wang and Zhang, 2025). For learners who lack formal music theory training, such tools can enhance musical self-efficacy and willingness to experiment as well as creative risk-taking (Bandura, 1997; Tierney and Farmer, 2002). By simplifying technically demanding production tasks, AI potentially enables a shift from cognitively constrained, knowledge-bound production towards broader creative exploration, thereby freeing cognitive resources for higher-order creative thinking (Sweller, 1988; Wang and Zhang, 2025).

Furthermore, AI systems can provide novel stimuli and multimodal sources of inspiration - for instance, transforming images, movement data, or textual input into musical structures - thereby functioning as a creative catalyst and expanding the palette of possibilities available to learners. However, the educational value of such tools depends critically on the degree to which learners actively shape, evaluate, and take ownership of the generative process. When AI systems produce near-complete outputs from minimal input, learners may experience reduced autonomy, diminished intrinsic motivation, and what Deci and Ryan (2000) identify as a weakening of the psychological conditions necessary for self-determined engagement. Rapid "click-to-compose" processes can bypass the iterative, effortful engagement that is essential for developing creative competence, aesthetic judgment, and a sense of personal authorship (Amabile, 1996).

This reinforces a broader principle in creativity education: the process carries greater developmental significance than the product. Creative competence emerges through cycles of exploration, failure, and revision - a view that aligns with the cognitive apprenticeship tradition (Collins et al., 1989) and with constructivist learning theories more broadly (Papert and Harel, 1991). From a neuroscientific perspective, this iterative engagement can be linked to predictive coding frameworks, in which aesthetic and creative experience arises precisely from the dynamic tension between expectation and surprise (Vuust et al., 2022). Creative engagement is neurologically rich when elements interact in partially unpredictable ways - a condition that requires active participation instead of passive consumption and end-to-end generation.

When AI is used as a source of suggestions that learners actively evaluate, modify, and integrate into their own artistic vision, it can strengthen creative ownership and meaningful engagement. The psychologically optimal configuration may therefore be one in which AI functions as a competence scaffold (Wood et al., 1976; Vygotsky and Cole, 1978), lowering the floor of entry without removing the challenge that makes creative work educationally meaningful. However, this scaffold carries a risk of dependency if successful creative outcomes are systematically attributed to the AI system rather than to the learner's own agency, potentially undermining the development of a robust creative self-concept (Beghetto, 2006) and creative self-efficacy over time.

Beyond individual processes, AI also reshapes the social dimension of creativity. In collaborative settings, AI can function as a "third participant" in creative dialogue, influencing peer interaction, shared decision-making, and collective creative identity. These dynamics remain undertheorised in the existing literature and represent an important frontier for future research (Sawyer, 2012).

4 AI MUSIC APPLICATIONS

4.1 Transformer Music Models

Transformer based diffusion music models currently dominate the market. As the model architectures and training data of "ElevenLabs", "Suno" and "Udio" are proprietary, we can only make educated assumptions about their exact model architectures. Heartmula (Yang et al., 2026a) is an open-source foundational music transformer model from which we can extract the potential underlying architectures of these models. In GenAudio transformers audio files are compressed by Encoders like "ENCODEC" (Défossez et al., 2022) to fit into a vector space where they can be treated like tokens in a LLM (Copet et al., 2023). Some GenAI music applications seem to rely on Generative Adversarial Networks (Goodfellow et al., 2014). The underlying principle of Generative Adversarial Networks (GANs) is simple. They consist of two sub-networks: a generator and a discriminator. The discriminator has access to a set of training audio files. It tries to distinguish between "real" audio from the training set and "fake" audio generated by the generator. Meanwhile, the generator attempts to produce audio that resembles the training set despite having no access to it. It begins by producing random sounds or noise, and receives feedback from the discriminator indicating whether these sounds are real or fake. At equilibrium, the discriminator should not be able to tell the difference between audio generated by the generator and actual audio in the training set, meaning that the generator succeeds in generating audio that comes from the same distribution as the training set. Some authors experimented with that approach, to maximise style ambiguity by adding another style classifier (Elgammal et al., 2017). This approach can lead to a more diverse output because the generator constantly tries to produce "art" while avoiding any stylistic labellings.

General purpose AI models like Google's Gemini, for example, are also capable of generating audio. Gemini is a multimodal large language model (MLLM) that can in- and output multiple data types (Team et al., 2025). Gemini's model family uses a mixture of expert (MoE) approach to enable multimodality (Comanici et al., 2025).

Generative MIDI (Musical Instrument Digital Interface) models, alongside GenAudio transformer models, are quite popular. Examples include the Google Magenta Suite and, more recently, AIVA. MIDI input and output is much easier to compute because it does not need to be compressed (Yang et al., 2026b; Roberts et al., 2019). Some models rely on Variational Autoencoders (VAE), which are also used in hybrid models with GAN (Kingma and Welling, 2013).

4.2 Diffusion Models

Diffusion models (Zhang et al., 2023) are typically used for music source separation, i.e. to separate individual tracks from a mix. These models primarily rely on U-Net architectures, which are also employed for audio denoising (Défossez et al., 2019). Diffusion models can also be used to generate text, images, and sound. Google's Lyria uses this latent diffusion architecture (Ho et al., 2020). Advantages of diffusion models over autoregressive transformers (AR) include better token-level control and improved coherence thanks to parallel generation. However, one important downside is that diffusion model output is often less interpretable, which reduces their usability in educational settings (Zou et al., 2023).

A promising recent development within the diffusion model family is *flow matching* and its conditional variant, which learns straighter probability paths between noise and data distributions, resulting in faster inference and improved sample quality compared to standard diffusion (Song and Wang, 2025). Conditional Flow Matching (CFM) has demonstrated particular potential for fine-grained, text-guided control over generated outputs (Lan et al., 2024) - an affordance that may prove relevant for future tool designs aiming to increase human agency in the creative process without sacrificing output diversity. Furthermore, CFM-based tools enable multimodal music generation, thereby addressing a key limitation of purely text-to-music systems, which rely on users providing music-specific descriptive prompts and therefore often require a certain degree of musical expertise (Song and Wang, 2025).

4.3 Autoencoder Models

Variational Autoencoder Models (VAE) like *RAVE* (Caillon and Esling, 2021) compress audio signals into latent spaces and reconstruct them into high-fidelity sound. VAEs offer a broad variety of open-source models, but are not widely used for school music purposes. So far, only DAW plug-ins and smaller standalone applications, such as some plug-ins from the Google Magenta Suite (Roberts et al., 2019) and its successors, are ready for use in classrooms where teachers often feel overwhelmed by tools offered on HuggingFace, Github and Co. One reason for this is that there is still room for improvement in the widespread implementation of computer science education in Germany.[2]

[2] https://informatik-monitor.de/2025-26

4.4 Non generative machine learning in Music Education

In addition to generative models, there is also a large body of work on non-generative models for music production. These aren't always labelled as AI in music because the impact of these models is more subtle. Some examples include computer vision methods where the movement of cameras controls sound parameters (Dobrian and Bevilacqua, 2003). Other use cases include converting sound into notation or MIDI. This process is known as automatic music transcription (AMT). Spotify released an open-source model for this, which many DAWs use to transform audio (Bittner et al., 2022). Other use cases include the encoding of music in two-dimensional spectrograms to make them machine readable and therefore interpretable.

A further non-generative application of particular relevance to creative music education is sonification, which is defined as the systematic, non-speech representation of data through sound parameters such as pitch, timbre, rhythm, or spatial positioning (Walker and Nees, 2011; Hermann et al., 2011). Sonification can be used to map structured data onto auditory dimensions in real time. This process requires active creative decision-making about which data dimensions to encode and how, thereby opening a diverse creative action-space that expands expressive possibilities and invites the learner to articulate motivational and emotional associations with the entity being sonified. Therefore, sonification naturally encourages an iterative, exploratory creative process: because the mapping between data and sound is configurable and reversible, learners are invited into a cycle of hypothesis, listening, and adjustment that closely resembles the trial-and-error structure identified as central to creative competence development (Amabile, 1996; Cropley, 2006).

5 HOW WE DETERMINE CREATIVE POTENTIAL OF APPLICATIONS

5.1 Evaluation of the Potential of AI Output for Creative Inspiration

To address the creative potential of prompt based GenAI audio outputs, we conducted a CAT test with seven researchers working in the first phase of music teacher education in Germany, i.e. at universities or universities of applied sciences. To ensure anonymity, no further information about the participants will be given. We provided two reference songs from different genres for this purpose. One was produced at home with minimal effort by one of the authors, to create a product that could realistically be produced by an advanced student in a two-hour songwriting session. The second song was by the Canadian avant-garde instrumental rock duo “Angine de Poitrine”. We used Google Gemini to analyse those tracks and obtain the prompts[3] for the following GenAI audio systems: (1) Suno, (2) Udio and (3) ElevenLabs. The artefacts are all one-shot prompted. To control for bias towards AI authorship (Horton Jr et al., 2023), the examples were not labelled. We used AI-generated prompts for song generation to avoid overspecifying the instructions, given our nuanced understanding of the reference songs. AI-generated prompts for creative artifact generation are also common in multimodal creative approaches. The models were chosen because they appear predominantly in the teacher professionalisation courses. Although we consider open-source solutions to be highly feasible in the classroom, they have yet to be widely adopted.

These eight songs were evaluated by the expert raters using 13 items (Creativity, Liking, Overall aesthetic appeal, Effort evident, Timbral interest, Form, Sectional variation, Texture change, Meaningfulness, Accuracy of Performance, Expression, Complexity, Theme). Only the creativity measures are analysed as primary outcome measures, whereas the remaining items serve as covariates in the Intraclass Correlation Coefficient (ICC) analysis.

Table 1: CAT scores for each product

author/tool	*mean*	*SD*
Reference track Indie	4.00	1.29
Udio	3.57	1.13
Suno	4.43	0.79
ElevenLabs	3.57	1.62
Reference track “Avant-Garde”	4.86	0.90
Udio	5.43	0.79
Suno	4.71	1.11
ElevenLabs	4.57	1.51

Looking at the data of this exploratory study presented in Table 1, we can infer that, if not explicitly labelled as such, AI output is not generally perceived as either less or more creative by raters. In each category, one AI-generated song is rated as more creative than the reference example, while the remaining outputs were not rated as significantly less creative. Since this is not a benchmarking study and the primary aim is merely to establish whether an AI-generated output can be considered 'creative', we do not discuss these comparisons in greater detail, as the data do not support more fine-grained conclusions. Therefore, it is reasonable to suggest that AI output could be valuable for the *Inspiration Room*. However, this may depend heavily on the skills and strategies employed in prompting. Therefore, we used multimodal AI to analyse the reference songs, which may represent a viable strategy for music classrooms, since students may not have the domain-specific vocabulary needed to describe their desired outcomes precisely.

Regarding the limitations of this study, it should be noted that the target number of raters ($n = 10$–15) was not reached. This is evident in the outcomes of the ICC test. Therefore the inter-rater reliability is low with a single random rater level of $ICC = 0.24$ (CI [0.16; 0.33]), while the average random rater is medium with $ICC = 0.69$ (CI [0.58; 0.77]). Therefore, a subsequent study with more raters who are recruited more carefully might be needed, because the correlation analyses showed that two raters also lowered the reliability of the ratings.

Although the CAT is only suitable for one of the four songwriting rooms and audio GenAI models, we have developed a broader framework that incorporates the remaining rooms as well as other GenAI tools.

5.2 GenAI framework for Music Education

Figure 2 positions the tools introduced in Section 2 according to their potential influence on learners' creative work. Our framework draws on a conceptualisation of AI's relation to human cognition in educational settings proposed by Cukurova (2025), which distinguishes three roles of AI in relation to human cognition: *externalisation*, in which cognitive tasks are delegated to AI systems;

[3](1) Dreamy Indie-Folk (Singer-Songwriter), 74 BPM, melancholic & intimate. Breathy female vocals with plate reverb. Clean electric guitars with heavy reverb and chorus, dampened vintage drums, warm bass. Lo-fi tape saturation, spacious and hazy atmosphere;(2) Instrumental Progressive Indie Rock, 135 BPM. Atmosphere: technical, restless, and avant-garde

internalisation, in which interaction with AI alters or extends human representations and processes; and *extension*, in which humans and AI form tightly coupled, synergistic systems. These conceptualisations are situated within Shneiderman's two-dimensional account of human control and automation (Shneiderman, 2020). We adapt this logic to the domain of creative music production by distinguishing two dimensions: the degree to which AI contributes material to the resulting artefact (*AI input*) and the degree to which the learner retains consequential control over its creative development (*creative agency*).

Building on these dimensions, we introduce four categories that operationalise different modes of GenAI-supported creative work in music education. *Stimulus* tools occupy the high-agency, low-AI-contribution quadrant: they provide small amounts of AI-generated material that can initiate or inspire further learner-driven creation. *Empowerment* tools occupy the low-AI-contribution, low-agency quadrant and primarily provide technical capabilities for manipulating or transforming musical material, rather than contributing substantial new creative content. In contrast, *end-to-end generation* occupies the low-agency, high-AI-contribution quadrant: these systems generate substantial portions of the final artefact with relatively little consequential creative control remaining with the learner. Finally, *Co-creation* occupies the high-agency, high-AI-contribution quadrant and represents the theoretical target of tightly coupled human–AI collaboration, in which AI contributes substantial creative material while the learner retains meaningful control over its development.

The positions in Figure 2 were derived from a closed card sort (Rugg and McGeorge, 1997) within the author team. The classification was based on four criteria, summarised in Table 2: input modality, output granularity (finished track, intermediate artefact, or fragment), access to model parameters and subsequent editing, and the proportion of the four-room songwriting process remaining with the learner. Together, these criteria capture both the extent of AI contribution and the degree of consequential creative control available to the learner. The resulting positions should be understood as a conceptual classification of the tools' current affordances and the position of a tool may vary depending on how it is configured and used in practice. Figure 2 therefore serves as a heuristic for organising the current design space and generating testable hypotheses about how different forms of AI assistance may shape creative agency in music education. The individual applications and their relevant characteristics are described in more detail in Table 2.

While GenAI-generated audio can provide creative inspiration, currently available tools do not yet appear to realise the tightly coupled synergistic human–AI collaboration represented by the *Co-creation* quadrant. Current interfaces provide limited opportunities for learners to express intentional and authentic creative decisions (Urmeneta and Romero, 2025), instead abstracting away many intermediate compositional decisions. While iterative prompting can provide some additional control (Choi and Chang, 2025), prompt-based interaction remains indirect and natural-language descriptions do not necessarily map onto specific musical properties or operations, particularly for learners whose musical vocabulary is still developing.

Suno, *Udio*, and *ElevenLabs* exemplify *end-to-end generation*, occupying the high-AI-input, low-agency quadrant of the design space. These systems generate substantial portions of the final musical artefact from relatively limited user input, leaving comparatively few consequential creative decisions with the learner. We therefore consider this configuration to be at risk of *creative suppression*: as the proportion of AI-generated material increases, opportunities for learners to shape the development of the artefact may decrease.

AIVA occupies a position closer to the high-agency side of the design space. It generates substantive musical material, but this material can subsequently be edited, rearranged, and combined with different virtual instruments. AIVA therefore provides learners with substantial AI-generated input while retaining opportunities for consequential creative decisions during subsequent development. Its contribution to creative learning ultimately depends on whether learners actively manipulate and transform the generated material.

Melodea and *DeepDrum* are positioned as *stimulus* tools. They provide relatively small musical elements that can serve as starting points for further learner-driven creation, leaving most of the songwriting process with the learner, while the generated fragments can inspire creativity and complement human progress.

StemRoller is categorised as an empowerment tool. It enables learners to separate mixed songs into parts and extract individual acoustical features, which can then be recombined and creatively transformed. This also facilitates the creative reuse of end-to-end generated material by allowing learners to work with individual elements, such as drums or harmonies.

Taken together, the framework extends the exploratory CAT findings from the creative evaluation of AI-generated audio to the broader question of how GenAI may shape the creative process. While the CAT results suggest that AI-generated outputs can be perceived as creative, the framework highlights that the creative potential of AI-supported songwriting also depends on how AI input is integrated into the process, how much consequential agency remains with the learner, and what possibilities the generated material affords for further exploration. These dimensions cannot be established from the generated artefact alone and therefore require further evaluation in the context of actual learner–AI interaction.

Table 2: Used AI tools

Tools	**Input**	**Output**	**anticipated Architecture**
SunoAI[4] (not studio)	Text	Audio	Transformer based diffusion
ElevenLabs[5]	Text	Audio	Transformer based diffusion
Udio[6]	Text	Audio	Transformer based diffusion
AIVA[7]	Text/ MIDI	Midi, Audio	Transformer
Melodea[8]	adjustable Filter	Audio, MIDI & Chord Notation	Symbolic (randomiser) with Midi Output
StemRoller[9]	Audio	Stems of Audio (Drums, Chords, Melody, Voice)	Diffusion Model (U-Net + Masking)
DeepDrum[10]	Grid (DrumMap)	WAV	Variational Autoencoder

6 TOWARDS A PROCESS-ORIENTED ASSESSMENT FRAMEWORK FOR AI-SUPPORTED CREATIVE LEARNING IN MUSIC EDUCATION

Given the complexity of human-AI creative interactions, the question of how creativity can be meaningfully assessed in hybrid contexts becomes particularly pressing. As established in Section 3, existing evaluation methods for AI creativity operate predominantly at the output level, and those developed for human creativity - psychometric instruments, divergent thinking tasks, self-report scales - face substantial validity challenges when applied to AI-mediated learning

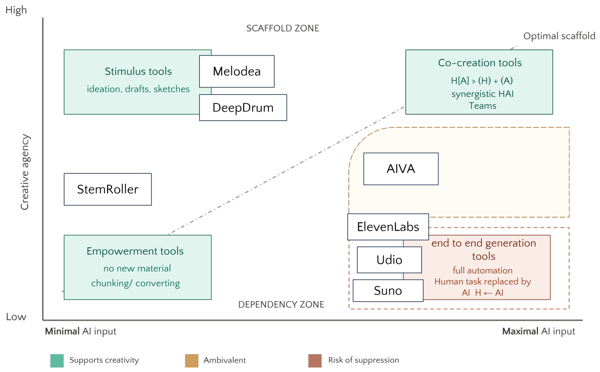

Figure 2: AI tools in music education: classification by creative influence.

contexts. The CAT, while methodologically defensible for product evaluation, provides no window into the generative process through which a creative outcome is reached. Yet it is precisely this process that carries the greatest developmental significance in music education (Amabile, 1996; Cropley, 2006). A rigorous assessment framework must therefore go beyond static product evaluation and address the cognitive, affective, behavioural, social, and neurophysiological dimensions of creative engagement as they unfold in real time.

Process-oriented assessment must also account for the breadth of the creative action-space made available through AI interaction. A broader action-space provides learners with more opportunities to generate, compare, and explore diverse alternatives, whereas systematic convergence may constrain these opportunities. GenAI may increase the creativity of individual outputs while simultaneously reducing their collective diversity, as demonstrated experimentally for AI-assisted story writing (Doshi and Hauser, 2024). Although this effect has not yet been established experimentally for music, generative music systems may similarly favour well-represented genres and styles because they learn statistical regularities from their training distributions. Substantial imbalances in music-generation datasets and disparities in model performance across musical traditions (Mehta et al., 2025), as well as evidence of mainstream stylistic convergence in commercial music-generation systems (Solak et al., 2025), suggest that this possibility warrants consideration. The generative objective itself may also shape this tendency: Creative Adversarial Networks, for example, were explicitly designed to encourage deviation from established style categories while remaining compatible with the learned distribution (Elgammal et al., 2017). Thus, generating plausible or creative-sounding material does not necessarily broaden the range of possibilities available for further creative exploration.

6.1 Design Principles

We propose that a principled assessment framework for AI-supported creative learning in music education should satisfy four core design principles. First, it must be *process-oriented*. Consequently, data collection must be continuous and longitudinal, capturing the trajectory of creative engagement rather than a single endpoint. Second, it must rely on a *multi-method approach* since no single measure can adequately reflect the multi-factorial nature of creativity as conceptualised in the 4P model (Rhodes, 1961). Therefore, we propose the comprehensive collection of convergent evidence across qualitatively distinct data streams. Third, it must be *ecologically valid*, which means that the assessment must be embedded in authentic songwriting tasks within genuine educational settings, not abstracted into laboratory conditions that strip away the social, motivational, and pedagogical context in which creative learning occurs. Fourth, it must be *experimentally controlled* to enable valid claims about affordances of AI tools on creative development.

6.2 A Four-Level Multi-Method Assessment Framework

We propose a four-level framework that provides a more principled assessment of creativity, designed to satisfy the multi-factorial nature of the creative process including its individual, cognitive, social, and environmental dimensions. The four levels are illustrated in Figure 3.

Level 1 - Psychological outcomes. Self-report instruments with established psychometric properties should be used to assess creative self-efficacy (Bandura, 1997; Tierney and Farmer, 2002), creative identity (Beghetto, 2006), intrinsic motivation (Deci and Ryan, 2000), and sense of ownership over the creative product. These are administered at pre-, mid-, and post-session timepoints to capture within-session dynamics as well as longer-term developmental trajectories across multiple sessions. Experience sampling methods (ESM; Csikszentmihalyi and Larson (1987)) can supplement fixed-timepoint measures by capturing momentary affective and motivational states at randomised intervals during the songwriting task, providing a more fine-grained picture of engagement fluctuations in response to AI interaction events.

Level 2 - Cognitive processes. Three complementary methods are proposed to operationalise central process indicators such as iteration depth, modification rate, and prompt diversity. First, *Think-aloud protocols* (Ericsson, 2017) provide concurrent verbal access to ideation, evaluation, and decision-making during the creative task, and can be coded for indicators of divergent thinking, metacognitive monitoring, and exploratory versus exploitative reasoning strategies. Second, *Interaction log data* from AI tools (including timestamps, prompt inputs, generated outputs, and learner modifications or rejections) provide a continuous, objective record of how learners navigate, accept, modify, or reject AI-generated suggestions (Cropley, 2006). Third, *Screen and audio recording* of the full songwriting session, combined with post-session stimulated recall interviews (Lyle, 2003), potentially enables retrospective access to decision points that learners may not have verbalised during the task itself.

Level 3 - Social interaction. In collaborative songwriting conditions, the social dimension of creativity requires dedicated assessment. Video recordings of group interaction could provide insight into how turn-taking patterns, idea elaboration, and creative conflict contribute to *collaborative emergence* (Sawyer, 2012). Sociometric measures and post-session peer assessments are proposed to complement observational coding by capturing perceived contribution equity and group creative climate. Where AI functions as a "third participant" in the creative dialogue, interaction logs could be analysed jointly with conversational data to examine how AI-generated suggestions are taken up, negotiated, or rejected within the group, and whether their introduction tends to broaden or constrain the group's creative action-space.

Level 4 - Creative outputs. Product-level assessment could be based on the CAT as the primary instrument, supplemented by computational measures of output diversity - including acoustic feature analysis across conditions to operationalise stylistic diversity and the presence or absence of diversity collapse (Yun et al., 2025). Additionally, learners' self-assessments of expressive quality and perceived authenticity of the final product could provide further insights at the level of individual subjective experience.

Critically, the proposed multi-level framework allows the product ratings to be interpreted in conjunction with process data from Levels 1-3, rather than in isolation. Thus, a high-quality output produced with minimal learner agency is evaluated differently from a comparable output that emerged from rich iterative engagement.

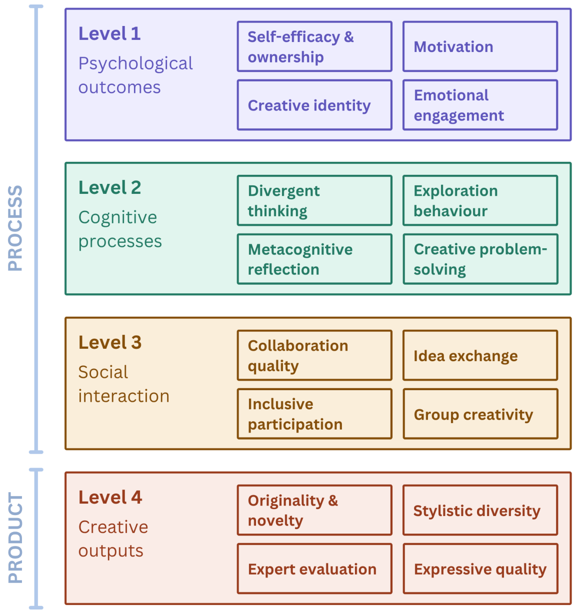

Figure 3: A four-level framework for assessing the creative potential of AI applications in music education.

6.3 Neurophysiological Measures as a Complementary Data Stream

In addition to the behavioural and self-report measures described above, emerging neurotechnological approaches offer a complementary window onto the neural substrates of creative engagement. Brain-computer interface (BCI) systems employing electroencephalography (EEG) have been used to investigate neural correlates of creative states, including frontal alpha oscillations and prefrontal engagement associated with divergent thinking and creative ideation (Fink and Benedek, 2014; Vanutelli et al., 2023). Alpha power over frontal and parietal regions has been consistently associated with internally directed, generative cognition and constitutes a plausible neural marker of the kind of open-ended ideation that creative scaffolding aims to support (Fink and Benedek, 2014).

Although still experimental, these approaches may allow for the assessment of both individual and collective creativity in the future - for instance, through measures of inter-brain synchrony in collaborative settings (Beaty et al., 2021). However, robust and replicable evidence linking these neural signatures to meaningful creative outcomes in ecologically valid educational settings remains limited, and the technical demands of mobile EEG in classroom environments must be carefully managed. Therefore, we position EEG as an *enrichment layer* within the proposed framework that could be valuable for explorative insight.

7 CURRENT STATUS AND FUTURE DIRECTIONS

We examined the relationship between AI tools and creativity in music education. This involved evaluating AI-generated output as a source of creative inspiration, analysing tools used in continuous teacher professional development, and proposing a multi-level theoretical framework to delve deeper into the underlying processes of songwriting. The findings converge on a central and nuanced assumption: AI tools can both enhance and suppress creativity, while this dual potential is most consequential not at the level of the creative product, but at the level of the creative process. In music education, this distinction is decisive. The development of creative competence depends not on the quality of the artefact that a learner produces, but on the cognitive, affective, and social processes through which it is brought into being.

Consequently, the most productive configuration for AI in music education is one characterised by recursive human-AI interaction rather than end-to-end generation. When learners are positioned as active agents who initiate, evaluate, modify, and redirect AI-generated suggestions, rather than as passive recipients of non-editable outputs, the conditions for genuine creative development are substantially more likely to be met. This human-in-the-loop principle is not merely a technical design preference; it is a pedagogical and ethical imperative, grounded in the extensive evidence linking creative ownership, intrinsic motivation, and creative self-efficacy to the sustainability of creative development over time.

Yet a significant gap remains. Despite the breadth of AI tools currently available for music creation, none adequately fulfils the conditions identified in this analysis as necessary for maximising both creative agency and creative breadth simultaneously. The tools that offer the widest creative action-space tend towards end-to-end generation, thereby reducing human agency and risking diversity collapse and homogenisation effects. Conversely, tools that preserve high levels of human agency typically do so by constraining the creative action-space to a degree that limits divergent thinking and exploratory behaviour. No existing tool sufficiently combines maximal human agency, support for divergent thinking and metacognitive reflection, and a genuinely wide, yet non-collapsing, creative action-space. The four-level assessment framework developed in this paper specifies precisely the criteria such a tool would need to satisfy across the psychological, cognitive, social, and output dimensions of creative potential.

This gap points towards two closely related priorities for future research. The first is the development of an (open source) AI tool that enables recursive, reflective, and genuinely open-ended creative interaction while keeping the human learner as the primary creative agent throughout. Such a tool would support rather than supplant divergent thinking, invite metacognitive reflection on creative choices, and resist the homogenising tendencies that characterise current large-scale generative systems. Adding to that, we assume that diffusion models might have more creative potential, but lack explainability. This is a trade-off worth monitoring, because auditability, reproducibility, and data provenance (especially for reasons of authorship and digital aesthetic sovereignty) are also critical.

The second priority is the development of corresponding educational (test) environments and teaching concepts that operationalise the psychological and social conditions identified at Levels 1 and 3 of the model - supporting learners' creative self-efficacy, ownership, and sense of agency within structured pedagogical frameworks. Tool design and pedagogical design must be developed in tandem; an optimally designed AI tool deployed within a pedagogically impoverished environment is unlikely to realise its creative potential, and vice versa. Therefore it is vital to establish transdisciplinary collaborations between school praxis, research and application developers.

A third priority for future research is the development of a validated measurement and assessment framework capable of evaluating AI tools and the teaching concepts in which they are embedded against the four-level model proposed here. Such a framework would need to address not only output-level indicators such as originality and stylistic diversity, but also process-level indicators including exploration behaviour, metacognitive reflection, and creative self-concept, as well as social indicators such as collaboration quality and inclusive participation. The development of such instruments represents a methodologically demanding but necessary step towards evidence-based recommendations for the integration of AI in creative music education at scale. In addition, an agile, subject-specific AI literacy framework is needed to guide learners and teachers through rapidly changing technological developments. Further research on these topics is essential to provide evidence of GenAI's actual impact on creative tasks.

Taken together, the most promising directions for AI in music education may lie in the principled expansion of the contexts, tools, and processes through which creativity can unfold.

REFERENCES

Abdulhai, M., White, I., Wan, Y., Qureshi, I., Leibo, J., Kleiman-Weiner, M., and Jaques, N. (2026). How LLMs distort our written language.

Acar, S., Lee, L. E., and Scherer, R. (2024). A Reliability Generalization of the Torrance Tests of Creative Thinking-Figural. *European Journal of Psychological Assessment*, 40(5):396–411.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology*, 43(5):997.

Amabile, T. M. (1996). *Creativity in context: Update to "The Social Psychology of Creativity."*. Creativity in context: Update to "The Social Psychology of Creativity.". Westview Press, Boulder, CO, US.

Baer, J. and McKool, S. S. (2012). The gold standard for assessing creativity. *International Journal of Engineering Innovation and Management*, 1(2):gold–standard.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W H Freeman/Times Books/ Henry Holt & Co, New York, NY, US.

Beaty, R. E., Benedek, M., Silvia, P. J., and Schacter, D. L. (2016). Creative Cognition and Brain Network Dynamics. *Trends in Cognitive Sciences*, 20(2):87–95.

Beaty, R. E., Zeitlen, D. C., Baker, B. S., and Kenett, Y. N. (2021). Forward flow and creative thought: Assessing associative cognition and its role in divergent thinking. *Thinking Skills and Creativity*, 41:100859.

Beghetto, R. A. (2006). Creative Self-Efficacy: Correlates in Middle and Secondary Students. *Creativity Research Journal*, 18(4):447–457.

Bittner, R. M., Bosch, J. J., Rubinstein, D., Meseguer-Brocal, G., and Ewert, S. (2022). A lightweight instrument-agnostic model for polyphonic note transcription and multipitch estimation. pages 781–785. IEEE.

Bringsjord, S., Bello, P., and Ferrucci, D. (2001). Creativity, the Turing test, and the (better) Lovelace test. In *Minds and Machines*, volume 11, pages 3–27.

Caillon, A. and Esling, P. (2021). Rave: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv preprint arXiv:2111.05011*.

Carnovalini, F. and Rodà, A. (2020). Computational creativity and music generation systems: An introduction to the state of the art. *Frontiers in Artificial Intelligence*, 3:14.

Choi, W. C. and Chang, C. I. (2025). Suno ai: Opportunities and challenges of ai-generated music and lyrics in secondary music education. *Authorea Preprints*.

Collins, A., Brown, J. S., and Newman, S. E. (1989). Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing, and Mathematics. In *Knowing, Learning, and instruction*. Routledge.

Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., et al. (2025). Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*.

Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., and Défossez, A. (2023). Simple and controllable music generation. *Advances in neural information processing systems*, 36:47704–47720.

Cropley, A. (2006). In Praise of Convergent Thinking. *Creativity Research Journal*, 18(3):391–404.

Csikszentmihalyi, M. and Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of nervous and mental disease*, 175(9):526–536.

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, 56(2):469–488.

Da Pelo, M. (2025). Artificial creativity: can there be creativity without cognition? *AI & SOCIETY*, pages 1–14.

Deci, E. L. and Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4):227–268.

Défossez, A., Copet, J., Synnaeve, G., and Adi, Y. (2022). High fidelity neural audio compression. *arXiv preprint arXiv:2210.13438*.

Défossez, A., Usunier, N., Bottou, L., and Bach, F. (2019). Music source separation in the waveform domain. *arXiv preprint arXiv:1911.13254*.

Dobrian, C. and Bevilacqua, F. (2003). Gestural control of music using the vicon 8 motion capture system.

Doshi, A. R. and Hauser, O. P. (2024). Generative ai enhances individual creativity but reduces the collective diversity of novel content. *Science advances*, 10(28):eadn5290.

Elgammal, A., Liu, B., Elhoseiny, M., and Mazzone, M. (2017). Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068*.

Endres, A. and Buchborn, T. (2026). Songwriting mit KI. Digitale Partner im kreativen Prozess?! *OpenMusicAcademy*.

Ericsson, K. A. (2017). Protocol analysis. *A companion to cognitive science*, pages 425–432.

Fink, A. and Benedek, M. (2014). EEG alpha power and creative ideation. *Neuroscience & Biobehavioral Reviews*, 44:111–123.

Gall, M., Buchborn, T., Hennessy, S., and Stumpfogger, M. (2025). *"Liberty - Equity - Creativity: Innovating and Inventing Music in the Classroom"*, volume 13 of *European Perspectives on Music Education*. Innsbruck: Helbling Verlagsgesellschaft mbH.

Godau, M. (2026). Vom „Push the Keys“ zum „Touch the Screen“? Eine Spurensuche zur Konkurrenz der Dinge im Musikunterricht. *Beiträge empirischer Musikpädagogik*, 18(1):1–52.

Godau, M., Weidner, V., and Hermann, K. (2025). (Post-)Digitale Songwritingpraktiken im Musikunterricht. Einblicke in die Erstellung eines Unterrichtsdesigns im Forschungsprojekt Musical Communities in the (Post)Digital Age (MusCoDA).

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Green, A. E., Beaty, R. E., Kenett, Y. N., and Kaufman, J. C. (2024). The Process Definition of Creativity. *Creativity Research Journal*, 36(3):544–572.

Hecht, B. and Krämer, O. (2025). Musikunterricht im Zeichen künstlicher Intelligenz. Der KI-Würfel als fachdidaktisches Modell. *Das Ästhetische im Zeichen künstlicher Verstehenssimulation*, page 159.

Hermann, T., Hunt, A., Neuhoff, J. G., et al. (2011). *The sonification handbook*, volume 1. Logos Verlag Berlin.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851.

Horton Jr, C. B., White, M. W., and Iyengar, S. S. (2023). Bias against ai art can enhance perceptions of human creativity. *Scientific reports*, 13(1):19001.

Jiang, L., Chai, Y., Li, M., Liu, M., Fok, R., Dziri, N., Tsvetkov, Y., Sap, M., Albalak, A., and Choi, Y. (2025). Artificial hivemind: The open-ended homogeneity of language models (and beyond). *arXiv preprint arXiv:2510.22954*.

Jung, R. E., Mead, B. S., Carrasco, J., and Flores, R. A. (2013). The structure of creative cognition in the human brain. *Frontiers in human neuroscience*, 7:330.

Kingma, D. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Krämer, O., Hüttmann, R., and Ziegenmeyer, A., editors (2026). *Diskussion Musikpädagogik: Sonderausgabe lernen:digital*, volume 109. Junker Verlag.

Krampen, G. (2019). *Psychologie der Kreativität: Divergentes Denken und Handeln in Forschung und Praxis*. Hogrefe Verlag GmbH & Company KG.

Krämer, Oliver and Hecht, Benjamin (2024). Anwendungsszenarien von KI im Musikunterricht.

Lan, G. L., Shi, B., Ni, Z., Srinivasan, S., Kumar, A., Ellis, B., Kant, D., Nagaraja, V., Chang, E., Hsu, W.-N., et al. (2024). High fidelity text-guided music editing via single-stage flow matching. *arXiv preprint arXiv:2407.03648*.

Lazarides, R., Richter, D., and Scheiter, K. (2024). Professionalisierung von Lehrkräften–digitalisierungsbezogen und digital? *Zeitschrift für Erziehungswissenschaft*, 27(3):607–612.

Liao, Q. V. and Vaughan, J. W. (2023). AI transparency in the age of LLMs: A human-centered research roadmap. *Harvard Data Science Review*, 5(3).

Luan, Y. L., Kim, Y. J., and Zhou, J. (2025). Augmented learning for joint creativity in human-genai co-creation. *Information Systems Research*.

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal*, 29(6):861–878.

Marcus, G. and Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon Books.

Mehta, A., Chauhan, S., Djanibekov, A., Kulkarni, A., Xia, G., and Choudhury, M. (2025). Music for all: Representational bias and cross-cultural adaptability of music generation models. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 4569–4585.

Merrotsy, P. (2013). A Note on Big-C Creativity and Little-c Creativity. *Creativity Research Journal*, 25(4):474–476.

Organisciak, P., Acar, S., Dumas, D., and Berthiaume, K. (2023). Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity*, 49:101356.

Palmer, C. (2015). Kreativität – praktische Messung einer schillernden Fähigkeit. *Personal Quarterly*, 4:29–35.

Papert, S. and Harel, I. (1991). Situating constructionism. *constructionism*, 36(2):1–11.

Phung, V. and Rolle, C. (2026). Lernziel Musical AI Literacy? Überlegungen zum Einsatz von KI beim Songwriting im Musikunterricht. *Diskussion Musikpädagogik*, 109(1):29–34.

Puccio, G. J., Burnett, C., Acar, S., Yudess, J. A., Holinger, M., and Cabra, J. F. (2020). Creative Problem Solving in Small Groups: The Effects of Creativity Training on Idea Generation, Solution Creativity, and Leadership Effectiveness. *The Journal of Creative Behavior*, 54(2):453–471.

Rausch, E (1952). Struktur und Metrik figural-optischer Wahrnehmung. *Frankfurt/Main: Kramer*.

Reich, Katharina & Phung, V. (2026). SongwrAIting. *lernen:digital Website*.

Rhodes, M. (1961). An Analysis of Creativity. *The Phi Delta Kappan*, 42(7):305–310.

Roberts, A., Engel, J., Mann, Y., Gillick, J., Kayacik, C., Nørly, S., Dinculescu, M., Radebaugh, C., Hawthorne, C., and Eck, D. (2019). Magenta studio: Augmenting creativity with deep learning in ableton live. In *Proceedings of the international workshop on musical metacreation (mume)*, page 11.

Rotsch, T. (2024). Write me a random song!? Musikmachen und Musiklernen im Zeitalter künstlicher Intelligenz. *TelevIZIon*.

Rotsch, T. and Werner, L. (2025). Künstliche Intelligenz im Musikunterricht. Musikbezogene Gestaltungsprozesse in Zukunftstechnologien. *OPUS-PHFR*, pages 197–211.

Rugg, G. and McGeorge, P. (1997). The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14(2):80–93.

Runco, M. A. (2021). Creativity and human flourishing 19. *The Oxford handbook of the positive humanities*, 109:252.

Runco, M. A. (2023). Ai can only produce artificial creativity. *Journal of Creativity*, 33(3):100063.

Runco, M. A. and Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, 24(1):92–96.

Sawyer, K. (2012). Extending sociocultural theory to group creativity. *Vocations and Learning*, 5(1):59–75.

Schleicher, A. (2019). OECD Future of Education and Skills 2030: OECD Learning Compass.

Schumacher, D., LaBounty Jr, F., and Schellekens, M. (2023). Enhancing suno's bark text-to-speech model: Addressing limitations through meta's encodec and pretrained hubert. *Available at SSRN 4443815*.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6):495–504.

Simonton, D. K. (2018). Creative Ideas and the Creative Process: Good News and Bad News for the Neuroscience of Creativity. In Jung, R. E. and Vartanian, O., editors, *The Cambridge Handbook of the Neuroscience of Creativity*, pages 9–18. Cambridge University Press, 1 edition.

Solak, A., Grötschla, F., Lanzendörfer, L. A., and Wattenhofer, R. (2025). Bias beyond borders: Global inequalities in ai-generated music. *arXiv preprint arXiv:2510.01963*.

Song, J. and Wang, Y.-Z. (2025). Musflow: Multimodal music generation via conditional flow matching. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 10200–10209.

Stevenson, B., Smal, I., Baas, M., Grasman, R., and De Dreu, C. K. W. (2022). Putting the creative in creative AI: Assessing the creativity of large language models. *arXiv preprint arXiv:2211.02649*.

Suhailudheen, P. and Sheena, K. (2025). Suno ai: Advancing ai-generated music with deep learning. *Authorea Preprints*.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285.

Tampu, E. M. and Lin, J. (2023). Deep Learning with Audio: An Explorative Syllabus for Music Composition and Production. *AIMC 2023*. https://aimc2023.pubpub.org/pub/fkmqnlfh.

Team, L., Caillon, A., McWilliams, B., Tarakajian, C., Simon, I., Manco, I., Engel, J., Constant, N., Li, Y., Denk, T. I., et al. (2025). Live music models. *arXiv preprint arXiv:2508.04651*.

Thakral, P. P., Yang, A. C., Addis, D. R., and Schacter, D. L. (2021). Divergent thinking and constructing future events: dissociating old from new ideas. *Memory*, 29(6):729–743.

Tierney, P. and Farmer, S. M. (2002). Creative Self-Efficacy: Its Potential Antecedents and Relationship to Creative Performance. *Academy of Management Journal*, 45(6):1137–1148.

Torrance, E. P. (1974). Torrance tests of creative thinking. *Educational and psychological measurement*.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.

Uecker, G. (2008). Schöpfung und Kreativität in der Kunst. In Klose, J. and Oehler, J., editors, *Gott oder Darwin?*, pages 349–360. Springer Berlin Heidelberg, Berlin, Heidelberg.

Urmeneta, A. and Romero, M. (2025). AI as a creative partner: a PRISMA review of AIs role in supporting creativity in education. *Frontiers in Education*, Volume 10 - 2025.

Vanutelli, M. E., Salvadore, M., and Lucchiari, C. (2023). Bci applications to creativity: Review and future directions, from little-c to c2. *Brain Sciences*, 13(4):665.

Vuust, P., Heggli, O. A., Friston, K. J., and Kringelbach, M. L. (2022). Music in the brain. *Nature Reviews Neuroscience*, 23(5):287–305.

Vygotsky, L. S. and Cole, M. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.

Walker, B. N. and Nees, M. A. (2011). Theory of sonification. *The sonification handbook*, 1:9–39.

Wang, T. and Zhang, J. (2025). Enhancing musical creativity through AI: a chain-mediation model of self-efficacy and emotional intelligence. *Frontiers in Psychology*, 16.

Wood, D., Bruner, J. S., and Ross, G. (1976). THE ROLE OF TUTORING IN PROBLEM SOLVING*. *Journal of Child Psychology and Psychiatry*, 17(2):89–100.

Yang, D., Xie, Y., Yin, Y., Wang, Z., Yi, X., Zhu, G., Weng, X., Xiong, Z., Ma, Y., Cong, D., et al. (2026a). Heartmula: A family of open sourced music foundation models. *arXiv preprint arXiv:2601.10547*.

Yang, M., McCormack, J., Llano, M. T., Su, W., and Lei, C. (2026b). Midi-llama: An instruction-following multimodal llm for symbolic music understanding. *arXiv preprint arXiv:2601.21740*.

Yun, L., An, C., Wang, Z., Peng, L., and Shang, J. (2025). The price of format: Diversity collapse in LLMs. *arXiv preprint arXiv:2505.18949*.

Zhang, C., Zhang, C., Zheng, S., Zhang, M., Qamar, M., Bae, S.-H., and Kweon, I. S. (2023). A survey on audio diffusion models: Text to speech synthesis and enhancement in generative ai. *arXiv preprint arXiv:2303.13336*.

Zou, H., Kim, Z. M., and Kang, D. (2023). A survey of diffusion models in natural language processing. *arXiv preprint arXiv:2305.14671*.