Politecnico di Torino

Master's degree in Engineering and Management
A.a. 2023/2024
Sessione di Laurea dicembre 2024

Exploring the impact of generative AI on the music composition market:

a study on public perception, behavior, and industry implications

Relatore:
Prof. Marco Cantamessa

Candidata:
Silvia Candusso

Summary

1. Introduction

In recent years, the intersection of technology and creativity has led to groundbreaking innovations in various artistic fields; when talking about music, one of the most profound and promising developments can be seen in the rising of generative artificial intelligence (AI) in music composition and production. Generative AI, powered by advancements in machine learning and neural networks, has proved the ability to analyse vast datasets of musical compositions, identify patterns, and generate new pieces of music that mimic specific styles or create entirely novel sounds. Tools like OpenAI's MuseNet, Google's Magenta, and Amper Music have empowered musicians and producers to explore uncharted musical territories, offering limitless creative possibilities to create music, reshaping the traditional processes of music creation and challenging the roles of human composers and producers.

The integration of generative AI into music composition and production raises critical questions about the nature of creativity and authorship. AI can be in fact seen as a revolutionary tool that democratizes music production and enhances human creativity or can be seen with concern about its potential to diminish the role of human artists and homogenize musical expression. The balance between human input and machine output is a pivotal area of investigation, as is the ethical consideration of authorship and intellectual property rights in AI-generated music.

This thesis aims to explore the multifaceted impact of generative AI on music composition and production, examining how AI technologies are transforming the creative process, analysing the implications for the music industry, and preliminarily assessing the potential future trajectories of AI-driven musical innovation. By critically evaluating the benefits and challenges posed by generative AI, and then through market research conducted with the use of a questionnaire that aimed to investigate people's will to approve the involvement of this technology in music creation and production process, this study seeks to contribute to a deeper understanding of how technology is reshaping one of humanity's oldest and most cherished art forms.

2. Music industry and AI at a glance

2.1 Music industry

2.1.1 Structure of the Music Industry

The music industry is a complex ecosystem composed by creation, performance, recording, promotion, distribution, and consumption of music. It is essential to understand the business structure of the music industry and how its different sectors are connected to each other and contribute to the overall ecosystem for navigating the music industry in a more effective way. There are four main sectors composing the industry: (Yellowbrick, 2023)

Recording Industry: Includes production and marketing of music tracks, albums and videos, and is mainly represented by record labels that are essential for financing and promoting artists, offering them resources needed for the record, production and distribution of their music and handling copyright and licensing issues for them. Another important aspect involves the promotion of the connection and interaction between the artists and their audience, which is essential for the artists success. Major labels such as Universal Music Group, Sony Music Entertainment, and Warner Music Group, are leading the industry, together with independent labels that play a crucial role in discovering and introducing new talents in the field. (Yellowbrick, 2023)

Publishing Industry: They are responsible for providing royalties to songwriters and composers by arranging licensing deals on different platforms such as radio, streaming services, films, television shows and visual media in general. Moreover, the publishing industry makes synchronization deals, for the usage of music into advertisements, movies, and TV programs. (Yellowbrick, 2023)

Live Performance: Includes concerts, music festivals and tours, where artists can make significant revenues and interact with their fans directly. Live performances also provide opportunities for brand partnerships and merchandising. Indeed, event production companies, promoters, and venue owners collaborate to create memorable experiences, handling everything from ticket sales and marketing to stage design and logistics. Booking agencies facilitate live performances for artists, including concerts, festivals, and tours, acting as a connection between artists and event organizers. (Yellowbrick, 2023)

Distribution: Distribution companies, both traditional and digital, work closely with record labels and artists to ensure their music is available on various platforms and to maximize revenue generation. The transition from physical formats like CDs and vinyl to digital platforms has revolutionized this sector, and streaming services such as Spotify, Apple Music, and YouTube have now become the primary means of music consumption, reshaping how music is accessed and monetized. (Yellowbrick, 2023)

Two other fundamental fields are found in Merchandising, which plays a crucial role in the music industry, allowing artists to monetize their brand and connect with fans on a deeper level and Music Media Outlets, such as radio stations, music magazines, blogs and podcasts, that play a pivotal role in promoting artists and shaping popular culture, providing exposure to both established and emerging artists, influencing trends and driving music consumption. (Yellowbrick, 2023)

2.1.2 Economic Value of the Music Industry

2.1.2.1 Global Economic Impact

The music industry, a major force in the global economy, has undergone profound transformations over the past decade, driven by the rise of digitization and the Internet. Between 1999 and 2003, U.S. retail sales of recorded music fell from $13 billion to $10.6 billion, reflecting the growing shift towards digital music. However, by 2004, recorded music sales in the U.S. saw a modest 1.4% increase, and global industry revenues began to stabilize, as noted by UK industry analyst Claire Enders. Despite this brief recovery, long-term projections still indicated potential declines in traditional revenue streams. (Bockstedt et al., 2006)

This shift towards digital music became increasingly evident as platforms like Apple iTunes expanded their customer base from 861,000 in July 2003 to 4.9 million by March 2004. The rising popularity of digital audio devices such as the Apple iPod and Dell JukeBox further fuelled demand for digitally formatted music. Apple capitalized on this trend by selling millions of iPods and launching multiple versions to increase its market share. By July 2005, Apple announced that iTunes had surpassed 500 million digital music file sales, underscoring the growing dominance of digital music formats as they quickly became the preferred choice for many consumers. (Bockstedt et al., 2006)

Digital transformation has been a key driver of the music industry’s expansion over the years. According to the International Federation of the Phonographic Industry’s (IFPI) Global Music Report 2024, global music sales experienced their ninth consecutive year of growth in 2023. Recorded music revenues increased across all markets and regions, and nearly all formats, pushing total revenues to $28.6 billion—a rise of just over 10% from the previous year. This marks the second-highest growth rate on record, following the 18.5% surge in 2021. (IFPI Global Music Report (2024 Edition), 2023)

Looking ahead, the industry's growth trajectory is expected to continue, with Goldman Sachs projecting that global music revenues will more than double to approximately $131 billion by 2030. This forecast highlights the increasing importance of digital music and streaming services in the industry’s future, as well as the ongoing potential for innovation and expansion within the global music market. The music industry’s resilience and adaptability in embracing digital formats have been pivotal to its sustained growth, and it remains poised for further evolution in the coming years. (Goldman Sachs, 2018)

Artists and industry stakeholders generate income through multiple streams; in 2023, streaming once again dominated global music revenues, yet nearly all music formats saw revenue increases, except for downloads and other digital formats. Notably, subscription streaming, performance rights, and physical formats like CDs and vinyl experienced accelerated growth compared to 2022. Physical formats, in particular, surged with a 13.4% increase, marking the highest growth rate of any format that year. (IFPI Global Music Report (2024 Edition), 2023)

Revenues in the music industry can be generated from sources such as:

Sales: Despite the overall decline in digital downloads and physical sales, these formats still contribute to revenue, with physical formats experiencing a 13.4% increase in 2023, driven by rising CD sales and growing vinyl interest, accounting for US$5.1 billion and 17.8% of the global market, with Asia leading at 49.2% of global physical revenues, largely due to strong K-Pop sales. (IFPI Global Music Report (2024 Edition), 2023)

Streaming: Subscription services and ad-supported platforms remain the primary revenue drivers, with streaming contributing to over two-thirds (67.3%) of the total global market. In 2023, global streaming revenues saw a 10.4% increase, reaching US$19.3 billion, which, despite being slightly lower than the previous year's 11.4% growth, included an acceleration in subscription revenue growth to 11.2%, up from 10.1% in 2022. (IFPI Global Music Report (2024 Edition), 2023)

Concerts and Touring: Live performances continue to be a crucial revenue stream, providing artists with significant income and opportunities for brand development. In 2023, performance rights revenues, which include the use of recorded music by broadcasters and public venues, saw a 9.5% increase, reaching US$2.7 billion and accounting for 9.5% of the global market, maintaining strong growth after surpassing pre-pandemic levels in 2022. (IFPI Global Music Report (2024 Edition), 2023)

Licensing and Synchronization: Licensing music for films, TV shows, and commercials generates extra income and visibility for artists. In 2023, synchronization revenues reached US$632 million, marking a 4.7% increase, though this growth was slower compared to the 23.9% rise seen in 2022. This category, covering the use of music in advertising, film, games, and television, made up 2.2% of total recorded music revenues. (IFPI Global Music Report (2024 Edition), 2023)

Download and other digitals: The only format to see a decline in 2023 was downloads and other non-streaming digital formats, with revenues dropping by 2.6%. Although this decline was less steep compared to the previous year's 11.8% decrease, these formats represented just 3.2% of global recorded music revenues, as streaming continues to dominate the digital market. (IFPI Global Music Report (2024 Edition), 2023)

2.1.2.2 Regional Markets

The music industry operates differently across various regions, each with unique characteristics.

North America is the largest market for recorded music, characterized by a high adoption of streaming services and significant investment in live events. According to the IFPI, music revenues increased across all 58 markets it monitors, with the U.S. maintaining its top position globally. Music sales in the U.S. grew by 7.2%, an improvement over the previous year's growth of 4.8%. IFPI reports that combined, the U.S. and Canada region accounts for almost 41% of global recorded music revenues. (IFPI Global Music Report (2024 Edition), 2023; Smirke, 2024)

Europe presents a diverse market with strong streaming growth and a rich tradition of live performances and festivals. Europe continues to hold its position as the second-largest market for music sales, contributing over a quarter (28%) of global revenues and experiencing an 8.9% increase compared to the previous year. Asia ranks third, with revenues climbing nearly 15% in 2023, fuelled by significant growth in both physical and digital sales. (IFPI Global Music Report (2024 Edition), 2023; Smirke, 2024)

Asia is emerging as a major player in the global music landscape, with rapid growth in streaming and a burgeoning pop music scene, particularly in countries like South Korea and Japan, that holds steady in second place with sales growing 7.6% in 2023 (the third and fourth-biggest markets for recorded music remain the United Kingdom (+8.1%) and Germany (+7%), respectively). The rest of the top 10 is made up of China (+25.9%), representing the fastest rate of increase in any top 10 market, followed by France (+4.4%), South Korea (percentage not provided), Canada (+12.2%), Brazil (+13.4%) and Australia (+11.3%). In Australia, which ranks among the top 10 music markets globally, revenue growth accelerated to 11.3%, up from 8.2% in the previous year. In New Zealand, revenues also saw a healthy increase of 8.4%. (IFPI Global Music Report (2024 Edition), 2023; Smirke, 2024)

Those cross-market gains are mirrored on a regional basis with revenues from the U.S. and Canada region up 7.4%. (IFPI Global Music Report (2024 Edition), 2023; Smirke, 2024)

Latin America, where streaming accounts for 86% of the market, experienced a remarkable 19.4% growth, significantly surpassing the global average. This marks the 14th consecutive year of revenue growth in the region. (IFPI Global Music Report (2024 Edition), 2023; Smirke, 2024)

Africa includes regions experiencing fast-paced growth in digital music consumption, driven by mobile technology and increasing internet penetration. Sub-Saharan Africa emerged as the fastest-growing market region, with a 25% increase in music sales, primarily fuelled by the growing adoption of paid subscription services and the booming South African music market, which expanded by nearly 20% and accounted for over three-quarters of the region's revenue. (IFPI Global Music Report (2024 Edition), 2023; Smirke, 2024)

Meanwhile, in the Middle East and North Africa, where streaming dominates with a 98% share of the recorded music market, revenues grew by almost 15%. (IFPI Global Music Report (2024 Edition), 2023; Smirke, 2024)

2.1.3 Music Composition and Production

Music composition is a creative process that depends on a large number of decisions, starting from inspiration, drawn from different sources, including composer's personal experiences, emotions, nature, and cultural influences. Inspiration leads the composer to a small unit of one or two bars or core progression called motif that is then developed to compose a melody or music phrase, from which the structure of each section of the song is formed. Each section has its own purpose so it can be written in different keys and its phrases usually follow different harmonic progressions than the other sections. (Hernandez-Olivan & Beltran, 2022)

2.1.3.1 Composition's Structure

Compositions are made of a melodic part, played by different instruments whose frequency range may or may not be similar, and an accompaniment or harmonic part, that gives the piece a deep and structured feel. Moreover, music is based on two dimensions, time dimension, represented by the note's duration or rhythm and harmony dimension related to the note values or pitch. (Hernandez-Olivan & Beltran, 2022)

Based on the ideas of Walton (Walton, 2005), we can identify some basic music principles or elements:

Harmony: It is the superposition of notes that form chords which compose a chord progression. The note-level could be considered as the lowest level in harmony, followed by the chord-level, while the highest-level can be considered as the progression-level which usually belongs to a certain key. (Hernandez-Olivan & Beltran, 2022)

Music Form or Structure: It is the high-level structure that of the composition and it is related with the time dimension. The smallest part of a music piece is the motif which is developed in a music phrase and the combination of music phrases form a section. Sections in music are ordered depending on the music style such as intro-verse-chorus-verse-outro for some pop songs (also represented as ABCBA) or exposition-development- recapitulation or ABA for Sonatas. The concatenation of sections which can be in different scales and modes gives us the entire composition. (Hernandez-Olivan & Beltran, 2022)

Melody and Texture: Texture in music terms refers to the melodic, rhythmic and harmonic contents that must be combined in a composition in order to form the music piece. Music can be monophonic or polyphonic depending on the notes that are played at the same time step, homophonic or heterophonic depending on the melody, if it has or not accompaniment. (Hernandez-Olivan & Beltran, 2022)

Instrumentation and Orchestration: These are music techniques that take into account the number of instruments or tracks in a music piece. Whereas instrumentation is related to the combination of musical instruments which compose a music piece, orchestration refers to the assignment of melodies and accompaniment to the different instruments that compose a determined music piece. In recording or software-based music representation, Instruments are organized as tracks, each of those containing the collection of notes played on a single instrument (a piece of music played by more than one instrument is called multi-track). Each track can contain one note (monophonic tracks) or multiple notes that sound simultaneously (polyphonic tracks). (Hernandez-Olivan & Beltran, 2022)

2.1.3.2 Roles and Collaboration in music composition

During the composition process, a few roles contribute to the final piece, here the main ones, which could also be embodied by the same person.

Composers: Derived from a Latin word meaning "one who puts together," a composer does just that, piecing together the various elements that comprise a piece of music, melodies and harmonies, rhythms and dynamics, structure and sensibility, to create an original work. Composers may have highly individual styles, methods, and goals, but all composers use music as a medium to express and evoke ideas, emotions, and sensibilities. (*Composer (Concert and Stage)*, n.d.)

Lyricists: Skilled lyricists were once highly valued in the music industry for their ability to craft beautiful lines and enhance songs, but today more songwriters are beginning to write their own lyrics, and that has reduced the demand for specialized lyric writing. Lyricists also have roles, such as top-line songwriting, staff writing at music publishing companies, freelancing and opportunities still exist in musical theatre and opera, where lyricists and librettists collaborate closely with composers to create complete works. (*Lyricist*, n.d.)

Arrangers: Musicians who adapt pre-existing compositions by altering elements such as instrumentation, orchestration, harmony, tempo, and genre to create a new sound for a piece of music. They reimagine the original composition to suit various performance settings or artistic interpretations. (*Arranger*, n.d.)

2.1.3.3 Music Production Process

Music production is the creative process of composing, recording, arranging, editing, mixing, and mastering audio, representing the multi-faceted journey of bringing a musical idea to life, ensuring it's ready for the audience to hear. This process is guided and managed by the music producer, who oversees the entire process, starting from the initial concept or idea, navigating through the songwriting, arranging, recording, and sound design phases, and finally culminating in the mixing and mastering stages, and can be done in a professional studio setting or at home with digital audio workstations (DAWs) and virtual instruments. (*Music Production: Guide to Producing & Releasing Tracks*, 2023) There are a few stages involved in music production, each critical to the creation of a polished piece of music, including:

Recording: The recording process begins with capturing the performance of vocals and instruments, doing multiple takes to ensure the best possible performance. The recording studio can be set either at home or in a professional facility; in both cases there are several factors to consider. Home studio setups are often more budget-friendly, relying on compact equipment and DIY acoustic treatment solutions, while professional recording studios require significant investment in high-end gear, acoustics, and infrastructure. On the other hand, professional recording studios typically offer larger, acoustically treated spaces with dedicated recording rooms, isolation booths, and control rooms, providing optimal conditions for recording and mixing, while home studios may be limited by space constraints, requiring creative solutions to optimize room acoustics and workflow. (*Music Production: Guide to Producing & Releasing Tracks*, 2023)

Mixing: It is the process of blending individual audio tracks together to create a balanced and cohesive final mix and involves different elements:

(I) Levels: balancing the volume levels of individual tracks ensures that each instrument and vocal sits well in the mix, preventing any one element from overpowering the others.
(II) Panning: determines the placement of audio signals within the stereo field, allowing for spatial separation and creating a sense of width and depth in the mix.
(III) EQ (Equalization): it is used to shape the frequency balance of each track, emphasizing or attenuating specific frequencies to enhance clarity and balance in the mix.
(IV) Compression: is a dynamic processing technique used to control the dynamic range of audio signals, reducing peaks and boosting quieter passages to achieve a more consistent and cohesive sound.
(V) Reverb and Effects: Reverb and other effects add depth, dimension, and atmosphere to the mix, enhancing the spatial characteristics and overall ambiance of the music.

(Samuels, 2024)

Mastering: in the final stage of the music production process the final mix is prepared for distribution and playback across various platforms and formats. Mastering engineers ensure that the final master is optimized for clarity, consistency, and fidelity by applying a range of techniques, including EQ and Compression (to fine-tune the overall balance, tonal characteristics, and dynamic range of the master, ensuring that it translates well across different playback systems), Stereo Imaging (to enhance the width, depth, and spatial separation of the mix, creating a more immersive listening experience), Loudness and Dynamics (ensuring that the master meets industry standards for volume levels and dynamic range while preserving the integrity of the music), Sequencing and Metadata (in addition to audio processing, mastering involves sequencing the tracks and adding metadata such as track titles, album artwork, and ISRC codes to prepare the master for distribution). (Samuels, 2024)

2.1.3.4 Technological Advancements

Digital Audio Workstations (DAWs): they are software platforms used by musicians, producers, and audio engineers to record, edit, mix, and produce music. There are several popular DAWs available on the market, each with its own unique features and workflow. Some of the most widely used DAWs include Avid Pro Tools (known for its industry-standard recording and editing capabilities, it is widely used in professional recording studios and post-production facilities), Apple Logic Pro (offers a comprehensive suite of tools for music production, including virtual instruments, MIDI sequencing, and audio editing features, making it popular among Mac users), Ableton Live (favoured by electronic music producers and performers for its intuitive session view, real-time audio manipulation, and extensive collection of built-in instruments and effects) and Steinberg Cubase (advanced MIDI editing capabilities, scoring features, and comprehensive mixing tools, make it a versatile choice for composers and producers). The choice of DAW often depends on personal preference, workflow requirements, and the specific needs of the project. (Samuels, 2024)

Virtual Instruments and Plugins: They are essential in modern music production, providing additional sound shaping and processing capabilities within DAWs. Plugins are software effects or instruments that can be added to a DAW to enhance its functionality, and some common types of plugins include EQs, compressors, reverbs, delays, and virtual synthesizers, each offering a unique set of controls and parameters for manipulating audio. Virtual instruments, in contrast, are software-based recreations of traditional acoustic instruments, synthesizers, and samplers, enabling musicians to produce realistic or innovative sounds directly within their DAW. The use of plugins and virtual instruments make it possible for producers to experiment with different sounds, textures, and effects, increasing the creative possibilities of music production. (Samuels, 2024)

Automation and AI Tools: Advancements in artificial intelligence (AI) and machine learning are revolutionizing the field of music creation, offering new tools and techniques for composers, producers, and musicians. AI-powered music composition software, such as Amper Music, AIVA, and OpenAI's MuseNet, can generate original compositions in different styles and genres, providing inspiration and starting points for creative projects. Machine learning algorithms can be used to analyse and extract musical patterns, enabling new approaches to music recommendation, synthesis, and production. AI and machine learning have the potential to streamline workflows, enhance creativity, and expand the possibilities of music production. (Samuels, 2024)

All these elements, processes, artists and tools together ensure that the final track resonates with listeners, conveying the intended emotions and messages. (Samuels, 2024)

2.2 Generative AI

2.2.1 History of Generative AI

Generative AI is an artificial intelligence technology with the ability to rapidly generate high-quality text, visuals, sounds and videos, and now is gaining more and more visibility. While the definition of Generative AI may seem recent, its roots trace back to the 1940s. (Kılınç & Keçecioğlu, 2024)

In 1956, AI was introduced as a science; two years later the perceptron, the world's first neural network, was proposed. The first chatbots of the 1960s can be considered as primitive versions of the advanced chatbots used today, starting from ELIZA, introduced as a chatbot simulating conversation and published in the 1960s as one of the significant works in human-computer interaction. During the 1960s and 70s, the research was focused on implementing computer vision and utilizing some fundamental recognition models; in the meanwhile, more advanced expert systems were developed. One of the early examples of generative artificial intelligence in the field of computer vision was Harold Cohen's AARON computer program, designed to create art. The field of machine learning typically employs statistical models, including generative ones, to model and predict data. Between the 1980s and the 1990s the Recurrent Neural Network and the more complex LSTM (Long Short-Term Memory) have been developed, and in particular the second one allowed efficient processing of long sequences of data and capturing patterns. Advancements in neural networks and the emergence of deep learning since the 2000s have led to accelerated progress and research in technology's ability to automatically analyze text, classify image elements, convert speech to text through learning models and other tasks. Modern generative artificial intelligence is primarily based on deep learning techniques, and as a result, generative AI has rapidly evolved in the 2010s. (Kılınç & Keçecioğlu, 2024)

In 2014, with the introduction of Generative Adversarial Networks (GANs), a type of machine learning algorithm, generative AI became capable of creating convincingly original images, videos, and sounds comparable to those produced by humans. (Kılınç & Keçecioğlu, 2024)

This deep learning technique, developed by Ian Goodfellow, introduced a new approach to adversarial neural networks that generate content variations and perform ranking. In this model, two different neural networks compete with each other, producing realistic human-like images, sounds, music, and text. (Kılınç & Keçecioğlu, 2024)

In 2017, with the introduction of transformer libraries and developments in generative network models in the subsequent years, there was a significant acceleration in progress. Transformers, a type of machine learning, enabled researchers to train larger models without the need to pre-label all data. This allowed new models to be trained on more extensive datasets, providing more realistic responses to text. Additionally, Transformers can make inferences by deciphering connections between sentences, pages, or chapters. The transformers library, introducing a new concept called attention, enables the establishment of these connections, providing opportunities for novel research not only in textual contexts but also in analyzing code, proteins, chemicals, and DNA, with the potential to assist in future endeavors such as coding, designing new drugs, product development, reengineering business processes, and transforming supply chains. (Kılınç & Keçecioğlu, 2024)

The Transformers library led to the emergence in 2018 of Generative Pre-trained Transformer (GPT), a type of large language model introduced by OpenAI. In 2021, the release of DALL-E, a pixel-generating model based on Transformers, followed by Midjourney and Stable Diffusion, has given rise to practical, high-quality artificial intelligence art stemming from natural language prompts. (Kılınç & Keçecioğlu, 2024)

Although generative artificial intelligence models can produce interesting texts and realistic images, the current years represent the early stages of the technology's development. Consequently, we may encounter products with lower accuracy. Among other techniques, there are variational autoencoders (VAE), long short-term memory (LSTM), transformers, and diffusion models. The developmental timeline of generative artificial intelligence is illustrated in Figure below. (Kılınç & Keçecioğlu, 2024)

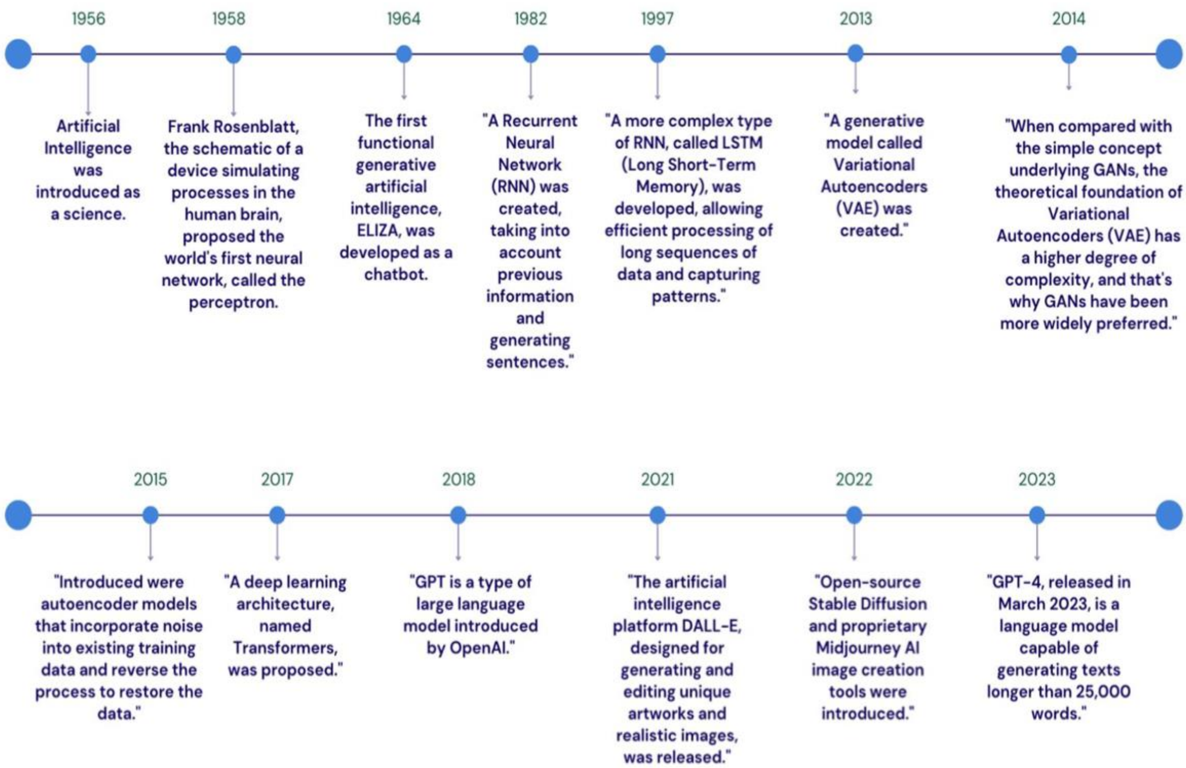


FIGURE 2.1 – Generative AI history timeline (Kılınç & Keçecioğlu, 2024)

2.2.2 Categorization of Generative AI

Generative AI encompasses a variety of models and techniques, each with distinct characteristics and applications. Multiple generative models have emerged with the capability of generating new data points like the training data inputs based on learning their distribution. The most used NN architectures in music composition task are Generative Models such as Variational AutoEncoders (VAEs) or Generative Adversarial Networks (GANs), and NLP-based models such as Long Short-Term Memory (LSTM) or Transformers. (Hernandez-Olivan & Beltran, 2022)

2.2.2.1 Types of Generative AI Models

Variational Auto-Encoders (VAEs): The original VAE model uses an Encoder-Decoder architecture to produce a latent space by reconstructing the input. A latent space is a multidimensional space of compressed data in which the most similar elements are located closest to each other. In a VAE, the encoder approximates the posterior and the decoder parameterizes the likelihood. The posterior and likelihood approximations are parametrized by a NN with λ and θ parameters for the encoder and decoder respectively. (Hernandez-Olivan & Beltran, 2022)

Generative Adversarial Networks (GANs): GANs are Generative Models composed by two NNs: the Generator G and the Discriminator D. The generator learns a distribution pg over the input data. The training is done in order to let the discriminator maximize the probability of assigning the correct label to the training samples and the samples generated by the generator. The generator and the discriminator can be formed by different NN layers such as Multi-Layer Perceptrons (MLP), LSTM or Convolutional Neural Networks (CNN). (Hernandez-Olivan & Beltran, 2022)

Transformers: Transformers are being currently used in NLP applications due to their well performance not only in NLP but also in Computer Vision models. Transformers can be used as auto-regressive models like the LSTMs which allow them to be used in generative tasks. The basic idea behind Transformers is the attention mechanism. The combination of the attention layer with feed forward layers leads to the formation of the

Encoder and Decoder of the Transformer, which differs from purely AutoEncoder models that are also composed by the Encoder and Decoder. Transformers are trained with tokens which are structured representations of the inputs. (Hernandez-Olivan & Beltran, 2022)

2.2.2.2 Applications Across Industries

Generative AI has applications across various creative fields:

3. **Image and visual content generation:** realistic works are created using models like GANs, Diffusion, Transformers and VAE. Systems trained with text, image or model-dependent sound inputs are commonly used, and tools like DALL-E, Stable Diffusion and Midjourney are frequently employed in this context. (Kılınç & Keçecioğlu, 2024)

4. **Text Generation:** Generative AI can produce text, articles, stories and poems using text-based data, relying on language models such as CPT-3, -4, LaMDA which are trained on words or tokens. This enables natural language processing, machine translation, text synthesis providing users with written contents. (Kılınç & Keçecioğlu, 2024)

5. **Music and Sound Production:** Generative AI can produce new music notes or sounds, including generating original composition, creating automatic music or combining different musical styles. They are usually trained on the sound waveforms of recorded music, and some examples can be models like MusicLM and MusicGen. (Kılınç & Keçecioğlu, 2024)

6. **Video and Animation Production:** Generative AI models with significant impacts on the digital publishing sector can be utilized for creating videos, 3D animations, commercials and game development. Advanced models used in filmmaking, animation and creating game characters have the potential to reshape industry dynamics. (Kılınç & Keçecioğlu, 2024)

3. Literature Review

A comprehensive literature review was conducted to investigate the state-of-the-art of the application of artificial intelligence in music composition and production and to understand how humans interact and collaborate with it. The research questions that were posed for this review aimed to uncover critical themes and challenges of the industry and to get a better understanding of the current research landscape.

- Which is the dominant technology?
- Which point of the innovation curve has this technology reached?
- Who are the major investors at the moment?
- How much interest is there around this technology? How is it currently used?
- How are the humans reacting to the technology? Are they willing to collaborate?

3.1 Methodology

The methodology for conducting the literature review adheres to the PRISMA framework (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), which suggests a 27-item checklist addressing the introduction, methods, results and discussion sections of a systematic review report. The framework was first developed in 2009 but was reviewed in 2020 to ensure its currency and relevance as described in detail by Matthew J Page et al. in *BMJ* 2021;372: n71 and n160.

As for this thesis, the analysis of the literature will follow four main steps, and the discussion will be centred on the division of the papers in clusters, allowing for a more organized and coherent presentation of findings. The four main stages are briefly described below:

1. *Collection of papers for review*: This first step involves the identification and use of research strings along with set inclusion and exclusion criteria to select relevant articles from an ample range of available literature.

2. *Analysis and category selection*: The selected papers go through a process of deductive and inductive categorization to assign them to different clusters, needed for a detailed content analysis.

3. *Descriptive analysis of literature*: This phase focuses on analysing demographic details such as publication year and the geographical focus of the studies to provide a contextual background for the research

4. *Content analysis:* In the final step the collected literature is analysed, and the papers are assigned to clusters according to the main scope or theme of the study. The analysis aims to identify and explore potential opportunities for enhancing the field of study.

3.2 Literature Review on Generative AI in music composition and production

3.2.1 Collection of papers

The first step of the research involved using three different databases in order to gather the wider number of articles as possible; the chosen databases were Scopus, ScienceDirect and Word of Science (WoS). The search string was based on the research objectives and main focus, which were the three main key terms Music, AI and Generation. Those words were complemented with alternatives to get the greatest number of papers that address the same topics. The specific query used for Scopus was TITLE-ABS-KEY (generative AND (ai OR "artificial intelligence") AND music AND (composition OR production OR generation)). For ScienceDirect and WoS it was necessary to broaden the search by removing the TITLE-ABS-KEY restriction. From this first unfiltered search, were obtained: 71 results from Scopus, 907 results from ScienceDirect and 35 from WoS, for a total of 1013 papers in June 2024.

As for the inclusion/exclusion criteria, only papers that met the following criteria were included:

1. The work must be a full text; where possible, the "open access" filter was applied.
2. The work must be published after 2000.
3. The work must be peer-reviewed; it could be published in a journal, from proceedings of a conference or from a book. In some cases, arXiv works were also accepted after a check on the author.
4. The work must deal with "music composition", "artificial intelligence", "music production" or any equivalent formulation.

Concerning the last criteria, articles related to the subject areas of Energy, Medicine & Dentistry, Neuroscience, Mathematics, Building & construction, Chemical engineering, Materials Science, and Physics & Astronomy were excluded as they fell outside the research scope.

After a thorough review of titles, research fields, keywords, abstracts, and after eliminating duplicates across databases, a set of 85 articles in total was selected to be the reference database for this literature review.

At the same time, thanks to a parallel search on the Internet, aimed at seeking out specifically articles that dealt with technological methods and human responses, 21 more articles were added to the final database that will contribute to the content analysis. However, they will not be considered for the demographic analysis since they would add a bias due to the search mode.

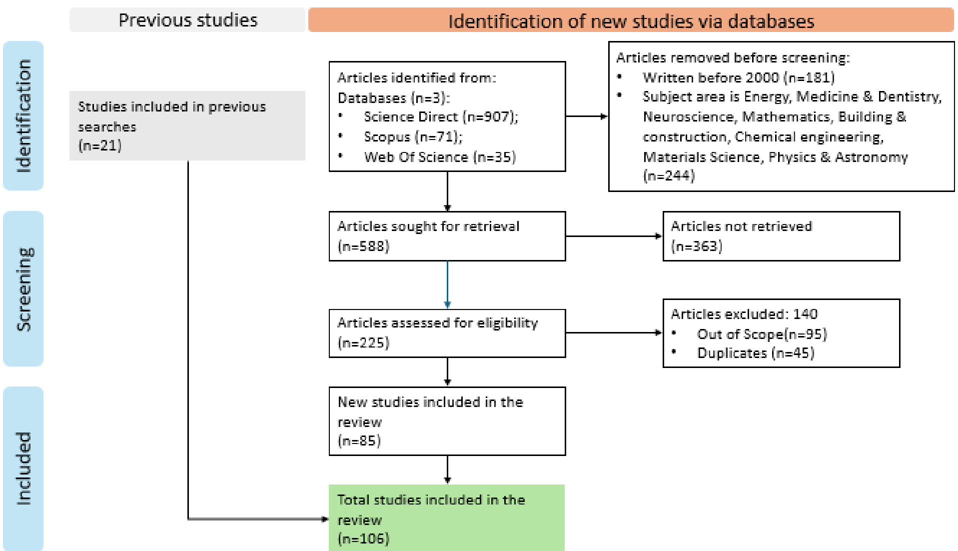


FIGURE 3.1 - PRISMA flow chart

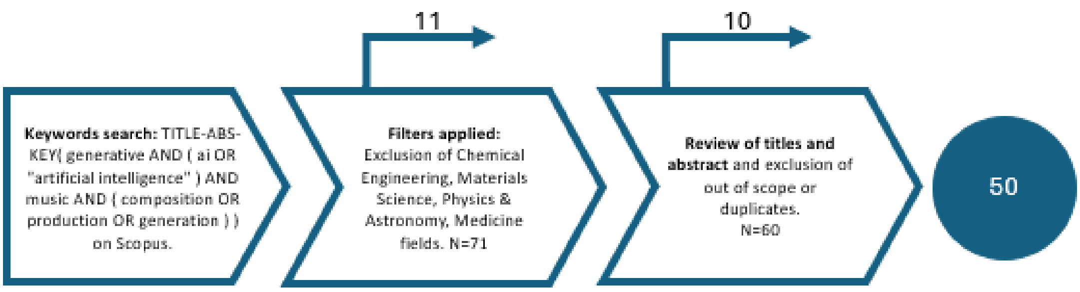


FIGURE 3.2 - Paper selection process on Scopus

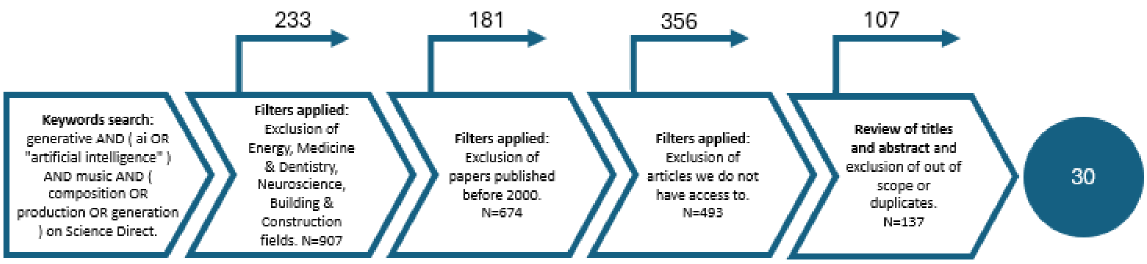


FIGURE 3.3 - Paper selection process on ScienceDirect

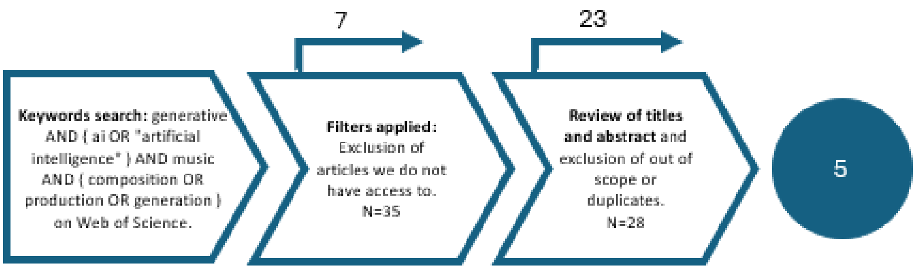


FIGURE 3.4 - Paper selection process on WoS

3.2.2 Analysis and Category selection

After the initial selection, the topic and contribution of each article were determined by reading and analysing their content. In order to address our review's research questions, we classified the papers into clusters depending on our objectives and what we discovered during the initial read. This strategic classification has assisted in determining the major themes and research dimensions within the stages of generative AI for music creation and production. The following clusters have been identified:

1. **Generative models and techniques:** These papers focus on various generative models used for music creation. For example, they explore techniques such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and Recurrent Neural Networks (RNNs) to synthesize music, but also they explore new methods that could be the new state-of-the-art. Overall, these articles highlight advancements in generative models and their application in creating diverse and culturally rich music.

2. **Creation of music generation systems:** This cluster includes articles that discuss the development of interactive music generation systems such as programs or platforms. These systems aim to bridge the gap between AI and human musicians, offering innovative tools for real-time and user-friendly music production. This cluster has two other subclusters, which are not exclusive and not mandatory:
   2a) Use of music theory: discussion of the importance of incorporating music theory rules to improve quality and coherence.
   2b) Creation for public use: creation of public platforms where the public can generate music. They primarily discuss the design and the interactive tools.

3. **Evaluation generative systems:** This cluster focuses on evaluating the quality of AI-generated music using both objective and subjective metrics. The papers propose various methods for assessing musical elements like repetition, structure, and originality. They also highlight the challenges in developing

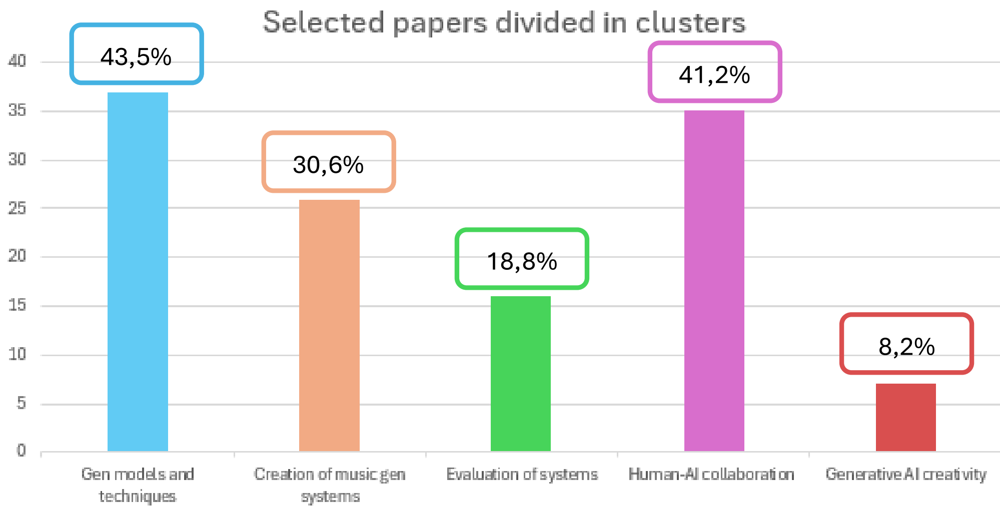
standardized evaluation frameworks that could comprehensively measure the effectiveness of different AI music generation models.

4. **Human-AI collaboration:** These articles explore the interaction between humans and AI in the context of music composition and production. They investigate how AI can assist and collaborate with human musicians to enhance creativity and emotional expression. This cluster is divided in two subclusters, it is mandatory to choose at least one of them, but they are not exclusive:
   4a) Human emotional reaction: how the general public and musicians react to AI, what is their sentiment and willingness to use it.
   4b) How to collaborate: broader applications and societal implications of AI in music. This subcluster also investigates the potential for AI to innovate within the music industry and its effects on human creativity.

5. **AI creativity:** This cluster focuses on the studies that had as primary goal to evaluate and enhance AI creativity. At the moment, generative models create music from a more or less restricted training pool of songs. However good the models may be, they are unlikely to invent anything new, with the risk that the quality of music will go flat and never evolve. The papers in this cluster address the problem and try to find a way to make the AI more creative.

The database considered when assigning papers to the clusters is the broader one, which includes 106 papers; however, in Graph 3.3 below, only the 85 articles coming from the last search are considered, since they are not biased by the author's internet search.

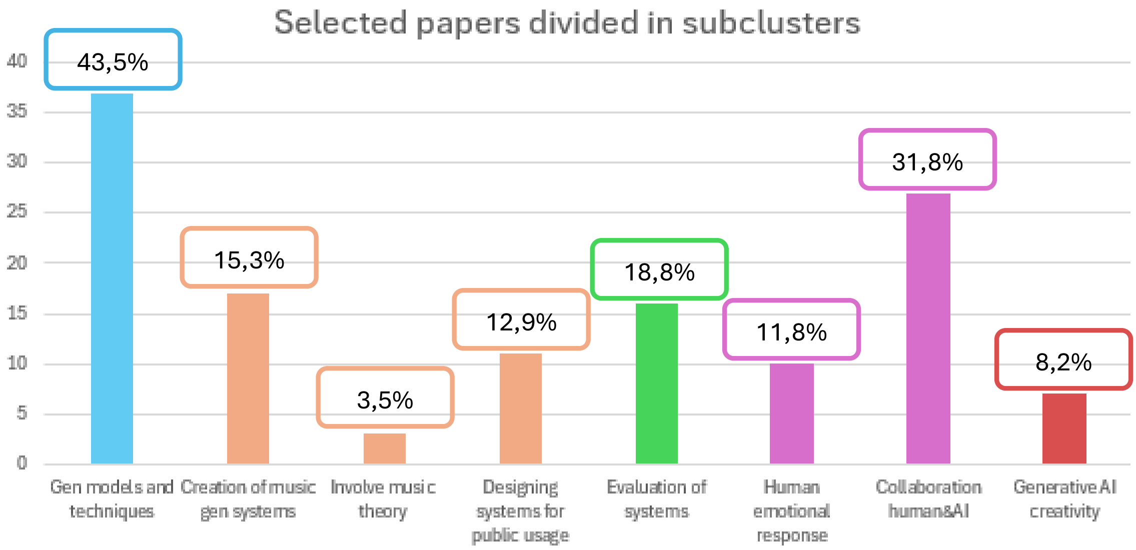
Moreover, the assignment to the clusters is not exclusive, therefore one paper can be attributed to more clusters at the same time.



GRAPH 3.1 - Selected papers divided into clusters.

First, for Graph 3.1 we can already get an understanding of which are the topics that, for the moment, have been studied more thoroughly. “Generative models and techniques” and “Human-AI collaboration” clusters have respectively 37 and 35 papers that address the matter, meaning that 43,5% and 41,2% of the selected papers have as one of their primaries focuses the exploration of AI techniques and how well they can collaborate with humans. This first insight could make us assume that, for this technology, we are still at a stage where we need to experiment and evaluate different techniques and do research on what use this innovation can have in society. This supposition is backed by the statistics of the other clusters: “Creation of music generation systems” has 26 papers, meaning that 30,6% of the selected papers addressed the matter as one of their primary focuses. The topic of creating a generative system is becoming more appealing to researchers, who are also starting to think of ways to evaluate the performances of those systems (topic addressed by 16 papers, 18,8% of the database) and to reflect on the AI’s ability to be creative (topics addressed by 7 papers, 8,2% of the database). Those last two clusters include topics that are based on research made for the other clusters (in order to evaluate the performance of a system, you need to study the technology and the goal and then try to build one). Therefore, the supposition is that, even though we are already moving steps towards assessing the AI’s creativity and quality, we are still looking for a dominant technology and design, as well as a way for Artificial Intelligence to effectively collaborate with humans towards the goal of creating new music.

In Graph 3.2, two clusters are broken down into subclusters that we got from the first skimming of abstracts.



GRAPH 3.2 - Selected papers divided into subclusters.

For the orange cluster, "Creation of music generation systems", two subclusters where identified: "Involve music theory" and "Designing systems for public usage". According to the graph, only 3 papers (3,5% of the database) took into account music theory, focusing on integrating it as much as possible into the system, while 11 papers (12,9%) focused on creating platforms or systems that could be used by the general public, as well as looking into the best features in terms of design and musical capabilities to add to the system. The third orange subcluster, titled "Creation of music gen systems", includes all publications (13, accounting for 15,3% of the database) whose study's intent was to create a music generator but did not fit into the preceding categories.

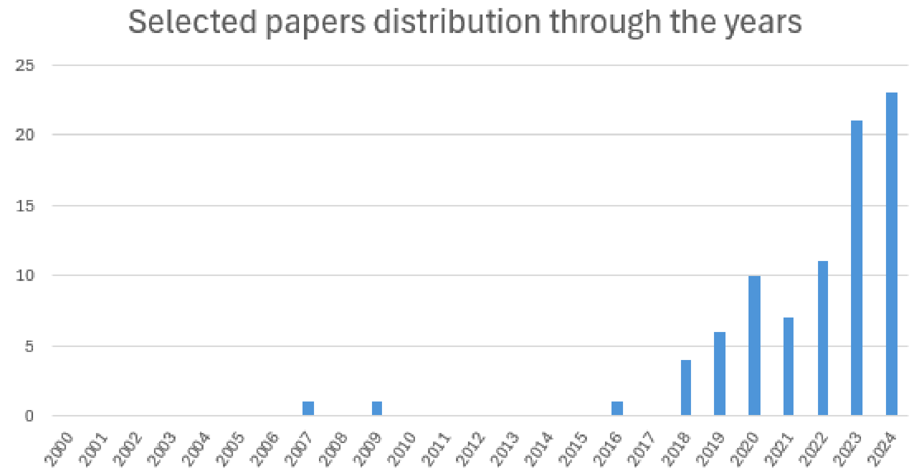
Switching to the pink cluster, "Human-AI collaboration", we were able to identify two subclusters: "Human emotional response" and "Collaboration human & AI". The first one counts 10 publications (11,8% of the database), while the second one comprises 27 publications (31,8% of the database).

3.2.3 Descriptive Analysis of literature

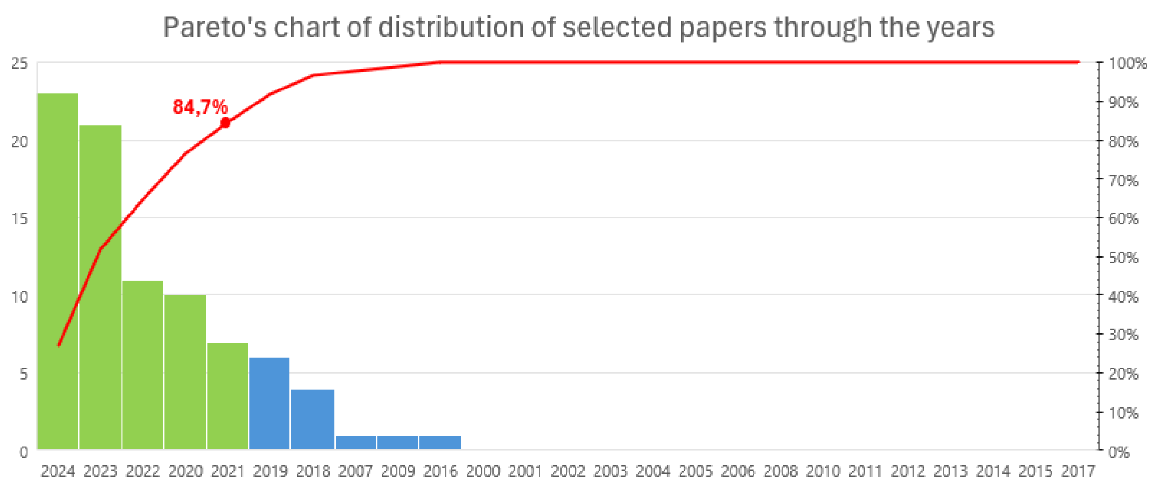
As stated before, the descriptive analysis only refers to the first database of 85 articles.

First of all, Graph 3.3 shows how the number of articles increases over time, reaching the peak in 2024 with 23 papers, despite the search being done in June. It can also be observed how relevant articles begin to be published after 2018, representing the 96,5% of the database and, as Graph 3.4 suggests, that 84,7% articles were written over the last 5 years.

remarking how much the interest in the AI for music generation has spiked lately. This information makes us think that, even though some researchers started studying the topic from 2000, we still are at an early stage for this technology.

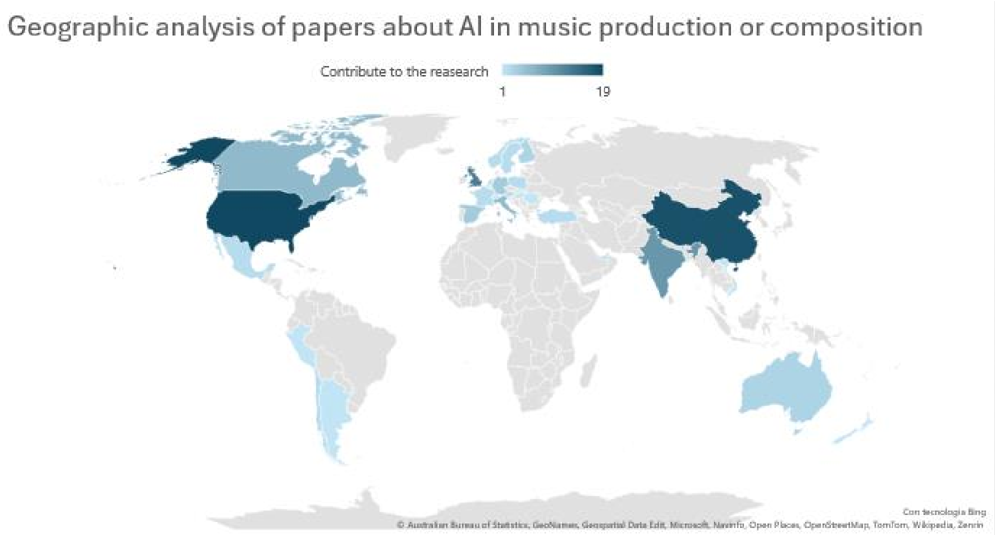


GRAPH 3.3 - Selected papers distribution through the years



GRAPH 3.4 – Pareto's chart of distribution of selected papers through the years

On the geographical spectrum, as showed in Graph 3.5, the US, China, the UK and India stand out as the most significant contributors with respectively 19, 18, 13 and 10 articles being funded and written in their countries' universities. This insight confirms the robust engagement on this topic from countries that lately are the top investors on cutting-edge technology and innovation. Following them, we can find Canada and Italy gave a contribution of 6 articles, Germany and Spain with 4 and Australia, Finland and Singapore with 3. The remaining articles come from countries mainly across Europe and South America, outlining a strong interest from all over the world in finding solutions for music generation with artificial intelligence.



GRAPH 3.5 – Geographic analysis of papers about AI in music production or composition

3.2.4 Content Analysis

A descriptive analysis of content, aimed at detecting the state-of-the-art, challenges and open points in every nuance of the AI generated music topic has been done. In this analysis, 21 publications from previous searches will be added to the database.

3.2.4.1 Generative models and techniques (44 articles)

In the introduction chapter we already had a brief overview of the main generative architectures used for composing music. This cluster will delve deeper into the technological applications, presenting the state-of-the-art that could be gleaned from our review.

First of all, we notice how 14 out of 44 papers (31,8%) of the cluster are systematic or comprehensive reviews, while the remaining papers experimented on new combinations of architectures, frameworks and use cases.

This work focuses on generative AI, which greatly differs from the traditional models, used for prediction, classification and regression. Some other differences between the two types are the learning approach and the data requirement, which can be supervised, unsupervised or semi-supervised and a better performance with medium-sized datasets for traditional models, and unsupervised or semi-supervised and a requirement of large amounts of data for effective learning for generative ones. (Rani et al., 2023)

Generative AI can be used for a handful of applications, which can include creative ones such as image, text, audio and music generation. As for now, music generation is the least developed application among them, primarily because of issues with time (a musical piece has to follow structural rules and have to reference what has been created before, therefore a generative system cannot only move forward) and with the complexity created by layers and elements (instruments, melody, voice, arrangement, lyrics, etc.). The composition process is complex and has many layers: the composer should write all the parts trying to create a harmonious piece which will have characteristics and rules based on the music theory and the music style chosen. The basic music elements are harmony, form and structure, melody and texture and finally instrumentation and orchestration, all of this without thinking about the potential lyrics and their connection to the melody, the sound design, mixing and mastering.

We follow Figure 3.5 as a structure to present the music composition process and the technologies that make it possible.

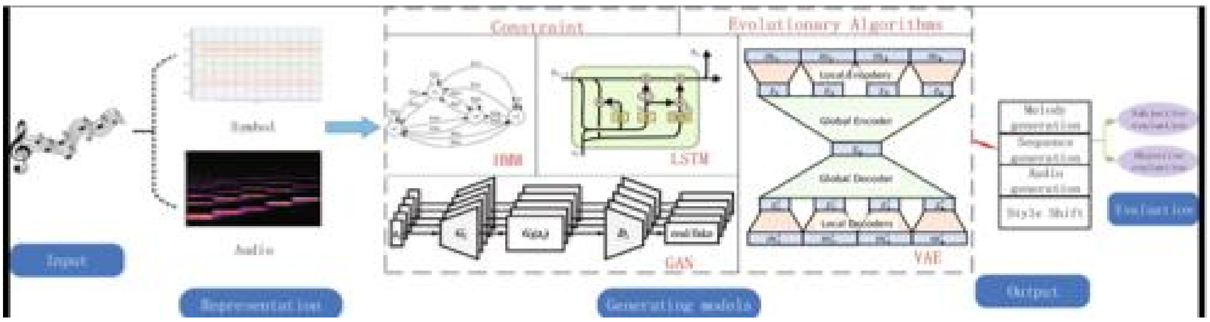


FIGURE 3.5 – Structure for the music composition process and technologies involved

1. Representation

   The input given to the generative system must have a format that can influence the training and the output. The formats are mainly two:

   - Audio (continuous signal). It can be waveform or spectrogram
   - Symbolic (Discrete signal). It can be MIDI, Piano-roll or ABC notation

   Moreover, for deep neural network-based generative systems, is equally important to select the appropriate dataset: some examples are Nsynth (Multitrack, WAV), Lakh MIDI (Multitrack, MIDI), JSB-Chorales (Harmonized chorale, MIDI), Groove MIDI (Drum, MIDI) and Nottingham (Folk tunes, ABC).

2. Music generation systems

   According to the purpose of the generation, the generators can be categorized as

   - Melody generation: can produce single melodies.
   - Arrangement generation: can generate harmonies, particularly in terms of chords and melodies that are musically pleasing and coherent.
   - Audio generation: generate sound segments.
   - Style transfer: tries to apply a style or an instrument to a musical composition.

Finally, we talk about the algorithms and architectures for music composition. Even though many applications actually use combinations and nesting, we first present them as follows:

I. Non-deep learning methods

   Contrary to deep learning, which learns patterns from data, these systems rely on algorithmic approaches, rule-based systems, and statistical models.

- Rule-based music generation: the composition process is based on a set of predefined rules, often derived from music theory. (L. Wang et al., 2024)
- Markov Models: probability-based generation models for processing time series data. It leverages the Markov chains, where the next note or musical element is chosen based on the current state or a short history, rather than the entire sequence of preceding notes. Since the generation is based on probability, it lacks

long-term coherence and usually leads to excessive repetition of segments. (L. Wang et al., 2024)

- Genetic Algorithms (Evolutionary Computation): they are inspired by natural evolution, and they can be optimized in the direction of high relative quality or adaptability. In fact, each algorithm has to have a problem domain, an individual representation, and a fitness measure. The fitness measure is what decides guides the selection process. (L. Wang et al., 2024)
- Grammar-based Methods: based on the linguistic grammar concept, consists of a set of production rules that recursively define how symbols (representing musical elements) can be combined and expanded to generate complex structures. Grammar-based systems are hierarchical and recursive. For example, a non-terminal symbol representing a musical phrase might expand into a sequence of chords, which then expand into individual notes. (Lopez-Rincon et al., 2018)

II. Deep learning methods

They involve neural networks with multiple layers that learn complex patterns from datasets through backpropagation and optimization. It is a subset of machine learning that can model highly non-linear relationships and, in order to do that, they need significant computational resources for training. Below are listed the most widely used architectures.

- Generative Adversarial Networks (GAN)

  It's almost safe to say that GANs are the state-of-the art method, and they surely are the most studied and praised at the moment. The architecture is based on two networks that play against each other: a generator and a discriminator, with the latter that has to determine if the generated data is fake or real while the former turns random noise into data. This adversarial strategy pushes the GAN to improve continuously. Despite being very successful and robust, it still suffers from some limitations such as mode collapse (where the generator can only produce a single type or a very limited number of outputs), non-convergence and instability (it could happen that the discriminator learns too quickly and can easily distinguish real and fake data; the process would therefore stall since the generator would not be able to learn from the feedback because the gradients passed to the generator during backpropagation may become very small (vanish); moreover, the high sensibility of GANs to hyperparameters means that even small changes could lead to drastically different results). (Bengesi et al., 2024; J. P. Briot, 2021)

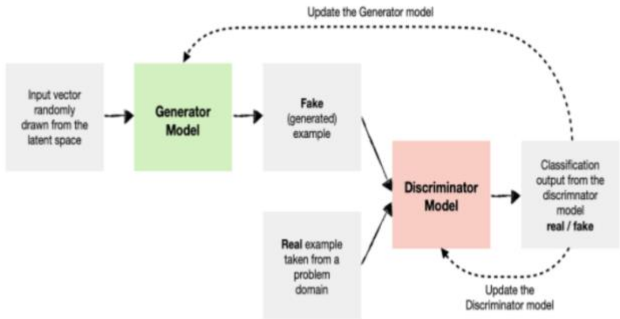


FIGURE 3.6 – GANs structure

- Variational Auto-Encoder (VAE)

  The VAE originates from the Autoencoder architecture, which is a refinement of a feedforward network with two constraints: one hidden layer and the number of output and input nodes has to be equal. The simple autoencoder suffers from discontinuity in the generation in the latent space, therefore the VAE is a refinement as it encodes an example as a probability distribution over the latent space, and not as a single point. This change ensures completeness and continuity. A latent space is a multidimensional space of compressed data where the most similar elements are placed next to each other. VAEs are good at creating outputs similar to the input and are based on an encoder-decoder structure. The encoder converts the input into the likelihood distribution in a latent space, capturing the data's fundamental structure; then the decoder, using samples from the distribution, produces an output similar to the input. (Hernandez-Olivan & Beltran, 2022; Pathariya et al., 2024). Generating music with VAEs is challenging, especially because the model can have problems generating music that stays coherent over time and capturing and reproducing long-term dependencies. Moreover, because of the complex nature of music, the model might produce something that is theoretically correct but lacks musical value.

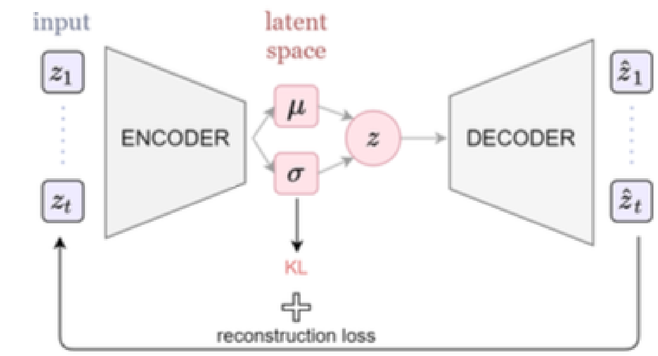


FIGURE 3.7 – VAE structure

- Transformer

  Transformers use the self-attention mechanism, which analyzes data sequentially. They consist of an encoder and decoder framework where each layer of the model incorporates feed-forward neural networks and multi-head self-attention mechanisms. They allow for parallel data computing, visualization of self-reference, and they solve the problem of long-term dependence and continuity better than RNN. e.g. MuseNet. (Pathariya et al., 2024; L. Wang et al., 2024)

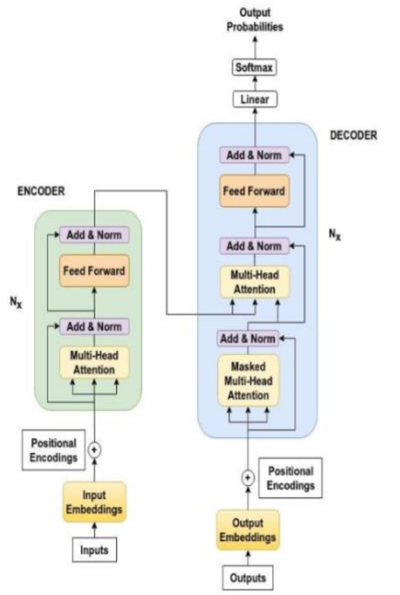


FIGURE 3.8 – Transformers structure

- Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)

  These two types of Neural Networks are two broader architectures, often used together with other architectures mentioned before. CNN is a neural network designed for processing and analyzing data with a grid-like topology, in 2 dimensions, therefore not the perfect solution for generating music (which is time-dependent and multi-dimensional). However, used together with other architecture can help create harmonies, doing style transfers or generating RAW audio.

RNN, on the other hand, are a class of neural networks for processing time series. However, they cannot solve the long-time dependency problem. A RNN variant, LSTM (Long Short-Term Memory) uses protected memory cells and gated mechanisms that control the flow of information to prevent issues like vanishing or exploding gradients during training. This design makes LSTM particularly powerful for tasks involving time series data, such as melody prediction, due to their strong temporal learning abilities. (W. Wang, 2023)

Apart from these most common architectures, for completeness we mention the Normalizing Flows (where a series of invertible transformations are applied to the input data and are trained to map simple distributions to the complex distribution of the target data. Moreover, since the transformations are invertible, NF can both generate data and estimate the likelihood of data) (Ji et al., 2023), the Diffusion Model (DM) which works by modeling the process of gradually adding noise to data and then learning to reverse this process to generate new data. The model learns a step-by-step denoising process, which allows it to generate new samples from noise. (Bengesi et al., 2024)

As the reader could already understand, these architectures alone are not sufficient to create good quality music, so researchers try to combine different architectures and models to overcome some limits and increase the performance of the system. They are called compound architectures, which comprehend many types:

- Composition
  Several architectures of the same or different type e.g. C-RNN-GAN or Music VAE
- Refinement
  One architecture is specialized through additional constraints e.g. the VAE itself
- Nesting

An architecture is nested into the other one e.g. Music VAE is a Recurrent autoencoder architecture where an RNN is nester within the autoencoder

- Pattern

  An architectural pattern is instantiated onto a given architecture(s) e.g. the MidiNet architecture, where GAN pattern is instantiated onto two convolutional feedforward architectures, on which a conditional pattern is instantiated.

(J. P. Briot, 2021)

Some of the articles from this cluster experimented on different compositions of architectures, trying to find new models to improve performance. Below we list some examples: the usage of LSTM networks with the adoption of a grey wolf optimizer(grey wolf), to better control the hyperparameters, the composition of Guzheng music using LSTM and Reinforcement Learning (guzheng music), the VAE “Latent Chords” that generated chords and chord sequences (latent chords), the zero-shot singing voice conversion method based on VITS model (text-to-speech) and Glow (Normalizing Flows) and many others (zero-shot). In the same way we can recognize some variants of GAN such as cGAN (Conditional Generative Adversarial Network) which incorporates extra information such as class labels or style attributes, DCGAN (Deep Convolutional GAN) (Bengesi et al., 2024), Creative Adversarial Networks (CAN) which is an extension of GAN: when getting feedback from the discriminator, the generator gets two signals instead of one, the first is the same of the normal GAN architecture, which specifies how the discriminator believes that the generated data is real or fake; while the second one is about how easily the discriminator can classify the generated item into established styles. The goal of the generator is to create original pieces that do not get discovered as fake while staying close to the distribution of existing art pieces. (J.-P. Briot & Pachet, 2020)

Finally, we can state that, although there is not a specific NN architecture that performs better than others, transformers and GANs are emerging as the best alternative. However, combinations of different models always work best as they bridge each other’s gaps, sometimes even combinations of DL with probabilities methods.

Moreover, best alternatives are still heavily linked to the output we want to obtain. (Hernandez-Olivan & Beltran, 2022)

On the side, two articles present the first proofs-of-concept of Unconventional Computing (UC) technologies in music composition. UC technologies explore non-digital ways of data storage, processing, input, and output with paradigms such as Biocomputing and Quantum Computing that delve into domains beyond the binary bit to handle complex non-linear functions. In their article on using quantum computing, Miranda et al. Present “Quanthoven”, which demonstrates that is possible to use a quantum computer to classify music and how this capability can be leveraged to develop a system that composes meaningful musical pieces. After, they show the techniques developed to encode musical compositions as quantum circuits and how to design a quantum classifier (quantum computing). On the other hand, Venkatesh et al. show how they harnessed *Physarum polycephalum* as a memristor to process and generate creative data for popular music. The organism works as a collaborator in the process of composing a song titled “Creep into my Lawn” (Venkatesh et al., 2020). The innovative contribution of these two articles is the fact that UC technologies explore analog forms of computation and storage. The traditional unit is the binary bit (0 or 1), while these methods use qubits and a plasmodial slime mold.

In the end, it would be interesting for us to know the readiness of GenAI to compose music, in order to know what to expect in the future and to start hypothesizing market trends and changes. The Technology Readiness Levels (TRL) is a system, first presented by Mankins in 1995, that assesses the maturity of a technology: the scale consists of 9 levels going from a mere idea (1) to the full deployment on the market (9) (John C. Mankins, 1995). Martinez-Plumed et al. in 2020 tried to assess the levels of various AI technologies considering different levels of capability: each technology is at a different level depending on the breadth of the application, therefore more niche applications will have higher TRL, which will decrease while it becomes more general. In the article the authors study 8 different categories for a total of 12 applications, but we are interested only in the “Knowledge Representation & Reasoning” category, and more specifically, in the “Recommender Systems” (Figure 3.9). The fourth level comprises innovations such as recommending new items that do not exist and should be created to fill missing needs. In 2020, the article stated that this application was relatively new but that it already included some validated proof-of-concept systems (TRL 2 to TRL 4), including music generation. (Martínez-Plumed et al., 2021) However, the past four years have substantially changed the rules of the game, with some big models such as Magenta, AIVA, Suno, Udio that are currently on the market. According to Table 3.1 and these last considerations, we must assign to AI music composition at least TRL 7.

Environment	Goal	Product/ Evaluation	Outputs	TRL	Description
			Scientific articles published on the principles of the new technology	TRL 1	Basic principles observed
			Publications or references highlighting the applications of the new technology.	TRL 2	Technology concept formulated
		Proof of concept	Measurement of parameters in the laboratory	TRL 3	Experimental proof of concept
Laboratory	Research		Results of tests carried out in the laboratory.	TRL 4	Technology validated in lab
			Components validated in a relevant environment.	TRL 5	Technology validated in relevant environment
Simulation	Development		Results of tests carried out at the prototype in a relevant environment.	TRL 6	Technology demonstrated in relevant environment
		Prototype	Result of the prototype level tests carried out in the operating environment.	TRL 7	System prototype demonstration in operational environment
		Commercial (certified) product	Results of system tests in final configuration.	TRL 8	System complete and qualified
Operational	Implementation	Deployed product	Final reports in working condition or actual mission.	TRL 9	Actual system proven in operational environment

TABLE 3.1 - Technology Readiness Levels

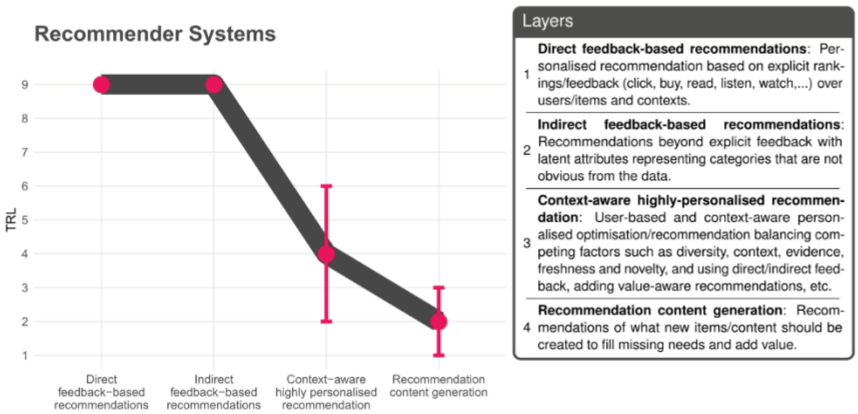
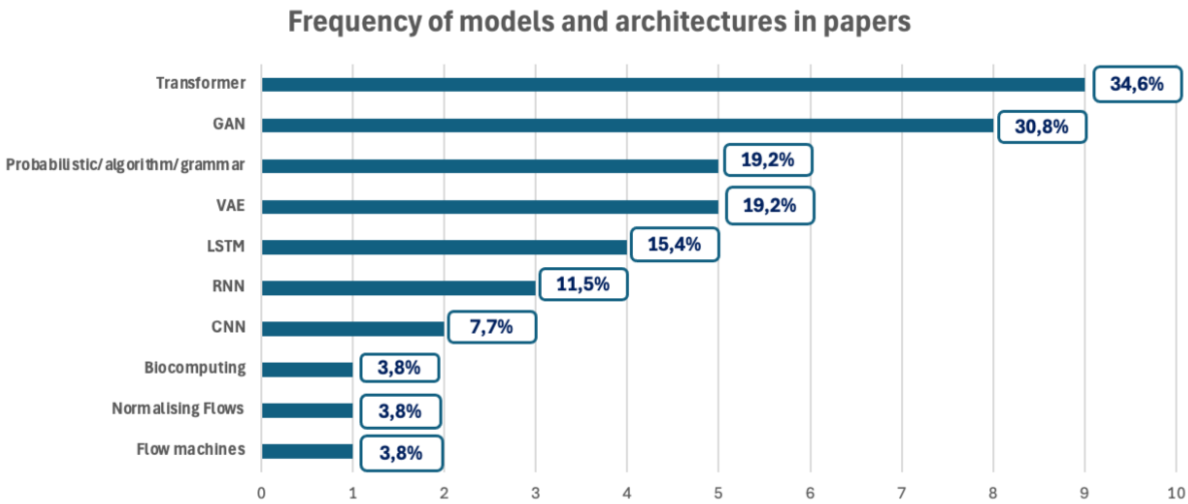


FIGURE 3.9 – Recommender systems

3.2.4.2 Creation of music generation systems (34 articles)

The analysis of this cluster will be conducted on the technology used to create the systems that belong to it. In order to try to answer the question "which are the most used technologies?" We will study the architectures and models used and the representation used for the datasets. Unfortunately, not every paper gave an explicit response to all three, therefore we will also include the number of missing data.

In the graph below, we can see how Transformers and GANs are the most utilized architectures, used respectively 34,6% and 30,8% of the time, reflecting the fact that they are the most advanced architectures at the moment, as we discovered during the analysis of the first cluster.



GRAPH 3.6 – Frequency of models and architectures in papers

Another element that stands out is the fact that 5 papers out of 26 (19,2%) have used non-deep learning models to create their generative systems. This data point is made of the sum of probabilistic or grammar models and algorithms and it hints at the fact that it is still one of the best methods to use due to their low data requirement, simplicity of programming and the fact that they are controllable. Throughout this cluster we came across with the concept of control various times, in particular how humans like the systems more when they can have control on every detail of the songs they are creating, as stated by Boryczka in their paper about ACO systems (Boryczka et al., 2023) where they emphasized the importance of configurability of the systems' parameters and allowed for adjustments. These traditional models allow for more control, given the fact that we know how they work and the rules that they have to follow, opposed to deep learning models that are unpredictable; therefore, for many applications, they are still favoured.

Moreover, from this analysis we could notice the large presence of hybrid models which, as seen in the previous cluster, are the best choice a generative system creator could make since the models fill each other's gaps and problems. For instance, combining

LSTM and GANs helps modelling long-term dependencies and generating new data, producing sequences that are more realistic and coherent, as seen in 4 papers: in fact, every time that LSTM was used, it was as generator or discriminator of a GAN architecture. (Do Quang & Hoang, 2023; Gowri Ganesh & Venkata Vara Prasad, 2023; Hou, 2022; J. Lu & Eirinaki, 2021)

In fact, Do Quang and Hoang demonstrated an effective method that uses a combination of GAN and LSTM models that yields competitive results in both qualitative and quantitative evaluations. With their method, the GAN manages the overall harmony, while the LSTM strengthens the connections between adjacent notes. Through the training process, the generator learns key features to embed throughout the composition rather than just within isolated groups of notes; meanwhile LSTM's sequential generation, enhanced by attention mechanisms that enhances the connections by providing extra memory storage, improves notes connections and continuity. As a result, the GAN generator integrates this linking capability through its interaction with the LSTM during training. (Do Quang & Hoang, 2023)

Finally, for the papers that described it, we analysed the type of dataset that the models were using: out of 12 papers that explicitly described their dataset, 10 had a MIDI-based one, while the others were XML-based and in GuitarPro format.

To continue the analysis on generative systems, we identified two subclusters (already included in the previous analysis) that focus on more specific topics: the involvement of music theory in the training of models and the study of models' interactions with humans through interfaces.

3.2.4.2.1 Involvement of music theory (5 articles)

In this cluster we analyse how integrating music theory into the generation models can enhance music generation in quality, coherence, and expressiveness of the generated music.

In their article on probabilistic models for melodic prediction, Paiement et al. affirm that understanding chord progressions is crucial for generating harmonically coherent music. This approach is grounded in set theory and tonal harmony and can improve the ability to generate long-term music structures, focusing on the relationships between chords and melodies, leading to more musically pleasing and theoretically sound output.
Delving a bit deeper, a chord is a group of three or more notes, while a chord progression is simply a sequence of chords. In probabilistic terms, a given chord can be seen as a latent variable that influences the probability of selecting certain notes in other musical components, such as melodies or accompaniments. Among the possible chord representations, according to Paiement et al.'s experiments, representing chords by their roots appears to be a good compromise (Paiement et al., 2009). The probabilistic chord progress model was also used by Bian et al. while building the MoMusic system: it features a partially randomized harmonic sequencing model based on a probabilistic analysis of tonal chord progressions, mathematically abstracted through musical set theory (Bian et al., 2023).

Apart from using chord progressions in order to have a better control on the output and a better performance, Guo et al. employed musically meaningful control tokens to add to the original input of MusIAC, even though they were thought for infilling. As track level controls: track's note density rate $\#notes/timesteps_{total}$, track's note polyphony rate $timesteps_{polynote}/timesteps_{anynote}$, track's note occupation rate $timesteps_{anynote}/timesteps_{total}$. On the other hand, as bar level controls: tensile strain of the notes in that bar: $\sum_{i=1}^{n}(note_{pos}[i] - key_{pos})/n$, cloud diameter of the notes in that bar: $max_{i\in[1\ldots n-1],j\in[i+1\ldots n]}(note_{pos}[i] - note_{pos}[j])$. Additionally, the following elements are added to the input: key of the song, tempo, time signature and track's instruments (Guo et al., 2022).

Finally, it should be noted that two out of the three generative systems that fall in this category (MoMusic and Blues for Gary) (Bian et al., 2023; Keller et al., 2007) were developed for real-time usage, respectively for motion-driven music composition and performance and to help students learn jazz improvisation; on the other hand, MusIAC was created to do music infilling (Guo et al., 2022).

3.2.4.2.2 Designing systems for public usage (14 articles)

In this section, we explore the design, interfaces and features of some generative systems, especially focusing on the models that have been made available to the public or that at least have been tested and evaluated.

The most recurring theme among the papers is the need for human-centered and user-friendly design approaches, which derive from a collaborative integration of UX and ML. One of the goals of leveraging AI is to democratize music generation, making music generative systems accessible and user-friendly so that even users how do not have adequate training, money or computational capacity can create music. Both "Composing computer-generated music using IGME" (2020) and "Blues for Gary" (2007) emphasize the need for interfaces that are intuitive and familiar, and IGME achieves this by integrating generative music techniques into a linear music sequencer, resembling traditional music software like Logic Pro, lowering the barrier for musicians to experiment with algorithmic composition. (Hunt et al., 2020; Keller et al., 2007)

Similarly, the ASCIML prototype discussed in the "Exploring the potential of interactive Machine Learning for Sound Generation" (2023) article highlights the value of interactive machine learning (IML) systems that allow users to design and train models without needing deep ML knowledge. This system allows you to create your own dataset and to get immediate auditory feedback for your creation, empowering users to iteratively refine their creations. Users claimed that getting more auditory information was beneficial and helped understanding and utilizing the tool better. The study was conducted on musicians, who claimed that they preferred using sing microphone recording and synthesis over loading pre-existing audio files to create their datasets. (Meza, 2023) An opposing view comes from the music creating by example study, which is focused on users that do not have any musical expertise, but simply need to use music for their videos and content. The study revealed that the users preferred to give to the system a selection of example songs, opposed to just one or to them humming a melody. This is because they still wanted to have creative control over the generation, also thanks to a grid interface that allowed for mixing and matching of different AI-generated tracks, empowering them to get personalized and customized sounds, without the necessity of having an idea or musical talent. (Frid et al., 2020)

The importance of providing users with control over the AI-generated content is also shown by MusIAC, MuseCoco, and the DADAGP dataset, who allow the users to manipulate some musical attributes like track density, polyphony, and tonal tension. MusIAC introduces control tokens within a generative framework for music infilling, enabling composers to fine-tune the AI-generated segments to fit seamlessly into existing compositions, enhancing the overall musical coherence. (Guo et al., 2022) MuseCoco takes this one step further by separating text-to-attribute understanding from attribute-to-music generation, offering precise control over musical elements such as tempo, key, and instrumentation. (P. Lu et al., 2023) Along the same lines, Sarmento et al. discuss how the tokenization approach in the DADAGP dataset, inspired by event-based MIDI encodings, allows for precise manipulation and generation of music using models like transformers. (Sarmento et al., 2021)

Another recurring theme for the system to be user-friendly, is the importance of integrating AI tools into existing music production workflows, ensuring that they complement and adapt to existing systems rather than disrupt established practices. Giuliani at al. do it with MusiComb, which focuses on sample-based music generation that integrates seamlessly with digital audio workstations (DAWs) and sample libraries (Giuliani et al., 2023). This approach reduces computational demands and hardware requirements, making the system more accessible and easier to adopt for professional music producers.

To conclude, developers who study and create these systems would highly benefit from multiple actors contributing to creative generative machine learning interfaces: in order for the interfaces to be user-friendly, ML scientists should collaborate in close contact with UX and design teams or get feedback from users like Jordà et al. did, to more effectively identify human needs (Jordà et al., 2016). The People and AI Research (PAIR) group instituted the PAIR Bungee program, a method that embeds three UXers into an ML research host team for three months. During that time, UXers receive training on basic

ML concepts and provide UX input and direction to jump-start the host team's ML projects. One of the first projects that were developed with this new method was the Magenta Project, which is now embedded in Ableton Live. (Kayacik et al., 2019)

3.2.4.3 Evaluation of systems (21 articles)

After building models and architecture that have the power of generating music from scratch, it is important to define a way to evaluate the systems and their output in order to assess how advanced the state of the art is and whether or not the generated music is of good quality. The ideal method of evaluation would be an autonomous one, where the machine could automatically evaluate itself and its output to find out the deficiencies and give valuable feedback, thereby improving the quality of the generated music. However, music is too subjective and full of emotion, and even though it may seem quite easy for a human, AI does not have a way to determine if something is musically appealing or is just following the rules. Moreover, there are no unified evaluation criteria established (Ji et al., 2023), in part also because they have to be specific for the model and purpose (Hernandez-Olivan & Beltran, 2022); therefore, in this section we will describe the most used and effective methods of evaluation.

As a general scheme, we can refer to figure 3.10 from "A comprehensive survey for evaluation methodologies of AI-generated Music" (Xiong et al., 2023), which identifies two primary classifications: subjective and objective evaluations, with a third category that comprehends the combination of the two and heuristic methods. All the articles that at one point have evaluated their model, have used either a subjective or an objective evaluation, with more than a half using the combined version. However, since there is no unified framework, each paper had to define its own method and parameters, which makes it hard to compare the results.

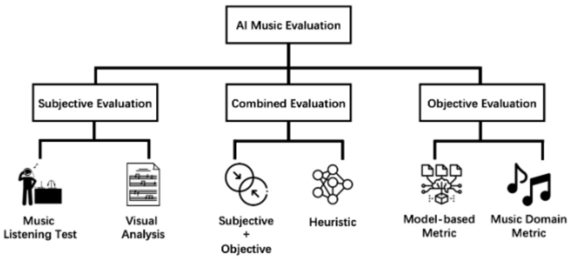


FIGURE 3.10 – AI music Evaluation

- **Subjective Evaluation**

Also called qualitative evaluation, it relies on assessments of human's satisfaction. Even though it is more resource-intensive and less reproducible, it is a fundamental part to evaluate quality. Among all the model, there are two most important categories:

<u>Music Listening tests</u>

These methods imply the participation to the test of subjects with evenly distributed music knowledge and the listening to musical pieces. There are different types of listening test, but we can identify two approaches: the Turing test (a test where it is asked to the subjects to state whether, according to them, the piece was AI or human generated; usually t-test or h-test are then performed), or subjective query metrics (where the researchers ask to evaluate or recognize different things such as the identification of music type, giving a score or a rank based on several evaluation metrics and parameters). The songs are usually rated using a Likert scale and, apart from the question "How much do you like this song?" researchers could ask to evaluate according to melody (pleasantness and naturalness), rhythm (accuracy and well-formedness), tonality (consonance of chords) or musicality (overall coherence). (Zhu et al., 2024). Another example of a subjective study is presented by Wu and Yang (Wu & Yang, 2020), who asked to rate the pieces in a five-point Likert scale on Overall quality (Does it sound good overall?), Impression (Can you remember a certain part or the melody?), Structuredness (Does it involve recurring music ideas, clear phrases, and coherent sections?) and Richness (Is the music diverse and interesting?).

Sometimes listening tests only invite professionals to ask them questions on music theory, which require a strong background; for example, Chen et al. even though they classified it as objective, made an assessment inviting music experts and asking them to evaluate the similarity between the musical instrument orbital note sequence and the chord and to analyse the interval relationship between notes. (H. Chen et al., 2019)

Visual Analysis

For these types of tests, the presence of music experts is mandatory since it is required to analyse the visual music representation of the generated music (could be musical score, piano roll, chord progression, waveform, spectrogram…)

When talking about subjective evaluation, although it is still indispensable because of the artistic subjectivity of music, the resources spent are enormous and the tests are hard to reproduce. However, apart from the parameters considered in the subjective query metrics tests, this evaluation is mostly standard.

- **Objective evaluation**

This is a quantitative evaluation of the model and the generated music, performed automatically. In this case the metrics of the assessment vary according to the papers and who created the test so we will report the most frequent. Generally, it is possible to identify two categories, based on what the parameters have to evaluate: the machine or the song.

Model-based Metrics

These metrics do not have a strong universality due to the fact that they have to be defined for the specific model that they will have to evaluate since it would be a problem to compare models trained for different purposes and with different datasets. However, we can present the most common metrics used in state-of-the-art models. Hernandez-Olivan and Beltran identify in their article some metrics used to compare different Deep Learning model built for the same purpose: Loss (difference between the inputs and outputs of a model from a mathematical point of view), Perplexity (generalization capability that a model has), the BLEU score, the precision, Recall or F-score. (Hernandez-Olivan & Beltran, 2021)

Music Domain Metrics

They refer to metrics that come from the music theory domain, also known as musical statistical descriptors. Many methods compare the real music's statistics and the AI-generated ones since, ideally, they should be very close.

Since the development of music generation, researchers have designed numerous methods and metrics, therefore here we will only point out the most common ones, organized according to the categories proposed by Ji et al. (Ji et al., 2023):

1 - 2. Pitch & Rhythm related Metrics

For the state-of-the art metric design, Yang and Lerch (Yang & Lerch, 2020) proposed a set of metrics that were validated in experiments and are reproducible. The proposed features include pitch counts, pitch category histograms, pitch shift matrices, pitch spans, average pitch intervals, note counts, average repetition intervals, note length histograms, and note length shift matrices.

3. Harmony related Metrics

These metrics focus on measuring harmonic consistency, chord histogram entropy, chord coverage, polyphony, tone span and others to measure the compatibility and musicality of generated outputs.

4. Style related Metrics

The most common metrics are "Style fit", that measures how well the generated music fits the desired style, and "Content preservation", which measures how much content was kept from the original.

Type	Metrics	Definition
Pitch	Empty Bar (EB) [30]	the ratio of empty bars (in %).
	Used Pitch Class (UPC) [30]	the number of used pitch classes per bar (from 0 to 12).
	Pitch Range (PR) [125]	the subtraction of the highest and lowest used pitch in semitones.
	Number of Unique Note Pitches/Durations [41]	the count of how many different pitches/durations are used in a music piece.
	Pitch Class Histogram Entropy [129]	construct the 12-dimensional pitch class histogram $\vec{h}$, normalized by the total note count in the period such that $\sum_i h_i = 1$. Then, calculate the entropy of $\vec{h}$: $$\mathcal{H}(\vec{h}) = -\sum_{i=0}^{11} h_i log_2(h_i)$$
	Qualified Note (QN) [30]	the ratio of "qualified" notes (in %). Dong et al. [30] consider a note no shorter than three timesteps (i.e., a 32nd note) as a qualified note.
	Polyphony [95]	the average number of pitches being played at the same time, evaluated only at time steps where at least one pitch is on.
	Scale Consistency [60]	the fraction of tones that are part of a standard scale, reporting the number for the best matching such scale.
	Repetitions [60]	the repetitions of short subsequences, giving a score on how much recurrence there is in a sample.
	Consecutive Pitch Repetitions (CPR) [68]	the frequency of occurrences of ℓ consecutive pitch repetitions for a specified length ℓ.
	Durations of Pitch Repetitions (DPR) [68]	the frequency of pitch repetitions that last at least d long in total for a specified duration d.
	Tone Span [60]	the number of half-tone steps between the lowest and the highest tone in a sample.
	Tone Spans (TS) [68]	the frequency of pitch changes that span more than d half-steps for a specified tone distance d.
	Average Pitch Interval (PI) [125]	the average value of the interval between two consecutive pitches in semitones.
Rhythm	Drum Pattern (DP) [30]	the ratio of notes in 8- or 16-beat patterns, common ones for Rock songs in 4/4 time (in %).
	Qualified Rhythm Frequency (QR) [68]	the frequency of note durations within valid beat ratios of {1,1/2,1/4,1/8,1/16}, their dotted and triplet counterparts, and any tied combination of two valid ratios.
	Rhythmic Intensity Score [95]	the percentage of sub-beats with at least one note onset: $$s^{rhym} = \frac{1}{B}\sum_{b=1}^{B} 1(n_{onset,\ b} \geq 1)$$ where B is the number of sub-beats in a bar and $1(\cdot)$ is the indicator function.
	Grooving Pattern Similarity [129]	the grooving pattern represents the positions in a bar at which there is at least a note onset, and the similarity between a pair of grooving patterns $\vec{g}^a,\ \vec{g}^b$ as: $$\mathcal{GS}(\vec{g}^a,\ \vec{g}^b) = 1 - \frac{1}{Q}\sum_{i=0}^{Q-1} XOR\left(g_i^a, g_i^b\right)$$ where Q is the dimensionality of $\vec{g}^a,\ \vec{g}^b$, and $XOR(\cdot,\ \cdot)$ is the exclusive OR operation. Note that the value of $\mathcal{GS}(\cdot,\ \cdot)$ would always lie in between 0 and 1.
Harmony	Tonal Distance (TD) [30]	the harmonicity between a pair of tracks. Larger TD [128] implies weaker inter-track harmonic relations.
	Chord Progression Irregularity [129]	the percentage of unique chord trigrams in the chord progression of an entire piece.
	Chord histogram entropy (CHE) [73]	given a chord sequence, create a histogram of chord occurrences with $\|C\|$ bins. Then, normalize the counts to sum to 1, and calculate its entropy: $$H = -\sum_{i=1}^{\|C\|} p_i log p_i$$ where p_i is the relative probability of the i-th bin. The entropy is greatest when the histogram follows a uniform distribution, and lowest when the chord sequence uses only one chord throughout.
	Chord coverage (CC) [73]	the number of chord labels with non-zero counts in the chord histogram in a chord sequence.
	Chord tonal distance (CTD) [73]	the average value of the tonal distance computed between every pair of adjacent chords in a chord sequence. The CTD is highest when there are abrupt changes in the chord progression (e.g., from C chord to B chord).
	Chord tone to non-chord tone ratio (CTnCTR) [73]	define CTnCTR as $\frac{n_c+n_p}{n_c+n_n}$, n_c is the number of the chord tones, n_n is the number of the non-chord tones, n_p is the number of a subset of non-chord tones that are two semitones within the notes which are right after them, where subscript p denotes a "proper" non-chord tone. CTnCTR equals one when there are no non-chord tones at all, or when $n_p = n_n$.
	Pitch consonance score (PCS) [73]	computed by averaging these consonance scores across a 16th-note windows, excluding rest periods, where the consonance score is set to 1 for consonance intervals including unison, major/minor 3rd, perfect 5th, major/minor 6th, set to 0 for a perfect 4th, and set to -1 for other intervals, which are considered dissonant.
	Melody-chord tonal distance (MCTD) [73]	the average of the tonal distance between every melody note and the corresponding chord label calculated across a melody sequence, with each distance weighted by the duration of the corresponding melody note.

Some other metrics that are worth mentioning are the Overlapping Area (OA) and the Kullback-Leibler Divergence (KLD) which were adopted to assess the degree of harmony between melody and accompaniment or to capture local self-similar patterns of generated melody sequences. (Ji et al., 2023)

Other methods

There are also some other methods that evaluate more general aspects that could include the structure or the originality (referred to plagiarism). While we will talk later about evaluating originality, we mention the LZ ratio as a metric to evaluate structural repetition. Defined by Chen et al. (A. Chen & Greer, 2023), Lempel-Ziv Dictionary LZ(c) represents the dictionary generated by the IPMotif algorithm when applied to a composition. The dictionary length is L(c)= |LZ(c)|, where |LZ(c)| is the number of entries in the LZ dictionary. As the composition grows in length, the piece could repeat existing material or always present something new, making a small or big L(c). In order to compare the metric across different compositions L(c) has to be normalized to the length of the composition, thus defining the LZ ratio as R(c)= L(c) / |c| where |c| is the number of tokens in the composition. This metric can be used for semi-supervised model tuning, where if a system generates a composition with LZ ratio A and a human listener thinks it is lacking structural repetition (large scale), the system can continue generating until it produces a composition with LZ ratio <A. On the other hand, if the system creates a composition with LZ ratio B and the human listener thinks it is too repetitive (small scale), the system can continue generating until it reaches a ratio >B.

- Combined Evaluation

Considering all that was said before about the advantages and challenges of both subjective and objective evaluation, combining the two methods seems the best possible approach to do the assessment. However, the problem of the diversity of the databases still remains and there is no uniformity in the interpretable migration of the quantitative assessment compared to the qualitative one. We can take as an example the evaluation of the GAN created by Liang for their paper on Xi'an Drum music: they made an objective evaluation using Empty Bars rate, used Pitch Classes and proportion of "Qualified" Notes with 32 diaeresis as the basic unit as indicators (Liang et al., 2023). They calculated the indicators from the samples created by their model and compared them with the indicators from the models that they wanted to compare it against (real music, CNN-RNN, MidiNet). Moreover, they adopted a subjective evaluation asking the respondents to conduct a ten points evaluation from four perspectives: melody, rhythm, pleasure and fluency.

Heuristic Evaluation

In their article Xiong et al. present a heuristic framework(Xiong et al., 2023), first proposed by Dervakos et al. (Dervakos et al., 2020), to calculate the frequency of different features by using a tool called "five-degree circle" and to get a quantitative score of each metric. The heuristic attributes were based on intuition and empirical observations as musicality but due to interpretability limitations, the author still had to rely on a subjective assessment.

All of the previous discussion prevalently focuses on the instrumental part or to the compositions in general, never lingering on the lyrics. In the article "Can a machine win a Grammy? An evaluation of AI-generated song lyrics", Lu and Eirinaki conducted both a quantitative and a qualitative analysis of the lyrics generated by GPT-2 and LSTM-based deep learning models. In order to define who was the best model, they analysed three different genres (pop, hip-hop and rock) and identified three categories to evaluate the generated songs: Quality (number of words and syllables in a song), Rhyme density and Sentiment Analysis. The final score will be determined by the sum of these three parameters: the perfect score was set at 100 with a max of 20 for lyrics quality and 40 for rhyme density and sentiment. The interesting part is that the parameters were defined in a way that it is possible for the machine to give a score autonomously and. The higher the score, the closer the model is to human-composed songs. (J. Lu & Eirinaki, 2021)

As we briefly mentioned before, trying to evaluate the creativity, to which we will dedicate an entire cluster, is crucial. Some of the most used criteria are novelty, originality and value, which at the moment are very difficult aspects to capture, but they are not a standard definition. Therefore, developing methods to evaluate creativity effectively is a significant challenge.

To conclude, the evaluation of models and compositions is complicated and not standardized because of the many studies that define different parameters from each other, making it difficult to compare music generated in different studies. Also, there is a lack of correlation between quantitative metrics and subjective evaluation (since objective measures are based on mathematical concepts that may not necessarily correlate with human perception of quality). It would be beneficial to achieve a set of standardized metrics, both subjective and objective, that could be applicable across different genres and AI models. (Xiong et al., 2023)

3.2.4.4 Human-AI collaboration

This cluster is composed of two subclusters that analyse, on one hand the sentiment and the acceptance of AI-generated music, investigating the prejudices and the concept of artistry, on the other the ways in which humans and AI con collaborate.

3.2.4.4.1 Human emotional response (15 articles)

While studying the generative AI systems, we still haven't addressed the human sentiment towards them. The major questions that seek a resolution are whether the human mind is willing to consider AI-created music as creative, as art worth listening to and if it has a prejudice associated with it. Many authors tried to answer these questions to try to define the human standing and possible acceptance of its applications.

Not all the papers that fall under this category address the topic only from a musical point of view, but some of them studied humans' attitude towards creative art in general. We decided to keep those papers as part of our analysis since they could give us some indication of what the general sentiment and practices are and to try and make a comparison between general creative arts and music.

Generally speaking, the tests most used to study the sentiment are the tests already described as subjective evaluation in the previous chapter: Turing test, giving a score of liking, stating whether or not they see AI as an artist… the researchers then add some other variables to study more precisely the behaviour. However, to this day we do not have a unified opinion on this matter.

There is a lot of scepticism and prejudices against AI in music and creative art in general, where people have significantly more negative attitudes towards using AI compared to other fields like security, defence forces, medicine, etc. (Latikka et al., 2023). In particular, people tend to have a bias regarding AI's ability to replicate human emotional depth and creativity. The problem then shifts towards the humans' capability of accepting AI as something that can create art: Kelly et al. resumed the Technology Acceptance Model (TAM) first postulated by Davis in 1985 and 1989 and identified perceived usefulness, performance expectancy, attitudes, trust, and effort expectancy as significant and positive predictors of behavioural intention and willingness of use of AI. (Kelly et al., 2023)

Some articles argue that the bias against the capabilities of AI has an impact on how AI-generated music is evaluated, showing that people that have a higher acceptance of AI, give a higher score in the evaluation of its art (Hong et al., 2022), while some other articles claim that there is no significant difference between the evaluations of the two groups (Zlatkov et al., 2023). When expanding the research to creative art in general, Millet et al. support the claim that more negative scores come from people with anthropocentric creativity beliefs. (Millet et al., 2023)

Similarly, when studying the impact on the evaluation of the belief of listening to AI-generated music, there is a discordance. The articles on creative arts found out that people who believe they are looking to AI-generated art, even if they do not have particular prejudices about it, give lower ratings (Messer, 2024; Millet et al., 2023), even though the strongest emotions came from human-made art (Demmer et al., 2023). On the other hand, articles on music creation state that knowledge about the creator of the piece, or having an anthropomorphized AI creating the piece, does not influence the rating, even

though the latter have influence on the acceptance of it as an artist (Hong et al., 2022; Zlatkov et al., 2023).
At the same time, however, Tubadji et al. discovered that respondents reveal lower valuations towards music generated by AI and will moderate their evaluations of quality away from AI- and towards human-generated compositions when the type of composer is known. This effect is probably connected to the concepts of cultural proximity and utility function, meaning that in the authors' opinion, the overall value of a product is a sum of its economic and cultural values, therefore the demand for creative goods is sensitive to consumers' perceptions of cultural proximity to humanness that determine the acceptability of AI products. They support the opinion that people would defend humanness, therefore AI technology has the potential to be diffused invisibly and penetrate human life without triggering the cultural proximity preference. (Tubadji et al., 2021). Similarly, Latikka et al. identified relatedness and more experience with the technology as positively impacting attitudes towards the creative art, while they did not find correlation between perceived competence and positive attitudes (like Zlatkov et al.) nor between autonomy and positive attitudes (like Hong et al.) (Latikka et al., 2023)

3.2.4.4.2 Collaboration human & AI (31 articles)

The articles reviewed for this subcluster analyse the transformative potential of artificial intelligence (AI) in music composition and production, emphasizing both the advantages and challenges of human-AI collaboration.

From the artists' point of view, AI should become a creative partner that enhances human creativity by generating musical elements such as melodies, harmonies, and lyrics. Studies like "I Keep Counting" (Micchi et al., 2021) and "ReStyle-MusicVAE" (Prvulovic et al., 2022) illustrate how AI can suggest new creative directions while leaving the final decision-making to human composers, offering fresh ideas and helping overcome creative blocks, particularly in the ideation phase. The ideal purpose should be to serve as a tool to expand creative possibilities rather than replace human input, allowing musicians to refine and personalize the AI-generated content. For many artists having AI helping them in the process can be time and cost saving, in addition to the fact that it can help them do and learn new skills: exactly as when it is used for other purposes, AI has the potential to democratize music creation, making sophisticated tools accessible to non-experts. This is evident in "El futuro de la industria musical en la era de la inteligencia artificial", which discusses how AI platforms are lowering the entry technical barrier for a broader range of artists (Apolo Valdivia, 2022). Similarly, "Towards Intelligent Music Production" (Giuliani et al., 2022)shows how sample-based AI tools like MusiComb enable producers to create music efficiently and in a way that is accessible to everyone. Reducing the complexity of music production could be beneficial to the industry since it fosters a more inclusive and diverse music landscape. Moreover, the papers in this cluster reiterate the need for maintaining creative agency while collaborating, highlighting tools that allow musicians to manipulate AI outputs through semantic controls. Customization and personalization are key to fostering a collaborative relationship between humans and AI, ensuring that the latter serves as a supportive partner rather than dominating the creative process. (Louie et al., 2020)

Apart from creating songs for recording, musicians also expressed their interest in interactive systems like MoMusic (Bian et al., 2023) and Drumming with Style (Jordà et al., 2016), which allow for real-time interaction and collaboration. These systems allow musicians to control sound synthesis or rhythmic patterns through intuitive interfaces, enhancing live performances by fostering spontaneity and exploration, and turning AI into a dynamic co-creator that responds to human input during performances.

However, the biggest challenge of using AI as a collaborator is posed by ethical and ownership problems: in order to exploit the benefits of generative AI, those systems have to undergo several sessions of training based on existing songs. The problem comes up when the pieces used to train the models do not get recognition, raising significant questions around authorship, ownership, and authenticity. As AI takes on more prominent roles in music composition, determining who owns the rights to AI-generated music becomes increasingly complex. Studies like "El futuro de la industria musical" (Apolo Valdivia, 2022) and "Generative artificial intelligence, human creativity, and art" (Zhou & Lee, 2024) emphasize the need for clear guidelines on intellectual property.

To conclude, AI's role enhances creativity by generating new ideas, facilitating collaboration, and making music production more accessible; on the other hand, it also brings up issues about authorship, emotional depth, and originality. With the increasing integration of AI into the music industry, it will be crucial to balance its capabilities with human creativity, ensuring that AI only serves as a tool for enhancing the artistic process rather than replacing it.

3.2.4.5 Generative AI creativity (9 articles)

The exploration of AI's role in music generation has led to rethink the concept of creativity from a human-centric view to a more inclusive one, which considers autonomy, combinational synthesis, emotional responsiveness, and adaptability as key indicators of creative potential in AI systems. The rising of AI pushed the boundaries of the concept and brought up the question of what can be considered creative and whether Artificial Intelligence could be considered an artist.

As common view, creativity and the ability to convey emotions are part of what makes humans unique (DiPaola et al., 2018) and that is also why making creative forms of art with the aid of AI has been, at least until today, very challenging and sort of a failure. Novelli and Proksch argue that generative music AIs must experience, or robustly simulate, something akin to the interoceptive processes that underlie emotional states and that should be able to be aware of the musical and historical context, getting exteroceptive information that go beyond a mere probabilistic distribution of notes, harmonies or rhythm (Novelli & Proksch, 2022).

All the analysed articles agree that, without human intervention, the systems' output was not original or coherent with itself. For example, when writing about the attempt of composing the Beethoven's Tenth symphony with AI, Brandt stated that the AI model did not grasp crucial features of his creative process. The algorithm was uncapable to think non-linearly, to do world- building making local decisions that responded to the larger context, to go for something less likely than the average, and to revise. All these capabilities are essential of deliberate creativity, where human inspiration and decision-making become gradually circumscribed by the work itself, whereas AI continues to draw on its complete dataset. Brandt states that what AI can do at the moment is more similar to what we call spontaneous creativity, which is more linear and chronological (Brandt, 2023) while some other authors suggest that some new models, specially auto-regressive models, can be similar to the way humans compose. Auto-regression (AR) consists in predicting the future values from past events, and the fact that these new models may be able to generate longer sequences by taking information from past steps is a progress towards a system that can resemble the human composing process.

Apart from the limitations of Artificial Intelligence in this sector, we should also analyse what being creative means to us.

Gowri Ganesh et al. proposed a model where the generated music can be considered creative in the sense that the structures must (1) resemble actual music, (2) are novel, and meaningful musical structures, (3) are abstract enough or ambiguous enough to not truly be classified as an existing piece of music or conforming to a particular style of music. However, they also mention the Wundt Curve, a relation between the hedonistic value (a measure of interestingness) and the novelty of an idea (Gowri Ganesh et al., 2023). In Figure 3.11 we can observe how an idea becomes more enjoyable as the novelty increases until a point where it becomes ambiguous and not pleasant.

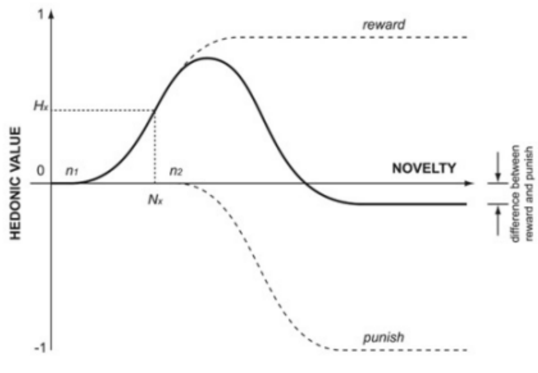


FIGURE 3.11 – Wundt Curve

The concept of novelty was important also to Briot and Pachet, who looked into the creativity issue not only from an artistic way, but also a legal and economical one raising a copyright issue: with an approach *a posteriori* we would make sure that the generated music is not too similar to others, while with *a priori* approach we would try to solve the problem from the source, making sure that the music generated would not copy entire portions on music from the training pool (J.-P. Briot & Pachet, 2020).

When talking about AI creativity, we also have to mention what humans think about machines creating art: in their work Hong et al. studied creative machine heuristics, in this study referring to the public perception of machine's creativity. The study is based on the assumption that there are two ways a person could be sceptical about machines being creative: (1) whether they could be creative and (2) to what extent. The purpose of the study was to study anthropomorphism and autonomy, which are key factors to understand better the public perception. The results reveal that, while there was a statistically significant effect of anthropomorphism on accepting an AI music generator as a musician, no significant results were found from the autonomy; however, the AI music generator's characteristics affected the AI musician evaluation but did not lead to the change of music evaluation (Hong et al., 2022).

On the other hand, Moruzzi also argued that the creative process adopted by AI in music generation should not only be assessed by the final music produced but also by its autonomy and intentionality during this process. In fact, she decided to study machine's creativity from a different perspective and shift the focus of the inquiry to the autonomy possessed by the software since creativity evaluation is mostly subjective. The real challenge for the machines would then be to be perceived as subject-independently intentional, or autonomous entities indented as they would not need external guide in each passage. Moruzzi then suggests an alternative definition of minimal creativity (CREATIVITYm), which focuses on the autonomy needed by a system to produce an output that can be recognized as creative: it focuses on the reception, selection, elaboration and production of stimuli and how autonomous the machine is. In her work Moruzzi argues that most softwares for music generation such as Jukedeck and Flow Machines cannot be considered creative since the output mimics the training corpus and matches the constraints indicated. However, GANs are producing unexpected results since they do not require the human presence apart from the initial inputs, which they do not mimic (Moruzzi, 2018).

Finally, it can be worth mentioning some psychology of creativity concepts that could inspire future works, even if not focused on creating music: (1) conceptual combination or blending, (2) contextual focus and dual processing modes, (3) the Honing theory (creativity arises via cognitive restructuring interaction between an individual's worldview and the conceptual space of a task or creative problem. According to the HT notion of art-creation, we might obtain more cognitively valid, or artistically interesting results by incorporating "training inside the loop," where the neural network structure and/or weights would be modified even within the generation process of each individual image artwork), (4) personal style, (5) intrinsic motivation, (6) The sense of completion, (7) Therapeutic impact of creativity. While for the first four concepts there has been at least a way of incorporating them in a model (even though not in music), the last three concepts have yet to find a place in computational models' creativity, but they could inspire future works. DiPaola et al. have demonstrated in their work that cross-fertilization with the psychological cognition of creativity can be beneficial (even though the work focused on visual creativity, we can take inspiration for an application in music) (DiPaola et al., 2018).

4. Research Methodology

We decided to investigate some of the gaps brought to light by the literature review, which is currently filled with studies focused on researching the most performing generative AI systems, how to optimise them and which technology could be the most usefully developed; moreover, some studies also focused on people's reactions and the emotional impact of generative AI. Methods, architectures, models, and possible collaborations with humans have been deeply explored, while the literature is lacking market research, focused on eventual effects on real everyday life and what genAI could bring to the music market. We are still early in the process, given that the best technology is still to be defined, but this thesis's purpose is to preliminarily explore this market. Moreover, other studies currently disagree on the influence of prejudice against AI-generated music on people's liking of a song, and on the influence of believing that a piece of music is made by AI has on people's liking of a song. We will also investigate these topics with the intent of contributing to the discussion.

Our main questions emerging from the review can be summarised in 2 and to each of them we gave a hypothesis as an answer to be confirmed afterwords in this research:

- Does prejudice have a relevant impact on the way people respond to this technology? Are people diffident?
  Our hypothesis is that people are diffident prejudice have a relevant impact, especially because they do not have enough information about the topic, and they see music as something exclusively human.
- Which type of products or services can be appealing to the public?
  Apart from personalised playlists that adapt to people's liking and mood, businesses could use generative AI as a background to some activities. We want to test whether the application to activities where less attention is paid is more tolerated.

To answer these questions, we prepared a questionnaire to be distributed in order to investigate people’s openness to the usage of AI in the music industry. The questionnaire is anonymously completed in approximately 10-15 minutes.

The first section contains demographic questions, made to categorize the population and investigate eventual differences that can influence the perception of this new technology.

The second section is dedicated to musical knowledge, habits and dedication, including questions about the personal relationship with music, such as the most important aspect in a song between lyrics, genre, artist, instrumental, articulate melody, production, and general mood of a song; the aim of this question was to understand whether they create or not an emotional connection with the song (if the main aspect they are interested in is lyrics or artist, they probably want create an emotional link with the song, on the other hand listening to a song for the general mood does not leave so much space for an emotional connection). It is also asked whether they have a perception of music as an art only destined to humans to create and how aware they are of the applications of AI in music creation. With these questions we wanted to investigate if the idea that people have is only a prejudice or if it is backed by their knowledge in the field, to make an accurate comparison between the prejudices and informed opinions.

The third section is for the personal perception and thoughts of generative AI applications in music, meaning the process of creating a new song, asking about the personal feeling towards AI generated music, and the quality of music they think it would result. Those questions aim is once again to investigate the preconceptions about AI-generated music. This section was structured to make people think twice and clarify their opinion about generative AI as a preparation for the following section, the listening section.

The fourth section, the listening section, contains four songs extracts: 3 of them completely made using a generative AI platform called Suno and one human-made song, named “Monkey Moves” by Nelwards (Third extract). The first and the fourth extract have been developed with AI by defining a general mood for the song, describing the situation the lyrics would have to be about and then selecting some musical genres: Suno produced the whole songs, including melody, sounds and lyrics. For the second song we took an already created text, uploaded on Suno and selected the musical genres; the AI then generated the song. For each extract it is asked to define the music quality, whether it was made by human or generative AI, and the reason why they chose that source. Later in the same section the answer about the origin of the four extracts was given and we asked the public their thoughts about generative AI quality and their feelings about it after having heard its product. This section was made with the purpose of identifying people's prejudices on this technology, also considering that ignorance can play an important role in it. Analysing their answers before and after the listening section, we tried to get more accurate data of the people perception and if it changes after getting to know the products. Moreover, it may lead to a more self-aware answer in the last section.

The last section represents a first market analysis about AI generated music and its applications, such as personalised soundtracks in video games, advertisements, social media, for therapy, for staying awake while driving and many more.
The respondents had to identify how much value added AI-generated music would bring in particular contexts and then it was asked how much attention they pay to music while doing some activities (such as driving, going to events, into shops, restaurants, listening to podcasts...), since we suppose that the opinion of people paying more attention should be taken under more consideration than the one of people not considering music during those activities at all. Lastly, it was asked their supposed reaction when discovering that AI created the music they are listening to during the same activities.

This questionnaire, made using Google Forms, was than shared using the snowball method, reaching a total of 113 people.

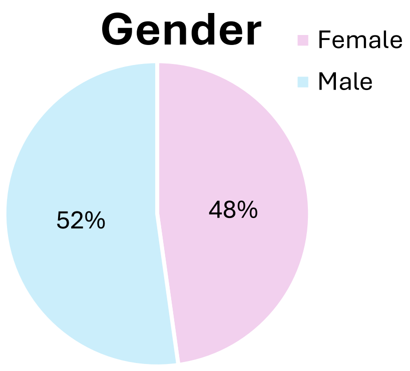
The questionnaire was initially organised in 6 different sections, the sixth especially directed to professionals to investigate possible market applications, identifying features they may be interested to and a willingness to pay in order to begin market research. Unfortunately, time was not enough to submit the questionnaire to enough professionals to make this analysis statistically relevant, but this is an aspect that may be interesting to investigate more in a following study.

4.1 Demographics

We managed to get 113 answers to the questionnaire, and since it posed questions about the demographics of the participants, we will report them below. Among the purely demographic questions about gender, age, nationality, highest level of education, current occupation, and average income, we include the questions about the self-assessed level of expertise on generative AI and about the respondent's relationship with music.

Gender		Percentage
Female	54	47.79%
Male	59	52.21%
Total	113	100.00%

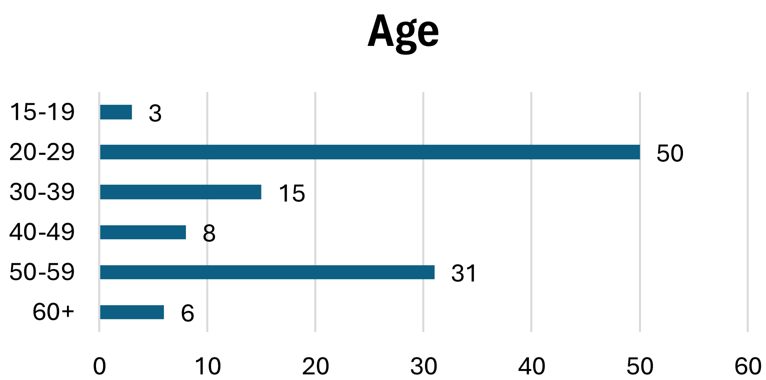
TABLE 4.1 – Gender



GRAPH 4.1 – Gender

Age		Percentage
60+	6	5.31%
50-59	31	27.43%
40-49	8	7.08%
30-39	15	13.27%
20-29	50	44.25%
15-19	3	2.65%
Total	113	100.00%

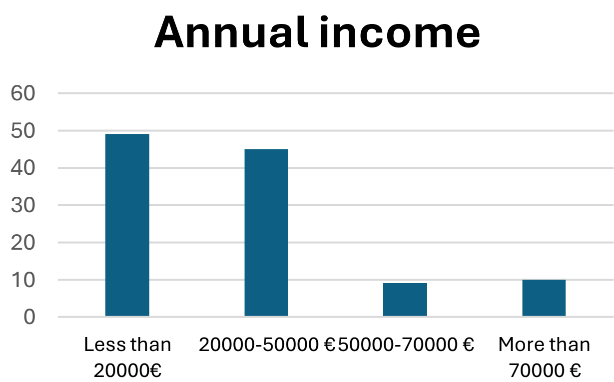
TABLE 4.2 – Age



GRAPH 4.2 – Age

Annual income		Percentage
Less than 20000€	49	43.36%
20000-50000 €	45	39.82%
50000-70000 €	9	7.96%
More than 70000 €	10	8.85%
Total	113	100.00%

TABLE 4.3 – Annual income



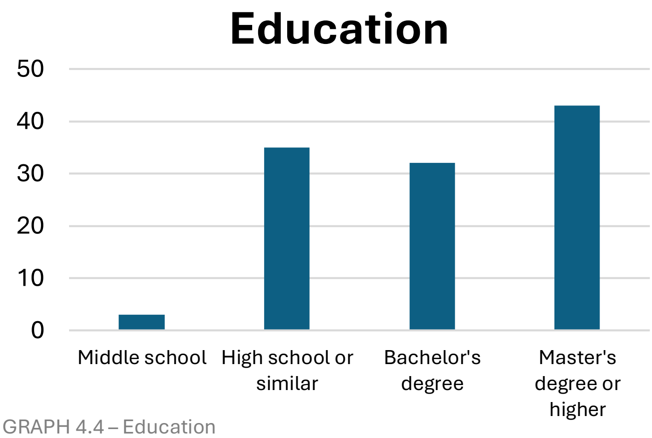
GRAPH 4.3 – Annual income

Nationality		Percentage
African American	1	0.88%
Indian	2	1.77%
Italian	109	96.46%
Italian & British	1	0.88%
Total	113	100.00%

TABLE 4.4 – Nationality

Education		Percentage
Middle school	3	2.65%
High school or similar	35	30.97%
Bachelor's degree	32	28.32%
Master's degree or higher	43	38.05%
Total	113	100.00%

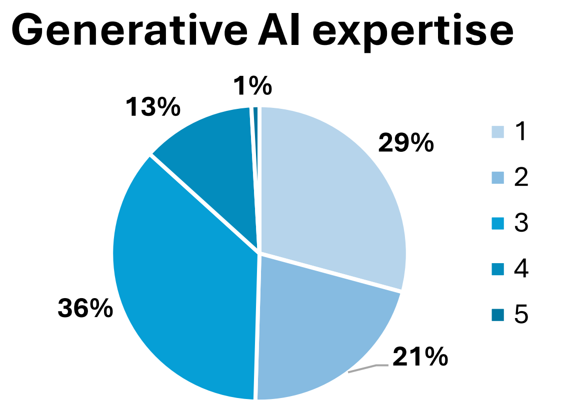
TABLE 4.5 – Education



GRAPH 4.4 – Education

Expert in generative AI		Percentage
1	33	30.28%
2	24	22.02%
3	41	37.61%
4	14	12.84%
5	1	0.92%
Total	113	103.67%

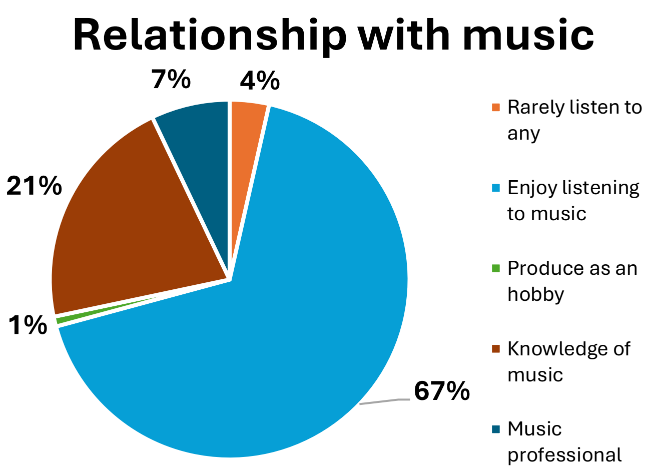
TABLE 4.6 – Expert in generative AI



GRAPH 4.4 – Generative AI expertise

Relationship with music		Percentage
I don't care/ I rarely listen to any	4	3.54%
I enjoy listening to music	76	67.26%
I produce as an hobby	1	0.88%
I have musical knowledge/ I have studied music	24	21.24%
I am a music professional	8	7.08%
Total	113	100.00%

TABLE 4.7 – Relationship with music



GRAPH 4.5 – Relationship with music

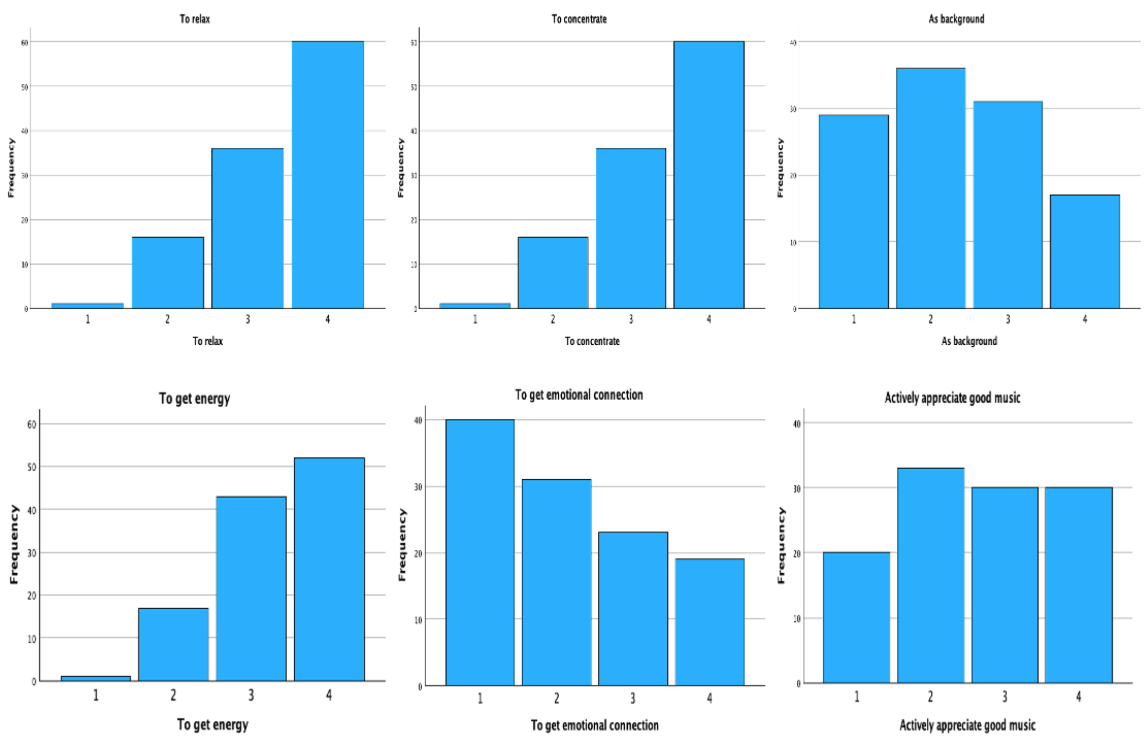
For what it concerns the occupation of the respondents, the majority of them were students, employees, consultants; only 2 people described themselves as freelancers in the music industry (the only ones that addressed it as their primary occupation) while about 10 people work in highly technological environments where at least some tech skills are required.

5. Results

After analysing the demographics statistics, we focused on the outcomes that seemed more interesting and valuable. Here we present some descriptive results and then analyse some correlations that we think might be significant done with the aid of the statistical software SPSS.

5.1 Purposes of listening to music

The question that we analyse is "When listening to music, how much do the following statements sound like you? "I listen to music...""



GRAPH 5.1 – Purposes of listening to music

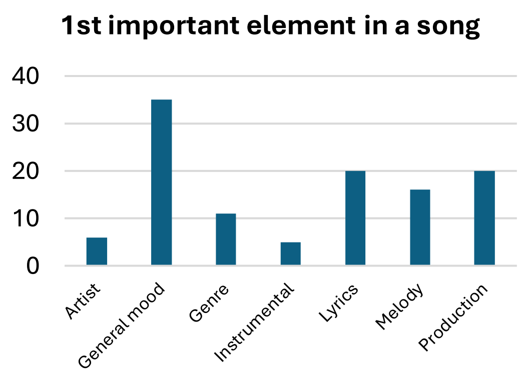
In the graph 5.1 we represent different situations and reasons why people listen to music, the respondents had to rate from 1 to 4 all of them, depending on whether they reasoned with the statement.

	Relax	Concentrate	Background	Get energy	Emotional connection	Appreciate good music
Mean	3.37	3.37	2.32	3.29	2.19	2.62
Std. Deviation	.758	.758	1.020	.752	1.098	1.063
1 (%)	.9	.9	25.7	.9	35.4	17.7
2 (%)	14.2	14.2	31.9	15.0	27.4	29.2
3 (%)	31.9	31.9	27.4	38.1	20.4	26.5
4 (%)	53.1	53.1	15.0	46.0	16.8	26.5

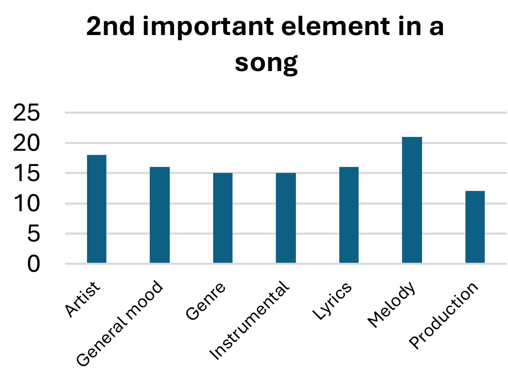
TABLE 5.1 – Purposes of listening to music

We can see that the purpose of relaxing, concentrate and getting energy have very high means, and also the distribution of values is very skewed to the right: relaxing and concentrate even have the highest possible score already from the 50% mark. As for the others, apart from the emotional connection that seems to decrease, we cannot define trends.

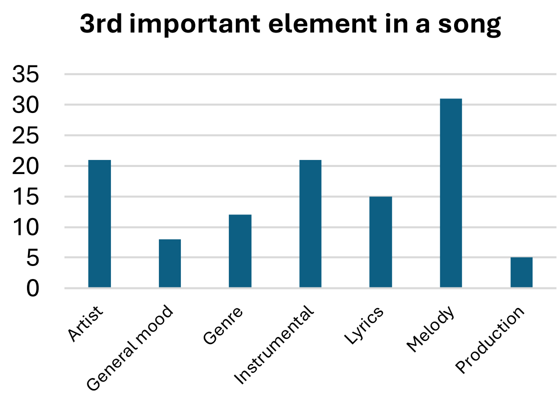
5.2 Relevance elements in songs



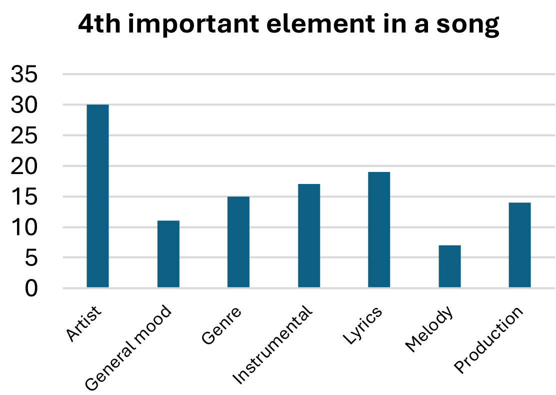
GRAPH 5.2 – 1st important element in a song



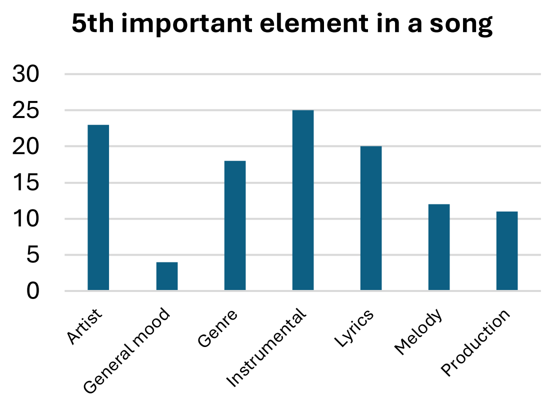
GRAPH 5.3 – 2nd important element in a song



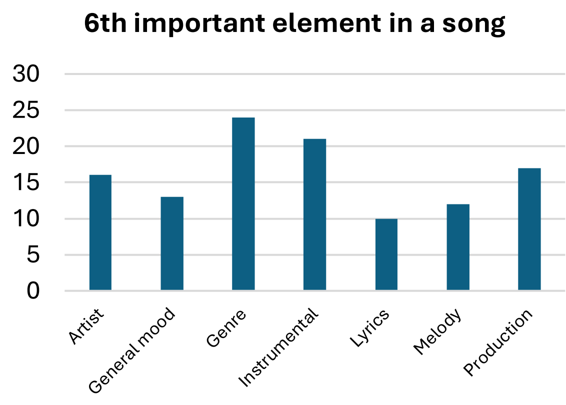
GRAPH 5.4 – 3rd important element in a song



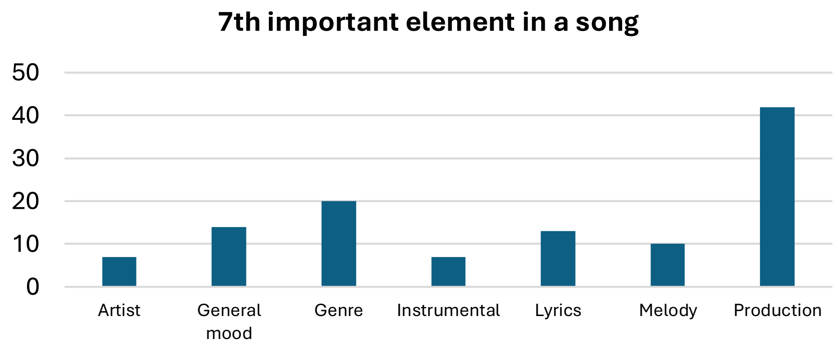
GRAPH 5.5 – 4th important element in a song



GRAPH 5.5 – 5th important element in a song



GRAPH 5.6 – 6th important element in a song



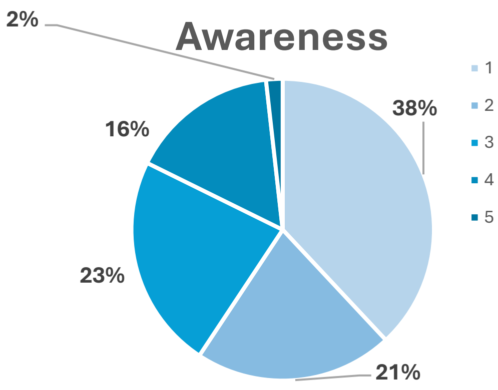
GRAPH 5.7 – 7th important element in a song

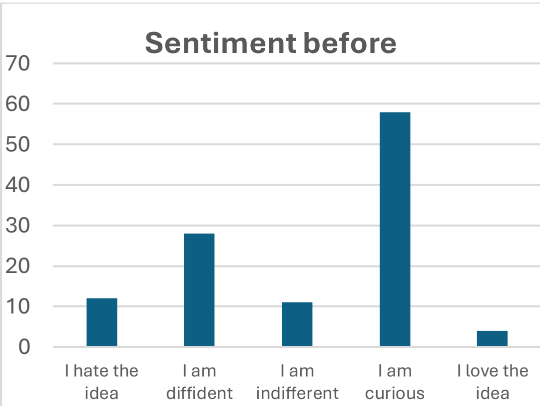
From these graphs we can determine that the general mood is considered the most important element in a song when listening to it and deciding whether we like it or not. As for the other positions, there is not a clear element that is higher than the others, however we could safely claim that melody occupies one of the first positions while the artist is somewhere near the fourth or fifth position. Interestingly, the lyrics results are scattered among the positions, always averaging 10-15 votes. In the last position we find the production, which people claim as the last element they consider when listening to a song.

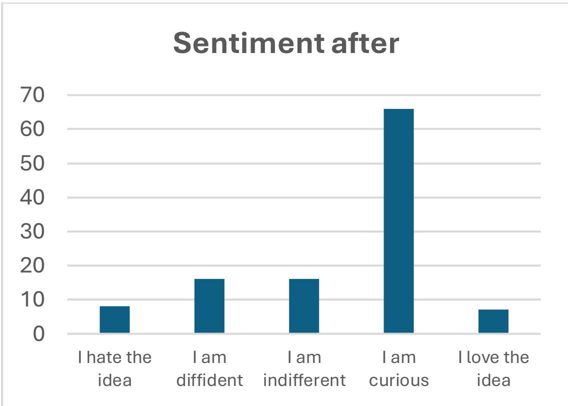
5.3 Awareness and its impact on sentiment

In order to assess how people respond to this innovation we decided to test their sentiment towards it (expressed with a scale from 1 to 5 where 1=I hate the idea, 2=I am diffident 3= I am in different 4 = I am curious 5= I love the idea) before and after they listened to the extracts. Moreover, we decided to test whether the awareness of the

possible AI applications in music composition (tested with a scale from 1 to 5) had an influence on the sentiment.
Here the graphs of the distributions of every category.







We can already see how the majority of people (59%) attribute to themselves a low awareness score (1-2) and that half of the respondents said to be curious about the technology, while a about a quarter of them said to be diffident about it. It seems that, after listening to the extracts the sentiment has changed, but we have to verify if it is statistically significant.

With the first test we want to study the correlation between the awareness and the before sentiment, which we did with a Spearman test that came out statistically significant with a correlation value of 0,281 and p-value=0,003. The correlation, however, is weak.

We then decided to investigate more the differences between the sentiment before and after, using the Wilcoxon signed-rank test, which confirmed that we have to reject the null hypothesis that the median differences between Sentiment before and Sentiment after equals 0 (with significance <0.001)

To further investigate the relationship between awareness and change of sentiment we went to see which awareness groups changed opinion the most and we discovered that groups 1 and 2 are the ones that made the biggest change. In total, 75 people did not change opinion, 8 people gave a more negative opinion and 30 gave a more positive one.

5.4 Turing test and addressing prejudice

For this part, we asked the participants to listen to 4 distinct song extracts and asked to give a score from 1 to 10 depending on how much they liked the song and then to try to identify if it was created using generative AI. In the table below we present the mean and the standard deviation for the scoring of each song.

	Mean	Std. Deviation
Score extract 1	5.88	2.021
Score extract 2	6.06	2.041
Score extract 3	4.73	1.932
Score extract 4	6.22	1.926

TABLE 5.2 – Extracts evaluation scores: means and standard dev.

To go deeper into the analysis, we tried to find out if supposing that a song was composed by a human or by a machine, influences the way people evaluate the piece. Unfortunately, we were unable to do the analysis taking into account also the nuances of being sure or just supposing, so we had to aggregate the results of the options "Supposedly AI" and "Surely AI" in the category "AI" and the results from "Supposedly human" and "Surely human" in the category "human". In the table 5.3 you can find the mean and standard deviation of these groups and, in the last column, the percentage of people who were right about the origin in the same row of the correct one. A one-way ANOVA test with dependent variable the score and the supposition as factor (and a post hoc Tukey test) was conducted, which results are reported in table 5.4; we can determine that for extract 1 the scoring associated to human is significantly higher than the one associated to AI or not sure. Also, in extracts 2 and 4 we assist at the statistically significant higher score given by people who believe that a human created the extract. On the other hand, extract 3, which is the only one made by a real human, has no significant difference. For all the tests, the increasing of the scoring if the respondents believe that they are listening to a human-created song is statistically significant, however it only explains a small part of it, as we can see from the eta squared value.

	Supposition	N	Mean	Std. Deviation	Correct source
EXTRACT 1	AI	52	5.56	2.081	46.0%
	Not sure	32	5.47	1.759	
	Human	29	6.93	1.870	
EXTRACT 2	AI	34	5.06	2.131	30.1%
	Not sure	27	5.89	2.044	
	Human	52	6.81	1.681	
EXTRACT 3	AI	59	4.34	2.039	
	Not sure	38	5.05	1.627	
	Human	16	5.44	1.965	14.2%
EXTRACT 4	AI	37	5.57	1.980	32.7%
	Not sure	36	5.97	1.964	
	Human	40	7.05	1.552	

TABLE 5.3 – Extracts evaluation based on Turing test and evaluation score

	Supposition		Mean difference	Std. error	
EXTRACT1	AI	Not sure	-0.52003	0.25348	0.105
		Human	-1.04697*	0.26148	<.001
	Not sure	AI	0.52003	0.25348	0.105
		Human	-0.52694	0.28925	0.167
	Human	AI	1.04697*	0.26148	<.001
		Not sure	0.52694	0.28925	0.167
	Eta-squared	0,094			

	Supposition		Mean Difference	Std. Error	Sig.
EXTRACT2	AI	Not sure	-0.83	0.493	0.216
		Human	-1.749*	0.422	<.001
	Not sure	AI	0.83	0.493	0.216
		Human	-0.919	0.454	0.111
	Human	AI	1.749*	0.422	<.001
		Not sure	0.919	0.454	0.111
	Eta-squared	0,137			

	Supposition		Mean Difference	Std. Error	Sig.
EXTRACT3	AI	Not sure	-0.71	0.395	0.172
		Human	-1.099	0.536	0.105
	Not sure	AI	0.714	0.395	0.172
		Human	-0.385	0.566	0.776
	Human	AI	1.099	0.536	0.105
		Not sure	0.385	0.566	0.776
	Eta-squared	0,05			

	Supposition		Mean Difference (I-J)	Std. Error	Sig.
EXTRACT4	AI	Not sure	-0.405	0.429	0.615
		Human	-1.482*	0.418	0.002
	Not sure	AI	0.405	0.429	0.615
		Human	-1.078*	0.421	0.032
	Human	AI	1.482*	0.418	0.002
		Not sure	1.078*	0.421	0.032
	Eta-squared	0,110			

TABLE 5.4 – One-way ANOVA test for four extracts Supposition/Score

In order to try and eliminate the bias that could be introduced by some people giving generally lower scores and some giving generally higher we introduced a new variable called "Delta" which is the difference from the mean score of a person. The mean is also weighted in order to take into account the fact that only one out of four extracts was made by a human.

$$\boldsymbol{mean} = \frac{\frac{1}{3}(\boldsymbol{score1} + \boldsymbol{score2} + \boldsymbol{score4}) + \boldsymbol{score3}}{2}$$

$$\boldsymbol{Delta1} = \boldsymbol{mean} - \boldsymbol{score1}$$

Then, we completed the same analysis we did previously.

	Supposition		Mean Difference (I-J)	Std. Error	Sig.
EXTRACT1	AI	Not sure	-0.52003	0.25348	0.105
		Human	-1.04697*	0.26148	<.001
	Not sure	AI	0.52003	0.25348	0.105
		Human	-0.52694	0.28925	0.167
	Human	AI	1.04697*	0.26148	<.001
		Not sure	0.52694	0.28925	0.167
	Eta-squared	0,130			

	Supposition		Mean Difference (I-J)	Std. Error	Sig.
EXTRACT2	AI	Not sure	-1.33497*	0.36667	0.001
		Human	-1.36275*	0.31372	<.001
	Not sure	AI	1.33497*	0.36667	0.001
		Human	-0.02778	0.33742	0.996
	Human	AI	1.36275*	0.31372	<.001
		Not sure	0.02778	0.33742	0.996
	Eta-squared	0,164			

	Supposition		Mean Difference (I-J)	Std. Error	Sig.
EXTRACT3	AI	Not sure	-.38388*	0.14669	0.027
		Human	-.72599*	0.19878	0.001
	Not sure	AI	.38388*	0.14669	0.027
		Human	-0.34211	0.21017	0.238
	Human	AI	.72599*	0.19878	0.001
		Not sure	0.34211	0.21017	0.238
	Eta-squared	0,127			

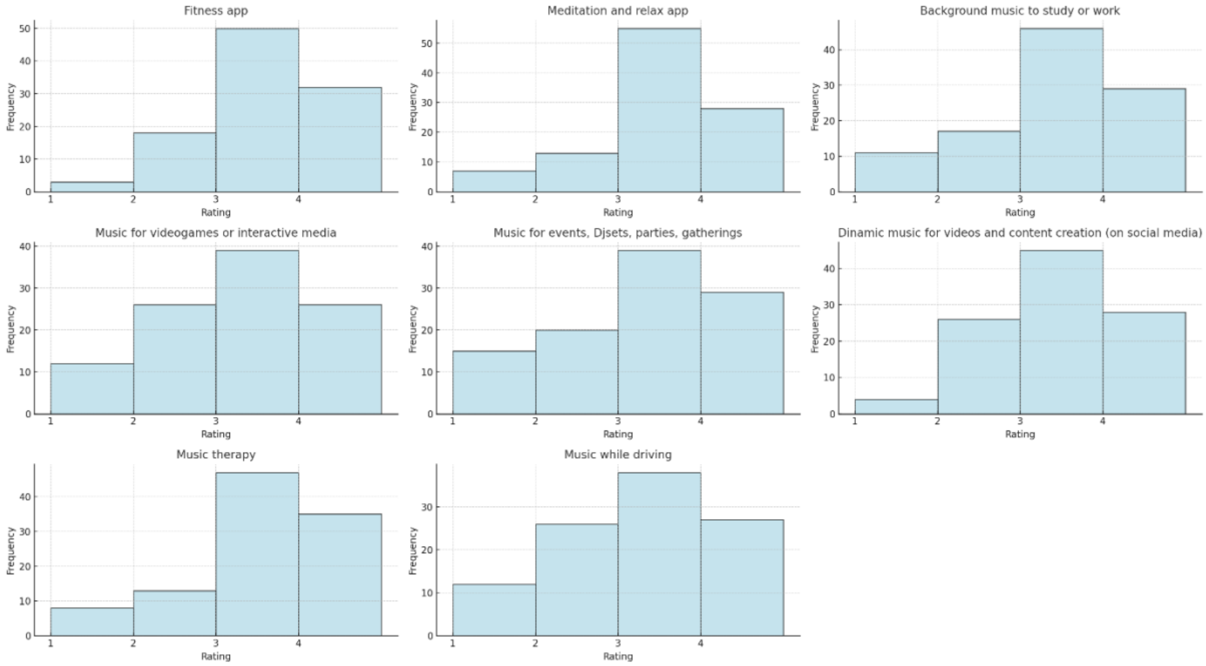
	Supposition		Mean Difference (I-J)	Std. Error	Sig.
EXTRACT4	AI	Not sure	-0.23311	0.29689	0.713
		Human	-1.48727*	0.28927	<.001
	Not sure	AI	0.23311	0.29689	0.713
		Human	-1.25417*	0.29135	<.001
	Human	AI	1.48727*	0.28927	<.001
		Not sure	1.25417*	0.29135	<.001
	Eta-squared	0,219			

TABLE 5.5– One-way ANOVA test for four extracts Supposition/Delta

The results show again the difference in mean especially between the ones that suppose it is AI and the ones that suppose it is human-made. Eliminating some of the biases now we see this tendency also in extract 3 and, also, the eta-squared gets higher.

5.5 Value added from personalized playlists made with generative AI

First, we wanted to analyse if people attribute the same value to personalised playlists across the different applications. From the resulting graphs we can see how the majority perceives at least some value added; however, there is no recognizable trend or application that behaves differently from the others.

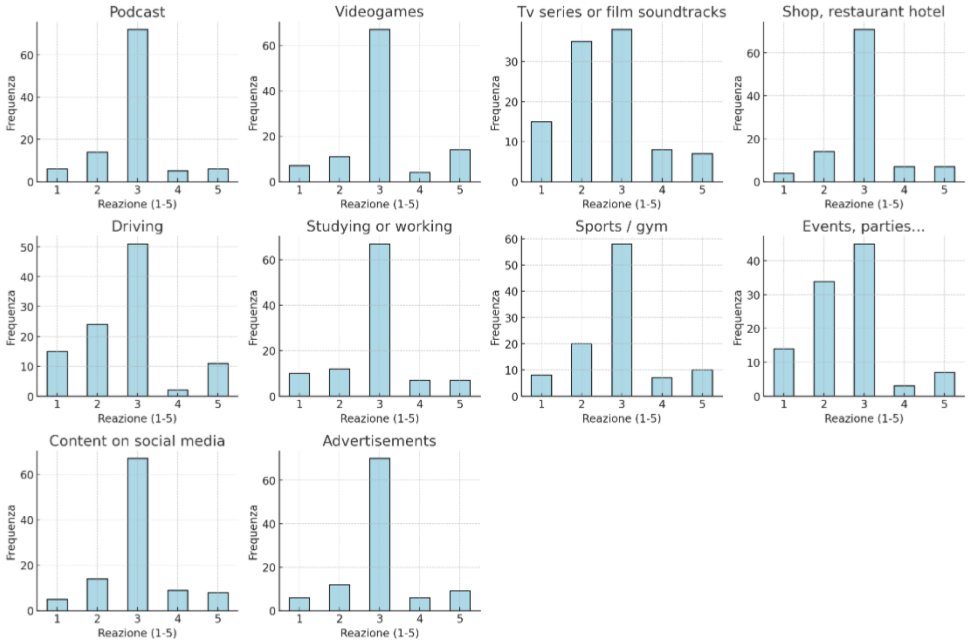


GRAPH 5.8 – Personalized playlists made with generative AI added value

5.6 Market applications, reactions and attention paid

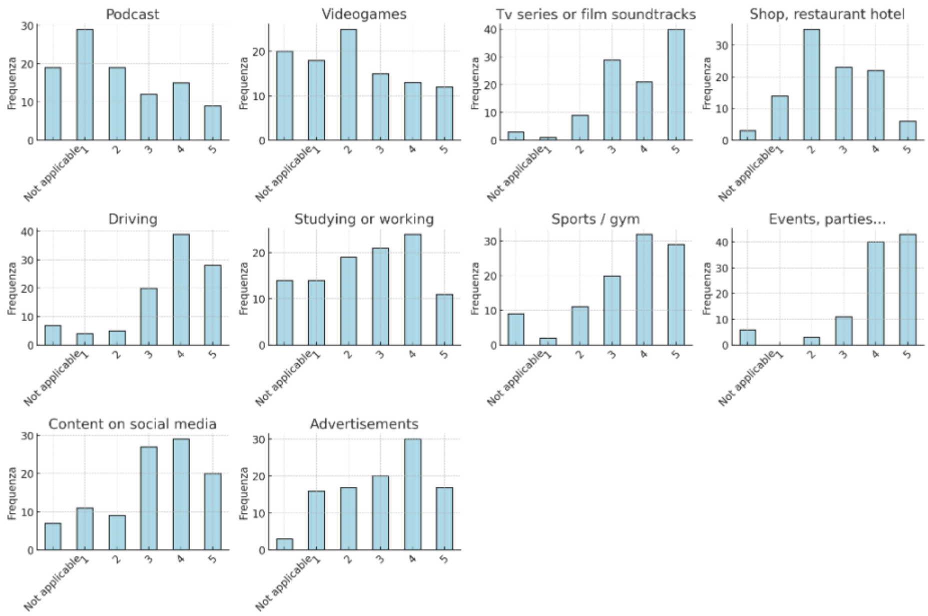
For this analysis we wanted to study the reaction of people to knowing that AI-generated music is being played while they are doing something that is not actively listening to it.

The scores range from 1 to 5 as Negative, Slightly Negative, Indifferent, Slightly Positive, Positive. From the graphs we can already see that the majority of people would be indifferent to AI-generated music in these contexts; however, we decided to conduct a statistical analysis to further prove it. Another thing that we can see from the graphs, is the fact that for TV series or film soundtracks, Events and parties, and driving, there seems to be a skew towards left. In order to verify if the trend is statistically significant, we measured the difference of the median from the neutral 3. The test results confirm that in those three cases the median is significantly different from the neutral value 3, with a p-value of 0,00045 for "TV series and film soundtracks", 0,000197 for "Events, parties…" and, even though less significant than the first two, a p-value of 0,0325 for "Driving".



GRAPH 5.9 – Reaction to eventual AI generated music in different fields

To go on with the analysis, we need to introduce another variable, which is the attention that people pose to the music being played while they are doing specific activities. We kept the same activities so that we could analyse the interaction and possible correlations between attention and reaction. From the graphs below we can see that in events and parties, while driving, and while watching TV people pay the most attention to music.



Graph 5.10 – Attention to music in different situations

First, we wanted to verify whether there is a correlation between the attention that a person pays to music during an activity, and the reaction that they have when they know that that said music was generated by AI. In order to get rid of unvaluable data coming from people that do not do some of the activities described before, we asked the participants to answer "not applicable" if that was the case. Thanks to this, we could eliminate the corresponding answer they gave to the reaction question. Using the Pearson test we analysed the correlation between attention (independent) and reaction (dependent), but it came out to be very weak and not significant (-0,045, p-value 0,164). Consequently, we decided to verify whether there was some specific activity that had a significant correlation, since from the graphical representation in driving, TV series and films soundtracks and events they seemed to have some sort of correlation.

ACTIVITY	PEARSON CORRELATION	P-VALUE
PODCAST	-0.1284	0.2444
VIDEOGAMES	-0.0981	0.3773
TV SERIES OR FILM SOUNDTRACKS	-0.0706	0.4853
SHOP, RESTAURANT, HOTEL	-0.225	0.0275
DRIVING	-0.1639	0.1105
STUDYING OR WORKING	0.0181	0.8660
SPORTS / GYM	0.0106	0.9199
EVENTS, PARTIES	-0.0974	0.3423
CONTENT ON SOCIAL MEDIA	0.0801	0.4377
ADVERTISEMENTS	-0.2473	0.0131

Table 5.7 –Pearson correlation between an attention during an activity and reaction to AI music

We can see that the only significant correlations are negative and come from music in shops, restaurants and hotels and from advertisements; however, even in these cases, the correlations are weak.

To better analyse the way people react, we wanted to give more power to the people that listen to music with attention while doing the activities, therefore we decided to weight the reactions for the attention paid. If the respondent declared to pay attention as "1-not at all", we counted their contribution as 1 and we did the same for all the possible scores of the attention. If the respondent declared it was "not applicable", it was not counted. Later, we compared the weighted and not-weighted reaction frequencies for all the activities to see whether there is a significant difference using chi-squared test.

For podcasts, the chi-squared was 4,335 with p-value of 0,363. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	19	41	126	14	8
not weighted	6	14	72	5	6

For video games, the chi-squared was 4,418 with p-value of 0,352. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	20	34	119	14	38
not weighted	7	11	67	4	14

For TV series and films soundtracks, the chi-squared was 4,421 with p-value of 0,981. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	56	136	135	30	33
not weighted	15	35	38	8	7

For shops, restaurants and hotels, the chi-squared was 2,831 with p-value of 0,586. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	18	47	163	23	20
not weighted	4	14	71	7	7

For driving, the chi-squared was 1,791 with p-value of 0,774. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	66	92	157	8	47
not weighted	15	24	51	2	11

For studying or working, the chi-squared was 4,049 with p-value of 0,399. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	27	43	144	26	26
not weighted	10	12	67	7	7

For sports/gym, the chi-squared was 0,339 with p-value of 0,987. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	25	78	196	25	33
not weighted	8	20	58	7	10

For events, parties…, the chi-squared was 0,407 with p-value of 0,982. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	52	149	171	13	29
not weighted	14	34	45	3	7

For content on social media, the chi-squared was 4,657 with p-value of 0,324. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	5	55	203	33	30
not weighted	5	14	67	9	8

For advertisements, the chi-squared was 1,514 with p-value of 0,824. It is therefore not significant.

Reaction	1	2	3	4	5
weighted	24	49	201	17	24
not weighted	6	12	70	6	9

The distribution of the reactions for the analysed activities did not significantly change after considering the reactions of the ones paying more attention as more important.

6. Discussion

Through the previous analysis, we wanted to conduct a study on the possible future applications of generative AI in the music industry. Since we did not have time to deliver the questionnaire to music professionals so that we could study the way they could collaborate, we focused on the other side of the industry: the general public.

This study could be tricky since many people do not have musical knowledge or interest; therefore we tried to at first get a glimpse of their consumption habits.

From what we have seen, people usually listen to music to relax and to get energy; the other four situations did not have a particular distribution trend, but it is interesting to notice how the field “getting an emotional connection with the artist” has a decreasing trend, meaning that the majority of people answered to this question saying that it does not resonate with them at all. From a market search point of view, our results indicate a strong interest of the public for using music to relax and motivate (concentration or energy), areas where generative AI, as of now, is performing well; on the other hand, an area where AI is lacking is creating an emotional connection with an artist, which from our study appears not to be as much relevant as we might have expected. Already from this first evaluation, we can get a first direction for the application of genAI in music composition.

This first point is backed by the information collected from the second evaluation we made about what elements are considered more important in a song. As we said in the results, except for the general mood that is considered by far the most important one, the other positions seem to have a uniform trend with similar frequencies for all the elements. Circling back to the first position, the artist and the instrumental part are the two elements less chosen to be most important, while in the last position the production was chosen the most times. These last outcomes back our case that people do not listen to music only for the emotions they can feel or for a specific artist, but they primarily listen to songs that, in a utilitarian way, makes them feel good, whether to relax or to get motivated, people first look at the mood they need in the moment they have to choose a song. The author supposes that the meaning behind the production being in the last position is the fact that it is a practice that feels distant to the listeners, which do not have the education and interest necessary to notice it.

Another point of the research was to assess how people feel towards AI, what is their general sentiment and awareness of the possible applications. We decided to pair the questions about the awareness and the sentiment because sometimes low awareness can be a seed that grows negative sentiments. 60% of the respondents had low awareness of AI's capabilities and we proved that it did have an impact on how they felt about it. With our research we even proved that the awareness did not have significant correlation with the sentiment after the listening of the extract, sustaining the point that people that do not have enough information about this technology and have not seen what is capable of, tend to develop negative feelings towards it. In fact, when comparing the sentiment levels of before and after, we found a significant positive change and, even when studying the change by levels of awareness, the two groups that significantly changed sentiments were the lowest one, proving that discovering more about the technology can help people changing their ideas. Generally speaking, the majority of people is curious about the potential of AI, but there is still a large percentage of people who feel negatively or at least diffident about it (49,5%), which was lowered (21,2%) after the listening session. Therefore, if the industry was interested in developing a customer base and arouse interest, they should dedicate their resources towards increasing awareness, this becomes even more important since, as we will see below, prejudice that could come from low awareness can have a negative impact on the evaluation of quality of the system's output.

Next, we addressed an issue that was already covered by many other researchers in the literature: is there some prejudice against AI that has a negative impact on the evaluation of the outcomes? In order to try to give our answer to this question, we first analysed the evaluations given to the extracts, interestingly, the third extract, which was also the only piece composed by a human, got a significantly lower evaluation compared to the others composed by AI. This outcome could have many explanations, starting from the fact that people did not like the song or that the song was in fact of a lower quality, but it does not support a hypothesis of people getting an unconscious emotional bond with human-made songs and therefore giving them higher evaluations. This is proof, together with the very low percentages of people recognizing the AI-created songs, that the technology has come to a point where it can pass the Turing test, or at least put in serious doubt the people that try to assess it. However, the public still have serious prejudices about artificial intelligence, which can be seen from the same experiment: for all of the extracts, the mean evaluation from people who thought that the piece was AI-created was the lowest one, followed by the mean from people who were not sure, and to end with the people who believed that it was human-made. After the change of dependent variable to Delta (less subject to biases of people), all the extracts register a significant difference between supposition AI and supposition human, with also a moderate eta-squared that suggests that this correlation can explain a good part of the variance. These differences prove the existence of an unconscious prejudice against AI-generated music.

After these first evaluations, we focused on the possible applications in a real market.

First of all, we asked people whether they would find at least some value added to products or services if they came with the possibility to have personalized playlists generated on-demand and in real-time based on the users' mood and preferences. The survey showed there is a significantly higher number of people who believe they would get at least some value added from those products, which is really encouraging for the industry. However, the kind of technology that some of them might have thought of while filling the questionnaire does not exist at the moment: what we presented here are products that would find their best application when connected with human emotional and physical responses in order to adapt to them and generate songs consequently. At the state of the art, there is no such thing on the market or even as a functioning prototype, but these results can serve as a first attempt to evaluate their possible reception on the market.

Finally, we get to the part of the applications that could be implemented as of now. In fact, from a music business point of view, creating and producing songs that are not for artists (and therefore with their own business plan to support them), can be quite costly and time consuming with not much payback. Our goal is to study which areas or activities would be less negatively impacted if businesses decided to adopt generative AI to save. From the distributions of the first graph, we see that for podcasts, videogames, shops, restaurants and hotels, studying or working, sport/gym, content on social media and advertisements, people would be prevalently indifferent if they knew that the music they were listening to was made by AI. Conversely, for TV series and films soundtracks, driving and events and parties, they tend towards a negative reaction, which means that they would not be the best industries to start to invest in.

In addition, we decided to add another variable to expand the research. We identified in the attention that people pose to music while doing certain activities a variable that could influence the reactions. Activities such as Tv series and films, Events and parties, Driving, doing Sports / gym and even Content on social media have a distribution of reactions that tend towards a medium or high attention, while for the others it seems distributed more uniformly if not a bit low for podcasts and shops, restaurants and hotels. Moreover, many people reported not to listen to music while working or studying and that they do not listen to podcasts or play videogames. This means that, for these three activities, the number of reactions considered will be lower than the others. Seeing that generally there is a negative correlation between attention and reaction, even if weak, means that the more people pay attention to music while doing a certain activity, the less they will enjoy listening to AI-generated music. We could than take this rule and generalize it so that, for the activities that have lower attention, industries could more safely implement the technology. However, we wanted to verify the correlation for each activity and in fact we found that the only significant ones where negative and coming from music in shops, restaurants and hotels, driving and from advertisements. We then suggest that, for these activities in particular, but also for the others even though in a smaller part, industries look out for the people that likely pay more attention to the music they are listening to and do not use generative AI with them since it could lead to negative reactions. This finding does not mean that business should not use genAI in these situations, but that they should analyse better their customer base particularly for those three activities.

However, when analysing the distribution of the weighted reactions, we could see that, even though we gave more power to the ones that pay more attention, the distributions did not change from the non-weighted ones, meaning that, even though there is a correlation, the attention do not have that much power on the general public reactions and sentiments towards hearing AI-generated music during certain activities.

To summarize, the main reactions for all the activities is indifference, which already gives us an answer to the question of whether people would accept to listen to music generated by AI in certain environments: business who think they would save money and time using this technology would likely just be met with indifference, which is not a bad outcome if music is just used as a background and not as the main purpose. However, some activities (events, driving and soundtracks) have also a big tail of negative responses, meaning that they should not be the first one to try the application of generative AI. Similarly, taking into account the considerations made for attentions should not be taken as a deterrent since the correlations are significant but weak and the distribution of weighted reactions remained the same. The suggestion is just to be more aware of the customers habits and types and, when dealing with a customer base who pays attention to the music being played, consider more carefully the options. Also, human behaviour is complex and sometimes difficult to comprehend, and all the significant correlations found in this research were low. Therefore, even though we cannot overlook their contribution to the behaviours observed, we have to keep in mind that they are not the only elements involved.

7. Conclusions

Generative AI has the potential to greatly impact the music industry, which it already did for some areas. When talking about the composition process, it can help democratizing the art by making it more accessible and efficient. Even though there are still many challenges such as the lack of a recognised dominant architecture and/or model that does not have problems, the non-standardized evaluation methods, which do not allow for fair comparisons between models, and ethical and authorships concerns, generative AI can be a great ally for both musicians and listeners. From our findings, even though people still have negative prejudices towards AI-generated music, they are curious about the technology, and they would be indifferent if they had to listen to AI-generate music while listening to podcasts, playing videogames, going into shops, restaurants and hotels, studying or working, doing sport/gym, seeing or creating content on social media or while seeing advertisements. Music and non-music businesses, if they consider it profitable, should invest in generative AI music for these activities, which reflect the search of the public for relax or motivation while listening to music, and study their customer base. It is in fact crucial to understand whether they pay a lot of attention to music while doing specific activities since for them, listening to AI-generated music in those contexts can turn into a negative experience. Our research has also shown that when people are not aware of the capabilities and possibilities of AI in music composition, they have more negative feelings, which can be easily changed after listening to a few audios. It is important then that the industry contributes to increasing the public awareness on the topic, or at least waits until it happens through other ways, if they do not want negative prejudice to grow from an initial diffidence and therefore have negative evaluations of music's quality.

The industry is already moving towards this innovation, with a lot of research and first products already on the market (even though not of really good quality), but it is important that they correctly dimension the expectations from the public; otherwise the risk is that, because of negative prejudice and low awareness on one hand, and too high expectations on the other, there will be no market to serve.

7. Bibliography

Apolo Valdivia, P. R. (2022). El futuro de la industria musical en la era de la inteligencia artificial. *Artnodes*. https://doi.org/10.7238/artnodes.v0i30.399485

Arranger. (n.d.). Berklee.

Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2024). Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers. *IEEE Access*, *12*, 69812–69837. https://doi.org/10.1109/ACCESS.2024.3397775

Bian, W., Song, Y., Gu, N., Yan Chan, T., To Lo, T., Sun Li, T., Chak Wong, K., Xue, W., & Alonso Trillo, R. (2023). MoMusic: A Motion-Driven Human-AI Collaborative Music Composition and Performing System. *The Thirty-Seventh AAAI Conference on Artificial Intelligence*. www.aaai.org

Bockstedt, J. C., Kauffman, R. J., & Riggins, F. J. (2006). The move to artist-led on-line music distribution: A theory-based assessment and prospects for structural changes in the digital music market. *International Journal of Electronic Commerce*, *10*(3). https://doi.org/10.2753/JEC1086-4415100301

Boryczka, U., Boryczka, M., & Chmielarski, P. (2023). ACO and generative art - artificial music. *Procedia Computer Science*, *225*, 2624–2633. https://doi.org/10.1016/j.procs.2023.10.254

Brandt, A. K. (2023). Beethoven's Ninth and AI's Tenth: A comparison of human and computational creativity. *Journal of Creativity*, *33*(3). https://doi.org/10.1016/j.yjoc.2023.100068

Briot, J. P. (2021). From artificial neural networks to deep learning for music generation: history, concepts and trends. *Neural Computing and Applications*, *33*(1), 39–65. https://doi.org/10.1007/s00521-020-05399-0

Briot, J.-P., & Pachet, F. (2020). Deep learning for music generation: challenges and directions. *Neural Computing and Applications*, *32*(4), 981–993. https://doi.org/10.1007/s00521-018-3813-6

Chen, A., & Greer, R. (2023). Quantifying Repetition in Symbolic Music using Lempel-Ziv Compression. *2023 IEEE 14th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2023*, 387–393. https://doi.org/10.1109/UEMCON59035.2023.10316147

Chen, H., Xiao, Q., & Yin, X. (2019, May 10). Generating Music Algorithm with Deep Convolutional Generative Adversarial Networks. *2nd International Conference on Electronics Technology (ICET)* .

Composer (Concert and Stage). (n.d.). Berklee.

Demmer, T. R., Kühnapfel, C., Fingerhut, J., & Pelowski, M. (2023). Does an emotional connection to art really require a human artist? Emotion and intentionality responses to AI- versus human-created art and impact on aesthetic experience. *Computers in Human Behavior*, *148*. https://doi.org/10.1016/j.chb.2023.107875

Dervakos, E., Filandrianos, G., & Stamou, G. (2020). Heuristics for evaluation of AI generated music. *Proceedings - International Conference on Pattern Recognition*. https://doi.org/10.1109/ICPR48806.2021.9413310

Dipaola, S., Gabora, L., & McCaig, G. (2018). Informing artificial intelligence generative techniques using cognitive theories of human creativity. *Procedia Computer Science*, *145*, 158–168. https://doi.org/10.1016/j.procs.2018.11.024

Do Quang, T., & Hoang, T. (2023). An efficient method to build music generative model by controlling both general and local note characteristics. *Journal of King Saud University - Computer and Information Sciences*, *35*(9). https://doi.org/10.1016/j.jksuci.2023.101761

Frid, E., Gomes, C., & Jin, Z. (2020, April 21). Music Creation by Example. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376514

Giuliani, L., Ballerini, F., De Filippo, A., & Borghesi, A. (2023). MusiComb: a Sample-based Approach to Music Generation Through Constraints. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*. https://doi.org/10.1109/ICTAI59109.2023.00036

Giuliani, L., De Filippo, A., & Borghesi, A. (2022, November 6). Towards Intelligent Music Production: A Sample-based Approach. *CREAI- Workshop on Artificial Intelligence and Creativity*. http://ceur-ws.org

Goldman Sachs. (2018). *Streaming Turns Up the Beat for the Music Industry*.

Gowri Ganesh, N. S., & Venkata Vara Prasad, D. (2023). GENERATING CREATIVE CLASSICAL MUSIC BY LEARNING AND COMBINING EXISTING STYLES. *4th International Conference on Communication, Computing and Industry 6.0, C216 2023*. https://doi.org/10.1109/C2I659362.2023.10431294

Guo, R., Simpson, I., Kiefer, C., Magnusson, T., & Herremans, D. (2022). MusIAC: An extensible generative framework for Music Infilling Applications with multi-level Control. In *ArXiv*. http://arxiv.org/abs/2202.05528

Hernandez-Olivan, C., & Beltran, J. R. (2021). Music Composition with Deep Learning: A Review. In *ArXiv*. http://arxiv.org/abs/2108.12290

Hernandez-Olivan, C., & Beltran, J. R. (2022). Music Composition with Deep Learning: A Review. *Advances in Speech and Music Technology*, 25–50. http://arxiv.org/abs/2108.12290

Hong, J. W., 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*. https://doi.org/10.1016/j.chb.2022.107239

Hou, Y. (2022). AI Music Therapist: A Study on Generating Specific Therapeutic Music based on Deep Generative Adversarial Network Approach. *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information, ICETCI 2022*, 1277–1281. https://doi.org/10.1109/ICETCI55101.2022.9832398

Hunt, S. J., Mitchell, T. J., & Nash, C. (2020, July 21). Composing computer generated music, an observational study using IGME: the Interactive Generative Music Environment. *NIME'20, Royal Birmingham Conservatoire*.

IFPI Global Music Report (2024 edition). (2023).

Ji, S., Yang, X., & Luo, J. (2023). A Survey on Deep Learning for Symbolic Music Generation: Representations, Algorithms, Evaluations, and Challenges. *ACM Computing Surveys*, *56*(1). https://doi.org/10.1145/3597493

John C. Mankins. (1995). TECHNOLOGY READINESS LEVELS: A White Paper. *Journal of Chemical Information and Modeling*, *53*(9).

Jordà, S., Gómez-Marín, D., Faraldo, Á., & Herrera, P. (2016, July 11). Drumming with style: From user needs to a working prototype. *NIME '16*. http://www.redbullmusicacademy.com/

Kayacik, C., Chen, S., Noerly, S., Holbrook, J., Roberts, A., & Eck, D. (2019, May 2). Identifying the intersections: User experience + research scientist collaboration in a generative machine learning interface. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3299059

Keller, R., Hunt, M., Jones, S., Morrison, D., Wolin, A., & Gomez, S. (2007). Blues for Gary: Design Abstractions for a Jazz Improvisation Assistant. *Electronic Notes in Theoretical Computer Science*, *193*, 47–60. https://doi.org/10.1016/j.entcs.2007.10.007

Kelly, S., Kaye, S. A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, *77*. https://doi.org/10.1016/j.tele.2022.101925

Kılınç, H. K., & Keçecioğlu, Ö. F. (2024). Generative Artificial Intelligence: A Historical and Future Perspective. *Academic Platform Journal of Engineering and Smart Systems*, *12*(2), 47–58. https://doi.org/10.21541/apjess.1398155

Latikka, R., Bergdahl, J., Savela, N., & Oksanen, A. (2023). AI as an Artist? A Two-Wave Survey Study on Attitudes Toward Using Artificial Intelligence in Art. *Poetics*, *101*. https://doi.org/10.1016/j.poetic.2023.101839

Liang, T., Li, P., Cao, Y., Wu, X., Zhao, H., & Qi, C. (2023). Research on Generating Xi'an Drum Music Based on Generative Adversarial Network. *2023 IEEE International Conference on Sensors, Electronics and Computer Engineering, ICSECE 2023*, 1398–1402. https://doi.org/10.1109/ICSECE58870.2023.10263414

Lopez-Rincon, O., Starostenko, O., & Ayala-San Martín, G. (2018). Algoritmic Music Composition Based on Artificial Intelligence: A Survey. *CONIELECOMP : 28th 2018 International Conference on Electronics, Communications and Computers : Universidad de Las Américas Puebla*.

Louie, R., Coenen, A., Huang, C. Z., Terry, M., & Cai, C. J. (2020, April 21). Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376739

Lu, J., & Eirinaki, M. (2021). Can a machine win a Grammy? An evaluation of AI-generated song lyrics. *Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021*, 4896–4905. https://doi.org/10.1109/BigData52589.2021.9671431

Lu, P., Xu, X., Kang, C., Yu, B., Xing, C., Tan, X., & Bian, J. (2023). MuseCoco: Generating Symbolic Music from Text. In *ArXiv*. http://arxiv.org/abs/2306.00110

Lyricist. (n.d.). Berklee.

Martínez-Plumed, F., Gómez, E., & Hernández-Orallo, J. (2021). Futures of artificial intelligence through technology readiness levels. *Telematics and Informatics*, *58*. https://doi.org/10.1016/j.tele.2020.101525

Messer, U. (2024). Co-creating art with generative artificial intelligence: Implications for artworks and artists. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100056. https://doi.org/10.1016/j.chbah.2024.100056

Meza, G. (2023, May 31). Exploring the potential of interactive Machine Learning for Sound Generation: A preliminary study with sound artists. *NIME'23*. https://asciml.github.io/ASCIML/.

Micchi, G., Bigo, L., Giraud, M., Groult, R., & Levé, F. (2021). I Keep Counting: An Experiment in Human/AI Co-creative Songwriting. *Transactions of the International Society for Music Information Retrieval*, *4*(1), 263–275. https://doi.org/10.5334/tismir.93

Millet, K., Buehler, F., Du, G., & Kokkoris, M. D. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior*, *143*. https://doi.org/10.1016/j.chb.2023.107707

Moruzzi, C. (2018). Creative AI: Music Composition Programs as an Extension of the Composer's Mind. *Philosophy and Theory of Artificial Intelligence. Springer, Cham.*, *32*(4), 69–72. https://doi.org/10.1007/978-3-319-96448-5_8

Music Production: Guide to Producing & Releasing Tracks. (2023, September 27).

Novelli, N., & Proksch, S. (2022). Am I (Deep) Blue? Music-Making AI and Emotional Awareness. *Frontiers in Neurorobotics*, *16*. https://doi.org/10.3389/fnbot.2022.897110

Paiement, J. F., Bengio, S., & Eck, D. (2009). Probabilistic models for melodic prediction. *Artificial Intelligence*, *173*(14), 1266–1274. https://doi.org/10.1016/j.artint.2009.06.001

Pathariya, M. J., Basavraj Jalkote, P., Patil, A. M., Ashok Sutar, A., & Ghule, R. L. (2024). Tunes by Technology: A Comprehensive Survey of Music Generation Models. *2024 International Conference on Cognitive Robotics and Intelligent Systems, ICC - ROBINS 2024*, 506–512. https://doi.org/10.1109/ICC-ROBINS60238.2024.10534029

Prvulovic, D., Vogl, R., & Knees, P. (2022). ReStyle-MusicVAE: Enhancing User Control of Deep Generative Music Models with Expert Labeled Anchors. *UMAP2022 - Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, 63–66. https://doi.org/10.1145/3511047.3536412

Rani, G., Singh, J., & Khanna, A. (2023). Comparative Analysis of Generative AI Models. *2023 International Conference on Advances in Computation, Communication and Information Technology, ICAICCIT 2023*, 760–765. https://doi.org/10.1109/ICAICCIT60255.2023.10465941

Samuels, J. I. (2024). *Exploring Contemporary Trends and Techniques in Music Production*.

Sarmento, P., Mary, Q., Kumar, A., Barthet, M., Carr, C. J., Zukowski, Z., & Yang, Y.-H. (2021). DadaGP: A Dataset of Tokenized GuitarPro Songs for Sequence Models. *22nd Int. Society for Music Information Retrieval Conf*.

Smirke, R. (2024). *IFPI Global Report 2024: Music Revenues Climb 10% to $28.6 Billion*. Billboard.

Tubadji, A., Huang, H., & Webber, D. J. (2021). Cultural proximity bias in AI-acceptability: The importance of being human. *Technological Forecasting and Social Change*, *173*. https://doi.org/10.1016/j.techfore.2021.121100

Venkatesh, S., Braund, E., & Reck Miranda, E. (2020, July 21). Composing Popular Music with Physarum polycephalum-based Memristors. *NIME'20, Royal Birmingham Conservatoire*.

Walton, C. W. (2005). Basic forms in music. *Alfred Music*.

Wang, L., Zhao, Z., Liu, H., Pang, J., Qin, Y., & Wu, Q. (2024). A review of intelligent music generation systems. In *Neural Computing and Applications* (Vol. 36, Issue 12, pp. 6381–6401). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00521-024-09418-2

Wang, W. (2023). Research on AI composition based on deep learning techniques. *2023 IEEE 3rd International Conference on Data Science and Computer Application, ICDSCA 2023*, 1486–1489. https://doi.org/10.1109/ICDSCA59871.2023.10393552

Wu, S.-L., & Yang, Y.-H. (2020). THE JAZZ TRANSFORMER ON THE FRONT LINE: EXPLORING THE SHORTCOMINGS OF AI-COMPOSED MUSIC THROUGH QUANTITATIVE MEASURES. *21st International Society for Music Information Retrieval Conference*. https://drive.google.com/drive/folders/

Xiong, Z., Wang, W., Yu, J., Lin, Y., & Wang, Z. (2023). A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music. *ResearchGate*. https://doi.org/10.13140/RG.2.2.22173.90084

Yang, L. C., & Lerch, A. (2020). On the evaluation of generative models in music. *Neural Computing and Applications*, *32*(9). https://doi.org/10.1007/s00521-018-3849-7

Yellowbrick. (2023). *Understanding the Music Industry's Business Structure*.

Zhou, E., & Lee, D. (2024). Generative artificial intelligence, human creativity, and art. *PNAS Nexus*, *3*(3). https://doi.org/10.1093/pnasnexus/pgae052

Zhu, Q., Shankar, A., & Maple, C. (2024). Grey wolf optimizer based deep learning mechanism for music composition with data analysis. *Applied Soft Computing*, *153*. https://doi.org/10.1016/j.asoc.2024.111294

Zlatkov, D., Ens, J., & Pasquier, P. (2023). Searching for Human Bias Against AI-Composed Music. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13988 LNCS*. https://doi.org/10.1007/978-3-031-29956-8_20

9. Appendix

9.1 Questionnaire

Here we attach the questionnaire administered to the public, including the section only for professionals that was not part of this research.

Exploring the future of generative AI in the music composition market

Welcome!

Thank you for taking the time to participate in this research survey. I'm a master's student currently conducting research on the impact of Generative AI in music generation, now exploring its possible market applications.
The purpose of this survey is to gather insights from music experts as well as the general public and explore which products or services could be appealing to and satisfy needs of different customer segments.

What to expect:
This survey consists of 5/6 sections and it should take approximately 10-15 minutes to complete. Your responses will stay anonymous and all data will be kept confidential and used solely for academic research purposes.

If you have any questions about the research or the survey, please feel free to contact me at silvia.candusso@edu.escp.eu

Thank you for your time and valuable insights!

* Indicates required question

1. Gender *

Mark only one oval.

- Female
- Male
- Prefer not to say
- Other: ____________________

2. Age group *

Mark only one oval.

- 15-19
- 20-29
- 30-39
- 40-49
- 50-59
- 60-69
- 70+

3. What is your nationality? *

____

4. What is your highest level of education? *

Mark only one oval.

- Middle School
- High School diploma or equivalent
- Bachelor's degree
- Master's degree or higher
- Prefer not to say
- Other: _______________________________

5. What is your current profession? *

____

6. What is your annual income? *

Mark only one oval.

- Less than 20000€
- 20000-50000€
- 50000-70000€
- More than 70000€

7. How much do you know about generative AI? *

Mark only one oval.

	1	2	3	4	5	
Notl						I am an expert in the field

Music habits assessment

In this section you will answer to questions on your musical knowledge, habits and dedication

8. What is your relationship with music? *

Mark only one oval.

- I don't care/ I rarely listen to any
- I enjoy listening to music
- I have musical knowledge/ I have studied music
- I am a music professional (Whoever works in the music industry or is trying to make a career)
- Other: ______________________

9. If you are a music professional, what is your role? *

Mark only one oval.

- Not a professional
- I produce beats/ compose music/ write lyrics
- I do not have experience in writing songs, I only perform
- I do not have experience in writing songs, I have an auxiliary role
- Other: ____________________

10. What genres do you usually listen to? *

____

11. When listening to music how much do the following statements sound like you? "I listen to music..."
1(Not at all) - 4(Absolutely yes)

Mark only one oval per row.

	1 Not at all	2	3	4 Absolutely yes
to relax				
to concentrate				
as a background, I do not pay much attention				
to get energy				
to get an emotional connection with the singer				
actively and to appreciate good artistry				

15. Can you name specific applications or tools? *

____

16. If you have experience in composing and producing music, how many songs have you contributed to?

Mark only one oval.

- No experience
- Less than 5
- 5-20
- 20-50
- 50-100
- 100+

17. If you have experience in composing and producing music, what tools/software do you us

____

18. Do you make money from the music you create? *

Mark only one oval.

- No
- Yes, but not enough for making it a living
- Yes, it is my primary source of income
- I do not create music

12. StrumentaleWhat makes you like a song? What do you feel more connected to when listening to it?
Put them in order from the most important (1) to the least (7) ---> only one per column

Mark only one oval per row.

	Lyrics	Genre	Artist	Instrumental	Articulate melody	Production	General mood
1	◯	◯	◯	◯	◯	◯	◯
2	◯	◯	◯	◯	◯	◯	◯
3	◯	◯	◯	◯	◯	◯	◯
4	◯	◯	◯	◯	◯	◯	◯
5	◯	◯	◯	◯	◯	◯	◯
6	◯	◯	◯	◯	◯	◯	◯
7	◯	◯	◯	◯	◯	◯	◯

13. Do you see music as an art only destined to humans? *

Mark only one oval.

	1	2	3	4	
Not	◯	◯	◯	◯	Categorically yes

14. How aware are you about the applications of AI in music creation? (How it is used and could be)

Mark only one oval.

	1	2	3	4	5	
I do	◯	◯	◯	◯	◯	I am an expert

Music and AI

Let's now explore what do you think of AI applications in music generation.

When we talk about AI in music generation, we talk about the process of creating a new song from scratch.

The process can happen with the aid of AI to get some ideas or the AI could do the work but itself. However, at the state of the art, someone has to give some kind of input to the AI to let it know the requirements and the direction to take.

19. How you feel towards AI generated music? *

Mark only one oval.

- ◯ I hate the idea
- ◯ I am diffident
- ◯ I am indifferent
- ◯ I am curious
- ◯ I love the idea
- ◯ Other: ____________

20. Do you think AI can produce good quality music as of now? *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Real	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Exceptional quality (better than humans)

21. Why? *

____

22. Would you listen to music knowing that it was completely made by AI? *

Mark only one oval.

	1	2	3	4	
No r	◯	◯	◯	◯	Yes if I like it

23. Would you listen to music knowing that the artist was helped by AI? *

Mark only one oval.

	1	2	3	4	
No r	◯	◯	◯	◯	Yes if I like it

24. What is the maximum "human/AI ratio" you think you would tolerate? *

Mark only one oval.

- ◯ 0% AI - 100% human
- ◯ 25% AI - 75% human
- ◯ 50% AI - 50% human
- ◯ 75% AI - 25% human
- ◯ 100% AI - 0% human

25. If you have experience in composing and producing music, have you ever tried using AI to help you in the process?

Mark only one oval.

- Yes
- No
- No experience in composing or producing

26. If yes, do you think it helped and how?

Check all that apply.

- [ ] Time saving
- [ ] Cost saving
- [ ] New ideas when I was stuck
- [ ] It didn't really help
- [ ] Did something that I didn't know how to do
- [ ] Other: ______________________

27. If no, why?

Check all that apply.

- [ ] I do not know the technology well enough
- [ ] I do not want to use it
- [ ] It costs too much
- [ ] I tried to use it but think it still needs development before I can use it for a song
- [ ] Other: ______________________

Listening section

Please listen to the following audio, it will present an extract from 4 different songs.

Provide a score for how much you think the **song is well-made and good quality**, please try to put aside your personal preferences.
Then try to define whether it is **human-made or AI-made**.

Listening

http://youtube.com/watch?v=76_lQXNEJdw

28. First extract, quality *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Real	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Excellent

29. First extract *

Mark only one oval.

- ◯ Surely human-made
- ◯ Supposedly human-made
- ◯ Not sure
- ◯ Supposedly AI-made
- ◯ Surely AI-made

30. Why? *

31. Second extract, quality *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Real	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Excellent

32. Second extract *

Mark only one oval.

- ◯ Surely human-made
- ◯ Supposedly human-made
- ◯ Not sure
- ◯ Supposedly AI-made
- ◯ Surely AI-made

33. Why? *

____

34. Third extract, quality *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Real	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Excellent

35. Third extract *

Mark only one oval.

- Surely human-made
- Supposedly human-made
- Not sure
- Supposedly AI-made
- Surely AI-made

36. Why? *

____

37. Fourth extract, quality *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Real											Excellent

38. Fourth extract *

Mark only one oval.

- Surely human-made
- Supposedly human-made
- Not sure
- Supposedly AI-made
- Surely AI-made

39. Why? *

Results

Thank you for coming this far!

Results from before
First, I'll let you know what are the answers from the previous exercise.
The first, second and fourth were AI generated. (Suno)

For the first and fourth one, I gave a 100 words description of the theme, the mood and the topic of the song and the AI generated lyrics, instrumental and vocals.

For the second I gave the lyrics to the AI and it only generated instrumental and vocals.

The third one was "Monkey Moves" by an artist named Nelwards.

40. Now that you have listened to AI generated music, what do you think about its quality? *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Real	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Exceptional (better than humans)

41. How do you feel now towards AI generated music? *

Mark only one oval.

- I hate the idea
- I am diffident
- I am indifferent
- I am curious
- I love the idea
- Other: ____________________

42. Do you create, compose or produce music? *

Mark only one oval.

- Yes *Skip to question 43*
- No *Skip to question 50*

Composing with AI

Only for people that create music

43. Rate the importance that the following features and functionalities would have on a product, based on generative AI, that helps you in your music composition process.
1= Not important 4=Essential

Mark only one oval per row.

	1 Not important	2	3	4 Essentia
Compatible to owned systems (DAW, hardware, softwares, ...)				
Output is a song with no need to edit (with also vocals if part of the song)				
Output is editable audio tracks (opposed to only having MIDI)				
Output also has lyrics written by AI (not with the voice already synthesized)				
Give lyrics as an input				
Inputs are sounds from a microphone				
Define key, tempo, instruments, structure...				
Define mood, a genre, a theme				
Music-infilling or expand feature (fill the gap between two pieces of already existing music or expand tracks with new sections)				
Real-time composition assistance (composing while playing)				

44. Would you like to add some features or services that you deem as important but were not in the previous grid?

45. State a price that you would consider too low for it to be a good product (for your ideal product) ?
Annual payment €

____

46. State a price that you would consider a "good deal" (for your ideal product) *
Annual payment €

____

47. State a price that you would consider too high (for your ideal product) *
Annual payment €

____

48. How likely are you to use generative AI tools in the future? *

Mark only one oval.

	1	2	3	4	5	
Very	◯	◯	◯	◯	◯	Very Likely

49. If you are, what will you use it for?

____

AI-generated music, applications

This is the last section, it's almost over!

We are almost at the end of the questionnaire, where you will need to answer to questions about your willingness to use and/or pay for services or products that use generative AI to compose music

For everyone

The use of generative AI in music production can have many applications that go further the use from music professionals.
It could be used for personalized soundtracks, in video games, advertisements, for social media, for therapeutic reasons or even to help you stay awake while driving.
The possibilities are broad and almost unexplored.

50. How much value is added to services with personalized playlists that generate music based on you? (what you like, how you are feeling...)
4 options

Mark only one oval per row.

	Counterproductive	No value added	Some value added	Exceptional value added
Fitness app				
Meditation and relaxation apps				
Background music for studying or working				
Music in video games or interactive media				
Personalized soundtracks for events, DJsets, parties, gatherings				
Dynamic music for videos or content creation (music adapts to the content you are posting on social media)				
Music therapy				
Music while driving				

51. How much attention do you pay to music when doing these activities? *
1=I don't pay attention at all 5=I pay a lot of attention

Mark only one oval per row.

	1 I don't pay attention	2	3	4	5 I pay a lot of attention	Not applicable/ I don't listen to music during it
Listening to podcasts (music of the podcast)						
Playing video games (music of video game)						
Watching a film or TV show (soundtrack)						
Music in a shop, restaurant or hotel						
Driving						
Studying or working						
Doing sports / going to the gym						
Going to events, parties...						
Social media content						
Advertisements						

52. What would be your reaction if you knew the following music was made by AI? *

Mark only one oval per row.

	Negative	Slightly negative	Indifferent	Slightly positive	Positiv
Listening to podcasts (music of the podcast)					
Playing video games (music of video game)					
Watching a film or TV show (soundtrack)					
Music in a shop, restaurant or hotel					
Driving					
Studying or working					
Doing sports / going to the gym					
Going to events, parties...					
Social media content					
Advertisements					