UCLouvain

LOUVAIN
School of Management

Louvain School of Management

The human touch and the algorithmic mind: A study on the perception of AI-generated music

CONFIDENTIAL

Author(s): Anass Housni
Supervisor(s): Vincenzo Verardi

Academic year 2025-2026

Dissertation for the Master of Management Sciences

Daytime schedule

Statement of the use of AI

In producing this dissertation, I used artificial intelligence tools (ChatGPT by OpenAI) to support the research, reformulate and improve the flow of certain parts of the text. However, it is important to note that this report is the result of my work and research: AI was not used to produce the content, but only as a tool for editorial optimization and support.

Statements on responsible use of generative AI	**Answer**
The content of my master's thesis is my original output, even if I used generative AI to help prepare it.	**Yes**
I master the theories, tools and methods that I use for the thesis.	**Yes**
I verified that the output presented in the thesis is accurate and valid, and that it contains no hallucination, no false reference, etc.	**Yes**
I acknowledge that I master the final output and that I am able to explain it and answer questions about it.	**Yes**
I made sure not to provide generative AI tools with access to confidential data or information.	**Yes**

Author last name, first name	**Date**	**Signature**
Housni, Anass	28/05/2026	

Supervisor: Vincenzo Verardi

Acknowledgement

This dissertation is the culmination of my final year of Master's studies, but it also marks the closing chapter of an important period of my academic journey that began several years ago. It is with great gratitude and emotion that I would like to acknowledge the people who have accompanied and supported me throughout these years, from the beginning of my studies until the completion of this work.

First and foremost, I would like to sincerely thank my thesis supervisor, Vincenzo Verardi, for his invaluable guidance, insightful advice, and continuous availability throughout the development of this dissertation. His expertise, constructive feedback, and support were essential in helping me carry out this research project and complete it with confidence.

I would also like to express my gratitude to the musician Elliott Tzouros, whose contribution played an important role in this exploratory experimental study. His professionalism and involvement in providing one of the key elements of this research were greatly appreciated.

I would equally like to thank the 93 participants who agreed to take part in this study and placed their trust in this experiment. Their participation and responses were essential to the realization of this research and provided valuable insights for this dissertation.

I would also like to thank my classmates and friends whom I had the chance to meet throughout my academic journey. Their presence, encouragement, and support over the years made this experience particularly meaningful and unforgettable.

Finally, I would like to express my deepest gratitude to friends, my family and loved one for their unwavering moral and emotional support, not only during the writing of this dissertation, but throughout all these years of study. Their encouragement and confidence in me have been a constant source of strength and motivation.

Table of contents

1. Introduction

1.1. Context

Artificial Intelligence ("**AI**") is increasingly becoming a creator in the music industry. From songs fully generated by AI to human & AI co-creations and algorithmic playlists, the boundary between human artistic ability and machine output is shifting. Despite the technological progresses, a central question remains: do listeners care who or what creates the music? Consumer behaviour research suggests that the origin of a product and its perceived authenticity influence how it is evaluated (Grayson & Martinec, 2004). In the music sphere, authenticity, emotional connection and the human "touch" have long been key drivers of value and loyalty. However, recent studies show listeners are more and more exposed to music with unknown or machine-mediated origins (Rigole et al., 2025; Lecamwasam & Chaudhuri, 2025). This thesis investigates how disclosure of origin (AI vs human) impacts perceptions of music pieces, situating the issue within marketing, consumer behaviour and music technology.

1.2. Research question

This thesis investigates how the disclosure of the origin of the music composition (AI vs human) has an effect on listeners' preferences, as well as their perceived authenticity, creativity, emotional connection and audio quality.

Specifically, it will answer the question:

"How does the disclosure of the origin (AI-generated vs. human) of musical compositions influence listeners' preferences and perceptions of authenticity, creativity, quality and emotional value?"

1.3. Objectives of the study

Main objective:

This study will first analyse whether the information about the source of a musical work (AI vs human) influences how Brussels-based listeners of indie-pop-rock music respond to the same piece of music. In other words, it will determine whether listeners prefer AI or human music, and how this preference changes under different disclosure conditions. It will quantitatively analyse how the framing (true disclosure, false disclosure, or no disclosure) affects individual preference and perception for the same music piece as measured by authenticity, creativity, audio quality and emotional connection.

Secondary objectives:

- Statistically compare listener preferences between AI-generated and human-made music under different disclosure conditions (true, false, or no disclosure) using A/B testing logics.
- Analyse how disclosure framing affects perceptions of authenticity, creativity, emotional connection, and audio quality, based on Likert scale responses, using t-tests.
- Detect whether perceived origin (AI vs human) has a significant effect on consumers evaluation, regardless of the actual content.
- Provide empirical insights on consumer bias or trust toward AI-generated art (specifically music), contributing to the literature on AI perception in creative industries.

In addition, I will attempt to address the following hypotheses:

- Listeners have a negative perception of AI-generated music
- Listeners prefer human artists to AI artists
- Music (whether AI-generated or created by humans) is more appreciated if it is stated to be created by humans and is less appreciated if it is stated to be AI-generated
- The same music that listeners enjoy without any label will be less appreciated if it is labelled as AI-generated
- The framing effect is confirmed and influences responses
- Authenticity, creativity, and emotional connection are perceived more positively if the same music is said to be created by a human
- Authenticity, creativity, and emotional connection are perceived less positively if the same music is said to be created by AI

1.4. Rationale and relevance

Music created by AI solutions such as AIVA, Suno and Amper is now becoming essentially indistinguishable from music produced by human musicians; therefore, it will be very important to have an understanding of how consumers view and perceive AI music, particularly in terms of whether or not having an indication of authorship affects their response to it, as platforms begin incorporating these types of solutions into their production and recommendation processes.

The intersection of music marketing, ethics of AI, and consumer behaviour makes up the essence of this study. The findings of the current research contribute to an ongoing academic discussion regarding consumer trust in algorithmic creativity as well as provide useful information and knowledge for marketers. By rigorously analysing listeners' reactions in a controlled experiment, the study bridges theory and practice, while highlighting evolving attitudes toward artificial creativity.

2. Literature review

2.1. Similar works

The rapid development of generative AI has transformed debates surrounding creativity, artistic authorship, and music production. AI systems are now capable of composing melodies, generating lyrics, producing instrumental arrangements, and synthesizing human-like singing voices with increasing realism. As these technologies become integrated into commercial music production and streaming platforms, questions regarding authenticity, emotional resonance, creativity, and audience perception have become central concerns within both academic research and public discourse.

The emergence of AI-generated music challenges longstanding assumptions that artistic creativity is uniquely human. Previous literature in aesthetics and consumer psychology has emphasized the importance of intentionality, emotional depth, and lived human experience in artistic appreciation (Samo & Highhouse, 2023; Merz, 2025). However, recent advancements in machine learning and generative systems increasingly blur the perceptual distinction between human-created and machine-generated music (Figueiredo et al., 2025).

At the same time, studies suggest that listeners continue to associate human-created music with greater authenticity and emotional sincerity, even when they cannot accurately distinguish human compositions from AI-generated ones (Lecamwasam & Chaudhuri, 2025; Rigole et al., 2025). These tensions reveal that judgments about music are not based solely on auditory perception but are also shaped by cognitive framing, contextual information, and broader cultural beliefs about human creativity.

This literature review synthesizes the uploaded studies to examine how consumers perceive AI-generated music as compared to human-created music, with particular emphasis on authenticity, emotional connection, creativity, framing effects, objective musical quality, and evaluative biases.

Theoretical background

Theoretical discussions surrounding AI-generated music are grounded primarily in aesthetics, authenticity theory, technology acceptance research, music perception, and behavioral decision-making frameworks.

A central theoretical issue concerns the definition of creativity itself. Merz argues that generative AI systems do not possess genuine creativity because they lack consciousness, vulnerability, intentionality, and lived experience (Merz, 2025). According to this perspective, AI systems statistically recombine existing patterns without producing meaning through emotional or existential engagement. Human artistic creation is therefore distinguished by fragility, embodiment, and subjective experience.

This debate aligns closely with authenticity theory developed by Grayson and Martinec (2004). Their framework distinguishes between indexical authenticity and iconic authenticity. Indexical authenticity refers to an authentic connection to an original creator or historical origin, whereas iconic authenticity refers to resemblance to what consumers expect authenticity to feel or look like. Applied to AI-generated music, this distinction suggests that AI compositions may achieve iconic authenticity by sounding emotionally convincing while simultaneously lacking indexical authenticity because they are not rooted in human lived experience.

Research in empirical aesthetics further supports the importance of contextual and psychological factors in artistic evaluation. Samo and Highhouse (2023) argue that aesthetic judgments are influenced by objective properties of the artwork, individual characteristics of listeners, and contextual framing. Emotional responses to art therefore depend not only on sensory perception but also on beliefs regarding authorship and intentionality.

Behavioral decision theory also provides an important theoretical foundation for understanding listener perception. Tversky and Kahneman (1986) demonstrate that judgments are highly sensitive to framing effects and contextual presentation rather than purely objective evaluation. In the context of AI-generated music, framing effects help explain why identical musical excerpts may receive different evaluations depending on whether listeners believe the composition was created by a human or an AI system.

Technology acceptance theory further contributes to this discussion. Bagratuni et al. (2025) argue that traditional acceptance frameworks such as TAM and UTAUT2 are insufficient for evaluating AI-generated music because creative AI systems involve emotional and identity-related dimensions absent from purely functional technologies. Their work emphasizes the role of anthropomorphism and perceived humanness in shaping audience acceptance.

Finally, music perception theory emphasizes that musical evaluation depends on cognitive expectations, auditory organization, rhythm perception, memory, and emotional anticipation (Pearce, 2023). Listener judgments are therefore influenced by both perceptual mechanisms and culturally learned schemas.

Authenticity, human creativity, and emotional meaning

One of the most dominant themes across the literature concerns the relationship between authenticity and human creativity. Multiple studies suggest that listeners continue to associate human-created music with emotional depth, intentionality, and artistic sincerity.

Lecamwasam and Chaudhuri (2025) found that participants consistently rated human-composed music as more emotionally effective even when quantitative emotional

responses did not significantly differ between AI-generated and human-created music. Participants of that study associated human-created music with qualities such as “soul,” imperfection, expressive flow, and emotional vulnerability.

Similarly, Rigole et al. (2025) found that participants expressed cautious openness toward AI-human collaboration while still maintaining concerns regarding authenticity and emotional depth in fully AI-generated music.

Merz (2025) provides a philosophical interpretation of these findings by arguing that creativity is inseparable from human fragility and embodied experience. According to this perspective, AI-generated music may imitate stylistic patterns without reproducing the existential conditions underlying authentic artistic expression.

However, some studies complicate this narrative. Chia et al. (2025) found that participants occasionally reported more positive emotional reactions toward music labeled as AI-generated than toward music labeled as human-created. This finding suggests that perceptions of authenticity may vary depending on genre, listener expectations, and contextual framing.

The literature therefore reveals an unresolved debate regarding whether authenticity is fundamentally tied to human intentionality or whether audiences may gradually accept emotionally convincing AI-generated music as authentic in its own right.

Biases, framing effects, and algorithm aversion

Another major theme concerns the influence of labeling and framing effects on listener evaluations.

Horton et al. (2023) found strong evidence of algorithm aversion in artistic contexts. Across six experiments involving nearly 3,000 participants, artworks labeled as AI-generated were consistently evaluated less favorably than identical artworks labeled as human-made. Importantly, this negative bias persisted even when participants acknowledged that the artworks were visually indistinguishable.

Similar findings emerge in music research. Lecamwasam and Chaudhuri (2025) manipulated labels by presenting music as correctly labeled, incorrectly labeled, or unlabeled. Their findings showed that participants systematically perceived music labeled as human-created as more emotionally effective regardless of the actual origin of the composition.

These findings align closely with framing theory proposed by Tversky and Kahneman (1986). Being that, listener evaluations are not purely objective responses to acoustic stimuli but are rather shaped by contextual information regarding authorship and perceived intentionality.

However, Chia et al. (2025) challenge the assumption that AI labeling always produces negative bias. Contrary to expectations, participants sometimes rated AI-labeled pop songs higher on dimensions such as happiness, energy, awe, and interest. This suggests that framing effects may differ according to genre, emotional context, and listener expectations.

Collectively, the literature demonstrates that audience evaluations of AI-generated music are partially socially constructed and heavily influenced by cognitive framing.

Emotional resonance and aesthetic experience

A third recurring theme concerns emotional engagement and the capacity of AI-generated music to evoke affective responses comparable to human-composed music.

Fišer et al. (2025) used biometric and self-report measures to compare emotional reactions to AI-generated and human-composed music in audiovisual contexts. Their findings revealed that AI-generated music produced similar emotional valence ratings while generating higher levels of physiological arousal, including greater pupil dilation and cognitive load indicators.

Pearce (2023) emphasizes that emotional perception in music depends on learned cultural structures, memory systems, auditory expectations, and affective communication processes. From this perspective, emotional engagement with music may not necessarily depend on whether the creator is human or artificial, but rather on how successfully the music activates perceptual and cognitive expectations.

Wilson and Fazenda (2016) similarly demonstrate that listener judgments of audio quality and liking are influenced by distinct perceptual dimensions, including loudness, dynamic range, familiarity, and timbral characteristics.

Meanwhile, qualitative findings suggest that listeners continue to associate emotional authenticity with human imperfection and expressive irregularities (Lecamwasam & Chaudhuri, 2025). This creates an important distinction between emotional stimulation and perceived emotional sincerity.

Objective measures of musical quality in musicology

An important but less explored dimension within the literature concerns objective measures of musical quality. Several uploaded sources contribute theoretical or empirical insight into how musical quality may be evaluated independently from subjective preference.

Wilson and Fazenda (2016) argue that perceived audio quality and preferential judgement represent distinct constructs. Their study identifies measurable acoustic variables associated with perceived quality, including loudness, dynamic range compression, timbral balance, distortion, and spatial properties. The authors show that familiarity

influences preferential judgments, whereas quality ratings are more strongly associated with signal-processing characteristics.

Pearce (2023) further explains that music perception relies on multiple objective perceptual dimensions, including pitch organization, rhythmic structure, timbre, loudness, harmonic relationships, timing, memory, and expectation formation. These perceptual mechanisms provide a cognitive basis for evaluating musical coherence and complexity.

Nguyen et al. (2022) contribute additional insight through research on beat perception and synchronization. Their findings demonstrate that rhythm processing, beat synchronization accuracy, and movement coordination vary according to musical expertise and task type. This work suggests that temporal precision and rhythmic consistency may constitute measurable dimensions of musical quality perception.

Sarmento et al. (2024) similarly identify specific evaluation criteria used by listeners when distinguishing AI-generated from human-created progressive metal music. Participants referred to consistency, playability, structural coherence, genre congruence, and perceived technical quality when evaluating compositions.

Figueiredo et al. (2025) found that listeners relied heavily on vocal realism, production artifacts, repetitive structures, and technical irregularities when identifying AI-generated songs. These findings imply that objective sonic characteristics influence perceived humanness and quality.

Collectively, these studies suggest that music evaluation is not exclusively subjective. Certain measurable musical dimensions such as rhythmic coherence, timbral richness, dynamic variation, structural consistency, and vocal realism appear to systematically influence listener judgments. However, the literature also demonstrates that objective quality does not automatically determine emotional preference or perceived authenticity.

Perceived humanness and detection of AI music

Several studies investigate whether listeners can reliably distinguish AI-generated music from human-created music.

A Turing-like listening experiment using commercial AI-generated songs from Suno alongside human-made music showed that participants performed no better than random guessing when evaluating randomly paired songs (Figueiredo et al., 2025). However, detection accuracy increased significantly when AI and human songs were stylistically similar (Figueiredo et al., 2025).

Listeners relied heavily on vocal quality, production artifacts, repetitive structures, and perceived technical irregularities when attempting to identify AI-generated songs (Figueiredo et al., 2025).

Similar results emerged in progressive metal music research. Although some AI-generated excerpts received ratings comparable to human compositions, listeners still preferred human-created works overall (Sarmento et al., 2024). Thematic analyses revealed that listeners used genre congruence, consistency, playability, and perceived humanness as evaluation criteria (Sarmento et al., 2024).

Co-created (hybrid human-AI collaboration) tracks were also found to be the most difficult for participants to identify correctly (Rigole et al., 2025). This finding suggests that hybrid human-AI collaboration may blur perceptual boundaries more effectively than fully autonomous AI generation.

Overall, these studies indicate that listeners are imperfect detectors of AI-generated music and frequently rely on contextual heuristics rather than objective perceptual accuracy.

Technology acceptance and anthropomorphism

Another significant debate concerns anthropomorphism and audience acceptance of AI-generated music.

Hong et al. (2022) found that AI systems displaying human-like characteristics were more likely to be accepted as legitimate musicians. Humanlike traits, such as perceived creativity, perceived intentionality, emotional expressiveness, etc., increased perceptions of artistic legitimacy even when the actual musical evaluation remained unchanged.

Perceived animacy, sociability, and humanlike fit significantly increase the likeability of AI-generated singing (Bagratuni et al., 2025). Curiosity and word-of-mouth were also identified as major drivers of acceptance intentions (Bagratuni et al., 2025).

The same study argues that traditional technology acceptance frameworks such as TAM and UTAUT2 are insufficient for evaluating AI systems in creative domains because creative AI engages emotional and identity-related dimensions absent from utilitarian technologies (Bagratuni et al., 2025).

These findings complement other studies showing that listeners evaluate AI-generated music not only according to sonic quality but also based on perceived humanity, intentionality, and emotional sincerity (Lecamwasam & Chaudhuri, 2025; Rigole et al., 2025).

Interestingly, the literature does not suggest outright rejection of AI music. Instead, audiences appear more accepting of collaborative or assistive AI systems than fully autonomous artistic replacement.

Order effects and sequential biases in music evaluation

Another important issue concerns order effects and sequential biases in evaluative judgment.

Ginsburgh and van Ours (2003) demonstrate that performance order significantly influenced rankings and subsequent economic outcomes in the Queen Elisabeth musical competition despite the fact that the order of appearance was assigned randomly. Their findings suggest that judgments presented as objective evaluations of artistic quality may in fact be influenced by contextual sequencing effects unrelated to actual performance quality.

This work is highly relevant for studies investigating AI-generated music perception because many experimental designs expose participants to multiple musical excerpts in sequence. Listener evaluations may therefore be unintentionally influenced by comparison effects, fatigue, contrast effects, adaptation, or memory biases.

The theoretical framework proposed by Tversky and Kahneman (1986) further supports this interpretation. Their work demonstrates that human judgments are highly sensitive to contextual framing and cognitive heuristics. Applied to music perception experiments, this implies that the order in which AI-generated and human-created songs are presented may shape listener evaluations independently from the actual musical characteristics.

These findings have important methodological implications for AI music perception research. Randomization procedures, crossover designs, and counterbalancing become essential to minimize order-related biases.

Figueiredo et al. (2025) partially address this concern through a randomized controlled crossover design intended to control for pairwise similarity effects and contextual influence during AI music identification tasks.

Overall, the literature suggests that evaluations of AI-generated music may be influenced not only by the music itself but also by the broader structure and sequencing of the listening experience.

Comparison of authors' perspectives

The literature reveals substantial convergence regarding the importance of authenticity and emotionality in music perception. Most studies agree that listeners continue to value human-created music more highly in terms of emotional sincerity and artistic authenticity (Merz, 2025; Lecamwasam & Chaudhuri, 2025; Rigole et al., 2025).

However, disagreement emerges regarding the extent of negative bias toward AI-generated music. Horton et al. (2023) and Lecamwasam and Chaudhuri (2025) identify strong algorithm aversion effects, whereas Chia et al. (2025) report more positive emotional evaluations for AI-labeled pop songs.

Differences in methodology may partially explain these contradictions. Horton et al. (2023) focus primarily on visual art, whereas Chia et al. (2025) examine commercially oriented pop music. Additionally, some studies use deceptive labeling paradigms while others rely on blind listening experiments.

Theoretical perspectives also diverge. Merz (2025) adopts a philosophical position emphasizing the irreplaceability of human fragility in artistic creation, whereas technology acceptance researchers such as Bagratuni et al. (2025) focus more pragmatically on conditions that facilitate audience acceptance of AI-generated music.

At the same time, studies focused on objective quality measures emphasize measurable sonic and structural dimensions of music evaluation (Wilson & Fazenda, 2016; Nguyen et al., 2022), whereas authenticity-focused research prioritizes emotional intentionality and symbolic meaning.

Despite these differences, the literature consistently suggests that perceptions of AI-generated music depend not only on objective musical properties but also on symbolic beliefs regarding humanity, creativity, intentionality, and contextual framing.

Methodological approaches in the literature

The literature employs several major methodological approaches. Blind listening experiments are among the most common methodologies and are used to examine whether participants can distinguish AI-generated music from human-created music (Figueiredo et al., 2025; Sarmento et al., 2024; Rigole et al., 2025).

Label manipulation experiments investigate framing effects by varying whether music is described as AI-generated or human-created regardless of its actual origin (Lecamwasam & Chaudhuri, 2025; Chia et al., 2025).

Mixed-methods designs combine quantitative ratings with qualitative thematic analysis to explore emotional reactions and listener interpretations (Sarmento et al., 2024; Lecamwasam & Chaudhuri, 2025).

Survey-based acceptance models examine anthropomorphism, perceived humanness, and technology acceptance using statistical modeling approaches (Bagratuni et al., 2025).

Biometric methods are also increasingly used. Fišer et al. (2025) combined self-report questionnaires with physiological indicators such as pupil dilation and skin impedance to investigate emotional responses to AI-generated music.

Music perception studies additionally employ synchronization tasks, beat production experiments, and auditory processing measures to evaluate objective perceptual mechanisms (Nguyen et al., 2022).

Finally, several studies adopt conceptual or philosophical methodologies focused on defining creativity, authenticity, and artistic legitimacy (Merz, 2025; Grayson & Martinec, 2004).

Research gaps and limitations

Several important limitations emerge across the literature:

- First, most studies rely on relatively short listening excerpts presented in controlled experimental settings. This limits ecological validity because real-world music consumption often occurs over repeated exposures and emotionally meaningful personal contexts.
- Second, most research focuses on Western audiences, limiting understanding of cross-cultural perceptions of AI-generated music and authenticity.
- Third, there is limited exploration of genre-specific differences. Existing studies focus primarily on pop music, ambient music, and progressive metal, leaving many genres unexplored.
- Fourth, many studies rely on self-report measures despite the increasing availability of biometric and behavioral methodologies.
- Fifth, the distinction between fully AI-generated music, AI-assisted composition, and human-AI co-creation remains theoretically underdeveloped.
- Sixth, although several studies discuss objective measures of musical quality, there is still no consensus regarding which acoustic, structural, or perceptual variables best predict listener evaluations of AI-generated music.
- Finally, few studies systematically address order effects, fatigue effects, or sequential biases despite evidence suggesting that presentation order may significantly influence artistic evaluation.

Conclusion

The literature demonstrates that consumer perceptions of AI-generated music emerge from a complex interaction between auditory perception, emotional expectations, cognitive framing, objective musical characteristics, and cultural beliefs regarding human creativity.

Although listeners frequently struggle to reliably distinguish AI-generated music from human-created compositions, they continue to associate human authorship with greater authenticity, emotional depth, and artistic legitimacy. At the same time, evidence suggests that AI-generated music can successfully evoke emotional responses and, in some contexts, may even generate positive audience reactions.

Framing effects, anthropomorphic perceptions, and order effects all play important roles in shaping evaluations of AI-generated music. Listener judgments are therefore not

determined solely by acoustic characteristics but are strongly influenced by contextual assumptions regarding intentionality, creativity, and humanness.

Research on objective musical quality further suggests that measurable dimensions such as rhythmic coherence, vocal realism, dynamic variation, timbral richness, and structural consistency systematically influence listener perception. However, objective quality alone does not fully explain perceived authenticity or emotional connection.

Overall, the literature suggests that AI-generated music does not merely challenge technical boundaries in music production. It fundamentally challenges how audiences define creativity, authenticity, emotional expression, and artistic value.

2.2. Theoretical model

AI mechanism in content generation

Iterative Model Refinement

Modern AI systems rely on iterative model refinement to improve performance through repeated cycles of training. This process involves adjusting model parameters based on the difference between predictions and actual outputs, known as the loss function. Over multiple iterations, the model learns to minimize this loss by updating internal weights using algorithms such as gradient descent (Goodfellow, Bengio, & Courville, 2016). This dynamic learning mechanism allows the system to converge on patterns that reflect the statistical structure of its training data. Iterative refinement is the foundation of how neural networks become increasingly accurate in tasks such as classification, translation, or generation.

Generative Pretrained Transformer (GPT) as a Neural Network

One of the most prominent examples of such neural networks is the Generative Pretrained Transformer ("**GPT**"), a model based on the transformer architecture. Simply put, GPT is trained to predict the next word in a sequence, allowing it to generate coherent and contextually appropriate text (Radford et al., 2019; Vaswani et al., 2017). This transformer-based model uses attention mechanisms to process and relate words over long sequences, capturing linguistic structure more effectively than previous architectures. Although powerful, its generative capability is rooted entirely in statistical prediction: it selects the most probable next token based on training data rather than drawing on intention, consciousness, or purpose.

Regression to the Mean in AI-Generated Outputs

This leads to a central limitation of generative AI: its tendency to produce outputs that approximate the average of its input data. This phenomenon, often described as a form of regression to the mean, arises because models like GPT are optimized to generate the most statistically likely outputs (Bender et al., 2021). As a result, while they can

recombine elements from large datasets in novel ways, they generally fail to produce truly original or disruptive content. Instead, they replicate dominant trends and stylistic norms present in the training data. This makes generative AI effective for imitation but fundamentally limited in its ability to create works that challenge conventions or introduce radical innovation.

Application to AI Music Platforms like SUNO

In combination, these three principles explain how platforms such as SUNO generate AI-composed music. These systems do not invent from scratch. Rather, they draw on statistical regularities in vast musical datasets to produce songs that sound plausible and familiar. While this allows for efficiency and stylistic mimicry, it also restricts the output to variations of existing musical patterns. Unlike human composers who create based on lived experience, emotion, and intentional disruption (Merz, 2025), AI models lack the cognitive and existential grounding necessary for innovation. As a result, based on current technology, AI-generated music remains a reflection of what already exists, rather than a source of genuine artistic evolution.

Statistical models for comparisons

To analyse how listeners respond to musical excerpts under different labelling conditions, two statistical models can be considered. The first is the Bradley-Terry model, often used in ranking and preference research. The second is the A/B testing framework, widely used for testing differences between two groups or conditions. Each model offers a different approach to interpreting paired comparisons.

The Bradley-Terry model is a probabilistic method used to estimate the likelihood that one item is preferred over another based on pairwise comparisons. Each item is assigned a latent score, and the probability that item A is chosen over item B is calculated as the ratio of their scores. This model is useful when analysing structured preferences across multiple items and conditions. It has been widely applied in psychology, tournament ranking, and marketing.

A/B testing is a simpler but robust method used to compare two versions of a stimulus (A and B) to see if there is a statistically significant difference in response. It is commonly used in experimental and behavioural sciences to test how different conditions influence decisions or perceptions. In its most basic form, it compares the means (or proportions) of two samples using t-tests (for continuous data) or z-tests (for proportions).

A/B testing was selected for this study because it allows direct and transparent comparisons between conditions that mirror the structure of the experiment. The research question investigates whether labelling (AI versus human) influences preference and perceived qualities of the same musical excerpt. A/B testing supports this by comparing means across sessions and identifying framing effects. It is easy to implement

using Stata, interpretable without complex modelling, and appropriate for the available sample size. Although the Bradley-Terry model provides a richer analytical framework, its implementation would require more participants, deeper statistical expertise, more inputs, more budget and different experiment conditions and more time. A/B testing offers a practical and reliable alternative that meets the objectives and constraints of this research.

2.3. Conceptual hypotheses

1. **Listeners hold a negative perception toward AI-generated music.**

This hypothesis is grounded in empirical studies showing that AI-generated creative work tends to be evaluated more critically than human-made art, especially when the AI origin is disclosed. Horton et al. (2023) demonstrated a systematic bias against AI-labelled artworks, even when quality was held constant. Listeners may perceive AI compositions as lacking depth or emotional intent, especially given the absence of human experience behind the creation (Merz, 2025). This scepticism aligns with the theory of authenticity in consumer evaluation (Grayson & Martinec, 2004), where AI lacks both iconic and indexical markers of authenticity.

2. **Listeners prefer human artists over AI composers.**

The literature consistently shows that audiences value human creators because they attribute intentionality, emotional labour, and vulnerability to their work. Horton et al. (2023) and Merz (2025) both demonstrate that music labelled as human-made tends to be perceived as more creative and emotionally engaging. The human origin of a piece contributes to a stronger sense of artistic merit, leading listeners to prefer music when they believe it was created by a person rather than a machine.

3. **Music is rated higher when labelled as human-made and rated lower when labelled as AI-generated, regardless of the actual source.**

This idea is rooted in framing theory (Tversky & Kahneman, 1986) and supported by multiple studies showing that labelling a stimulus changes its perceived value. Lecamwasam and Chaudhuri (2025) found that identical music clips received significantly higher ratings when labelled as human-created compared to AI-generated. The label shapes expectations, which in turn influence how the music is interpreted and evaluated.

4. **A track enjoyed under blind conditions will be rated lower if later disclosed as AI-generated.**

This hypothesis follows from post-evaluation framing effects discussed by Lecamwasam and Chaudhuri (2025). They show that when listeners discover that a previously enjoyed piece was generated by AI, they tend to reevaluate it more negatively. The initial

appreciation becomes tempered by cognitive dissonance or a feeling of being “fooled,” leading to a downward revision of earlier ratings.

5. **Framing effects significantly influence listener evaluations.**

Building from Hypotheses 3 and 4, this broader hypothesis examines whether changes in labelling (truthful or misleading) affect preference and perception. The idea that the same musical piece can be perceived differently based solely on how it is presented is well-supported. Framing effects alter the cognitive lens through which the music is experienced. Horton et al. (2023) and Chia et al. (2025) show that music labelled as AI triggers lower ratings even when content is held constant, while Merz (2025) finds the reverse effect when music is labelled as human-made. These findings confirm that framing is a strong determinant of listener response.

6. **Authenticity, creativity, audio quality and emotional connection are rated higher when the music is labelled as human made.**

Several sources in the literature review show that human labels are associated with higher scores across key dimensions of perception. Wilson and Fazenda (2016) note that audiences link human creation with greater nuance in emotional delivery, while Merz (2025) finds that authenticity is strongly tied to perceived authorship. This pattern holds across multiple evaluative categories, suggesting that listeners not only prefer human-made music overall but also score it more favourably in detailed assessments.

7. **Authenticity, creativity, audio quality and emotional connection are rated lower when the music is labelled as AI-generated.**

This final hypothesis mirrors the previous one and emphasizes the inverse effect. AI-labelled music is systematically disadvantaged across core evaluative dimensions. Studies show that listeners question whether AI can produce meaningful emotional content or originality (Samo & Highhouse, 2023). Even when musical quality is technically high, the label "AI" introduces doubt and leads to lower scores for creativity, emotional engagement, and perceived authenticity.

3. Methodology and data requirements

3.1. Overview of the experimental design

To examine how the disclosure of the origin of a musical piece (AI vs human) influences listeners’ preferences and perceptions of authenticity, emotional resonance, emotional connection and quality, this study implemented a within-subject experimental design based on paired comparisons analysed using the A/B testing logic. Each participant listened to short musical clips under three different framing conditions to determine whether labelling affects preferences and perception independently of actual musical content.

3.2. Participants and sampling

Sample size and statistical power

This study used a within-subject experimental design in which each participant evaluated music under three framing conditions: blind, true disclosure, and false disclosure. Because the same participants contributed responses across conditions, observations are not treated as independent across sessions. Instead, the design produced paired, correlated observations, which increased control over between-participant variability but must be accounted for in the analysis.

The target statistical power was set at 0.80 with a significance level of $\alpha = 0.05$, following conventional standards in behavioural research. Based on prior studies on framing effects and AI perception, an effect size corresponding to an approximate 0.20 was expected. Power estimations for paired binary comparisons indicated that approximately 90 participants would provide sufficient power to detect such effects in a within-subject design.

Importantly, the three sessions are not pooled as independent observations. Comparisons between framing conditions were analysed as within-subject comparisons, using statistical methods appropriate for paired or repeated-measures data.

Participants profile and sampling

The experiment involved 93 participants, all self-declared indie-pop-rock music fans and artistic people, currently living in Brussels, Belgium.

This study adopts an exploratory approach focused on a specific community with the goal of understanding their behaviour in relation to AI-generated music rather than aiming for a statistical representativeness of the general population.

Consequently, a convenience sampling method was adopted to reach participants who meet participation criteria:

- Indie-rock music
- Personal or professional involvement with art
- Residence in Brussels, Belgium

Participants were selected from my personal network to ensure alignment with these characteristics. Recruitment was conducted in person during local cultural events, as well as online via social media platforms.

In other words, this sampling strategy does not aim to extrapolate to the broader Belgian or Brussels population. The intent is not statistical generalization, but rather an in-depth exploration of how individuals from this community perceive and respond to AI-generated music.

3.3. Stimuli and experiment procedure

Music creation

Two music excerpts were composed for this experiment:

- One composed by a human independent artist named **Elliott Tzouros**. Elliott Tzouros is a seasoned professional musician currently based in Hasselt, Belgium. He studied at the Conservatory of Music in Ghent and has built a career spanning over 15 years as a touring and session guitarist, producer, and sound engineer. His work includes songwriting, production, and live performance collaborations with a wide range of Belgian artists. Elliott has contributed as a recording engineer and co-producer for Iskander Moon, also serving as a live guitar technician and in-house sound engineer for the artist's Living Room Concerts. He has written songs for Ibe and PAAK, and played live and/or in studio for acts like M1TCH, Bordoo, Maxime, B!NO, Abraham Blue, and various blues ensembles. He has also worked on songwriting and production with Josephine Rioda, Wavery, and Alioth, and taken on broadcast engineering roles for Sevens. His versatile skill-set and deep technical expertise make him a highly sought-after figure in Belgium's music scene.

- One generated by an AI music system called **Suno**. Suno is an AI tool developed by a team of engineers and researchers based in Cambridge, Massachusetts. The team includes former employees from tech and AI-focused companies like Meta, TikTok, and Google, with expertise in machine learning, audio processing, and music generation. The platform allows users to create original songs simply by writing a short text prompt and can generate full music tracks with instruments and vocals that sound like real human voices, all based on the style, mood, or theme prompted by the user. Suno works by using powerful generative models trained on large datasets of music and lyrics. These models have learned how to combine melodies, harmonies, rhythms, and words in a way that sounds coherent and expressive.

The exact same prompt was given to both the human artist and the AI music platform in order to create the music excerpts used in the experiment:

"Create a 30-seconds-long instrumental chorus (no lyrics) with folk, rock, blues and indie influences with a BPM of 70. Be creative."

Each excerpt are 20-30 seconds long and carefully matched in tempo, structure, and genre. Audio normalization was applied to ensure loudness parity. No vocals or lyrics was used to avoid confounds related to voice identity.

Experiment

The experiment was conducted online using a Google Form survey (appendix I, page 42) consisting of several sections and listening sessions. The experiment lasted between 5 and 7 minutes, which is the optimal duration for maintaining participants' attention and concentration at a consistent level throughout the experiment, thereby avoiding biases caused by fatigue and loss of focus. Each participant went through three sequential listening sessions, in the same order:

- ***Session 1 - Blind Condition***: Both excerpts were labelled as "Music A" and "Music B," with no information on their origin. Participant selects preferred excerpt.
- ***Session 2 - True Disclosure***: Participants were told: "Music A was composed by a human", "Excerpt B was generated by AI". Participant selected preferred excerpt and rated (on a 7-point Likert scale) each track on: perceived authenticity; emotional connection; perceived creativity; perceived audio quality
- ***Session 3 - False Disclosure:*** The labels were reversed from the true origin:
  "Music A was generated by AI" (when it's actually human)
  "Music B was composed by a human" (when it's actually AI)
  Participant again selected their preferred excerpt and rated each track on the same Likert items.

Breaks of 20 seconds separated each session to avoid fatigue and anchor bias.

To ensure that all participants had the same understanding of the four dimensions authenticity, emotional connection, creativity, audio quality), a definition of each dimension was provided before moving on to the rating step. This helps avoid any bias, ensuring that participants rate the different dimensions based on the same criteria and definitions. The given definitions are the following:

1. Authenticity: How genuine or sincere the music feels to you? Does it seem personal, meaningful, or true to a real artistic intention?
2. Emotional connection: How much the music makes you feel something? Does it move you, touch you, or evoke any emotion?
3. Creativity: How original or inventive the music seems? Does it sound fresh, unique, or imaginative?
4. Audio quality: How clear and pleasant the sound is? Is it well-produced, clean, and comfortable to listen to?

3.4. Data collection and inputs

Each participant provided:

- Three binary preference decisions (Session 1, 2, 3);
- 16 Likert ratings (2 excerpts * 4 dimensions * 2 sessions); and
- Responses to the final attitude question.

In addition, the survey collected:

- Basic socio-demographic data (age, gender, municipality, occupation, frequence of music consumption); and
- A final multiple-choice question on their general opinion about AI-generated music.

The independent variables are:

- The actual origin of the music clips (human or AI);
- The displayed label of the music clip (human-generated, AI-generated, or undisclosed); and
- The informational framing condition (blind, true disclosure, false disclosure).

The dependent variables are:

- Music clip preference (binary: preferred A or B); and
- Perceived authenticity, creativity, audio quality and emotional connection (Likert scale, 1-7).

3.5. Statistical modelling: A/B testing approach

The analysis followed a simple A/B testing logic to examine whether the way a piece of music is presented affects how people react to it. Each participant listened to the same two music clips under three different situations. This setup helped to identify if people react differently depending on what they are told about the origin of the music.

I compared how often each track was preferred depending on the condition, using mean tests when the outcome is a clear choice between two clips. For the perception scores such as authenticity, creativity, audio quality and emotional connection, paired t-tests were used to compare how the same track was rated under different labels. Since all participants gave repeated answers across the three sessions, we will account for this to avoid misleading effects related to individual preferences. This method makes it possible to clearly test if the framing of the music, rather than the music itself, has an impact on how it is judged.

Formula (z-test for proportions):
$\mathbf{z = (p_1 - p_2) / \sqrt{[p(1-p)(1/n_1 + 1/n_2)]}}$, where $\mathbf{p = (x_1 + x_2)/(n_1 + n_2)}$ is the pooled proportion.

The statistical tests and analysis will be conducted using the statistical software Stata.

The main statistical hypotheses $\mathbf{H_1}$ that are tested are the following:

- $\mathbf{H_1.1}$: Participants show a statistically significant preference for a track they are told are composed by human artists over another one labelled as AI-generated.
- $\mathbf{H_1.2}$: Scores for perceived authenticity, creativity, emotional connection and audio quality are significantly higher when participants are told the music was created by a human.

- **H_1.3**: Human-labelled music consistently receives higher scores than AI-labelled music, across all subjective dimensions

Concretely, the procedure is as follows:

For musical preference (binary variable):

- Comparing the proportion of participants who preferred Music A across:
  - ➔ Session 1 and Session 2
  - ➔ Session 1 and Session 3
  - ➔ Session 2 and Session 3
- Same for Music B

This logic applies to both excerpts and helps identify whether labelling influences preferences, even when the musical content remains constant.

For perception ratings (Likert scales):

- Focus on Sessions 2 and 3, where participants evaluated each excerpt on four dimensions
- For both Music A and Music B, conduct paired t-tests between Session 2 and Session 3 on the four dimensions

4. Results

The dataset consists of responses from 93 participants, successfully meeting the target sample size, which was determined based on a statistical power set at 0.80. The data will be analysed through both visual observation of descriptive graphs and statistical testing performed using Stata, allowing for a rigorous evaluation of the experimental hypotheses.

It is important to keep in mind that the statistical results presented in this study are not intended to be generalized to the broader population. They reflect the responses of a convenience sample, composed of participants who share specific socio-demographic characteristics. This approach allowed for an experimental exploration of perceptions within a defined group, without aiming to represent the general public or extend the findings beyond this specific context.

No separate statistical analyses were conducted on socio-demographic variables in the main comparisons because the study relies on a within-subject paired design. Since the same participants responded to all three experimental sessions, everyone effectively serves as their own control. This repeated-measures structure substantially reduces the influence of inter-individual variability and neutralises potential moderating effects linked to socio-demographic characteristics such as age, gender, or educational background across conditions. As a result, the observed differences between sessions can be

interpreted primarily as effects of the framing manipulation rather than differences between participant profiles.

4.1. First observations

A first observation from the comparison of the two tracks, both composed from the same prompt, is the noticeable difference in musical complexity and stylistic interpretation. The human-composed piece stood out by incorporating additional stylistic nuances, notably a Motown influence layered over the core genres mentioned in the prompt. It also featured a more sophisticated chord progression with jazzy inflections, and the mixing included stereo elements that added spatial depth. The arrangement showcased a diverse palette of instruments not found in the AI-generated version, including chimes, an organ-like synthesizer, and three distinct layers of guitar harmonies, alongside more conventional genre instruments such as bass, drums, acoustic and electric guitars. In contrast, the AI-generated track created using Suno adhered closely to standard rock folk blues conventions, employing a typical instrumental setup of acoustic drums, acoustic and electric guitars, bass, and banjo. The mix was monophonic, although stereo output can be achieved with higher-end AI tools. Both tracks offered comparable audio quality. However, this early analysis suggests that the human composer exercised more creative freedom, while the AI track remained stylistically faithful and conventional, aligning with expectations of genre norms.

4.2. Descriptive analysis

4.2.1. Music clip preferences under three different conditions

The pie charts from the three sessions (Figure 1) provide an initial overview of participant preferences between Music A (human-composed) and Music B (AI-generated). In the blind session (Session 1), 65% of respondents preferred Music A, indicating a general inclination even without knowing the origin. In Session 2, where the origins were truthfully disclosed, this preference increased sharply to 81%, suggesting that disclosure enhances appreciation for human-composed music. Conversely, in Session 3, when the labels were reversed (false disclosure), preference for Music A dropped to 57%. These differences in proportions across sessions suggest a framing or labelling effect on participants' choices. While these descriptive observations already indicate potential influences of disclosure, the validity of these differences will be tested through inferential statistics such as t-tests in the analytical section.

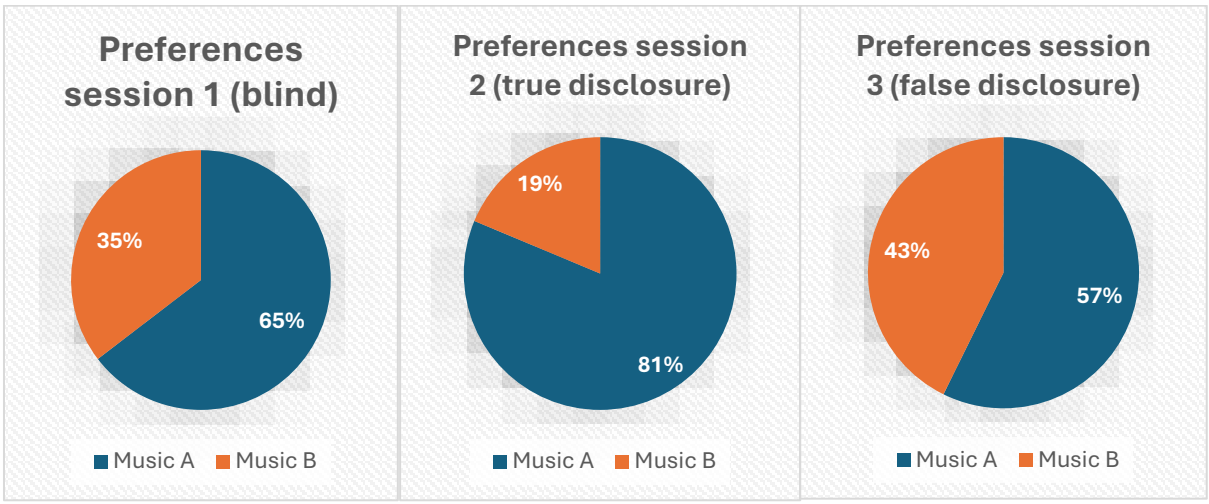

Figure 1: Proportion of preferences for Music A and Music B during the Session 1(blind), Session 2 (true disclosure) and Session 3 (false disclosure)

4.2.2. Sentiment towards AI-generated music

This pie chart (Figure 2) indicates a strong reluctance among respondents to engage with AI-generated music. A significant 67% of participants answered "No" when asked if they would listen to music knowing it was generated by artificial intelligence. Only 14% responded positively with "Yes," while 19% selected "Maybe," suggesting some uncertainty or conditional openness. These results demonstrate a general scepticism or negative bias toward AI-generated music, which aligns with the first hypothesis of the study. This initial descriptive statistic reinforces the relevance of further inferential testing to explore how such perceptions influence actual preferences in different labelling conditions.

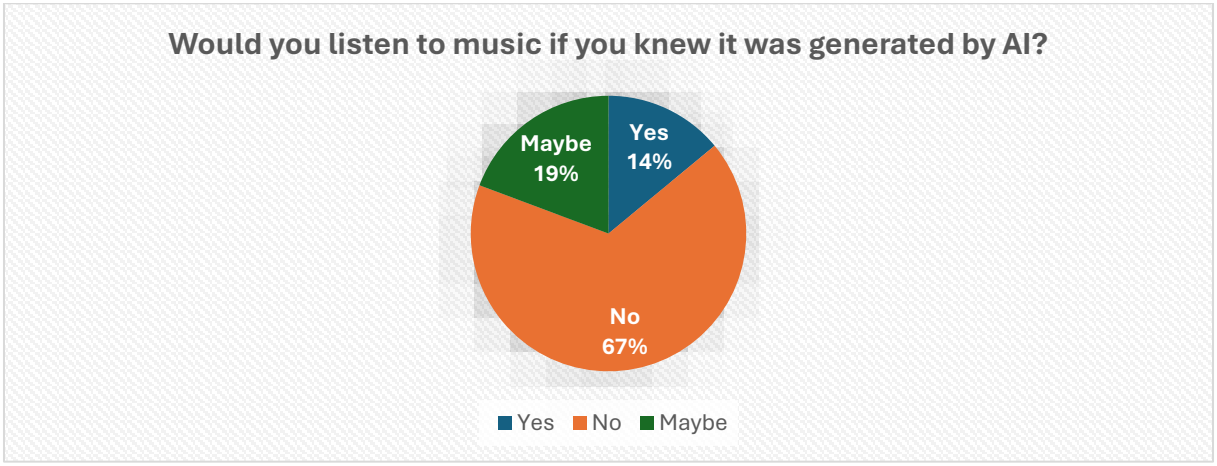

Figure 2: Proportion of answers to the question "Would you listen to music if you knew it was generated by AI?"

The responses to the question “What do you think of AI-generated music?” (appendix II, page 56) reveal a predominantly negative or ambivalent sentiment. Out of all responses, 45 explicitly state “I don't like it,” which is by far the most common answer. While a smaller group (13 respondents each) say “I like it,” “I can't really tell the difference,” or “I don't

mind it as long as the result sounds good," many others provided nuanced opinions. These include moral or ideological objections (e.g., "AI in art is a shame," "I prefer ethical music by humans"), political stances, or reluctant appreciation (e.g., "I hate that I liked it").

Several responses combine multiple sentiments, reflecting internal conflict or conditional acceptance, such as liking it only if it supports real artists or being troubled by liking AI-generated content. A few respondents also noted difficulties distinguishing AI from human compositions or expressed surprise upon discovering their preferred track was AI-generated.

4.2.3. Perception ratings (7-point Likert scales) for authenticity, creativity, emotional connection and audio quality

To make the analysis easier to interpret, we compare how participants rated authenticity, creativity, emotional connection, and audio quality across Session 2 (true disclosure) and Session 3 (false disclosure) for both Music A and Music B. The Likert scale results (1 to 7) displayed in the graphs show consistent differences in ratings between the two sessions (Figures 3,4,5,6,7,8,9 and 10).

For both Music A and Music B, perceived authenticity, creativity, and emotional connection tend to be rated higher in Session 2, when participants are truthfully informed about the source of the music, compared to Session 3, where the source is falsely labelled. This suggests that the label indicating human authorship positively influences listener perception. However, a much smaller difference is observed for the audio quality dimension. Ratings for audio quality remain more stable between the two sessions, indicating that this dimension may be less influenced by framing and more tied to intrinsic properties of the music itself.

As with the comparison of preferences, the validity of these differences will be tested through inferential statistics such as t-tests in the analytical section.

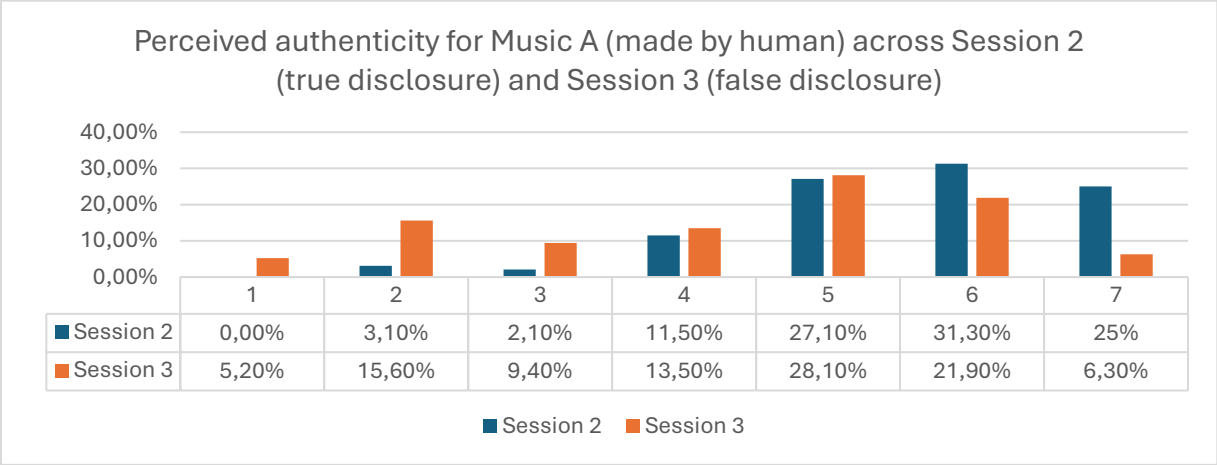

Figure 3: Perceived authenticity for Music A (made by human) across Session 2 (true disclosure) and Session 3 (false disclosure) on a 7 points Likert scale

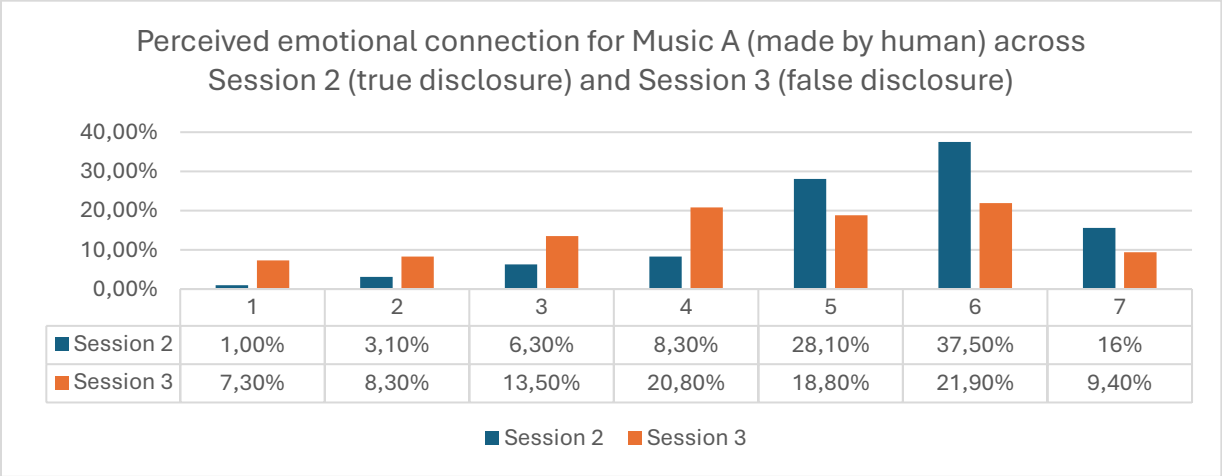

Figure 4: Perceived emotional connection for Music A (made by human) across Session 2 (true disclosure) and Session 3 (false disclosure) on a 7 points Likert scale

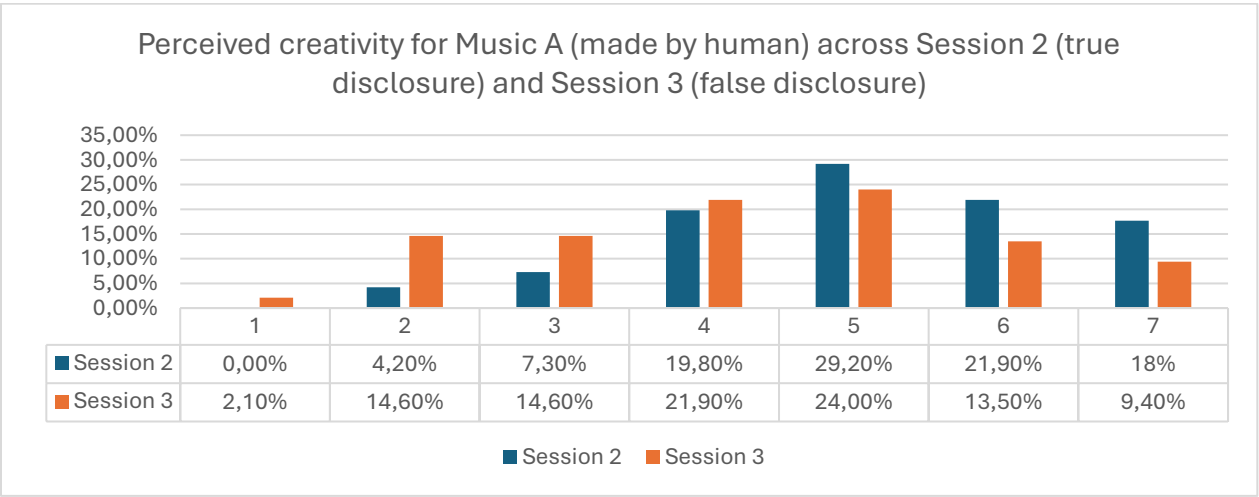

Figure 5: Perceived creativity for Music A (made by human) across Session 2 (true disclosure) and Session 3 (false disclosure) on a 7 points Likert scale

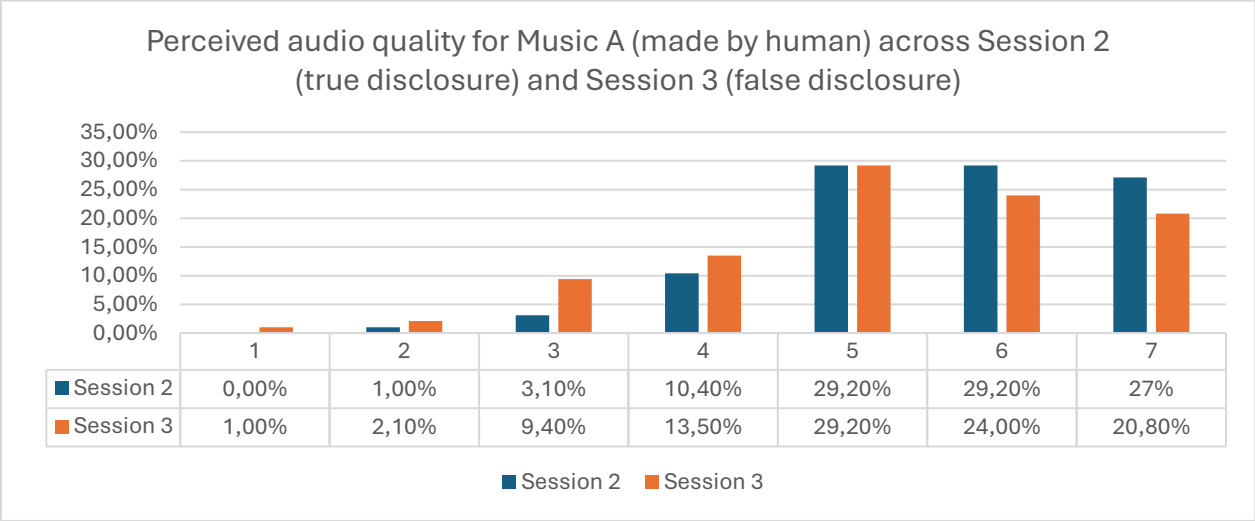

Figure 6: Perceived audio quality for Music A (made by human) across Session 2 (true disclosure) and Session 3 (false disclosure) on a 7 points Likert scale

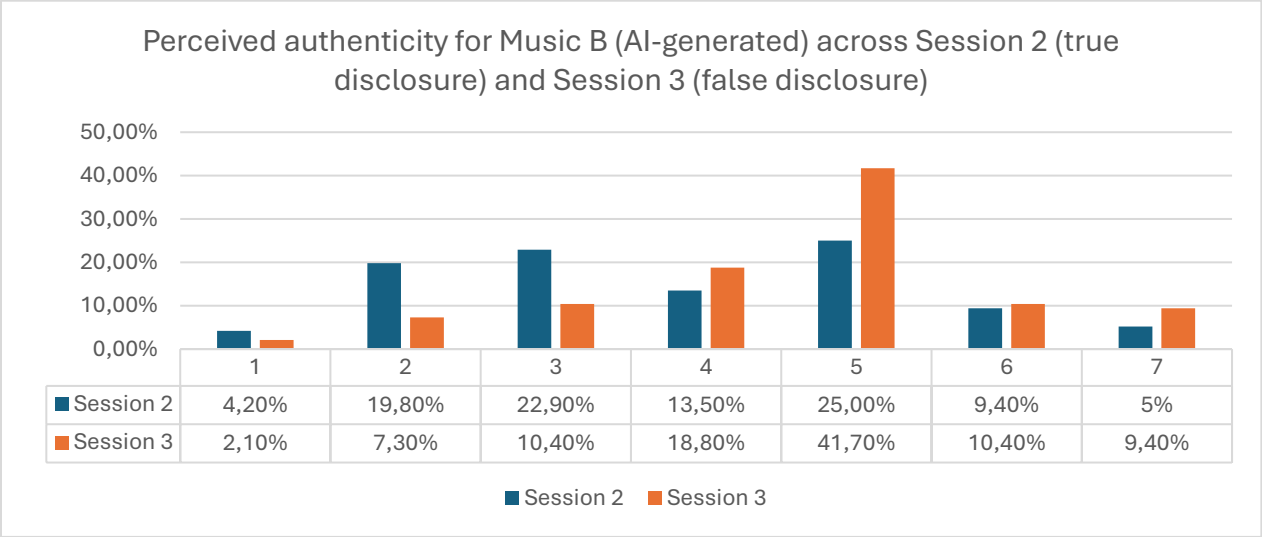

Figure 7: Perceived authenticity for Music B (AI-generated) across Session 2 (true disclosure) and Session 3 (false disclosure) on a 7 points Likert scale

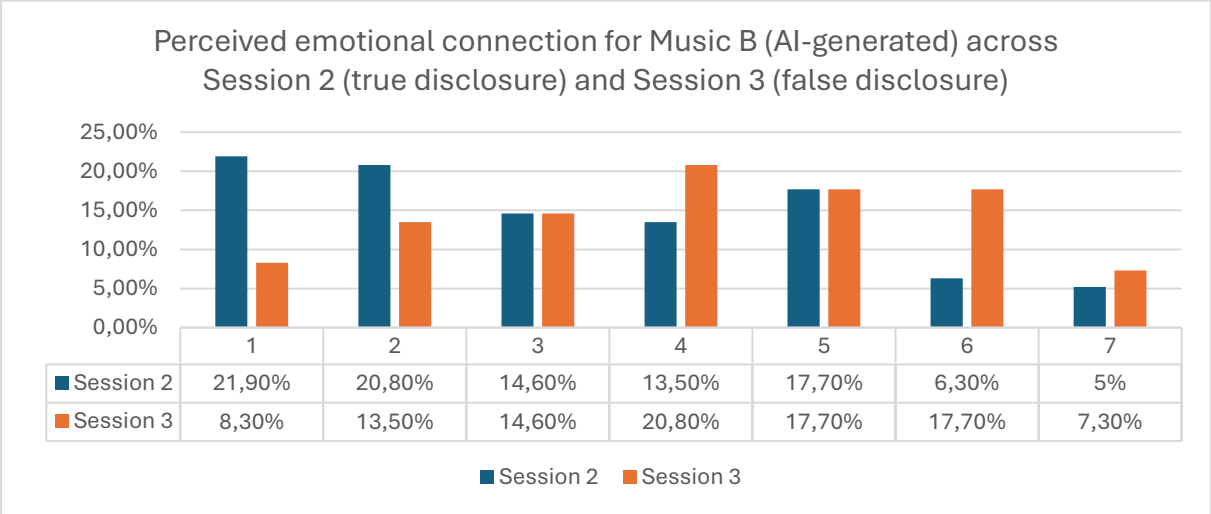

Figure 8: Perceived emotional connection for Music B (AI-generated) across Session 2 (true disclosure) and Session 3 (false disclosure) on a 7 points Likert scale

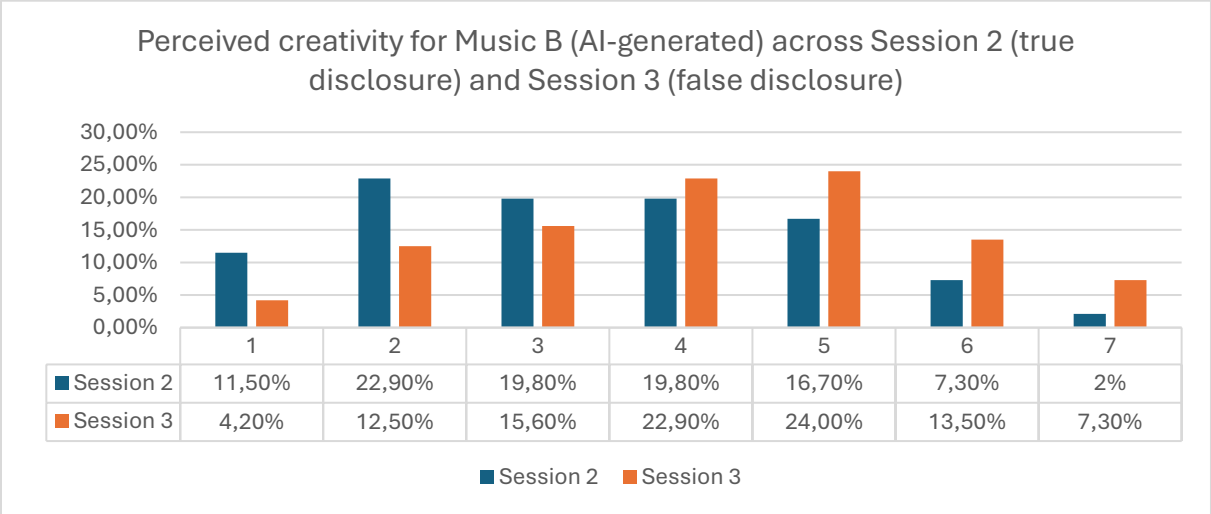

Figure 9: Perceived creativity for Music B (AI-generated) across Session 2 (true disclosure) and Session 3 (false disclosure) on a 7 points Likert scale

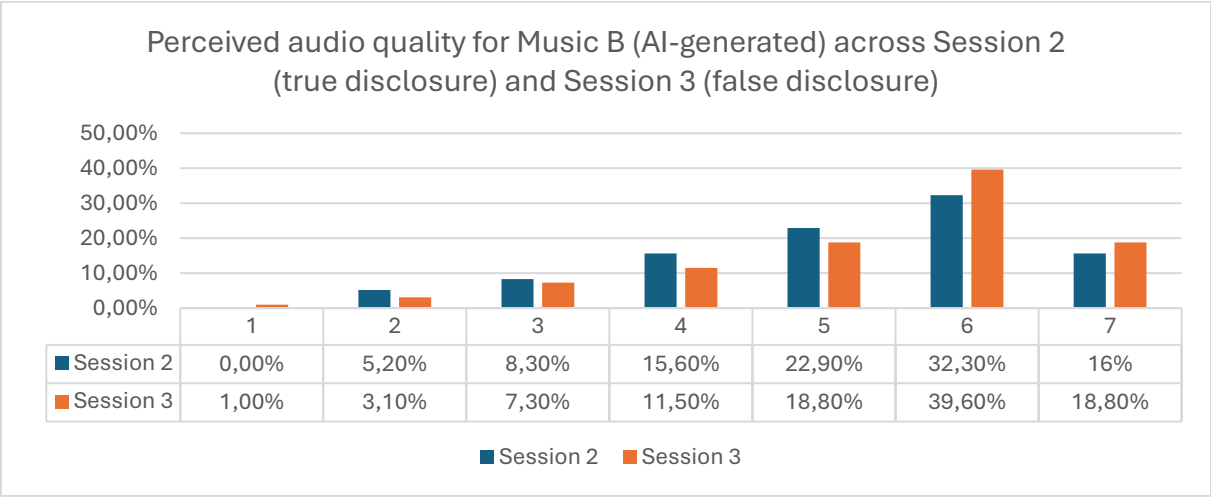

Figure 10: Perceived audio quality for Music B (AI-generated) across Session 2 (true disclosure) and Session 3 (false disclosure) on a 7 points Likert scale

Overall, the data shows that while some participants acknowledge the technical quality or potential utility of AI in music, the prevailing reaction includes scepticism, discomfort, or outright rejection, particularly tied to authenticity, authorship, and emotional value.

4.3. Statistics analysis

All statistical tests were performed using Stata. The output from the tests used to draw statistical conclusions can be found in Appendix III, page 60.

The statistical analyses using both paired t-tests and Wilcoxon signed-rank tests led to consistent results. Because normality and homoscedasticity could not be guaranteed, the Wilcoxon tests were used alongside the t-tests to confirm the robustness of the results. Both methods confirmed that differences between sessions were statistically significant in key comparisons, especially when the label given to the music changed the perception of its origin.

Because the study relies on multiple paired comparisons across the three experimental framing conditions (blind, true disclosure, and false disclosure), a Bonferroni correction was applied to control for the inflation of Type I error associated with repeated statistical testing. Although the design is based on paired observations, conducting multiple comparisons increases the probability of obtaining significant results by chance alone. Since three pairwise comparisons were performed between conditions (S1 vs. S2, S1 vs. S3, and S2 vs. S3), the significance threshold was adjusted according to the Bonferroni procedure: adjusted $\alpha = 0.05/3 = 0.0167$

Therefore, statistical results were considered significant only when p-value < 0.0167. This approach is consistent with previous studies using repeated-measures or paired comparison methodologies in music perception research, where Bonferroni-adjusted pairwise comparisons were applied to reduce family-wise Type I error rates.

In addition to statistical significance testing, effect sizes were systematically calculated using Cohen's d in order to assess the magnitude of the observed differences between experimental conditions. Cohen's d is determined by: $d = \frac{t}{\sqrt{n}}$

4.3.1. Statistical Conclusions for music preferences

The Table 1 shows that there is a clear framing effect when music is labelled as either human-made or AI-generated, regardless of its true origin. This effect is strongest in the comparison between Session 2 (true disclosure) and Session 3 (false disclosure)**.** The same piece of music was more appreciated when labelled as human and less appreciated when labelled as AI. This suggests that listeners are heavily influenced by how the music is presented, even if the content is identical.

On the other hand, comparisons between Session 1 (blind) and the disclosure conditions (Session 2 or 3) produced lower effects. Some comparisons were statistically significant

(blind vs. true disclosure), but others were not (blind vs. false disclosure). This indicates that the framing effect is more pronounced when expectations are directly manipulated, rather than when participants listen first without any label.

It should also be noted that the Bonferroni correction would not have changed the statistical results or conclusions in this case.

The observed effect sizes ranged from small ($d = 0.17$) to moderate ($d = 0.55$), suggesting that framing manipulations produced meaningful changes in participants' musical preferences.

Table 1: Summary table of statistical results comparing the mean of preferences of participants between Music A and Music B in Session 1, Session 2 and Session 3

Mean comparison	Variable	t-test (p-value)	Signrank (exact p-value)	Cohen's d (effect size)	Conclusion
Session 1 vs Session 2	Music A	0.0002	0.0004	-0.40 (moderate)	**Significant difference**
Session 1 vs Session 3	Music A	0.1092	0.1796	0.17 (small)	Not significant
Session 2 vs Session 3	Music A	> 0.0001	> 0.0001	0.55 (moderate)	**Significant difference**
Session 1 vs Session 2	Music B	0.0002	0.0004	0.40 (moderate)	**Significant difference**
Session 1 vs Session 3	Music B	0.1092	0.1796	-0.17 (small)	Not significant
Session 2 vs Session 3	Music B	> 0.0001	> 0.0001	-0.55 (moderate)	**Significant difference**

These findings support the hypothesis that the perceived origin of music strongly influences listener preferences. The framing of music as human or AI-made plays a central role in shaping how it is received, even when the music itself remains unchanged. It is also worth mentioning that the actual source of the music does not affect the results. The label has a significant effect on both human-made music and AI-generated music.

H_1.1

- **H_0:** There is no significant difference in preference between tracks labelled as human-composed and those labelled as AI-generated.
- **H_1:** Participants show a statistically significant preference for a track labelled as composed by human artists over another one labelled as AI-generated.

  Conclusion: **Reject H_0.** Preferences were significantly higher for music labelled as human over the ones labelled as AI, regardless the actual origin.

We can therefore confirm with 95% confidence that a music is preferred when it is labelled as having been created by a human, and less preferred when it is labelled as having been generated by AI, regardless of the music's actual origin.

4.3.2. Statistical Conclusions for authenticity, creativity, emotional connection and audio quality ratings

Comparison of results from Session 2 and Session 3 across the four dimensions for Music A:

The Table 2 shows that all four tests conducted on Music A reveal statistically significant differences between the true disclosure (Session 2) and false disclosure (Session 3) conditions. The results consistently show that participants rated the same music higher on all four dimensions (authenticity, emotional connection, creativity, and audio quality) when they were told it was composed by a human, and lower when told it was generated by AI. These findings support the hypothesis that the framing of authorship strongly influences subjective evaluations. While all dimensions show significant effects, the difference for audio quality is comparatively smaller. A plausible explanation for this weaker effect is that audio quality, being more technical and less emotionally or socially loaded, may be less susceptible to cognitive framing biases than more interpretive dimensions like authenticity, creativity or emotional impact.

Effect size analyses revealed that the framing manipulation had the strongest impact on perceived authenticity ($d = 0.76$) and emotional connection ($d = 0.64$), followed by creativity ($d = 0.52$). In contrast, audio quality showed a smaller effect size ($d = 0.38$). This pattern suggests that disclosure primarily affected symbolic and emotional evaluations of the music rather than purely technical auditory characteristics.

Table 2: Summary table of statistical results comparing the mean of the ratings between Session 2 (true disclosure) and Session 3 (false disclosure) for Music A.

Dimension	Mean session 2	Mean session 3	p-value t-test	p-value Wilcoxon	Conclusion	Cohen's d (effect size)
Authenticity	5.56	4.33	> 0.0001	> 0.0001	**Significant difference**	0.76 (large)
Emotional connection	5.33	4.37	> 0.0001	> 0.0001	**Significant difference**	0.64 (moderate)
Creativity	5.08	4.28	> 0.0001	> 0.0001	**Significant difference**	0.52 (moderate)
Audio quality	5.6	5.2	0.0004	0.0003	**Significant difference**	0.38 (moderate)

Comparison of results from Session 2 and Session 3 across the four dimensions for Music B:

Table 3 shows that the results for Music B closely mirror those observed for Music A. Across all four evaluated dimensions (authenticity, emotional connection, creativity, and audio quality), there is a statistically significant difference between the blind condition (session 2) and the human-labelled condition (session 3). In each case, the dimension receives higher ratings when the music is labelled as human made. This pattern strongly supports the presence of a framing effect influencing participants' evaluations. As with Music A, the difference is smaller for the audio quality dimension, suggesting that more technical aspects may be less sensitive to framing compared to more subjective dimensions such as emotional impact or perceived authenticity.

For Music B, the effect size analyses revealed a similar pattern across the evaluated dimensions. Moderate effects were observed for authenticity ($d = 0.55$), emotional connection ($d = 0.65$), and creativity ($d = 0.52$), indicating that the framing manipulation substantially influenced participants' subjective evaluations of the musical excerpt. In contrast, audio quality produced only a small effect size ($d = 0.21$), suggesting that participants' perception of the technical sound quality remained comparatively stable across disclosure conditions.

These findings reinforce the idea that framing effects associated with AI-generated music primarily operate at a symbolic and emotional level rather than at a purely acoustic level. Once the musical piece was presented under a different informational context, participants appeared to reassess its emotional value, authenticity, and perceived creativity more strongly than its objective sonic characteristics. This further supports the interpretation that disclosure affects more higher-order evaluative dimensions rather than listeners' perception of the intrinsic audio properties of the music.

Table 3: Summary table of statistical results comparing the mean of the ratings between Session 2 (true disclosure) and Session 3 (false disclosure) for Music B.

Dimension	Mean session 2	Mean session 3	p-value t-test	p-value Wilcoxon	Conclusion	Cohen's d (effect size)
Authenticity	3.81	4.6	> 0.0001	> 0.0001	**Significant difference**	-0.55 (moderate)
Emotional connection	3.22	4.08	> 0.0001	> 0.001	**Significant difference**	-0.65 (moderate)
Creativity	3.38	4.2	> 0.0001	> 0.0001	**Significant difference**	-0.52 (moderate)
Audio quality	5.18	5.4	0.0449	0.0299	**Significant difference**	-0.21 (small)

H_1.2

- **H_0:** There is no significant difference in ratings for authenticity, creativity, emotional connection or audio quality between music labelled as human and music labelled as AI-generated.

- **H_1:** Scores for perceived authenticity, creativity, emotional connection and audio quality are significantly higher when participants are told the music was created by a human.

  Conclusion: **Reject H_0.** Ratings (authenticity, creativity, emotional connection and audio quality) were consistently higher under human labelling.

We can therefore confirm with 95% confidence that the aspects of authenticity, emotional connection, creativity, and audio quality are perceived more favourably when music is labelled as having been created by a human, and less favourably when it is labelled as having been generated by AI, regardless of the music's actual origin.

An additional observation worth considering is whether any dimension (such as authenticity, creativity, or emotional connection) of a musical piece labelled as human is consistently rated more favourably than any other dimension of a piece labelled as AI, regardless of whether the music was actually composed by a human or by AI. The "Audio Quality" dimension was intentionally excluded from this test because of its technical and more objective nature, which is not present in the other categories.

For this purpose, we compared the lowest average score for the most preferred dimensions (Session 2 for Music A and Session 3 for Music B) with the highest average score for the least preferred dimensions (Session 3 for Music A and Session 2 for Music B). If both tests are significant, the hypothesis is confirmed.

The results presented in Tables 4 and 5 show that for Music A, the comparison between the lowest-rated dimension of the human-labelled condition (Creativity in Session 2) and the highest-rated dimension of the AI-labelled condition (Emotional Connection in Session 3) revealed a statistically significant difference, with a small-to-moderate effect size. This suggests that even the least favourably evaluated subjective dimension of the human-labelled music remained significantly higher than the most positively evaluated dimension of the AI-labelled version.

In contrast, the same comparison conducted for Music B did not produce a statistically significant result, and the observed effect size remained small. This indicates that the superiority of the human-labelled condition was not consistently replicated across both musical excerpts. As a result, the hypothesis can only be partially supported. Overall, these findings suggest that the framing effect favouring human-labelled music may exist, but its intensity appears to vary depending on the musical piece and the perceptual dimension being evaluated.

Table 4: Summary table of statistical results comparing the ratings between Creativity in Session 2 (true disclosure) and Emotional Connection in Session 3 (false disclosure) for Music A.

Mean Creativity Music A session 2	Mean Emotional Connection Music A session 3	p-value t-test	p-value Wilcoxon	Conclusion	Cohen's d (effect size)
5.08	4.37	0.0008	0.0023	**Significant difference**	0.36 (small-moderate)

Table 5: Summary table of statistical results comparing the ratings between Authenticity in Session 2 (true disclosure) between Emotional Connection in Session 3 (false disclosure) for Music B.

Mean Authenticity Music B session 2	Mean Emotional Connection Music B session 3	p-value t-test	p-value Wilcoxon	Conclusion	Cohen's d (effect size)
3.81	4.08	0.1324	0.2336	Not significant	-0.16 (small)

H_1.3

- **H_0:** There is no consistent difference in scoring patterns between human-labelled and AI-labelled music across subjective dimensions.
- **H_1:** Human-labelled music consistently receives higher scores than AI-labelled music, across all subjective dimensions.

  Conclusion: **No reject of H_0.** The effect of human versus AI labelling is not systematic and may vary across dimensions or tracks.

We cannot therefore confirm with 95% confidence that any subjective aspect (authenticity, creativity, emotional connection) of music labelled as human is preferred over any other aspect of music labelled as AI.

5. Limitations and constraints

The most important limitation of this study is its use of only one human-composed piece and one AI-generated piece as musical stimuli. This restricts the scope of the conclusions, as participants' perceptions could be shaped by characteristics specific to those two tracks. Ideally, the experiment would have included a broader selection of compositions, such as five tracks composed by humans and five generated by AI. However, due to resource and budget constraints, this was not feasible. The aim was not to generalize to the Belgian population but to test a focused experimental framework within a specific musical community. Future research could expand the dataset with a larger, randomized sample through recruitment panels with quotas. This would allow for

more advanced statistical models such as the Bradley-Terry model and the inclusion of multiple musicians and AI-generated tracks using higher-performance tools. In this study, only one AI track was used to ensure consistency and allow a clear comparison of matched observations in an A versus B format.

Another constraint was the use of Google Forms to administer the survey. Although it offered the best available free option, it limited control over participants' listening environments. Variables such as audio quality, device type, and background noise could not be standardized. Additionally, participant attention could not be verified. A custom-built platform would have provided more advanced features, such as randomization of track order, but exceeded the available resources.

All participants went through the same sequence of sessions: blind, true disclosure, and false disclosure. This fixed order could introduce potential order effects, anchoring effects, or learning effects, as earlier sessions may have influenced responses in later stages of the experiment. In experimental psychology, randomizing the order of presentation is generally recommended to reduce this type of bias. However, the present study was designed as an exploratory experiment with a fixed and controlled progression intended to expose all participants to the same informational sequence. Although an order effect cannot be completely excluded, it is unlikely to fully explain the observed results. The core objective of the study was not to determine which musical excerpt was objectively superior, but rather to examine whether participants' preferences changed when the labels associated with the music were manipulated. Importantly, despite Music A always being presented first, statistically significant differences in mean preferences were still observed across conditions following the disclosure manipulations. Future studies could improve the robustness of the design by randomizing the order of sessions and musical excerpts across participants.

Participants only listened to two short audio clips, each about 20 to 30 seconds long. This format was selected to maintain participant focus throughout the online experiment, but it does not reflect typical listening behaviour in natural settings. The study was also conducted as a single-session experiment, without repeated or longitudinal exposure. This limits the ability to observe how familiarity or repeated listening might affect preferences.

Another limitation concerns the final attitudinal question asking participants about their general opinion of AI-generated music. Because this question was presented at the end of the experiment, after participants had already been exposed to the label manipulations, a potential post-experimental contamination effect cannot be excluded. In other words, participants' responses may have been partially influenced by the experimental procedure itself rather than exclusively reflecting their pre-existing attitudes toward AI-generated music. Consequently, the measured attitudes could represent, at least in part, a consequence of the framing manipulation. Nevertheless, even considering

this limitation, the overall perception of AI-generated music within this specific niche audience remained relatively negative. Furthermore, the significant changes observed in participants' preferences following the disclosure manipulations suggest the existence of a certain level of resistance or scepticism toward AI involvement in musical creation. This reinforces the interpretation that the informational framing associated with AI-generated music can influence listeners' evaluations beyond the intrinsic characteristics of the music itself.

These limitations should be considered when interpreting the findings. While they do not undermine the validity of the experimental design, they highlight several areas for improvement in future studies, particularly regarding randomization procedures, ecological validity, and the diversification of musical stimuli.

6. Conclusion

This thesis explored how the disclosure of musical authorship influences listeners' perceptions and preferences toward AI-generated and human-composed music within a specific artistic and indie-pop-rock community based in Brussels. Through a within-subject experimental design involving blind, true disclosure, and false disclosure conditions, the study examined how informational framing affects musical preference as well as perceived authenticity, creativity, emotional connection, and audio quality.

The results provide evidence supporting the existence of a framing effect in the evaluation of music. Across several comparisons, the same musical excerpts received significantly different evaluations depending on whether participants believed the music was created by a human or generated by artificial intelligence. Human-labelled music generally obtained higher preference rates and more positive evaluations on authenticity, creativity, and emotional connection, while AI-labelled music tended to receive lower scores on these subjective dimensions.

Importantly, the effect size analyses showed that framing effects were strongest for symbolic and emotional dimensions such as authenticity and emotional connection, while the impact on audio quality remained comparatively weaker. This distinction suggests that participants were not necessarily reacting to purely technical sonic properties, but rather to the symbolic meaning attached to human versus artificial creation. The findings therefore reinforce the idea that perceptions of AI-generated music are strongly influenced by contextual beliefs regarding creativity, intentionality, and artistic legitimacy.

At the same time, the results also revealed important nuances. Participants were not consistently able to reject AI-generated music under blind conditions, and the AI-generated track still obtained substantial levels of appreciation before disclosure manipulations occurred. In several cases, participants expressed surprise after

discovering the actual origin of the excerpts. This suggests that the rejection of AI-generated music may stem less from intrinsic musical inferiority than from broader cognitive, emotional, and cultural perceptions associated with artificial creativity.

However, the conclusions of this study must be interpreted with caution. The research relies on a highly specific convenience sample composed primarily of indie-pop-rock listeners and artistically involved individuals from Brussels. The objective of the study was not to statistically represent the Belgian population or general music consumers, but rather to conduct an exploratory investigation within a niche community particularly sensitive to questions of artistic authenticity and musical creation. Consequently, the findings cannot be generalized beyond this specific context.

Additionally, the experiment was conducted using only one human-composed track and one AI-generated track, which limits the ability to separate framing effects from track-specific characteristics. The fixed order of the sessions may also have introduced sequential or order-related biases, although the observed framing effects remained statistically significant despite these constraints. Furthermore, the attitudinal question regarding AI-generated music was administered after the experimental manipulations, meaning that participants' responses may have been partially influenced by the experience itself.

Despite these limitations, the study contributes to the growing literature on AI-generated art by providing empirical evidence that listener evaluations are shaped not only by musical content itself, but also by beliefs surrounding authorship and human creativity. More broadly, the findings highlight how artificial intelligence challenges traditional conceptions of authenticity, emotional sincerity, and artistic legitimacy in music consumption.

Rather than demonstrating whether AI-generated music is objectively better or worse than human-created music, this dissertation shows that the perceived origin of a musical work can significantly influence how listeners emotionally and symbolically interpret it. In this sense, the research contributes to a broader discussion about the evolving relationship between technology, creativity, and cultural perception in contemporary music environments.

7. Next steps

This study represents an exploratory effort to understand how listeners respond to AI-generated music and how labelling influences perception. While it provides important insights, several avenues can be pursued to build on this foundation.

A first step would be to expand the dataset. The current study used only one human-composed track and one AI-generated track, which limits the generalizability of the findings. Future research could include a broader range of musical genres, styles, and excerpts. A balanced sample of at least five compositions per origin would help control for track-specific variation and support stronger statistical inferences. The initial choice of just one excerpt per type was intentional to ensure pairwise comparison logic and limit variability, but scaling up would allow broader conclusions to be drawn.

Improving the experimental design is another priority. The current use of Google Forms was driven by resource limitations. While it allowed for basic survey implementation, it lacked the ability to control for sound quality, participant environment, and response conditions. A future study could use a purpose-built platform that enables randomization of session order, better sound control, and behavioural tracking. Such an upgrade would also address the current design's fixed session sequence, which may have introduced order effects.

Recruitment improvements could further enhance the quality of findings. Instead of relying on a convenience sample, future studies could use randomized sampling and demographic quotas to reach a more representative population. This would enable researchers to compare how different subgroups respond to AI-generated music and framing. Future research should also investigate longitudinal changes in listener attitudes toward AI-generated music as exposure to generative systems increases over time. Cross-cultural studies would provide valuable insight into how cultural norms and artistic traditions shape perceptions of authenticity, creativity, and emotional legitimacy in music. It would also be particularly interesting to replicate this experiment with non-artistic or non-musician populations in order to compare whether framing effects differ between creative communities and more general audiences. Going further, future research could compare several artistic communities, multiple cultures, and various musical genres simultaneously in order to better understand how socio-cultural environments influence the acceptance or rejection of AI-generated music.

With a larger dataset, more powerful statistical models could be implemented. Advanced techniques such as the Bradley-Terry model or linear mixed-effects models would enable a more nuanced analysis of listener preferences and perceptions while accounting for participant and track-level variability. Additionally, future work could include longitudinal or repeated exposure to assess whether framing effects persist or evolve over time. Further investigation into objective measures of musical quality would also strengthen

the field by clarifying which perceptual and acoustic dimensions most strongly influence listener judgments. Future studies could integrate musicological and computational approaches involving tonal structure, harmonic progression, rhythmic complexity, timbre analysis, loudness, and other acoustic descriptors in order to better distinguish between subjective framing effects and intrinsic musical characteristics.

The study could also extend beyond music. By testing AI-generated works in other creative domains such as visual art, literature, or film, researchers could explore whether similar framing effects occur across artistic media. Within music, future experiments could assess how co-creation and transparency affect perception, such as by testing reactions to pieces labelled as "human-AI collaboration." Additional research is also needed to distinguish more clearly between fully AI-generated music and collaborative human-AI creative processes. As generative systems increasingly become tools assisting artists rather than replacing them entirely, future studies should explore how different levels of human involvement affect perceived authenticity, creativity, and emotional engagement.

A further expansion could involve physiological and behavioural measures to complement survey responses. Metrics such as listening duration, replay frequency, or biometric data like eye tracking or heart rate could provide deeper insight into emotional engagement and cognitive processing. Future work could therefore integrate biometric, behavioural, qualitative, and objective acoustic methodologies simultaneously in order to develop a more multidimensional understanding of how listeners emotionally and cognitively engage with AI-generated music.

If this research were to be extended into a larger-scale project, a preregistration process on the Open Science Framework (“**OSF**”) would also be implemented prior to data collection. Following an OSF preregistration, future studies could establish predefined hypotheses, analysis plans, exclusion criteria, and statistical procedures before launching the experiment. Such an approach would strengthen transparency, reproducibility, and methodological rigor while reducing risks associated with analytical flexibility and post hoc interpretation.

This initial research served as a first step to understand how one specific community responds to AI-generated music and confirmed the existence of a framing effect. The experimental nature of the study was shaped by feasibility constraints, but the findings lay the groundwork for more ambitious work. Broader studies with more tracks, participants, and resources could generalize insights to a national or even global scale. In addition to academic inquiry, these results could inform applied strategies in music marketing and branding. Future campaigns could explore how to best communicate the origin of musical content, whether for artists, labels, or advertisers integrating AI-generated music in jingles and soundtracks. These directions show the real-world relevance and ongoing potential of this field.

8. Bibliography

Bagratuni, M., Müller, P., & Planing, P. (2025). Innovation in tune : An empirical investigation of user acceptance of artificial intelligence-generated music. *Computers In Human Behavior Reports*, *18*, 100660. https://doi.org/10.1016/j.chbr.2025.100660

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots : Can Language Models Be Too Big ? *ACM Digital Library*, 610-623. https://doi.org/10.1145/3442188.3445922

Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs : I. The Method of Paired Comparisons. *Biometrika*, *39*(3/4), 324. https://doi.org/10.2307/2334029

Chia, S., Hartanto, A., & Tong, E. M. (2025). Do listeners devalue AI-generated pop music ? Exploring negative biases in listeners' responses to AI-labelled vs human-labelled pop music. *Computers In Human Behavior Artificial Humans*, *6*, 100217. https://doi.org/10.1016/j.chbah.2025.100217

Figueiredo, F., Martinelli, G., Sousa, H., Rodrigues, P., Pedrosa, F., & Ferreira, L. N. (2025). Echoes of Humanity : Exploring the Perceived Humanness of AI Music. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2509.25601

Fišer, N., Martín-Pascual, M. Á., & Andreu-Sánchez, C. (2025). Emotional impact of AI-generated vs. human-composed music in audiovisual media : A biometric and self-report study. *PLoS ONE*, *20*(6), e0326498. https://doi.org/10.1371/journal.pone.0326498

Ginsburgh, V. A., & Van Ours, J. C. (2003). Expert opinion and compensation : Evidence from a musical competition. *The American Economic Review*, *93*(1), 289-296. https://www.jstor.org/stable/3132175

Goodfellow, I., Bengio, Y., & Courville, A. (2015). *Deep learning, Cambridge, Massachusetts : MIT*. https://www.deeplearningbook.org/

Grayson, K., & Martinec, R. (2004). Consumer Perceptions of Iconicity and Indexicality and Their Influence on Assessments of Authentic Market Offerings. *Journal Of Consumer Research*, *31*(2), 296-312. https://doi.org/10.1086/422109

Hong, J., Fischer, K., Ha, Y., & Zeng, Y. (2022). Human, I wrote a song for you : An experiment testing the influence of machines' attributes on the AI-composed music evaluation. *Computers In Human Behavior*, *131*, 107239. https://doi.org/10.1016/j.chb.2022.107239

Horton, C. B., Jr, White, M. W., & Iyengar, S. S. (2023). Bias against AI art can enhance perceptions of human creativity. *Scientific Reports*, *13*(1), 19001. https://doi.org/10.1038/s41598-023-45202-3

Lecamwasam, K., & Chaudhuri, T. R. (2025). Exploring listeners' perceptions of AI-generated and human-composed music for functional emotional applications. *ArXiv.org*. https://doi.org/10.48550/arxiv.2506.02856

Merz, J. (2025). Artificial Creativity and Human Fragility. *Modern Theology*. https://doi.org/10.1111/moth.70048

Nguyen, T., Sidhu, R. K., Everling, J. C., Wickett, M. C., Gibbings, A., & Grahn, J. A. (2022). Beat Perception and Production in Musicians and Dancers. *Music Perception An Interdisciplinary Journal*, *39*(3), 229-248. https://doi.org/10.1525/mp.2022.39.3.229

Pearce, M. T. (2023). Music perception. *Oxford Research Encyclopedia Of Psychology*. https://doi.org/10.1093/acrefore/9780190236557.013.890

Quin, F., Weyns, D., Galster, M., & Silva, C. C. (2024). A/B testing : A systematic literature review. *Journal Of Systems And Software*, *211*, 112011. https://doi.org/10.1016/j.jss.2024.112011

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI*. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Rigole, N., Sandoval, Z., Beasley, S., Lingelbach, K., Atif, U., Beato, R., Beato, R., Beato, R., Chan, J., Lee, S., Kim, H., Clarivate, Coeckelbergh, M., Colton, S., Cook, M., Raad, A., Culliton, J., Fernando, P., Mahanama, T.,. . . Zavada, I. (2025). Perceptions of AI-generated and co-created music among university students and social media users. *Issues In Information Systems*. https://doi.org/10.48009/4_iis_2025_110

Samo, A., & Highhouse, S. (2023). Artificial intelligence and art : Identifying the aesthetic judgment factors that distinguish human- and machine-generated artwork. *Psychology Of Aesthetics Creativity And The Arts*, *19*(5), 1084-1098. https://doi.org/10.1037/aca0000570

Sarmento, P., Loth, J., & Barthet, M. (2024). Between the AI and Me : Analysing Listeners' Perspectives on AI- and Human-Composed Progressive Metal Music. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2407.21615

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal Of Business*, *59*(4), 251-278. https://www.jstor.org/stable/2352759

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*. https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

Wilson, A., & Fazenda, B. (2016). Perception of Audio Quality in Productions of Popular Music. *Journal Of The Audio Engineering Society*, *64*(1/2), 23-34. https://doi.org/10.17743/jaes.2015.0090

9. Appendices

I. Google Forms survey, **page 42**
II. Answers to “What do you think of AI-generated music?”, **page 56**
III. Stata results printouts, **page 60**

Appendix I. Google Forms survey

The human touch and the algorithmic mind: A study on the perception of music - Listener Experience Survey

Hello and welcome,

My name is Anass Housni and I am currently completing my final year of the Master's program in Management at the Louvain School of Management (UCLouvain). This survey is part of my master's thesis, which explores how listeners perceive music when it is composed by artificial intelligence (AI) versus human musicians.

In this study, you will be invited to listen to a few short music clips and indicate which ones you prefer. You will not need any prior musical knowledge, all that's required is your opinion as a music listener. The listening conditions may vary slightly from one session to another, and we ask that you respond instinctively and honestly.

The entire study should take around **5 to 7 minutes**, and I kindly ask that you complete it in a **quiet environment** or with **headphones (if possible)** for the best listening quality. Your participation is anonymous, and there are no right or wrong answers. I am simply interested in how people experience different types of musical content today.

Thank you very much for your time and support. Let's begin with a few quick questions about you.

Disclaimer: Your data will only be used for the purposes of this academic study and will **not** be shared or used for any other reason. All responses will be **anonymized and permanently deleted** once the research is completed.

* Indique une question obligatoire

Section 1 - About You (Demographics)

1. 1. Age *

Une seule réponse possible.

- ◯ Under 18 (not eligible)
- ◯ 18-24
- ◯ 25-34
- ◯ 35-44
- ◯ 45+

Appendix I. Google Forms survey

2. 2. Gender *

Une seule réponse possible.

- ◯ Woman
- ◯ Man
- ◯ Non-binary/other
- ◯ Prefer not to say

3. 3. Occupation *

Une seule réponse possible.

- ◯ Student
- ◯ Employed
- ◯ Unemployed
- ◯ Self-employed
- ◯ Other

Appendix I. Google Forms survey

4. 4. Zip code *

Une seule réponse possible.

- Anderlecht - 1070
- Auderghem - 1160
- Berchem-Saint-Agathe - 1082
- Bruxelles Ville - 1000
- Laeken - 1020
- Neder-over-Heembeek - 1120
- Haren - 1130
- Etterbeek - 1040
- Evere - 1140
- Forest - 1190
- Ganshoren - 1083
- Ixelles - 1050
- Jette - 1090
- Koekelberg - 1081
- Molenbeek-Saint-Jean - 1080
- Saint-Gilles - 1060
- Saint-Josse-Ten-Node - 1210
- Schaerbeek - 1030
- Uccle - 1180
- Watermael-Boitsfort - 1170
- Woluwé-Saint-Lambert - 1200
- Woluwé-Saint-Pierre - 1150

Appendix I. Google Forms survey

5. 5. How often do you listen to music? *

Une seule réponse possible.

- ◯ Daily
- ◯ Several times a week
- ◯ Occasionally (Once a week or less)
- ◯ Rarely (Once a month or less)

6. 6. What genres of music do you enjoy most? *(Multiple selections allowed)* *

Plusieurs réponses possibles.

- ☐ Indie
- ☐ Rock
- ☐ Pop
- ☐ Electronic
- ☐ Classical
- ☐ Jazz
- ☐ Disco/funk
- ☐ Hip-hop/rap
- ☐ Autre : _____

Please make sure your sound is on, and click "Next" to begin the first session.

Main Experiment - Listening and Preferences

In the following sections, you will go through **three short listening sessions**. Each session will present **two short music clips** (about 20-30 seconds each), labeled "Music A" and "Music B".

After listening to both excerpts, you will be asked to choose the one you personally prefer. There are no correct answers, we are simply interested in your personal taste and impressions. Please try to listen to both excerpts carefully before making your choice. Please respond instinctively and honestly.

Appendix I. Google Forms survey

Session 1

Please listen to both music clips and select the one you prefer.

The audio mixes slightly vary, so please adjust the volume to a comfortable listening level for each music clip.

Music A

http://youtube.com/watch?v=ISP3X249eB4

Music B

http://youtube.com/watch?v=GiguovA2VLE

7. Which one did you prefer? (session 1) *

Une seule réponse possible.

- ◯ Music A
- ◯ Music B

Appendix I. Google Forms survey

Please take a 20 seconds break to reset your ears before moving on to the second session.

Once you are ready, click "Next" to begin the second session.

Session 2

In this session, please listen to both music clips and select the one you prefer.

This time, I'm going to reveal that one of the tracks was **generated by artificial intelligence (AI)** and the other one was **composed by a human.**

In addition, you will need to rate the level of **authenticity, quality, creativity,** and **emotional connection,** based on your opinion. Please see some short definitions:

🎵 **Authenticity**

How genuine or sincere the music feels to you - does it seem personal, meaningful, or true to a real artistic intention?

❤️ **Emotional Connection**

How much the music makes you feel something - does it move you, touch you, or evoke any emotion?

💡 **Creativity**

How original or inventive the music seems - does it sound fresh, unique, or imaginative?

🎧 **Audio Quality**

How clear and pleasant the sound is - is it well-produced, clean, and comfortable to listen to?

Music A - Composed by a human

http://youtube.com/watch?v=ISP3X249eB4

Appendix I. Google Forms survey

Music B - AI generated

http://youtube.com/watch?v=GiguovA2VLE

8. Which one did you prefer? (session 2) *

Une seule réponse possible.

- Music A
- Music B

Please rate the **MUSIC A**
Composed by a human

9. For the Music A, please rate **Authenticity** 🎵 *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

10. For the Music A, please rate **Emotional Connection** ❤️ *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

Appendix I. Google Forms survey

11. For the Music A, please rate **Creativity** 🤔 *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

12. For the Music A, please rate **Audio Quality** 🎧 *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

Please rate the **MUSIC B**
AI generated

13. For the Music B, please rate **Authenticity** 🎵 *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

14. For the Music B, please rate **Emotional Connection** ❤️ *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

Appendix I. Google Forms survey

15. For the Music B, please rate **Creativity** 🧠 *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

16. For the Music B, please rate **Audio Quality** 🎧 *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

Please take a 20 seconds break to reset your ears before moving on to the second session.

Once you are ready, click "Next" to begin the last session.

Appendix I. Google Forms survey

Session 3 (last one)

This session is similar to the previous one, please repeat the exercise.

This time, I'm going to reveal that I actually lied in the previous session... The **Music A was AI generated** and the **Music B was composed by a human.**

Please don't go back to modify your previous answers.

Please see some short definitions:

🎵 **Authenticity**

How genuine or sincere the music feels to you - does it seem personal, meaningful, or true to a real artistic intention?

❤️ **Emotional Connection**

How much the music makes you feel something - does it move you, touch you, or evoke any emotion?

🎨 **Creativity**

How original or inventive the music seems - does it sound fresh, unique, or imaginative?

🎧 **Audio Quality**

How clear and pleasant the sound is - is it well-produced, clean, and comfortable to listen to?

Music A - AI generated

http://youtube.com/watch?v=ISP3X249eB4

Appendix I. Google Forms survey

Music B - Composed by a human

http://youtube.com/watch?v=GiguovA2VLE

17. Which one did you prefer? (session 3) *

Une seule réponse possible.

- ◯ Music A
- ◯ Music B

Please rate the **MUSIC A**
AI generated

18. For the Music A, please rate **Authenticity** 🎵 (session 3) *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

19. For the Music A, please rate **Emotional Connection** ❤️ (session 3) *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

Appendix I. Google Forms survey

20. For the Music A, please rate **Creativity** 🤔 (session 3) *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

21. For the Music A, please rate **Audio Quality** 🎧 (session 3) *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

Please rate the **MUSIC B**
Composed by a human

22. For the Music B, please rate **Authenticity** 🎵 (session 3) *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

23. For the Music B, please rate **Emotional Connection** ❤️ (session 3) *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

Appendix I. Google Forms survey

24. For the Music B, please rate **Creativity** (session 3) *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

25. For the Music B, please rate **Audio Quality** (session 3) *

Une seule réponse possible.

	1	2	3	4	5	6	7	
Very	◯	◯	◯	◯	◯	◯	◯	Very good

Click "Next" to conclude the experiment.

Last questions

26. What do you think of AI-generated music? *

Plusieurs réponses possibles.

- [ ] I like it
- [ ] I don't like it
- [ ] I have no opinion
- [ ] I don't mind it as long as the result sounds good
- [ ] I can't really tell the difference
- [ ] Autre : _____

Appendix I. Google Forms survey

27. Would you listen to music if you knew it was generated by AI? *

Une seule réponse possible.

- ◯ Yes
- ◯ No
- ◯ Maybe

Final message

Thank you for participating in this research!

Your answers have been recorded. If you have any questions or wish to learn more about the study's findings once it concludes, or if you want to know which music was ACTUALLY composed by a human, feel free to contact me at:

anass.housni@student.uclouvain.be

Your help contributes to understanding how people relate to emerging technologies in music creation.

28. If you have any comments or feedback about this form, please let me know!

Appendix I. Google Forms survey

Appendix II. Answers to “What do you think of AI-generated music?”

I don't like it, I was ashamed upon finding out that the track I preferred was AI!
I don't like it
I hate the fact that I like it
I don't like it
I don't like it
I don’t like the idea that IA can do better or same things as humains
I don't like it, I can't really tell the difference
I don't like it
I don't mind it as long as the result sounds good
I like it
I don't like it
It’s not bad quality wise but I wouldn’t listen to it
I don't like it
I don't like it
I would rather someone composing himself a music than generated by Ai.
I have no opinion
I don't mind it as long as the result sounds good
I don't like it
I like it, I don't mind it as long as the result sounds good, I can't really tell the difference
I can't really tell the difference
I don't like it
Depend on the purpose
I don't like it
I don't like it
I can't really tell the difference
I like it, I don't mind it as long as the result sounds good
I don't like it
Can be interesting as a music production tool. As a listener, i don't find it relevant
I don't like it

Appendix II. Answers to “What do you think of AI-generated music?”

I can't really tell the difference

I can't really tell the difference

I don't like it

I don't like it

I like it

C'est vraiment le morceau A qui était généré par IA??? Alors je ne sais visiblement pas dire la différence 😅

I prefer ethical music by human

I don't mind it as long as the result sounds good

I have no opinion

I have no opinion

I like it

I don't mind it as long as the result sounds good

I don't like it

Ça me dérange l'utilisation de l'IA dans ce contexte, pas sure par contre que mon oreille fasse la différence

I don't like it

I don't like it

I don't like it

I hate that I liked it

I don't mind it as long as the result sounds good, Le son généré par IA était un peu particulier. Il faut chercher les accompagnements. Ce que j'entends par là c'est qu'il faut se concentrer pour bien percevoir chacun des sons qui composent l'accompagnement. Le 2e son lui était beaucoup plus clair et s'écoute plus naturellement.

It's no that I don't like it, I don't like the concept of it

I don't like it

I liked it more because it s the type of music that i listen too, i consider it has a nice background music

I like it

Meme si c'est de la bonne musique, je veux pas donner de la force à ca

I have no opinion

I don't like it

I don't like it, I don't like it because it is IA generated

I don't like it

I don't like it, I can't really tell the difference

Appendix II. Answers to "What do you think of AI-generated music?"

I like it, I don't like it
I have no opinion
I’m politically against it, but unfortunaltely, there are some good AI-generated sounds around
I don't like it
I accept it as long as it helps real artist doing stuff
I don't mind it as long as the result sounds good
I can't really tell the difference
I don't like it
I don't like it
I don't like it
I don't like it
I like it, I don't mind it as long as the result sounds good, I can't really tell the difference
I don't like it, I don't mind it as long as the result sounds good
I don't like it, I can't really tell the difference
I have no opinion
I don't like it
I have no opinion
I don't like it, AI in art is a shame
I don't like it, It shouldn’t exist
I don't like it
I don't mind it as long as the result sounds good
I like it, I don't mind it as long as the result sounds good, I can't really tell the difference
I like it
I don't like it
I don't like it
I like it, I can't really tell the difference
I like it
I like it, I can't really tell the difference
I like it, I don't mind it as long as the result sounds good
I don't like it, I can't really tell the difference

Appendix II. Answers to “What do you think of AI-generated music?”

I have no opinion
I don't like it
I don't like it
I don't like it
I can't really tell the difference

Appendix II. Answers to "What do you think of AI-generated music?"

Appendix III. Stata results printouts

To better understand Stata commands and codes used, here is a list of the codes used and the variables they correspond to.

Variable code in Stata	Variable name
Prefer_A_S1	Proportion of preferences for Music A in Session 1
Prefer_A_S2	Proportion of preferences for Music A in Session 2
Prefer_A_S3	Proportion of preferences for Music A in Session 3
Prefer_B_S1	Proportion of preferences for Music B in Session 1
Prefer_B_S2	Proportion of preferences for Music B in Session 2
Prefer_B_S3	Proportion of preferences for Music B in Session 3
Aut_A_S2	Mean of perceived authenticity for Music A in Session 2
Emo_A_S2	Mean of perceived emotional connection for Music A in Session 2
Crea_A_S2	Mean of perceived creativity for Music A in Session 2
Qual_A_S2	Mean of perceived audio quality for Music A in Session 2
Aut_B_S2	Mean of perceived authenticity for Music B in Session 2
Emo_B_S2	Mean of perceived emotional connection for Music B in Session 2
Crea_B_S2	Mean of perceived creativity for Music B in Session 2
Qual_B_S2	Mean of perceived audio quality for Music B in Session 2
Aut_A_S3	Mean of perceived authenticity for Music A in Session 3
Emo_A_S3	Mean of perceived emotional connection for Music A in Session 3
Crea_A_S3	Mean of perceived creativity for Music A in Session 3
Qual_A_S3	Mean of perceived audio quality for Music A in Session 3
Aut_B_S3	Mean of perceived authenticity for Music B in Session 3
Emo_B_S3	Mean of perceived emotional connection for Music B in Session 3
Crea_B_S3	Mean of perceived creativity for Music B in Session 3
Qual_B_S3	Mean of perceived audio quality for Music B in Session 3

Statistical tests for music preferences

For Music A

```
. ttest Prefer_A_S1 == Prefer_A_S2

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Pre~A_S1	93	.6451613	.0498834	.4810577	.5460886	.744234
Pre~A_S2	93	.8172043	.0402953	.3885938	.7371744	.8972342
diff	93	-.172043	.0448978	.432979	-.261214	-.0828721

```
    mean(diff) = mean(Prefer_A_S1 - Prefer_A_S2)                  t =  -3.8319
H0: mean(diff) = 0                                     Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.0001         Pr(|T| > |t|) = 0.0002          Pr(T > t) = 0.9999
```

```
. signrank Prefer_A_S1 = Prefer_A_S2

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	2	167	835
Negative	18	1503	835
Zero	73	2701	2701
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -166.25
Adjustment for zeros   -33087.25
                       ----------
Adjusted variance       34861.25

H0: Prefer_A_S1 = Prefer_A_S2
                 z = -3.578
Prob > |z|        = 0.0003
Exact prob        = 0.0004
```


Appendix III. Stata results printouts

```
. ttest Prefer_A_S1 == Prefer_A_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Pre~A_S1	93	.6451613	.0498834	.4810577	.5460886	.744234
Pre~A_S3	93	.5806452	.0514461	.4961281	.4784688	.6828215
diff	93	.0645161	.0398878	.3846639	-.0147045	.1437367

```
    mean(diff) = mean(Prefer_A_S1 - Prefer_A_S3)                  t =   1.6174
H0: mean(diff) = 0                              Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.9454         Pr(|T| > |t|) = 0.1092          Pr(T > t) = 0.0546
```

```
. signrank Prefer_A_S1 = Prefer_A_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	10	865	605.5
Negative	4	346	605.5
Zero	79	3160	3160
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties       -56.88
Adjustment for zeros   -41870.00
----------
Adjusted variance       26187.88

H0: Prefer_A_S1 = Prefer_A_S3
            z =   1.604
Prob > |z| = 0.1088
Exact prob = 0.1796
```

```
. ttest Prefer_A_S2 == Prefer_A_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Pre~A_S2	93	.8172043	.0402953	.3885938	.7371744	.8972342
Pre~A_S3	93	.5806452	.0514461	.4961281	.4784688	.6828215
diff	93	.2365591	.0443061	.4272727	.1485634	.3245549

```
    mean(diff) = mean(Prefer_A_S2 - Prefer_A_S3)                  t =   5.3392
H0: mean(diff) = 0                              Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

```
. signrank Prefer_A_S2 = Prefer_A_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	22	1815	907.5
Negative	0	0	907.5
Zero	71	2556	2556
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -221.38
Adjustment for zeros   -30459.00
----------
Adjusted variance       37434.38

H0: Prefer_A_S2 = Prefer_A_S3
            z =   4.690
Prob > |z| = 0.0000
Exact prob = 0.0000
```

For Music B

```
. ttest Prefer_B_S1 == Prefer_B_S2

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Pre~B_S1	93	.3548387	.0498834	.4810577	.255766	.4539114
Pre~B_S2	93	.1827957	.0402953	.3885938	.1027658	.2628256
diff	93	.172043	.0448978	.432979	.0828721	.261214

```
    mean(diff) = mean(Prefer_B_S1 - Prefer_B_S2)                  t =   3.8319
H0: mean(diff) = 0                              Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.9999         Pr(|T| > |t|) = 0.0002          Pr(T > t) = 0.0001
```

```
. signrank Prefer_B_S1 = Prefer_B_S2

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	18	1503	835
Negative	2	167	835
Zero	73	2701	2701
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -166.25
Adjustment for zeros   -33087.25
----------
Adjusted variance       34861.25

H0: Prefer_B_S1 = Prefer_B_S2
            z =   3.578
Prob > |z| = 0.0003
Exact prob = 0.0004
```


Appendix III. Stata results printouts

```
. ttest Prefer_B_S1 == Prefer_B_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Pre-B_S1	93	.3548387	.0498834	.4810577	.255766	.4539114
Pre-B_S3	93	.4193548	.0514461	.4961281	.3171785	.5215312
diff	93	-.0645161	.0398878	.3846639	-.1437367	.0147045

```
    mean(diff) = mean(Prefer_B_S1 - Prefer_B_S3)                  t =  -1.6174
H0: mean(diff) = 0                                     Degrees of freedom =     92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.0546         Pr(|T| > |t|) = 0.1092          Pr(T > t) = 0.9454
```

```
. signrank Prefer_B_S1 = Prefer_B_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	4	346	605.5
Negative	10	865	605.5
Zero	79	3160	3160
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties       -56.88
Adjustment for zeros   -41870.00
----------------------------------
Adjusted variance       26187.88

H0: Prefer_B_S1 = Prefer_B_S3
            z = -1.604
Prob > |z|  = 0.1088
Exact prob  = 0.1796
```

```
. ttest Prefer_B_S2 == Prefer_B_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Pre-B_S2	93	.1827957	.0402953	.3885938	.1027658	.2628256
Pre-B_S3	93	.4193548	.0514461	.4961281	.3171785	.5215312
diff	93	-.2365591	.0443061	.4272727	-.3245549	-.1485634

```
    mean(diff) = mean(Prefer_B_S2 - Prefer_B_S3)                  t =  -5.3392
H0: mean(diff) = 0                                     Degrees of freedom =     92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

```
. signrank Prefer_B_S2 = Prefer_B_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	0	0	907.5
Negative	22	1815	907.5
Zero	71	2556	2556
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -221.38
Adjustment for zeros   -30459.00
----------------------------------
Adjusted variance       37434.38

H0: Prefer_B_S2 = Prefer_B_S3
            z = -4.690
Prob > |z|  = 0.0000
Exact prob  = 0.0000
```


Appendix III. Stata results printouts

Statistical tests for authenticity, creativity, emotional connection and audio quality ratings

For Music A

```
. ttest Aut_A_S2 == Aut_A_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Aut_A_S2	93	5.55914	.1283389	1.237655	5.304248	5.814032
Aut_A_S3	93	4.333333	.1765424	1.702513	3.982705	4.683962
diff	93	1.225806	.1668739	1.609274	.8943805	1.557232

```
    mean(diff) = mean(Aut_A_S2 - Aut_A_S3)                        t =   7.3457
H0: mean(diff) = 0                             Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

```
. signrank Aut_A_S2 = Aut_A_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	53	3503	1852.5
Negative	4	202	1852.5
Zero	36	666	666
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -518.62
Adjustment for zeros    -4051.50
                        ---------
Adjusted variance       63544.62

H0: Aut_A_S2 = Aut_A_S3
            z =  6.548
Prob > |z| = 0.0000
Exact prob = 0.0000
```

```
. ttest Emo_A_S2 == Emo_A_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Emo_A_S2	93	5.333333	.1361782	1.313255	5.062872	5.603795
Emo_A_S3	93	4.365591	.1775008	1.711756	4.013059	4.718123
diff	93	.9677419	.157745	1.521238	.6544467	1.281037

```
    mean(diff) = mean(Emo_A_S2 - Emo_A_S3)                        t =   6.1348
H0: mean(diff) = 0                             Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

```
. signrank Emo_A_S2 = Emo_A_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	46	3171	1795.5
Negative	8	420	1795.5
Zero	39	780	780
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -469.62
Adjustment for zeros    -5135.00
                        ---------
Adjusted variance       62510.12

H0: Emo_A_S2 = Emo_A_S3
            z =  5.582
Prob > |z| = 0.0000
Exact prob = 0.0000
```

```
. ttest Crea_A_S2 == Crea_A_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Crea_A~2	93	5.075269	.1394879	1.345173	4.798234	5.352304
Crea_A~3	93	4.27957	.1642169	1.58365	3.953421	4.605719
diff	93	.7956989	.1600309	1.543282	.4778637	1.113534

```
    mean(diff) = mean(Crea_A_S2 - Crea_A_S3)                      t =   4.9722
H0: mean(diff) = 0                             Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

```
. signrank Crea_A_S2 = Crea_A_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	40	2816	1734
Negative	11	652	1734
Zero	42	903	903
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -440.12
Adjustment for zeros    -6396.25
                        ---------
Adjusted variance       61278.38

H0: Crea_A_S2 = Crea_A_S3
            z =  4.371
Prob > |z| = 0.0000
Exact prob = 0.0000
```


Appendix III. Stata results printouts

```
. ttest Qual_A_S2 == Qual_A_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Qual_A~2	93	5.602151	.1185661	1.14341	5.366668	5.837633
Qual_A~3	93	5.204301	.1438786	1.387515	4.918546	5.490056
diff	93	.3978495	.1082607	1.044028	.1828344	.6128646

```
    mean(diff) = mean(Qual_A_S2 - Qual_A_S3)                      t =   3.6749
H0: mean(diff) = 0                              Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.9998         Pr(|T| > |t|) = 0.0004          Pr(T > t) = 0.0002
```

```
. signrank Qual_A_S2 = Qual_A_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	29	2199	1387.5
Negative	8	576	1387.5
Zero	56	1596	1596
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -412.50
Adjustment for zeros   -15029.00
                       ----------
Adjusted variance       52673.25

H0: Qual_A_S2 = Qual_A_S3
            z =   3.536
Prob > |z|    =   0.0004
Exact prob    =   0.0003
```

For Music B

```
. ttest Aut_B_S2 == Aut_B_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Aut_B_S2	93	3.806452	.16271	1.569118	3.483296	4.129608
Aut_B_S3	93	4.602151	.145157	1.399843	4.313856	4.890445
diff	93	-.7956989	.1494568	1.441309	-1.092533	-.4988648

```
    mean(diff) = mean(Aut_B_S2 - Aut_B_S3)                      t =  -5.3239
H0: mean(diff) = 0                              Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

```
. signrank Aut_B_S2 = Aut_B_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	13	709.5	1834
Negative	43	2958.5	1834
Zero	37	703	703
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -662.50
Adjustment for zeros    -4393.75
                       ----------
Adjusted variance       63058.50

H0: Aut_B_S2 = Aut_B_S3
            z =  -4.478
Prob > |z|    =   0.0000
Exact prob    =   0.0000
```

```
. ttest Emo_B_S2 == Emo_B_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Emo_B_S2	93	3.215054	.1852746	1.786723	2.847082	3.583025
Emo_B_S3	93	4.075269	.1810549	1.74603	3.715678	4.434859
diff	93	-.8602151	.1381026	1.331813	-1.134499	-.5859314

```
    mean(diff) = mean(Emo_B_S2 - Emo_B_S3)                      t =  -6.2288
H0: mean(diff) = 0                              Degrees of freedom =       92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

```
. signrank Emo_B_S2 = Emo_B_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	7	395.5	1795.5
Negative	47	3195.5	1795.5
Zero	39	780	780
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -835.50
Adjustment for zeros    -5135.00
                       ----------
Adjusted variance       62144.25

H0: Emo_B_S2 = Emo_B_S3
            z =  -5.616
Prob > |z|    =   0.0000
Exact prob    =   0.0000
```


Appendix III. Stata results printouts

```
. ttest Crea_B_S2 == Crea_B_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Crea_B~2	93	3.376344	.1610174	1.552795	3.05655	3.696138
Crea_B~3	93	4.204301	.1622071	1.564269	3.882144	4.526458
diff	93	-.827957	.163688	1.57855	-1.153055	-.5028586

```
    mean(diff) = mean(Crea_B_S2 - Crea_B_S3)                      t =  -5.0581
H0: mean(diff) = 0                              Degrees of freedom =       92

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

```
. signrank Crea_B_S2 = Crea_B_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	12	687.5	1775.5
Negative	41	2863.5	1775.5
Zero	40	820	820
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -405.75
Adjustment for zeros    -5535.00
                        ---------
Adjusted variance       62174.00

H0: Crea_B_S2 = Crea_B_S3
                z =  -4.363
Prob > |z|      =   0.0000
Exact prob      =   0.0000
```

```
. ttest Qual_B_S2 == Qual_B_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Qual_B~2	93	5.182796	.1409132	1.358918	4.90293	5.462661
Qual_B~3	93	5.397849	.1394068	1.344391	5.120975	5.674723
diff	93	-.2150538	.1057592	1.019905	-.4251007	-.0050069

```
    mean(diff) = mean(Qual_B_S2 - Qual_B_S3)                      t =  -2.0334
H0: mean(diff) = 0                              Degrees of freedom =       92

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0224         Pr(|T| > |t|) = 0.0449          Pr(T > t) = 0.9776
```

```
. signrank Qual_B_S2 = Qual_B_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	13	961	1470
Negative	27	1979	1470
Zero	53	1431	1431
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -515.00
Adjustment for zeros   -12759.75
                        ---------
Adjusted variance       54840.00

H0: Qual_B_S2 = Qual_B_S3
                z =  -2.174
Prob > |z|      =   0.0297
Exact prob      =   0.0299
```

Statistical tests for cross-dimensions ratings

```
. ttest Crea_A_S2 == Emo_A_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Crea_A~2	93	5.075269	.1394879	1.345173	4.798234	5.352304
Emo_A_S3	93	4.365591	.1775008	1.711756	4.013059	4.718123
diff	93	.7096774	.2043134	1.970327	.3038934	1.115461

```
    mean(diff) = mean(Crea_A_S2 - Emo_A_S3)                      t =   3.4735
H0: mean(diff) = 0                              Degrees of freedom =       92

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.9996         Pr(|T| > |t|) = 0.0008          Pr(T > t) = 0.0004
```

```
. signrank Crea_A_S2 = Emo_A_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	44	2798	2023
Negative	24	1248	2023
Zero	25	325	325
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -910.62
Adjustment for zeros    -1381.25
                        ---------
Adjusted variance       65822.88

H0: Crea_A_S2 = Emo_A_S3
                z =   3.021
Prob > |z|      =   0.0025
Exact prob      =   0.0023
```


Appendix III. Stata results printouts

```
. ttest Aut_B_S2 == Emo_B_S3

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Aut_B_S2	93	3.806452	.16271	1.569118	3.483296	4.129608
Emo_B_S3	93	4.075269	.1810549	1.74603	3.715678	4.434859
diff	93	-.2688172	.1770755	1.707654	-.6205044	.08287

```
    mean(diff) = mean(Aut_B_S2 - Emo_B_S3)                    t =  -1.5181
H0: mean(diff) = 0                             Degrees of freedom =      92

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.0662         Pr(|T| > |t|) = 0.1324          Pr(T > t) = 0.9338
```

```
. signrank Aut_B_S2 = Emo_B_S3

Wilcoxon signed-rank test
```

Sign	Obs	Sum ranks	Expected
Positive	28	1647	1953
Negative	35	2259	1953
Zero	30	465	465
All	93	4371	4371

```
Unadjusted variance     68114.75
Adjustment for ties      -634.38
Adjustment for zeros    -2363.75
                        ---------
Adjusted variance       65116.62

H0: Aut_B_S2 = Emo_B_S3
            z =  -1.199
Prob > |z|    =  0.2305
Exact prob    =  0.2336
```


Appendix III. Stata results printouts

Abstract:

This thesis investigates how listeners from a specific community perceive music when told it was created either by a human or by artificial intelligence (AI). As AI tools like SUNO become more advanced in generating music that imitates human creativity, it is crucial to understand whether listeners react differently based on the label attached to the music. The main question examined is whether a framing effect influences preferences and perceptions depending on whether the music is labelled as human-composed or AI-generated.

The study involved 93 participants who listened to two musical excerpts, one composed by a human and one by AI. These excerpts were presented across three conditions: blind (no label), true disclosure (accurate label), and false disclosure (misleading label). Participants selected which excerpt they preferred and rated each piece on four dimensions: authenticity, creativity, emotional connection, and audio quality, using a 7-point Likert scale. The experiment followed an A/B testing logic. Differences between conditions were assessed using paired t-tests and confirmed through Wilcoxon signed-rank tests, to ensure reliability despite possible non-normal data distributions.

Findings indicate a consistent and statistically significant framing effect. Music labelled as human was rated more positively across all perceptual dimensions, even when the actual composition source was AI. This effect was particularly strong for authenticity, creativity, and emotional connection. Audio quality showed a smaller but still significant difference. Moreover, the same musical piece received lower ratings when it was presented as AI-generated compared to when it was labelled as human-composed, regardless its true origin.

These results support the idea that listeners' perception from this community is influenced by framing and labelling, regardless of the actual origin of the music. The study contributes to ongoing discussions about trust, creativity, and value in AI-generated content. It also opens avenues for further research into public acceptance of generative AI in the arts, the role of transparency, and how attribution affects appreciation across creative domains.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique
www.uclouvain.be/lsm