Culminating Experience:

A Series of AI in Music Evaluations

In this paper, I will utilize and evaluate the following platforms: AIVA, SUNO, Udio, and Moises in their AI capabilities, limitations, and use cases.

AIVA Evaluation

I chose an epic cinematic style in Step by Step with “any minor” chord progression. I could have expanded my input customization options and/or uploaded a reference track to get a closer match to what I wanted the output to be. That being said, the uniqueness of AIVA compared to other tools is the ability to adapt MIDI notes and engage in more manual adjustments after the output is provided without always having to rely on prompts. AIVA appears to be better used by those familiar with other DAWs. I did find the MIDI editor comparable to the one in ProTools.

From a user experience standpoint, it took me a moment to get used to layers with instruments within as within many DAWs, these would be synonymous. The hierarchical dropdowns make sense like when you’re in another DAW and show automation. It would be good if the parent layer restricted the instrument selection/addition within. On the other hand, you might want to group instruments in an overarching bucket like “Percussion.” So there is a fine line between restricting and enabling the user to group how he or she sees fit, which may not fall into anticipated, standard grouping.

When you add a new instrument or layer, the system dynamically incorporates it into the final composition. This involves using AI algorithms that analyze the new instrument’s data, such as its tempo, key, and rhythmic structure. The AI then aligns these parameters with the existing layers, ensuring synchronization of timing, key, and tempo. The system employs neural music synthesis techniques to understand the musical context and blend the new instrument seamlessly into the overall composition. This results in a coherent piece where each layer contributes to the overall structure and sound.

Feature Additions to Enhance AIVA:

- **Retaining Instrument Selection:** When regenerating sections, retaining chosen instruments can maintain consistency and reduce redundant adjustments. This could involve state preservation algorithms to remember user settings across sessions.
- **Allowing Dissonance:** Providing options to override restricted note movements in the Harmony layer would cater to more experimental compositions. Implementing a toggle for algorithmic constraints could offer users greater creative freedom.

- **Web Playback:** Enabling playback on the web without requiring a desktop app would improve accessibility. This could be achieved through cloud-based audio rendering and streaming solutions.
- **DAW Integration:** Offering AIVA as a plugin for popular DAWs like ProTools, Logic Pro, and Ableton Live would streamline workflows. This integration could use MIDI mapping and plugin wrapper technologies to embed AIVA's AI capabilities within existing DAW environments. Additionally, allowing direct import/export with well-known DAWs would facilitate a more seamless user experience.

SUNO Evaluation

My initial prompt was "A psychological thriller piece with elements similar to Hans Zimmer. Includes strings, heavy drums, bass, and a choir." The platform's restriction on using specific references like "Hans Zimmer" may aim to avoid producing outputs too close to copyrighted material. However, one might also be limited in this feature because it is preferred to use a Reference feature instead.

The style description gave some suggested tags, but they seemed minimal. By prompt, for someone who may be less versed in music to describe certain elements they are seeking, this would be a harder route to go to get the desired output. The more tags suggested, the better. Because I might replace it with a tag that is a better match for what I am trying to get as output and what the platform is trained to identify.

Even though I put "a choir" in my prompt, because I marked it as an instrumental, it trumped that part of the prompt due to inconsistency with the toggle request. This indicates a hierarchy of rules or constraints within the AI model, where certain parameters override others based on preset priorities. In this case, the instrumental designation took precedence over inclusion of a choir, demonstrating the AI's adherence to hierarchical decision-making protocols.

Though it was in spinning mode for me, there might be an option to download a custom video to accompany my song which seemed like a value-added feature that some might appreciate.

Udio Evaluation

Given the prompt for custom lyrics, Udio's 350-character limit for lyrics suggests an optimization for initial content generation. It might seem odd to use lyrics for an instrumental. Even with Large Language Processing, if I didn't have the metatags for genre, the lyrics could easily be part of a pop or rock song. I would have ended up with a wider range of options, which works if you don't have a melody or genre in mind.

I used ChatGPT to generate lyrics to feed into a Udio prompt. I inputted the chorus and used metatags: pop, dance-pop, and vocal, though cross-genre music can complicate accurate categorization. One output, even though it was tagged as pop, sounded more electronic. Also, I

had put vocal but it seemed to restrict me to instrumental first. Enhancing genre classification algorithms using deep learning could improve accuracy.

Udio provided two options rather than one like SUNO. I found the “show generation tree” feature intriguing, as it would be cool to see how the output was generated, but it only showed checkboxes for deletion.

The platform provided options to extend sections with additional metatags or remixing features, which demonstrates flexibility in content generation. I remixed one output by 50% as I didn't plan to keep it with the given lyrics.

I had two remix options and the first one sounded like a decent remix based on the minimal info given. I could hear a girl chime in with a pop voice now and again throughout with the lyrics I inserted.

For the original, unselected output for remix, I added an “adult contemporary” metatag as an extension. But I wish it would have allowed a “not” or “except” option so I could say “not electronic”. In that case, I would have to know what inclusions to get it closer to my desired output. Some of the first extensions added genres not there before like Disco. It was odd that it chose to extend that way when I didn’t add a tag for “Disco”. Again, without an exemption, it looks for the inclusion tags and may pull in some additional if they are often accompanied by the tags I inserted.

The final output was less hard and more carefree, underscoring the variability in AI-generated music and the need for iterative refinement to meet user expectations.

Summary of Technologies

Signal Processing, MIR, and Computer-Based Composition Technologies: AIVA, SUNO, and Udio all utilize advanced signal processing and Music Information Retrieval (MIR) techniques to interpret musical inputs and text descriptions, extracting features such as tempo, key, and chord progressions. These technologies ensure the generated music aligns with the user's specifications, creating coherent and musically rich outputs. AIVA particularly benefits from MIR in generating MIDI sequences, while SUNO and Udio leverage these techniques to produce complete musical compositions directly from text.

Algorithmic Composition and Deep Music Generation: All three platforms employ algorithmic composition methods to guide the music generation process. AIVA uses predefined rules and models to generate music, further enhanced by deep music generation techniques that learn from large datasets of existing music. Similarly, SUNO and Udio use deep learning models and neural networks to create original music from text descriptions.

Deep Learning and LLM-Based Techniques: AIVA, SUNO, and Udio incorporate deep learning and Large Language Models (LLMs) to process and interpret user inputs. These models analyze vast amounts of data to learn musical patterns and structures, allowing the

platforms to generate coherent and musically valid compositions. LLMs play a crucial role in understanding and processing text descriptions, ensuring the generated music matches the user's specifications.

Real-Life Use Cases

I chose to use the song I generated with SUNO, because it sounded like it could be synced with a movie. Often, in music supervision, we find that clients want the STEMs of the songs. The client may, for instance, want vocal STEMs removed to use a song as underscore with dialogue over. Moises does STEM separation but AIVA, SUNO, and Udio do not.

For drums, keys and winds, there might have been a bit of phase cancellation with other tracks sharing a frequency space. I found that bass and strings sounded good. Then you had an “other” track which included some of the tracks already present. But this goes back to an initial analysis I gave of Moises regarding the more complex (e.g. more instrumentation) a track has, the harder it is for Moises to separate accurately. It is worth noting that SUNO is text-to-music.

My guess would be if I had used my music output file from AIVA, which is text-to-MIDI, since MIDI files represent individual notes and instruments in a clear, structured format, it’s more straightforward to separate tracks into individual stems. Each MIDI channel can represent a different instrument, so you can easily isolate or manipulate specific parts of the composition. Since MIDI is not audio but rather a set of instructions for synthesizing music, it is much easier to manipulate and separate without quality loss.

I also wanted to apply AI mastering because this song may be used in a movie soundtrack. Again, this is where Moises and some other tools may be better used. I used a Cinematic Drama track under Preferences from APM Music, and uploaded it as a wave file since Moises recommends this file type. The mastered output seemed more well-rounded in sound, more impactful, and present. In comparison, the original sounded more limited in range (not to be confused with using a Limiter).

In general, I find the AI tools can be grouped as such:

- Music Creation
    - AIVA
    - SUNO
    - Udio
    - Amer Music
- Music Processing (Mixing, Mastering)
    - LANDER
    - Moises
    - IZotope Ozone
- Music Extraction
    - Spleeter

    - Moises
- Music Analysis
    - Sonic Visualiser
    - IBM Watson Beat

Interestingly enough, a limitation of many AI platforms is they aren't all-in-one yet. I look forward to when all four bullets can be done in one platform, and even better if integrated with industry-standard DAWs.

Wrap Up

While the ethical stance on certain providers can be debated, it's undeniable that many use cases can be found for these tools. Even with their limitations, these AI music tools streamline and facilitate many workflows for both creatives and non-creatives. While what took hours now seems available at the click of a button, it's important to remember what goes into training these models and ensuring the user experience is an enjoyable and/or productive one.