


# Semantic Representation of Musical Identity: AI-Driven Cover Image Generation from Lyrics

Kyle Hoang and Pablo Rivas 

Department of Computer Science, Baylor University, Waco, Texas, USA  
{Kyle\_Hoang2,Pablo\_Rivas}@Baylor.edu

**Abstract.** This study presents a computational approach to generating playlist cover images by leveraging lyric-based semantic analysis. We employ natural language processing techniques, including part-of-speech tagging, lemmatization, and K-means clustering, to extract thematic and emotional features from song lyrics. These extracted features are then transformed into structured prompts for text-to-image models, enabling the automated generation of visually representative playlist cover art. To evaluate the effectiveness of the generated images, we conducted a user study that assessed aesthetic quality, thematic coherence, and user satisfaction. Our results indicate that AI-driven lyric analysis can produce compelling visual representations aligned with musical identity. This framework has potential applications in music streaming platforms, digital media curation, and personalized content generation, offering a novel intersection between computational creativity and AI-driven design.

**Keywords:** Natural Language Processing · Generative AI · Computational Creativity.

## 1 Introduction

Integrating artificial intelligence (AI) into music and visual media has led to advancements in computational creativity, with applications ranging from automated music composition to AI-generated artwork. As machine learning and natural language processing (NLP) techniques continue to evolve, new opportunities arise for AI to bridge auditory and visual experiences. This study investigates the application of AI in generating playlist cover art that semantically aligns with the thematic and emotional content of song lyrics. By employing computational techniques to analyze lyrical data, we introduce a framework that translates textual representations into structured prompts for image generation models, facilitating the creation of personalized and contextually meaningful cover art.

The use of AI in music generation dates back to early expert systems and rule-based models in the late 20th century, with more recent approaches incorporating deep learning and probabilistic modeling to enhance creative outputs [12]. However, AI has also been explored beyond generating music in understanding and interpreting musical semantics, emphasizing the importance of emotional

and cultural context in AI-driven creativity [6]. This semantic understanding is crucial for producing artistically resonant content.

Parallel to advancements in AI-driven music analysis, generative models have gained traction in visual domains, enabling the creation of artwork that complements musical themes [2]. While AI-generated art has seen widespread adoption in various applications, existing methodologies often lack the capability to generate personalized, context-aware visuals that reflect the deeper meaning embedded in musical compositions [9]. Addressing this limitation, our research leverages NLP techniques, including part-of-speech tagging, lemmatization, and clustering, to extract meaningful lyrical features that inform text-to-image models, thus generating cover art that encapsulates the emotional and thematic essence of curated playlists.

The methodology presented in this paper follows a structured multi-stage process: (1) lyrical content analysis for thematic and mood extraction, (2) transformation of extracted features into structured text prompts, and (3) generation of cover images using AI-based generative models [13]. To assess the efficacy of the proposed approach, we conducted a user study evaluating aesthetic quality, thematic coherence, and perceived relevance of the generated images. Our findings indicate that AI-driven lyric analysis can be effectively employed to produce visually compelling and contextually aligned playlist cover art.

Beyond artistic enhancement, the broader implications of this research extend to user personalization and engagement in digital music consumption. AI-driven content personalization has become increasingly relevant in media technologies, contributing to more immersive and user-centric experiences [7]. By tailoring visual representations to reflect individual musical preferences, this work aligns with the growing trend of leveraging AI for enhancing digital interactions.

This paper contributes to the discourse on AI’s role in creative industries by synthesizing insights from computational creativity, music informatics, and generative AI. Through the proposed framework, we highlight the potential of AI in augmenting artistic expression and facilitating new forms of multimedia synthesis. Our findings underscore the importance of semantic understanding in AI-driven visual generation, demonstrating how AI can enhance the relationship between music and visual representation in digital platforms.

This paper is organized as follows: Section 2 reviews related work on AI in art and playlist analysis. Section 3 outlines the methodology, including data pre-processing, lyric analysis techniques, and prompt generation. Section 4 presents the results of the image generation and user feedback. Section 5 discusses the findings, project limitations, and implications. Finally, Section 6 summarizes the conclusions and findings of the paper.

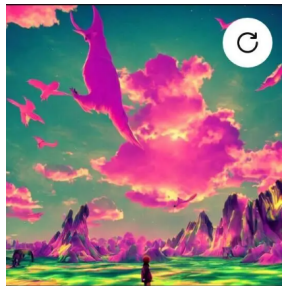
## 2 Background/Related Work

AI-generated album art has been an area of active experimentation, with various approaches demonstrating differing levels of success. Examples include the work by Purz [8] and NachoooCheezus [5].

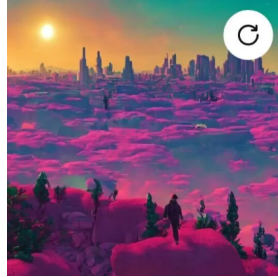
The album art generator by NachoooCheezus [5], which relies on an older version of the Stable Diffusion model, often produced unsatisfactory results. These images were characterized by limited detail and poor image quality. The model developed by Purz [8] performed well in creating jazz-themed album covers, showcasing a strong alignment with the stylistic elements of the genre. However, its utility for other genres was limited, as the generated images frequently retained the distinctly jazz-inspired aesthetic, regardless of the intended genre. These limitations highlight the need for more versatile models capable of adapting to diverse genres and details, as explored in this project.

Spotify has also previously explored the use of AI to generate playlist cover art, aiming to capture the "vibe" of the playlist. For example, in one demonstration, a playlist named 'One Last Summer Party With Justin' containing 50 songs by R&B, Rap, and Hip-Hop artists such as JAY-Z, Kendrick Lamar, Big Sean, and Usher was used to generate cover art three times. The process, described as requiring a few minutes, produces results that Spotify claims are tailored to the playlist's content [4].

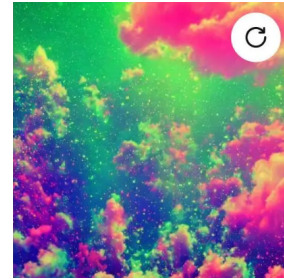
However, user feedback on these AI-generated images has been largely critical. Comments from individuals evaluating examples of Spotify's AI-generated covers often described them as "generic" and "uninspired." Many observed that the resulting visuals failed to align with the playlist's dominant genres, artists, or themes. Fig. 1a through Fig. 1c show three attempts made by Spotify's AI for the same previously mentioned playlist.



(a) A fantastical landscape that does not align with the playlist's genres.



(b) Another generic fantastical landscape, with little consideration for the playlist.



(c) A vivid abstract scene completely detached from the described playlist.

Fig. 1: Spotify's AI-generated cover art for the same playlist, showing attempts with varying results.

Although Spotify's efforts represent a step toward automated playlist visualization, they fall short of capturing the deeper meaning behind a playlist's creation. This project seeks to advance this concept by incorporating more detailed analyses of playlist characteristics, including:

1. **Playlist Name and Description:** To reflect the creator’s intent.
2. **Dominant Genres:** To provide genre-specific aesthetics.
3. **Mood and Themes:** To encapsulate the emotional tone of the songs.
4. **Common Lyrics:** To incorporate song-specific details

By addressing these aspects, the project aims to produce personalized cover art that resonates more closely with the user’s expectations and the playlist’s essence.

### 3 Methodology

This project began with identifying a suitable dataset for analysis. The Spotify Million Playlist Dataset was selected, as it provides metadata for one million playlists, including information such as playlist names, descriptions, creation and modification timestamps, track titles and artists, durations, and more. However, the dataset does not include song lyrics, which are critical for analyzing themes, moods, and including song-specific details. To address limitations in testing and feedback, the project later migrated from the dataset to Spotify playlist links, enabling improved evaluation and iteration.

#### 3.1 Data Collection

The project was implemented entirely in Python using Google Colab. Song metadata was obtained using the Spotify API. Song lyrics were retrieved using the lyrics.ovh and Genius APIs. Retrieved lyrics were stored in json files to reduce runtime and number of required API calls. To optimize data retrieval from the Genius API while avoiding rate limitations, adaptive timing mechanisms were implemented between API calls, ensuring uninterrupted data collection.

#### 3.2 Ethical Considerations and Copyright Issues

When collecting song metadata and lyrics from online sources, several ethical and legal considerations must be addressed. Lyrics are typically copyrighted material, and while APIs such as Genius and lyrics.ovh provide access to them, usage is subject to their respective terms of service. Unauthorized scraping or redistribution of lyrics may infringe upon copyright laws, potentially leading to legal consequences.

To mitigate these concerns, we adhered to the following ethical guidelines:

- **API Compliance:** Data was collected exclusively through officially provided APIs (Spotify, Genius, and lyrics.ovh) while adhering to their terms of service. This ensured that data retrieval remained within legally acceptable boundaries.
- **Minimization of Unnecessary Requests:** To reduce the burden on API providers and respect rate limits, lyrics were stored locally in JSON format, minimizing repeated requests and potential disruptions.



- **Fair Use Considerations:** The use of lyrics in this project was strictly for research and analytical purposes, without redistribution for commercial gain. However, fair use is a complex legal doctrine that varies by jurisdiction, and care was taken to avoid excessive reproduction of copyrighted material.

### 3.3 Text Analysis and Prompt Generation

Once the lyrics were collected, several NLP techniques were applied to analyze the text. The methodology included the following steps:

- **Word Filtering:** Common but contextually uninformative words (e.g., "ay," "na," "yeah") were filtered out to enhance the quality of the analysis.
- **POS Tagging and Lemmatization:** Each word was assigned its grammatical part of speech and reduced to its base form through lemmatization, ensuring uniformity and enhancing the accuracy of subsequent analyses.
- **Clustering:** Lyrics were grouped using algorithms such as K-means clustering to identify recurring themes and concepts.
- **Sentiment and Mood Analysis:** Pre-existing language models were employed to determine the sentiment and mood of each song, categorizing the results into moods such as "upbeat and fun" or "dark and moody."
- **Prompt Construction:** A comprehensive text prompt was generated for each playlist, incorporating the dominant genre, playlist description, name, and the results of the aforementioned analyses. For example, the playlist 'Throwbacks', which includes songs such as 'California Girls' by Katy Perry and 'Promiscuous' by Nelly Furtado, produced the following prompt:  
Generate an album cover for the playlist named 'Throwbacks': upbeat and fun mood, pop genre, themes: love, girl, tonight, time, heart, dance, life, feel, hey, and night. Integrate bright, playful text reading: 'Throwbacks' into the image. Helpful but not required lyrics: Lyrics clusters: ['floor,hands,dance,tonight', 'wake,die,heart', 'time,gonna,love', 'babe,stay,head', 'hey,boy,feel'].

### 3.4 Image Generation

The generated text prompts were input into an image generation model, initially a Stable Diffusion model, to create album cover art. However, the model's 77-token limit caused prompt truncation, leading to incomplete inputs. Additionally, user feedback indicated that the generated images were mediocre, as they did not sufficiently reflect playlist themes and lacked personalization due to the absence of playlist names.

To address these shortcomings, the project transitioned to using the Kandinsky Stable Diffusion model [3]. While this model improved upon the previous one, it exhibited significant difficulty in generating text. Additionally, it retained the same 77-token limit. To address this, embeddings were averaged from smaller chunks of the prompt. In an attempt to improve results, a new approach was adopted: passing the generated prompt through a prompt refinement language

model [1] before passing it into the Kandinsky model. However, the refined prompts often produced generic images, failing to meet user expectations for specificity and personalization.

Further iterations led to the use of the ShuttleAI model. While the ShuttleAI model proved too resource-intensive for Google Colab, its API enabled effective usage. ShuttleAI demonstrated superior performance with text, though occasional retries were needed to ensure accurate spelling. Prompt refinement was ultimately abandoned at this stage, as it contributed to the generation of overly generic images. The final approach relied directly on ShuttleAI, which consistently delivered improved results, delivering thematic relevance and playlist-specific personalization.

## 4 Experiments

### 4.1 Model Comparisons

Using the following prompt, we generated an image with our Stable Diffusion model:

Generate an album cover for the playlist named ‘Throwbacks’: upbeat and fun mood, pop genre, themes: love, girl, tonight, time, heart, dance, life, feel, hey, and night. Integrate bright, playful text reading: ‘Throwbacks’ into the image. Helpful but not required lyrics: Lyrics clusters: [‘floor, hands,dance,tonight’, ‘wake,die,heart’, ‘time,gonna,love’, ‘babe,stay,head’, ‘hey,boy,feel’]

Fig. 2a through Fig. 2c showcase the attempts of three different models to create cover art for the same "Throwbacks" playlist based on the given prompt.



(a) Using our Stable Diffusion model.



(b) Using the Kandinsky model.



(c) Using the ShuttleAI model.

Fig. 2: AI-generated cover art for the "Throwbacks" playlist using three different models.

The initial Stable Diffusion-generated images demonstrated limitations in thematic accuracy and aesthetic appeal, as indicated by user feedback. A survey was conducted to systematically evaluate model outputs. The ShuttleAI model received the highest average rating of 4.6/5 stars, demonstrating its ability to generate more visually appealing and contextually relevant images. In contrast, the initial Stable Diffusion model received an average rating of 2/5, indicating significant limitations in its ability to generate contextually relevant and visually coherent images. This survey included other images not shown here.

Having identified ShuttleAI as the most effective model, the project was expanded to include a broader user base for additional feedback and testing. When testing, it was found that the name of the playlist played a large role in the image generation. Fig. 3a and Fig. 3b show two images generated by the project, using the same playlist under different names.



Fig. 3: Two sample images generated with our project under two different names.

The user commented that the original image fit his playlist nicely. The second image garnered some confusion, as there would have been no other references to basketball besides the playlist's name.

## 4.2 User Results and Feedback

This section presents the generated images alongside user feedback. To enhance readability, the corresponding prompts are provided separately after the feedback. Users were asked to evaluate the generated images based on a star rating system (1 to 5), accompanied by optional comments. This approach allowed for both quantitative and qualitative assessments, capturing user preferences, criticisms, and suggestions. User feedback indicated a strong preference for integrating playlist names into generated images. To address this, the prompt generation process was refined to explicitly incorporate playlist titles with improved legibility.



(a) Rating: 5  
Feedback: very great, described my playlist perfectly! There is a double letter in the playlist name though.



(d) Rating: 4  
Feedback: This looks awesome. The playlist name is really cool. I expected something a little more aggressive or war-like, but I will definitely be using it!



(b) Rating: 4  
Feedback: This is really cool! I think it matched my mix of indie/granola and mainstream songs. The misspellings in back are funny, but not incredibly noticeable.



(e) Rating: 4  
Feedback: This is so pretty! It spelled my name right :) But not scorgo4000 lol



(c) Rating: 3  
Feedback: The image is cool, but I can tell the Genius API was unable to retrieve lyrics. It doesn't quite fit my playlist.



(f) Rating: 4  
Feedback: This really captures the grunge yet more laidback vibe I was going for. Not sure what the word is under 'Blind Melon' though

Fig. 4: AI-generated cover art for six different users and their playlists.

Fig. 4a through Fig. 4f show six images generated by the project, accompanied by the user feedback. The generated prompts given to the image model are provided here for completeness and better readability:

1. Fig. 4a – Generate an album cover for the playlist named 'Automatic': upbeat and fun mood, rap genre, themes: ich, beat, du, underground, tchê, b\*\*ch, eye, sound, and che. Integrate graffiti-style text reading: 'Automatic' into the image. Helpful but not required lyrics: Lyrics clusters: ['s\*\*t, che, money,

- b\*\*ch', 'rain, tattoo, care', 'foemen, mother', 'whiskey, gettin, bar, tipsy', 'der, aller, amos, ist']].
2. Fig. 4b – Generate an album cover for the playlist named 'Mix': upbeat and fun mood, pov: indie genre, themes: time, hey, love, feel, heart, stay, night, hand, mind, and body. Integrate clear, legible text reading: 'Mix' into the image. Helpful but not required lyrics: Lyrics clusters: ['love, hey, time', 'miles, moods, guru', 'thuggin, bouncin, girl, bounce, babe', 'fluorescent, play, head', 'mind, change, diane, time, ride']].
  3. Fig. 4c – Generate an album cover for the playlist named 'Defiant': upbeat and fun mood, metropolis genre, themes: empire, rise, fall, live, die, throne, stand, storm, coming, and desire. Integrate bright, playful text reading: 'Defiant' into the image. Helpful but not required lyrics: Lyrics clusters: ['stand, storm, throne, tides, war', 'throne, fall, rise, stand, empires', 'stand, storm, throne, tides, war', 'stand, storm, throne, tides, war', 'stand, storm, throne, tides, war']].
  4. Fig. 4d – Generate an album cover for the playlist named 'To Valhalla': dark and moody mood, antiviral pop genre, themes: pain, force, nature, leaf, shame, fury, fall, world, unseen, and truth. Integrate bright, playful text reading: 'To Valhalla' into the image. Helpful but not required lyrics: Lyrics clusters: ['energy, embrace, echoes, yo', 'fall, truth, nature, force', 'carry, fury, shame, pain', 'hoist, thieves, ho', 'energy, embrace, echoes, yo']].
  5. Fig. 4e – Generate an album cover for the playlist named 'Zita + scorgo4000': upbeat and fun mood, rock genre, themes: da, love, lie, time, boy, girl, feel, dance, casbah, and weezer. Playlist description: A Blend of music for scorgo4000 and Zita. Updates daily. Helpful but not required lyrics: Lyrics clusters: ['dirty, boy, heat, fool, linger', 'coalmine, tomorrow, casbah, girl', 'river, moonlight, strange', 'hope, remember, lies', 'time, guess, da, head']].
  6. Fig. 4f – Generate an album cover for the playlist named 'Blind Melon - No Rain': upbeat and fun mood, alternative rock genre, themes: time, sun, love, feel, day, heart, hole, eye, black, and life. Integrate clear, legible text reading: 'Blind Melon - No Rain' into the image. Helpful but not required lyrics: Lyrics clusters: ['hip, sun', 'gonna, difference, heart, time', 'release, understand, tomorrow, life', 'hearts, love', 'ground, head']].

## 5 Discussion and Analysis

The use of the ShuttleAI image generation model [10,11] yielded consistently favorable results, with no user rating an image below a 3. This remained true even in cases where text generation was imperfect or when the project failed to retrieve lyrics from Genius, leading to less specific prompt details. Based on user feedback, the project can be considered a success, as each user received an image they liked and intended to use within three image generation attempts. With each generation requiring a maximum of 30 seconds, the process proved efficient and user-friendly.

While the results are highly satisfactory, there are several aspects that could be further refined or expanded upon to improve the project:

### 5.1 Integrating Contextual Information

The current implementation focuses primarily on the themes, moods, genres, and lyrical details extracted from playlist metadata and lyrics. However, incorporating additional contextual information, such as the historical or cultural context at the time of playlist creation, could add depth and personalization to the generated images. For instance, if a playlist reflects a particular era or event (e.g., "Summer 2010 Hits"), this context could influence the visual style and elements of the album cover, aligning it more closely with the user's intent.

### 5.2 User Personalization

The project currently generates prompts and images based on playlist data. However, considering user-specific preferences or emotions could further enhance the relevance of the generated images. For example, including options for users to specify preferred visual styles (e.g., minimalist, abstract, retro). Surveys or additional feedback mechanisms could help identify these preferences. Additionally, the project could be extended to include an image-to-image generation feature, allowing users to refine the output further. For example, users who are satisfied with the overall design but wish to add or modify specific elements could input the generated image alongside a textual description to produce a revised version. This iterative process would enable greater customization and user satisfaction. Future implementations could integrate user-specific customization features to enhance engagement, particularly if deployed as an application or within a streaming service ecosystem. The current study focused on minimal user interaction, leaving this aspect for future exploration.

### 5.3 Improving Text Integration

One of the key challenges faced was ensuring accurate and visually appealing text integration in the generated images. While ShuttleAI performed better than previous models, issues with text clarity and accuracy persisted. Exploring alternative methods, such as overlaying text programmatically onto the images post-generation, could improve results. While preliminary attempts to integrate text programmatically did not yield optimal results, alternative strategies for seamless text incorporation remain an area for further investigation. Additionally, incorporating advanced techniques like Optical Character Recognition (OCR) could help validate and correct generated text.

### 5.4 Enhancing Prompt Generation

The quality of the generated images is inherently tied to the effectiveness of the prompts. While the current prompt generation captures themes, moods, and genres effectively, further refinement is possible. For instance, leveraging advanced NLP techniques to better understand lyrical metaphors or deeper semantic meanings could lead to richer, more nuanced prompts.

### 5.5 Addressing Limitations in Data Availability

The project’s reliance on scraped lyrics presents certain challenges, such as incomplete or missing data for some songs. Some queries yielded non-musical text segments, such as audiobook transcripts, introducing noise into the dataset. Future work could incorporate filtering mechanisms to enhance data quality. Expanding the dataset through partnerships with music platforms or licensed APIs could improve the coverage and quality of the lyrics used for analysis. Furthermore, creating a robust fallback mechanism to handle missing lyrics—perhaps by analyzing audio features—could ensure the prompt generation process remains effective even with incomplete data.

### 5.6 Future Applications and Scalability

This project demonstrates a proof of concept for creating personalized album covers based on playlist data. Future directions could include scaling the model for broader use cases, such as generating artwork for podcasts, audiobooks, or other audio-centric content. Integrating this tool into music platforms like Spotify or Apple Music could allow users to create custom visuals for their playlists directly within the platform.

## 6 Conclusions

This study demonstrates the potential of AI-driven lyric analysis for generating visually coherent and semantically relevant playlist cover art. By applying natural language processing techniques to extract thematic and emotional elements from song lyrics, we developed a framework that transforms textual representations into structured prompts for text-to-image models. The results indicate that this approach can successfully generate aesthetically compelling and contextually meaningful visual outputs, as validated through user evaluations.

Future improvements include refining prompt generation methods to enhance the consistency and interpretability of generated images, incorporating user-specific preferences for personalization, and integrating multimodal data sources beyond lyrics to provide richer contextual cues. Additionally, optimizing computational efficiency and exploring deployment on music streaming platforms could expand real-world applicability. This research highlights the intersection of artificial intelligence, computational creativity, and digital media, offering new possibilities for AI-assisted content generation in the music industry.

**Acknowledgments.** The authors thank the Rivas.AI Lab (<https://lab.rivas.ai>) for the support and helpful feedback throughout this project.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Baconnier: Prompt plus plus. <https://huggingface.co/spaces/baconnier/prompt-plus-plus> (2024), accessed: 2024-12-09
2. Carnovalini, F., Rodà, A.: Computational creativity and music generation systems: an introduction to the state of the art. *Frontiers in Artificial Intelligence* **3** (2020). <https://doi.org/10.3389/frai.2020.00014>
3. Community, K.: Kandinsky 2.2 prior model. <https://huggingface.co/kandinsky-community/kandinsky-2-2-prior> (2024), accessed: 2024-12-09
4. McDaid, J.: Spotify custom ai art for your playlist: How to do it. <https://thechainsaw.com/artificial-intelligence/spotify-custom-ai-art-for-your-playlist-how-to-do-it/> (October 2023), accessed: 2024-12-07
5. NachoooCheezus: Album art generator. <https://huggingface.co/spaces/NachoooCheezus/AlbumArt>, accessed: 2024-12-07
6. Novelli, N., Proksch, S.: Am i (deep) blue? music-making ai and emotional awareness. *Frontiers in Neurobotics* **16** (2022). <https://doi.org/10.3389/fnbot.2022.897110>
7. Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., Sra, M.: Ai-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence* **3**, 1013–1022 (2021). <https://doi.org/10.1038/s42256-021-00417-9>
8. Purz: Jazz album cover generator. <https://huggingface.co/Purz/jazz-album-cover>, accessed: 2024-12-07
9. Shank, D., Stefanik, C., Stuhlsatz, C., Kacirek, K., Belfi, A.: Ai composer bias: listeners like music less when they think it was composed by an ai. *Journal of Experimental Psychology Applied* **29**, 676–692 (2023). <https://doi.org/10.1037/xap0000447>
10. ShuttleAI: Shuttle-3.1 aesthetic model. <https://huggingface.co/shuttleai/shuttle-3.1-aesthetic> (2024), accessed: 2024-12-09
11. ShuttleAI: Shuttleai dashboard. <https://shuttleai.com/dashboard> (2024), accessed: 2024-12-09
12. Sturm, B., Ben-Tal, O., Monaghan, U., Collins, N., Herremans, D., Chew, E., Hadjeres, G., Deruty, E., Pachet, F.: Machine learning research that matters for music creation: a case study. *Journal of New Music Research* **48**, 36–55 (2018). <https://doi.org/10.1080/09298215.2018.1515233>
13. Sturm, B., Iglesias, M., Ben-Tal, O., Miron, M.: Artificial intelligence and music: open questions of copyright law and engineering praxis. *Arts* **8**, 115 (2019). <https://doi.org/10.3390/arts8030115>