SCHOLARLY REVIEW ONLINE

Scholarly Review Online - Spring 2025

How does an original classical piece perform against its AI-generated counterpart?

By Micah Chiang

AUTHOR BIO

Micah Chiang is a Sophomore at Valley Christian High in San Jose, California. With a background in Classical Violin and music composition, he's passionate about music and the growing presence of Artificial Intelligence in the music & entertainment field. Micah loves to play golf and listen to canto-pop in his free time.

ABSTRACT

This paper aims to address the impact of AI tools, specifically the Artificial Intelligence Virtual Assistant (AIVA), on musically-educated listeners. The study aims to compare a piece by Classical-era composer Franz Schubert, to three AIVA generated variations on the original. A survey was conducted on Reddit gauging respondents' opinions on metrics such as preference, complexity, and distinguishability after hearing each of the tracks. Respondents were unaware of the original track, making this study a single blind experiment. The survey consisted of Likert scaled questions as well as free-response questions that were based upon four music excerpts with one of them being the original human composition while the rest were AIVA - generated. It was circulated via multiple subforums within Reddit for a week, and twenty-six responses were collected. There were two primary findings. Over half (52%) the respondents correctly differentiated between the three AI-generated tracks and the original. Second, the most preferred track was an AI-generated one, not the original. Finally, using literature collected on the history of AI in music as well as the specific background of AIVA, this paper evaluates AIVA's compositional performance and listener perception when compared to an existing work.

Keywords: AIVA, Classical Music, Preference, Complexity, Distinguishability, Reddit, Tools, Ranking, Metrics, Survey

INTRODUCTION

Artificial Intelligence has taken the world by storm by embedding itself into a multitude of industries and aspects of our society. But because Artificial Intelligence hasn't even been around for a century, it is still a work in progress. In fact, AI is so new that the first bill to regulate it, the AI Act, was introduced in 2021 by the European Union, but only signed into action in March of 2024 (European Union, 2023). This paper aims to address the concerns of artificial intelligence's recent boom in the music composition industry, specifically the classical music genre. Music composition by definition, "can refer to an original piece or work of music, either vocal or instrumental, the structure of a musical piece or to the process of creating or writing a new piece of music" (Copyright.gov, n.d.). The study was built to answer the following question:

How does a known, reputable, publicly available AI composition tool (AIVA) perform compared to the human composer that influenced it, in terms of musical complexity and likability?

As the question suggests, this study will be based on one artificial intelligence tool, Artificial Intelligence Virtual Artist (AIVA). The study has two parts: a brief literature review of the history of AI and AIVA specifically, as well as an overview of the survey conducted and its findings.

A BRIEF HISTORY OF AI

Generative AI originated in the 1950s, with the term "Artificial Intelligence" being coined by American Computer Scientist John McCarthy (Zulic, 2021). Although it is unclear who made the first breakthroughs in this technology, many individuals contributed to the development of generative AI in the 1950s and onwards. Ever since that decade, engineers have been challenged to program algorithms that emulate the different elements of musical composition, like improvisation, musical recognition and notation among others. With the advent of this technology a debate surrounding the "creativity" of these algorithms and the ethics of using such programs would arise (Karpov, 2020). Alan Turing, the famous British computer scientist and mathematician, was involved in early music-generating computers. In 1951, Turing created one of the first computers that could generate simple melodies (Birk A., Chakrabarty, S. 2020). Apart from that, his famous Turing test was actively used as a standard to hold AI generators to when assessing their resemblance to humans. This standard asserted that a system is worthy of having human-like intelligence if it is indistinguishable from a human (McGuire, 2006).

Another prominent early invention took place at the University of Illinois Urbana-Champaign when Professors Lejarin Hiller and Leonard Issacson created the Illiac 1 computer. Included with the computer was the Illiac Suite, a four movement string quartet. The computer used Markov chains, which are a system in which future outcomes in a program are determined by the current state of the program (Hiller & Issacson, 1979). Markov chains among other algorithmic processes would continue to define how AI tools were created, and the different approaches to creating them. Fast forwarding to the 1980s, David Cope from the University of California, Santa Cruz developed EMI, which is short for Experiments in Music Intelligence (Adams, 2010). EMI was considered highly innovative because there was a "random" aspect to it. The program contained a small analytical engine that helped to randomize the melodies to make the pieces more story-like. Prior to the creation of EMI, the algorithmic process was very cut and dry. Although the pieces it created were modeled after Bach, the public showed a great amount of support for these initiatives.

CURRENT TRENDS & COPYRIGHT LAW

Nowadays, a term called "computational creativity" is rising in prominence as a measure of how creative a machine is when composing music (Frid et al., 2020). Most modern AI tools now incorporate "deep learning networks," which are composed of multiple layers of calculations, and "reinforcement techniques,"

where the algorithm is pre-fed a large number of existing compositions to learn from (Moolayil & Ketkar, 2021). It's clear that machine learning has come a long way, to the point where AI and human-made works are not easily distinguishable. Still, the multitude of ethical and technological problems that are presented with developing AI (especially in the area of art like music composition) reveal that machine learning has a long way to go. This study aims to both reinforce this assertion and gauge the extent to which AIVA - generated music is indistinguishable when put alongside a human's.

AIVA

AIVA, short for "Artificial Intelligence Virtual Artist," is a modern electronic music composer that is one, open and free to the public, and two, the first of its kind to be recognized by a music society - that is SACEM (Society of Authors, Composers, and Publishers of Music). It was created in 2016 by Luxembourgish company AIVA Technologies (AIVA, 2016). The electronic composer was even featured in a Ted Talk by CEO and Co-founder Piere Barreau, titled, "How AI could compose a personalized soundtrack to your life" (Barreau, 2018). Many know AIVA by its first studio album, Genesis, which was compared to the music of film composer Hans Zimmer. The comparison was made when both Zimmer's compositions and AIVA's Genesis shared many of the same qualities - masculinity, power, militaristic, and with a strong use of percussion (Zulic, 2019).

AIVA is composed of many systems, including deep learning which incorporates many layers of networks that process input and output data (Kaleagasi 2017). AIVA was created to specialize in classical orchestral music. It was designed with the influence of thousands of pieces from all eras of classical music (baroque, classical, romantic, etc), which is another reason why the survey conducted compared classical pieces. On top of that, AIVA wasn't just designed to assist human composers in creating soundtracks but is embodying the full essence of an independent AI composer (Karpov 2020). Of course, there are still many ethical concerns with AI tools like AIVA that are yet to be resolved - being recognized by a society is one thing, but using it in the music industry is raising a multitude of concerns. This is especially surprising when the concept of AI being able to create real art is an ethical concern. AIVA is still facing copyright issues. Users of the platform can only copyright generated work if they do so through the tool's developer, Pierre Barreau. He is currently the sole owner of AIVA until more laws are put into place governing the algorithm's ownership (Zulic, 2019).

It is important to note that previous studies, such as a study conducted examining the performance of AIVA vocals in music tracks compared to their human counterparts titled "Effects of Added Vocals and Human Production to AI-composed Music on Listener's Appreciation" concretely concluded that, even though, "AIVA, a commercial application that generates short songs in different genres, provided both the algorithmic compassion and the algorithmic production and seemed to offer the most reliably structured tracks that would be suitable for the addition of vocals… AI-created vocals are yet to believably replicate a human vocalist" (Dallas & Morreale 2020). This shows that AIVA is still not competitive at the complex level. Hence, this study aimed to compare simpler types of compositions like the 4-instrument quartet (which will be addressed in the Methodology), rather than much more complex compositions like symphonies and concertos which have much more instrumentation.

METHODOLOGY

As mentioned earlier, the only tool that was used in this study was AIVA. This along with other specifics in the study can introduce several types of biases. First, because AIVA was originally designed specifically for classical music composition, the majority of its training data consisted of music from the most famous composers in history (Wu, 2022). Therefore, the results of this study cannot be generalized to all genres of AIVA-composed music. Also, this study does not infer that its findings can be applied to other AI tools such as Soundraw or

Beatoveen which have been created using different techniques and training data. Apart from the fact that AIVA is the only artificial electronic composer to be recognized by a music society (SACEM), it is consistently ranked among the top publicly available AI tools for composing soundtracks of all genres. Furthermore, because this study used sources from Google Scholar, most of the literature found in the field of music and AI were written about AIVA. Other tools like Soundraw, Loudly, or Beatoveen.AI had limited research available.

AIVA provides users with free membership with several distinct methods of composing soundtracks. These are:

1. Generating From A Style: users can select pre-composed tracks that have already been generated by AIVA by searching using keywords like "spooky" or "fantasy."
2. Generating Using a Chord Progression: users still must select a "style" like the first method, but afterwards can modify the chord progression of the track by either selecting a pre-existing chord progression or make a new one.
3. Generating Step-By-Step: this method incorporates everything from methods 1 and 2, but then adds another layer of customization by allowing users to change the composition's layers.
4. Generating Using An Influence: this method does not include any of the elements from methods 1 through 3, but instead enables users to import external audio to "influence" what AIVA generates - which is not influenced by any pre-existing "style" mentioned above.

This study aimed specifically to compare a human's composition to the same composer's influenced AIVA track. To do this, researchers used method 4 exclusively. The human composition, Franz Schubert's "Quartet in G Major for Flute, Cello, Classical Guitar and Viola," was found in a MIDI file library and then imported into AIVA as an influence. To ensure there were no confounding variables, each of AIVA's compositions had the same key signature, time signature, and instrumentation as the original.

The survey consisted of posting a Google Form to multiple subgroups within Reddit, an American social forum network. Reddit is one of the most popular internet forums in the world with over 500 million users worldwide (Statista), and there are more than 3.4 million subforums, or subreddits on Reddit currently (BusinessDasher, 2024). Because there are so many niches in Reddit, users are virtually able to find specialists in almost any field of their choosing. The survey was posted to five subreddits, namely: r/composer, r/samplesize, r.surveycircle, r/surveyresearch, and r/surveyexchange. All five of these subreddits had user populations ranging from 6,000 to 250,000 daily users, ensuring that the survey would receive an adequate number of responses. The first subgroup, r/composer, is specifically geared towards the music composition society - which was the intended respondent audience. The other four subreddits were utilized specifically for gathering responses.Reddit limitations prohibited the posting of the survey to other subforums. Reddit is very specific with regulations regarding content, and each subgroup has rules that users must abide by in order to participate. Self-advertising/promotion or requesting services is usually banned from most forums.

Only the data mentioned below in the "survey" section was collected - none of the respondents' personal information of any kind were attained or accessible. To ensure data privacy, Reddit users are required to have a nickname that is not their email, and their publicly available profile does not contain any personal information. Also, Reddit approximates geolocation data to ensure location privacy among other info.

Survey Design

There were many genres from AIVA to choose from, but after sampling many AIVA - generated tracks, it was quite evident that genres like horror and fantasy simply weren't complex

enough and would be easily distinguishable from a human counterpart. Instead of picking an AIVA generated sample, another generating method was used. AIVA lets users import an "influence" track and let AIVA come up with similar tracks in a similar style - that way, the three AI-generated pieces in the survey would be consistent with each other and with the actual composition. Because AIVA only accepts certain file types such as MIDI files, a simple four-instrument classical quartet for cello, classical guitar, flute, and viola composed by Franz Schubert was selected.

In order to answer the research question, "How does a known, reputable, publicly-available AI composition tool (AIVA) perform compared to its human influencer?" survey respondents were asked to listen to four excerpts in the same setting. Three of these were AIVA - generated while one of them was the Schubert composition that influenced the three others. After listening, respondents were asked to rank each expert on a scale of 1-4 based upon a few qualities. They were also asked to briefly explain their choices. At the end, the survey revealed only one of the excerpts were composed by a human and gauged respondents' reactions to the new information.

The survey conducted was on Google Forms and consisted of two parts: a demographic section, and a listening section. Both sections consisted of both multiple-choice ranking and open-ended questions.

1. Background questions (demographics)
2. Four music excerpt samples (listening)
3. Ranking questions (reactions and opinions)
4. The four music samples' sources are revealed: 1 is human-generated while the rest are composed by AI (AIVA).
5. Two reaction questions based on #4.

RESULTS

The survey gathered twenty-six responses total over the course of one week. Participants were not required to answer any of the open-ended questions, but all of them did. It should be noted that because this was an experimental study in which the respondents were not randomly chosen, the study cannot generalize its findings to the public but can provide explanatory information helpful for future studies.

Demographics

Four demographic questions ranging in specificity were asked:

1. Do you listen to music?
2. Are you involved (in any way) with the music industry?
3. Do you compose music?
4. What age group/generation are you in?

(Chiang 2024)

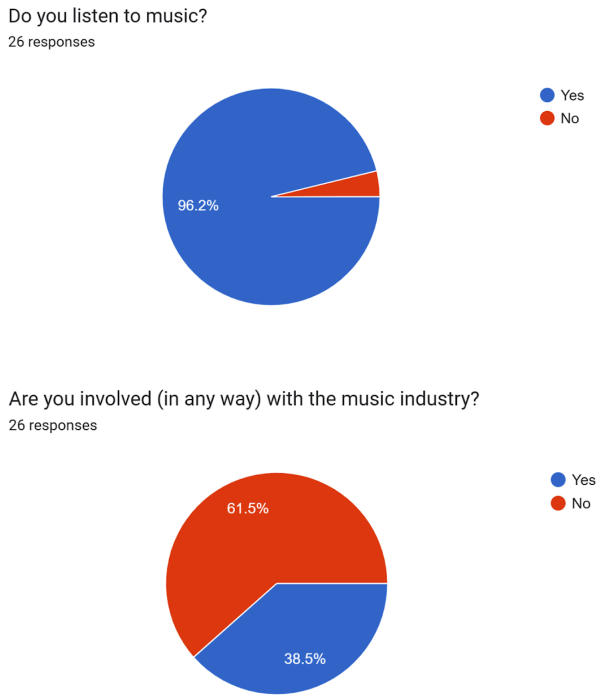


Figure 1: Demographic Data Part 1, Less Specific Questions

Evidently, it is to no surprise that only one respondent said they didn't listen to music. Although the conclusions from this experimental study cannot be generalized as the sample group was not chosen randomly, the data accurately reflects the worldwide percentage of people who

listen to music - that is, 90-93% of the world population (Nielsen Music 360). What is more interesting is the fact that more than sixty percent of respondents said they were actively involved in the music industry - a reflection of the intended audience, music experts (and why this survey was posted specifically to the five subreddits mentioned above), that in turn yields a highly specific demographic.

(Chiang 2024)

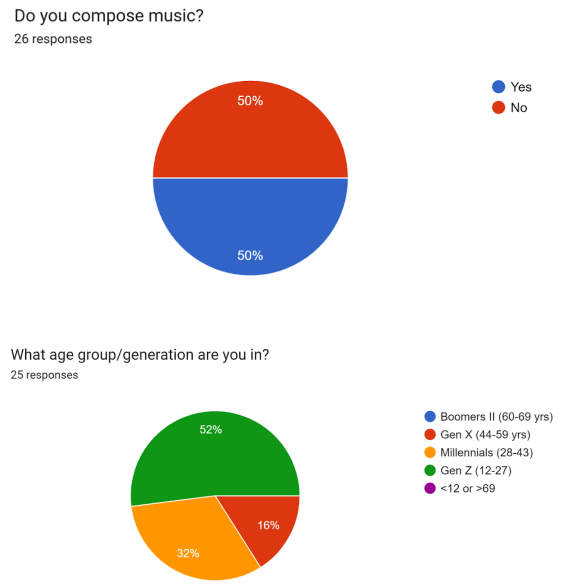


Figure 2: Demographic Data Part 2, More Specific Questions

The same theme of a specific demographic that is made up of involved music experts is one more reflected here - exactly 13 respondents said they composed music themselves, and half the respondents are part of Gen Z, or ages ranging from 12-27 approximately. According to Statista, the majority of classical music listeners across the globe are ages 55+ (Statista, 2024).

Listening

After responding to the four demographic questions, respondents were provided with four Google Drive audio files that were all titled in the same fashion, like this: https://drive.google.com/file/d/1OoZ-fF-4CA5Dgz9BBZodzNtZ9bhaWq1X/view?usp=drive_link (For reference only)

Each audio file contained a 0:30-1:00 excerpt from each of the four compositions, and respondents were asked to listen to each excerpt in the same manner as all the others to ensure consistency.

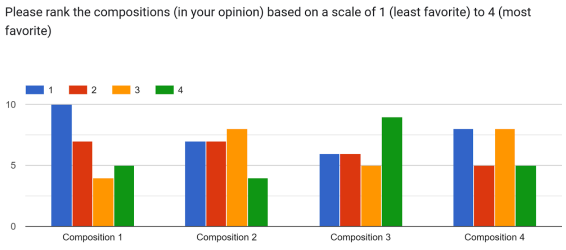


Figure 3: Ranking Data Part 1 - Preference

Similar to Dallas and Morreales' study on AIVA's vocal performance, the respondent group as a whole generally did not show any preference towards either AIVA or human compositions. The respondents were never told which of the four compositions was the human one - composition #4 was, but it is not obvious based on the graph as composition #3 was in fact the most preferred. Next, respondents were asked to rank the compositions according to their musical complexity, or in other words, how musically mature the excerpts sounded.

(Chiang 2024)

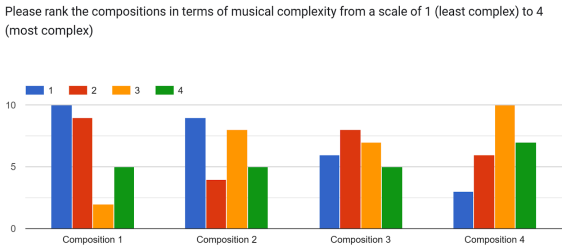
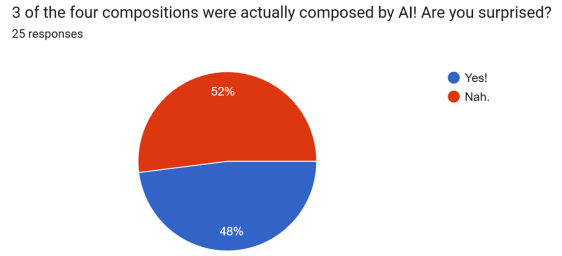


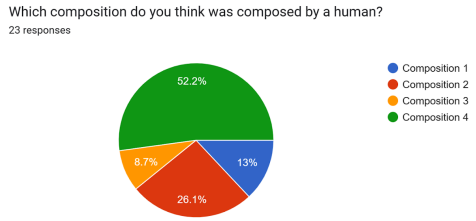
Figure 4: Ranking Data Part 2 - Musical Complexity

Here, the data more accurately reflects the general level of musical knowledge in the sample group - composition #4, the actual human piece, ranked higher on average than the other 3 compositions. While Schubert's quartet

was not the most liked, it was still the most complex piece out of the four compositions.

Finally, after revealing that three of the four compositions were indeed composed by AI not humans, two questions were asked gauging their reactions.





DISCUSSION

This study surveyed 26 people comparing a human composer (Franz Schubert)'s compositions to three AIVA compositions based on the original.. The comparison was based upon three metrics: 1. Preference/Likability, 2. Musical Complexity, and 3. Guessing (which one was the human composition). The survey was targeted towards the specific audience of musically educated people. To locate the appropriate population, the survey was posted to multiple subgroups on Reddit, one of the most popular online forums in the world.

As seen in the findings, there were two main conclusions:

1. The original human composition, excerpt #4, was in fact not the most popular piece when it came to raw preference and first impression.
2. Excerpt #4 however was voted to be the most complex excerpt, with the other excerpts trailing significantly.

It is important to recall that this study used a simple form of string-based classical music - a quartet (made up of four instruments), and over half the respondents were musically educated and even involved in the music industry. Therefore, it was not surprising that the original piece stood out and was verily distinguishable. The details/conditions of this study were very nuanced and specific, and do not accurately reflect the larger consumer group within the music industry.

As stated in the Results section, this study can only draw a cause-and-effect conclusion as it was an experimental study but did not implore random-selection when selecting respondents. Still, when comparing the demographic data to the ranking data, no specific correlation was found. THe findings showed more musically-inclined individuals incorrectly identified the human composition. They also excerpt preference and complexity differently throughout the survey. Even more musically-inclined individuals guessed wrong or ranked differently when it came to excerpt preference and complexity.

CONCLUSION

Through the survey conducted, this study identified three main findings:

1. When comparing three AIVA - generated tracks to their original composition influence, the original was not the most popular track for respondents.
2. When comparing the same three AIVA-generated tracks to the original, most respondents believed the original composition was the most musically complex.
3. When asked to guess which out of the four track excerpts was the original, over fifty percent guessed correctly.

As noted in the Discussion, this study was nuanced to the point that it can draw a

specific cause-and-effect conclusion but could not be generalized to the larger public (due to the sample group not being randomly chosen). For future studies of this type to be reflective of the larger population, the survey conducted should span over many weeks. Participants should be randomly selected individuals regardless of musical background to better represent the population as a whole.

Additionally, instead of simply comparing excerpts based on a quartet that is only made up of four instruments, future studies should compare more complex pieces such as symphonies, which are made up of many more instruments. Mainstream music in today's society is rarely simple, so studies on more complex music would add more to our knowledge of current society.

ACKNOWLEDGEMENT

I would like to thank Dr. Bridget Hamill for supervising the timeline of the paper and providing some inspiration as to the methodology of the research. Also, I would like to thank Dr. Christopher Brown for overseeing the completion and finalization of the paper. Most of all, I want to extend my gratitude to Dr. Roger Worthington for guiding the project from the start, and helping to navigate my research throughout the process.

REFERENCES

Adams, T. (2010, July 10). David Cope: 'You pushed the button and out came hundreds and thousands of sonatas'. *The Observer*. https://www.theguardian.com/music/2010/jul/10/david-cope-composer-virtual-sonatas

Aiva. (2024). AIVA - the AI composing emotional soundtrack music. https://www.aiva.ai/

Baggi, D. L. (1998). The role of computer technology in music and musicology. *lim.dico.unimi.it*. (Archived July 22, 2011).

Barreau, P. (2018, September 10). How AI could compose a personalized soundtrack to your life. *TED*. https://www.ted.com/talks/pierre_barreau_how_ai_could_compose_a_personalized_soundtrack_to_your_life?subtitle=en

Birk, J., & Chakrabarty, M. (2023). Manufacturing "hits": A data-driven AI approach to releasing a pop song in 2022. *Atlantic Marketing Journal, 12*(2), 2.

Bulayenko, O., Quintais, J. P., Gervais, D. J., & Poort, J. (2022). AI music outputs: Challenges to the copyright legal framework. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4072806

Dallas, L., & Morreale, F. (2020, October). Effects of added vocals and human production to AI-composed music on listener's appreciation. In *The 2020 Joint Conference on AI Music Creativity*.

European Parliament. (2024, June 18). EU AI Act: First regulation on artificial intelligence.

Nielsen. (2015, January). Everyone listens to music, but how we listen is changing. https://www.nielsen.com/insights/2015/everyone-listens-to-music-but-how-we-listen-is-changing/

Frid, E., Gomes, C., & Jin, Z. (2020, April). Music creation by example. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Kaleagasi, B. (2017). A new AI can write music as well as a human composer. *Futurism*. https://futurism.com/a-new-ai-can-write-music-as-well-as-a-human-composer

Karpov, N. (2020). Artificial intelligence for music composing: Future scenario analysis.

Ketkar, N., & Moolayil, J. (2021). *Deep learning with Python*. Springer.

Lima, A. S., & Blixt, A. (2021). Investigating the possibility of bias against AI-computer composed music.

McGuire, B., Smith, C., Hung, T., & Yang, G.

(2006). The history of artificial intelligence. University of Washington. https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf

U.S. Copyright Office. (n.d.). Musical composition. https://www.copyright.gov/prereg/music.html

Hiller, L. A., & Isaacson, L. M. (1959). *Experimental music: Composition with an electronic computer* (2nd ed.). McGraw-Hill. Reprinted 1979, Greenwood Press.

Statista. (2024). Reddit Inc. quarterly DAU by status 2024. *Statista*. https://www.statista.com/statistics/1453133/reddit-quarterly-dau-by-online-status/

Zulić, H. (2019). How AI can change/improve/influence music composition, performance, and education: Three case studies. *INSAM Journal of Contemporary Music, Art and Technology, (2)*, 100-114.

Wu, Y. (2022, November 29). AIVA Technology – Composing music using AI. *Digital Innovation and Transformation*. https://d3.harvard.edu/platform-digit/submission/aiva-technology-composing-music-using-ai/