

View Reviews

Paper ID

162

Paper Title

MetaScore: A Million-Scale Dataset for Tag- and Text-based Controllable Symbolic Music Generation

Track Name

Papers

Reviewer #2

Questions

2. I am an expert on the topic of the paper.

Disagree

3. The title and abstract reflect the content of the paper.

Disagree

4. The paper discusses, cites and compares with all relevant related work

Disagree

5. Please justify the previous choice (Required if “Strongly Disagree” or “Disagree” is chosen, otherwise write "n/a")

previous works on MuseScore datasets are not discussed (they exist for score difficulty analysis and inpainting)

6. Readability and paper organization: The writing and language are clear and structured in a logical manner.

Agree

7. The paper adheres to ISMIR 2024 submission guidelines (uses the ISMIR 2024 template, has at most 6 pages of technical content followed by “n” pages of references or ethical considerations, references are well formatted). If you selected “No”, please explain the issue in your comments.

Yes

8. Relevance of the topic to ISMIR: The topic of the paper is relevant to the ISMIR community. Note that submissions of novel music-related topics, tasks, and applications are highly encouraged. If you think that the paper has merit but does not exactly match the topics of ISMIR, please do not simply reject the paper but instead communicate this to the Program Committee Chairs. Please do not penalize

the paper when the proposed method can also be applied to non-music domains if it is shown to be useful in music domains.

Agree

9. Scholarly/scientific quality: The content is scientifically correct.

Disagree

10. Please justify the previous choice (Required if "Strongly Disagree" or "Disagree" is chosen, otherwise write "n/a")

Experimentation/explanation needed for how this specific prompt and the LLM is chosen for the task. The use of "copyright" as a "tag" has no justification. Tag and genre choices need justification. Since more than half of the data belongs to the artificially combined and non-stratified group of "classical traditional", the experiments regarding genre lack scientific meaning.

11. Novelty of the paper: The paper provides novel methods, applications, findings or results. Please do not narrowly view "novelty" as only new methods or theories. Papers proposing novel musical applications of existing methods from other research fields are considered novel at ISMIR conferences.

disagree

12. The paper provides all the necessary details or material to reproduce the results described in the paper. Keep in mind that ISMIR respects the diversity of academic disciplines, backgrounds, and approaches. Although ISMIR has a tradition of publishing open datasets and open-source projects to enhance the scientific reproducibility, ISMIR accepts submissions using proprietary datasets and implementations that are not sharable. Please do not simply reject the paper when proprietary datasets or implementations are used.

Disagree

13. Pioneering proposals: This paper proposes a novel topic, task or application. Since this is intended to encourage brave new ideas and challenges, papers rated "Strongly Agree" and "Agree" can be highlighted, but please do not penalize papers rated "Disagree" or "Strongly Disagree". Keep in mind that it is often difficult to provide baseline comparisons for novel topics, tasks, or applications. If you think that the novelty is high but the evaluation is weak, please do not simply reject the paper but carefully assess the value of the paper for the community.

Strongly Disagree (Well-explored topic, task, or application)

14. Reusable insights: The paper provides reusable insights (i.e. the capacity to gain an accurate and deep understanding). Such insights may go beyond the scope of the paper, domain or application, in order to build up consistent knowledge across the MIR community.

Strongly disagree

15. Please explain your assessment of reusable insights in the paper.

Web scraping without human verification is expected to solve the problem of symbolic music generation.

16. Write ONE line (in your own words) with the main take-home message from the paper.

Symbolic music has a potential to improve the current text-to-music generation based on audio.

19. Potential to generate discourse: The paper will generate discourse at the ISMIR conference or have a large influence/impact on the future of the ISMIR community.

Strongly disagree

20. Overall evaluation: Keep in mind that minor flaws can be corrected, and should not be a reason to reject a paper. Please familiarize yourself with the reviewer guidelines at <https://ismir.net/reviewer-guidelines>

Weak reject

21. Main review and comments for the authors. Please summarize strengths and weaknesses of the paper. It is essential that you justify the reason for the overall evaluation score in detail. Keep in mind that belittling or sarcastic comments are not appropriate.

This paper introduces a web-scraping based dataset for symbolic music generation built using the MuseScore platform. Alongside the scores (most of which include copyright infringement problems), the metadata entered by the uploader and the LLM-generated pseudo-captions are also provided. The contributions of the paper are unfortunately limited to the dataset itself and the novelty of the experiments are limited. Hence, the review is also based on the quality of the proposed dataset, which raises some concerns in my opinion:

- Some problems arising from the platform itself is not discussed: the score quality and accuracy, missing notes/ornaments, (sometimes intentional) wrong namings and copyright declarations, differences in multiple scores for the same musical piece etc. A MIR dataset based on MuseScore need to answer these platform driven problems.

Most importantly, the effect of the score content creator is not taken into account. In MuseScore the creators also have their follower counts and rankings by other members, which allow us to comment on the accuracy of the score. Some of these creators are supported through patreon and they produce high quality scores and tags. However, this is not the case for the majority, and calling this raw data as a "dataset" without any human filter has a potential to create more harm than good for the community. A dataset paper based on musescore is at least expected to have a filter based on the creator or user comments, statistics, score likes, views, or favourite counts.

- Use of composers and instrumentation as "tags": While some well regarded musescore users stick to the original scores and instrumentations by the composers, this is not the case for the majority. Furthermore, a big portion of the scores also include easier versions or adaptations of the famous pieces into other instruments or ensembles. A comparison

with IMSLP metadata or musicbrainz IDs might be needed for maintaining the accuracy of the acclaimed tags and labels.

- The distribution of "genre" tags: MuseScore is mainly a platform for music with strong music writing traditions, which explains why the majority of the scores belong to their "classical/traditional" tag. Given this fact, the tag classification chosen by the authors need explanation: why the scores are not categorized with respect to era? How do we explain tagging more than half of the pieces as merely classical-traditional, without giving any information on the era? Why traditional is merged with classical and not soundtrack?
- The quality of the pseudo tags and captions: With the F1 score of 62% and using copyright information as a tag, unfortunately, I cannot foresee how this dataset will be used for text based symbolic music generation. The examples in figure 2 speak for itself... (classical/traditional + programmatic + robot != A classical music piece)
- The copyright problems are already mentioned by the authors. If the paper proposes a new method other than the dataset, this wouldn't be an issue (since we could claim the main contribution is not affected by the copyright problems). However, in the current scenario, how do we distribute the main contribution?

In future, if the authors prepare a study on the quality of the community driven scores and choose the scores based on filtering with comments, likes, or creators, the contribution would be appreciated. Using a mix of LLMs (based on chain-of-thoughts reasoning or agent orchestration) and human feedback for such filters can be considered for the task.

Reviewer #3

Questions

2. I am an expert on the topic of the paper.

Agree

3. The title and abstract reflect the content of the paper.

Strongly agree

4. The paper discusses, cites and compares with all relevant related work

Strongly agree

5. Please justify the previous choice (Required if “Strongly Disagree” or “Disagree” is chosen, otherwise write "n/a")

n/a

6. Readability and paper organization: The writing and language are clear and structured in a logical manner.

Strongly agree

7. The paper adheres to ISMIR 2024 submission guidelines (uses the ISMIR 2024 template, has at most 6 pages of technical content followed by “n” pages of references or ethical considerations, references are well formatted). If you selected “No”, please explain the issue in your comments.

Yes

8. Relevance of the topic to ISMIR: The topic of the paper is relevant to the ISMIR community. Note that submissions of novel music-related topics, tasks, and applications are highly encouraged. If you think that the paper has merit but does not exactly match the topics of ISMIR, please do not simply reject the paper but instead communicate this to the Program Committee Chairs. Please do not penalize the paper when the proposed method can also be applied to non-music domains if it is shown to be useful in music domains.

Strongly agree

9. Scholarly/scientific quality: The content is scientifically correct.

Agree

10. Please justify the previous choice (Required if "Strongly Disagree" or "Disagree" is chosen, otherwise write "n/a")

n/a

11. Novelty of the paper: The paper provides novel methods, applications, findings or results. Please do not narrowly view "novelty" as only new methods or theories. Papers proposing novel musical applications of existing methods from other research fields are considered novel at ISMIR conferences.

disagree

12. The paper provides all the necessary details or material to reproduce the results described in the paper. Keep in mind that ISMIR respects the diversity of academic disciplines, backgrounds, and approaches. Although ISMIR has a tradition of publishing open datasets and open-source projects to enhance the scientific reproducibility, ISMIR accepts submissions using proprietary datasets and implementations that are not sharable. Please do not simply reject the paper when proprietary datasets or implementations are used.

Strongly agree

13. Pioneering proposals: This paper proposes a novel topic, task or application. Since this is intended to encourage brave new ideas and challenges, papers rated "Strongly Agree" and "Agree" can be highlighted, but please do not penalize papers rated "Disagree" or "Strongly Disagree". Keep in mind that it is often difficult to provide baseline comparisons for novel topics, tasks, or applications. If you think that the novelty is high but the evaluation is weak, please do not simply reject the paper but carefully assess the value of the paper for the community.

Disagree (Standard topic, task, or application)

14. Reusable insights: The paper provides reusable insights (i.e. the capacity to gain an accurate and deep understanding). Such insights may go beyond the scope of the paper, domain or application, in order to build up consistent knowledge across the MIR community.

Disagree

15. Please explain your assessment of reusable insights in the paper.

The two generations approaches presented are fairly standard. The ideas are already common in other fields. As a dataset paper, there are not many reusable insights. The main value is in the provided data.

16. Write ONE line (in your own words) with the main take-home message from the paper.

The paper presents a publicly available dataset of symbolic music (probably mscz or musicXML scores? See my review later). The main value derives from the dataset size (~1 million scores) and the rich metadata included. Text descriptions of the songs, generated with an LLM starting from the tags, and genre generated by a genre classifier, are also included. 2 models are trained and evaluated.

19. Potential to generate discourse: The paper will generate discourse at the ISMIR conference or have a large influence/impact on the future of the ISMIR community.

Strongly agree

20. Overall evaluation: Keep in mind that minor flaws can be corrected, and should not be a reason to reject a paper. Please familiarize yourself with the reviewer guidelines at <https://ismir.net/reviewer-guidelines>

Weak reject

21. Main review and comments for the authors. Please summarize strengths and weaknesses of the paper. It is essential that you justify the reason for the overall evaluation score in detail. Keep in mind that belittling or sarcastic comments are not appropriate.

I'm very concerned about the legal issues for this dataset. As far as I understand, redistributing (which is what the authors are doing) is not allowed under some of the licenses in the Musescore data, regardless of whether the user will use it for commercial purposes. The trick for audio is to redistribute the spectrograms, but this is not possible for symbolic music. Maybe sharing a pretrained model could be an idea? This is not an easy task in any of course, but it is very problematic since sharing the data is the main contribution of this paper.

Moreover, some analysis on the duplication of pieces could be very beneficial.

When looking at the example on the website (and in Figure 2), I found it problematic that the copyright information was used to generate the LLM-Caption. For example, let's suppose a user wants to generate the 4th track on the website. I doubt they will type

“Users must give credit to the creator when using this music”. Copyright should be an attribute of each song of the dataset, so people can decide in case to filter by it, but I don’t think it should be included in the LLM-generated captions.

What will be the format the scores will be released? This information is completely missing in the paper. I suppose MSCZ files, since this is the Musescore format? MusicXML obtained by a conversion with Musescore could be even better since it is more widely used in MIR, and natively supported by musical score libraries such as Music21 and Partitura. If it was up to me to decide, I would distribute both MIDI version and MusicXML version. This point is often oversimplified in music generation papers, since even the output of advanced tokenizer like REMI+ is more similar to a quantized MIDI, rather than a musical score. I invite the authors to not ignore this point, since having fully formatted musical scores instead of just MIDI would be extremely useful for a big variety of score typesetting tasks, such as pitch spelling, voice separation, staff separation, beaming and tuplet structuring, dynamic symbols and slurs generation.

If indeed the authors release fully formatted scores in some formats, I would add somewhere on the abstract, for example on line 9 in which format they are released.

62-66: this is still a bit weird to read since the feeling that I got was something like “We are the first to do text-to-symbolic, and we compare with an older text-to-symbolic”. I would suggest specifying why [3] is different for extra clarity. An extra sentence specifying why end-to-end training may be preferable compared to other approaches may also make the paper proposals stronger.

Is the time signature of all songs 4/4? Or are the models handling different time signatures? This needs to be stated explicitly. For the dataset, time signature could be another extremely useful tag to have, and it can be extracted automatically from the score, though some extra choices are needed for pieces with multiple time signatures.

Section 5.1: how are the different lengths of excerpts handled during training? Are they randomly truncated to have a maximum length? And which length? Is the standard sinusoidal positional encoding used? I suppose this means that the transformer during inference can only reliably generate a maximum number of tokens. How many notes does that correspond to more or less? I find such details useful to have in the paper.

334: I would change “we intend” to “we intended”

Reviewer #4

Questions

2. I am an expert on the topic of the paper.

Strongly agree

3. The title and abstract reflect the content of the paper.

Strongly agree

4. The paper discusses, cites and compares with all relevant related work

Agree

5. Please justify the previous choice (Required if “Strongly Disagree” or “Disagree” is chosen, otherwise write "n/a")

n/a

6. Readability and paper organization: The writing and language are clear and structured in a logical manner.

Agree

7. The paper adheres to ISMIR 2024 submission guidelines (uses the ISMIR 2024 template, has at most 6 pages of technical content followed by “n” pages of references or ethical considerations, references are well formatted). If you selected “No”, please explain the issue in your comments.

Yes

8. Relevance of the topic to ISMIR: The topic of the paper is relevant to the ISMIR community. Note that submissions of novel music-related topics, tasks, and applications are highly encouraged. If you think that the paper has merit but does not exactly match the topics of ISMIR, please do not simply reject the paper but instead communicate this to the Program Committee Chairs. Please do not penalize the paper when the proposed method can also be applied to non-music domains if it is shown to be useful in music domains.

Strongly agree

9. Scholarly/scientific quality: The content is scientifically correct.

Agree

10. Please justify the previous choice (Required if "Strongly Disagree" or "Disagree" is chosen, otherwise write "n/a")

n/a

11. Novelty of the paper: The paper provides novel methods, applications, findings or results. Please do not narrowly view "novelty" as only new methods or theories. Papers proposing novel musical applications of existing methods from other research fields are considered novel at ISMIR conferences.

Agree

12. The paper provides all the necessary details or material to reproduce the results described in the paper. Keep in mind that ISMIR respects the diversity of academic disciplines, backgrounds, and approaches. Although ISMIR has a tradition of publishing open datasets and open-source projects to enhance the scientific

reproducibility, ISMIR accepts submissions using proprietary datasets and implementations that are not sharable. Please do not simply reject the paper when proprietary datasets or implementations are used.

Agree

13. Pioneering proposals: This paper proposes a novel topic, task or application. Since this is intended to encourage brave new ideas and challenges, papers rated "Strongly Agree" and "Agree" can be highlighted, but please do not penalize papers rated "Disagree" or "Strongly Disagree". Keep in mind that it is often difficult to provide baseline comparisons for novel topics, tasks, or applications. If you think that the novelty is high but the evaluation is weak, please do not simply reject the paper but carefully assess the value of the paper for the community.

Agree (Novel topic, task, or application)

14. Reusable insights: The paper provides reusable insights (i.e. the capacity to gain an accurate and deep understanding). Such insights may go beyond the scope of the paper, domain or application, in order to build up consistent knowledge across the MIR community.

Agree

15. Please explain your assessment of reusable insights in the paper.

The curated dataset can be used in various applications. In addition, the finding that models trained on tags outperform those trained on captions is a reusable insight in the context of text2music generation

16. Write ONE line (in your own words) with the main take-home message from the paper.

The paper introduces MetaScore, a million-scale dataset for tag- and text-based controllable symbolic music generation, enabling models to generate music based on specified tags or text prompts with improved quality and control.

19. Potential to generate discourse: The paper will generate discourse at the ISMIR conference or have a large influence/impact on the future of the ISMIR community.

Agree

20. Overall evaluation: Keep in mind that minor flaws can be corrected, and should not be a reason to reject a paper. Please familiarize yourself with the reviewer guidelines at <https://ismir.net/reviewer-guidelines>

Weak reject

21. Main review and comments for the authors. Please summarize strengths and weaknesses of the paper. It is essential that you justify the reason for the overall evaluation score in detail. Keep in mind that belittling or sarcastic comments are not appropriate.

The paper introduces MetaScore, a large-scale dataset for tag- and text-based controllable symbolic music generation. It aims to improve the quality and control of generated music

by providing a million-scale dataset for training and evaluation. The dataset consists of symbolic music paired with tags and descriptions, allowing models to generate music based on specified tags or text prompts.

Strengths:

1. MetaScore's large size addresses the scarcity of data in symbolic music generation, offering a valuable resource for the research community
2. The dataset enables models to generate music based on specified tags or text prompts, enhancing control over the generated music

Weaknesses:

1. Releasing all data from MuseScore without permissions raises ethical and legal concerns regarding copyright infringement. It may be more appropriate to release only data with proper licenses (as plotted in Fig. 1e)
2. The study's reliance on 22 participants may not provide a robust subjective evaluation
3. The paper lacks a comprehensive evaluation of the dataset's quality and the extent to which the tags and captions provided in the dataset contribute to improved model performance. This includes not testing a condition without text prompts to understand the impact of tags and captions on music generation

Minor Issues:

In 86-87: ABC notation natively supports multitrack. WikiMusicText contains leadsheets only because it is based on Wikifonia

Fig. 1: The figure is too small and hard to read

In 206: Why copyright information is needed to generate pseudo captions?

Table 1: Based on In 305, it is suggested that the original format should be .mscz instead of .xml

Table 3: It's unclear why models that support free text prompts do not provide controls over certain attributes. It could be clarified whether they do not support free text prompts or if they support arbitrary attribute controls

View Meta-Reviews

Paper ID

162

Paper Title

MetaScore: A Million-Scale Dataset for Tag- and Text-based Controllable Symbolic Music Generation

Track Name

Papers

META-REVIEWER #1

META-REVIEW QUESTIONS

2. I am an expert on the topic of the paper.

Strongly agree

3. The title and abstract reflect the content of the paper.

Strongly agree

4. The paper discusses, cites and compares with all relevant related work.

Disagree

5. Please justify the previous choice (Required if “Strongly Disagree” or “Disagree” is chosen, otherwise write "n/a")

As one contribution of this paper is a dataset downloaded from MuseScore, it would be appropriate to discuss other work that uses MuseScore as a source of data, e.g., Edirisooriya et al. "AN EMPIRICAL EVALUATION OF END-TO-END POLYPHONIC OPTICAL MUSIC RECOGNITION", Proc. ISMIR 2021; and <https://github.com/fosfrancesco/asap-dataset> etc.

6. Readability and paper organization: The writing and language are clear and structured in a logical manner.

Agree

7. The paper adheres to ISMIR 2024 submission guidelines (uses the ISMIR 2024 template, has at most 6 pages of technical content followed by “n” pages of references or ethical considerations, references are well formatted). If you selected “No”, please explain the issue in your comments.

Yes

8. Relevance of the topic to ISMIR: The topic of the paper is relevant to the ISMIR community. Note that submissions of novel music-related topics, tasks, and applications are highly encouraged. If you think that the paper has merit but does not exactly match

the topics of ISMIR, please do not simply reject the paper but instead communicate this to the Program Committee Chairs. Please do not penalize the paper when the proposed method can also be applied to non-music domains if it is shown to be useful in music domains.

Agree

9. Scholarly/scientific quality: The content is scientifically correct.

Disagree

10. Please justify the previous choice (Required if “Strongly Disagree” or “Disagree” is chose, otherwise write "n/a")

The quality of the data assembled seems to be a secondary consideration to its quantity and usefulness. Some of the questions used in the subjective evaluation are vague and undefined, e.g., "Are the instruments arranged properly?"

11. Novelty of the paper: The paper provides novel methods, applications, findings or results. Please do not narrowly view "novelty" as only new methods or theories. Papers proposing novel musical applications of existing methods from other research fields are considered novel at ISMIR conferences.

Agree

12. The paper provides all the necessary details or material to reproduce the results described in the paper. Keep in mind that ISMIR respects the diversity of academic disciplines, backgrounds, and approaches. Although ISMIR has a tradition of publishing open datasets and open-source projects to enhance the scientific reproducibility, ISMIR accepts submissions using proprietary datasets and implementations that are not sharable. Please do not simply reject the paper when proprietary datasets or implementations are used.

Agree

13. Pioneering proposals: This paper proposes a novel topic, task or application. Since this is intended to encourage brave new ideas and challenges, papers rated “Strongly Agree” and “Agree” can be highlighted, but please do not penalize papers rated “Disagree” or “Strongly Disagree”. Keep in mind that it is often difficult to provide baseline comparisons for novel topics, tasks, or applications. If you think that the novelty is high but the evaluation is weak, please do not simply reject the paper but carefully assess the value of the paper for the community.

Agree (Novel topic, task, or application)

14. Reusable insights: The paper provides reusable insights (i.e. the capacity to gain an accurate and deep understanding). Such insights may go beyond the scope of the paper, domain or application, in order to build up consistent knowledge across the MIR community.

Agree

15. Please explain your assessment of reusable insights in the paper.

Text to symbolic music generation is a novel application and this work takes appropriate steps to creating it.

16. Write ONE line (in your own words) with the main take-home message from the paper.

Here's an application of text-to-symbolic music modeling.

19. Potential to generate discourse: The paper will generate discourse at the ISMIR conference or have a large influence/impact on the future of the ISMIR community.

Agree

20. Overall evaluation (to be completed before the discussion phase): Please first evaluate before the discussion phase. Keep in mind that minor flaws can be corrected, and should not be a reason to reject a paper. Please familiarize yourself with the reviewer guidelines at <https://ismir.net/reviewer-guidelines>.

Weak reject

21. Main review and comments for the authors (to be completed before the discussion phase). Please summarize strengths and weaknesses of the paper. It is essential that you justify the reason for the overall evaluation score in detail. Keep in mind that belittling or sarcastic comments are not appropriate.

It is clear that this paper has taken a lot of work to put together. The sound results are interesting as well, though perhaps with not as strong connection to the text prompts as suggested. I have several observations:

- The paper does not convincingly measure the quality of the assembled dataset, or show that it is useful for the given tasks.
- Choices taken in assembling the dataset, and even its origin (it doesn't seem to be from the MuseScore Forum as stated), are not clear from the text (How exactly was it acquired?)
- Since one of the main contributions is the dataset, and the legality of sharing the data itself (rather than links) is in doubt (How did the authors circumvent the limitations of the MuseScore resource? How would sharing this resource not violate the terms of service of the MuseScore resource?), the only contribution that would remain were the dataset not shared are the trained models, which themselves have unmotivated design decisions (e.g., the copyright tag?), and questionable evaluation (e.g., the "easy piano piece" in Fig 3 actually looks difficult because of it lacks simple patterns).
- The subjective listening is quite strange with questions like, "Are the instruments arranged properly?" What is a "proper arrangement"?
- The ethics statement is very weak and doesn't show much self-reflection, or even awareness of the problems with the data collection in the first place.

22. Final recommendation (to be completed after the discussion phase) Please give a final recommendation after the discussion phase. In the final recommendation, please do not simply average the scores of the reviewers. Note that the number of recommendation

options for reviewers is different from the number of options here. We encourage you to take a stand, and preferably avoid “weak accepts” or “weak rejects” if possible.

Reject

23. Meta-review and final comments for authors (to be completed after the discussion phase)

This paper generated a lot of discussion in the discussion phase. In addition to a general agreement with the points I raised in the review above, all reviewers noted very problematic aspects of the data collection, e.g., circumventing restrictions on the use of the MuseScore service (e.g., downloading no more than 300 scores a month), or violating the terms of service of the MuseScore resource (e.g., distributing all the data to third parties). The paper does not show any indication of critical self-reflection on these aspects, or give evidence that the authors approached MuseScore and its users for permission to use the data in this way. The ethics statement provided at the end only says the authors will release the data for research and non-commercial purposes, but provide no legal mechanism by which that can be accomplished.

Since I as metareviewer intend to only recommend accept or reject (to make the decisions of the paper chairs easier), I must choose reject principally because:

- the quality of the dataset is not convincingly measured
- how exactly the data was collected is not clear (the MuseScore forum <https://musescore.org/en/forum> does not contain scores, which appear to be from here instead <https://musescore.com?>)
- design decisions in the resulting models are not motivated
- the contribution that would remain given the dataset cannot be made available is not significant enough to warrant publication.

Furthermore, the online resource (<https://musescore.com/sheetmusic>) seems to prohibit what the authors appear to have done in assembling their dataset (<https://musescore.com/legal/terms>). The paper does not go into details how the authors got around the limits imposed by the website (max 300 scores downloaded/month, which would have required 416 years to collect this amount of data), and the paper does not substantively address the problematic aspect that were the authors to release the data to the public, it violates the terms of service in a major way:

"Except as expressly provided in this Terms of Use, you DO NOT have the right to:

Redistribute downloaded or printed Content to third parties by any means.

Rent, give or lend downloaded or printed Content to third parties.

Use automated systems (except for those that may be provided by MuseScore) for accessing or downloading Content, or

Attempt to circumvent, alter, or in any other way modify any copy-protection technologies placed

on downloads offered by MuseScore."

My suggestion is to tune this paper away from a "dataset paper" to a "model use paper" with focus on the use of the resulting system in _making_ music. There are lots of interesting questions here that can be investigated. But at the same time, such a paper still needs to consider the negative impacts of the data collection. I sincerely encourage the authors to attempt to work together with MuseScore to acquire data in a way that is not problematic.
