GREEN MIR? INVESTIGATING COMPUTATIONAL COST OF RECENT MUSIC-AI RESEARCH IN ISMIR

Andre Holzapfel **Anna-Kaisa Kaila** **Petra Jääskeläinen**
Division of Media Technology and Interaction Design, KTH Royal Institute of Technology, Sweden
holzap@kth.se, akkaila@kth.se, ppja@kth.se

ABSTRACT

The environmental footprint of Generative AI and other Deep Learning (DL) technologies is increasing. To understand the scale of the problem and to identify solutions for avoiding excessive energy use in DL research at communities such as ISMIR, more knowledge is needed of the current energy cost of the undertaken research. In this paper, we provide a scoping inquiry of how the ISMIR research concerning automatic music generation (AMG) and computing-heavy music analysis currently discloses information related to environmental impact. We present a study based on two corpora that document 1) ISMIR papers published in the years 2017–2023 that introduce an AMG model, and 2) ISMIR papers from the years 2022–2023 that propose music analysis models and include heavy computations with GPUs. Our study demonstrates a lack of transparency in model training documentation. It provides the first estimates of energy consumption related to model training at ISMIR, as a baseline for making more systematic estimates about the energy footprint of the ISMIR conference in relation to other machine learning events. Furthermore, we map the geographical distribution of generative model contributions and discuss the corporate role in the funding and model choices in this body of work.

1. INTRODUCTION

Interest in AMG and DL-based analytical models is increasing dramatically at conferences such as ISMIR [1]. Case studies in domains other than music [2–4] have established that the environmental impact of AI technologies can be massive, particularly when it comes to energy consumption. International Energy Agency predicts that the accumulated electricity consumption of data centers, AI, and the cryptocurrency sector will double, reaching the level of whole electricity consumption of Japan by 2026 [5]. In the US, a recent proposal for legislation (Artificial Intelligence Environmental Impacts Act) suggests that AI companies would be urged to start reporting the environmental impacts of their work [6]. With the increasing investment in AI and the general trend of high compute requirements for training state-of-the-art machine learning systems [7], we expect to see the accumulated energy footprint of the generative music industry and the surrounding research also growing. There is no reason why research around ISMIR would be isolated from these effects. For the community to gain an understanding of the scale of the problem and to identify solutions to avoid excessive energy use in AMG development, more knowledge is needed of the current energy cost of the research conducted. It is, hence, highly relevant and timely to investigate to what extent research at ISMIR acknowledges and documents the environmental impact of energy consumption.

Other research communities around music technology (e.g., NIME [8]) and machine learning technology (e.g., NeurIPS [9]) have shown increasing attention to various aspects of negative ethical impacts, among them environmental. Discussions of such topics continue, however, to be severely underrepresented in the generative music and audio research [10], and entirely absent from the ethics principles and guidelines for AI-music [11, p. 148]. In the context of the ISMIR community, Morreale [12] estimated that between 2011 and 2020, less than 0.5% of ISMIR submissions discussed issues related generally to ethics, of which sustainability could be seen as a subcategory. Our present scoping inquiry demonstrates this lack of concern and transparency in reporting the environmental impacts of AMG and other DL research, with a focus on ISMIR conferences. The title of our paper refers to Schwartz et al. [13], which proposes the concept of *Green AI* as "AI research that is more environmentally friendly and inclusive". While the concept of Green MIR should be used carefully, as it can lead to practices of greenwashing research, we use this term to raise questions about the current practices and energy impact of MIR.

This study advances a critical discussion in the ISMIR community around the ethical impacts of model development work and the responsibility of MIR research from the underexplored perspective of environmental sustainability. It documents, firstly, the level of transparency in reporting the environmental impact in terms of energy consumption and the computational resource use in the model training process in ISMIR papers in the seven years 2017–2023. Secondly, based on the information documented in these ISMIR publications, we provide preliminary estimates of the energy demands related to training individual AMG and other DL models, as well as an overall estimate of the total energy use and carbon footprint associated with DL training at an ISMIR conference. In this process, we also investigate if AMG models are related to higher energy consumption than other models at ISMIR. Our calculations establish a baseline to set the ISMIR community in relation to other machine learning communities and call for the issue of environmental impact to be addressed more systematically so that the conference can grow and evolve without compromising the environment. Thirdly, we map the geographical distribution of generative model contributions and discuss the role of corporate participation and the consequent political economies in this body of work.

Some limitations of the paper lie within (a) the many uncertainties in estimating the overall energy consumption of developing a DL model based on limited data related to the training process, (b) our focus on model training that leaves a focus on inference processes to future work, and, (c), the reduction of environmental sustainability to energy consumption. In the following two sections, we explain how these limitations emerge from the sparse amount of information and the complexity of the problem. We hope that this paper will motivate both individual authors and the ISMIR conference to take action toward more minute documentation of resource use in DL model development.

2. BACKGROUND

2.1 Environmental Impact of (Music-)AI

The soaring energy cost of AI technology is increasingly discussed in the academic literature [2, 3, 13–19]. Early work on the environmental impact of AI has introduced concepts, such as "green" and its opposite, "red" AI [13], problematized hidden environmental costs in AI [2], and provided methods for quantifying environmental impact [3]. We found only a few works that are related to music-AI (as a term to cover analytic and generative approaches based on – predominantly – deep learning). The first two concern the energy cost of AI models used in music information retrieval [20, 21], and the third focuses on the importance of studying sustainability in arts generally [4].

AI development has been increasingly steering towards Large Language Models (LLMs), which have a particularly high energy expenditure. The popularity of these models is attributed to their success in "generalizing" and performing better in tasks that have traditionally required human labor. But as a consequence, many research papers [22,23] take a pre-trained foundation model and adapt it. This results in a situation in which the energy cost can be estimated only partially, *i.e.* for the part that extended the pre-trained LLM. Furthermore, it raises questions about how research can account for the environmental cost of using LLMs. Arguably, the researchers who use those LLMs are somewhat responsible for the popularity and increased use of LLMs – through creating demand for their use – which can further aggravate the use of computational resources and energy in LLM development.

Many research works that focus on the environmental impact of AI take the assumption that energy (computational) cost is the core environmental problem of these technologies, and by reducing energy consumption, it is possible to work towards sustainability. However, this is a simplistic view because sustainability is a complex phenomenon that does not only concern electricity usage. For example, Jääskeläinen et al. [24,25] discussed the complex networks of environmental harm resulting from resource consumption and capitalistic colonialism that prevail in the case of generative AI. Strengers [26] has generally outlined how behavior change is central to change toward sustainability, and providing metrics such as energy consumption data is insufficient to address change toward sustainability. While keeping this in mind, energy use is a valid starting point for discussing the environmental impact of AI. In this paper, when we refer to *environmental impact*, we explicitly refer to the energy cost and leave out factors such as the life cycle of the technology and water usage of data centers [27], among others.

Technological advances in recent decades have entered the music industry with the promise of reduced material and energy demands. For instance, it was expected that the introduction of mp3, the digitalization of music production, and eventually the platformization of its consumption would diminish the environmental footprint of the industry. As Devine [28] and Brennan [29] have demonstrated, the opposite has historically been the case: while the demand for plastic dropped in the era of the mp3 to a fraction compared the previous music consumption models (CD, cassette, vinyl, etc.), the greenhouse gas emissions of the industry, on the contrary, *increased*. This increase was explained in sustainability research by Hilty's concept *rebound effects* [30], which describes indirect 2nd and 3rd order effects that result from adopting new technology.

Similar negative effects are emerging in the proliferation of AI in music. Calculations by Holzapfel [31] illustrate that creative applications of LLMs can amount to considerable levels of energy demand. Furthermore, results by Douwes [20, 21] and Ronchini and Serizel [32] indicate that the scale of energy consumption for audio generation and analysis tasks are not in a linear relation with the model performance, thus questioning the assumption that the growth in model complexity and resource demands are a prerequisite for better models. Even more importantly, Holzapfel [31] calls for the focus of inquiry to be expanded from the carbon footprint to the wider questions of political ecology, and to the perspectives of economic gain and power (see also [33]). In this view, we should not only measure the environmental impact but also form a more complete picture by looking at who is causing it, who is financing that work, and who benefits from it.

2.2 Timeliness of Addressing the Environmental Impact of Music-AI

Bringing energy concerns into research practices is still at an early stage in many communities [19], including MIR. In 2023, Morreale et al. [1] ran a systematic survey of the training datasets for AMG models presented at ISMIR

2013–2023. Their work illustrates a dramatic increase in the development of AMG models in the last decade, and especially since 2017. Taken together with the general lack of both breadth and depth of addressing environmental concerns in music and audio research contexts [10], this increase highlights the urgency of addressing the computational cost of the AMG models in ISMIR research.

Conferences such as NIME have already taken a proactive lead in promoting awareness of the environmental impacts of the research conducted around the conference [34], and by making resources available [35] for the research community to adopt more environmentally conscious research and development practices. NeurIPS [36] requires authors to disclose information on the training procedure as well as the amount and type of compute resources used in the development and research of AI models. [1] We argue that such practices of accessible documentation should be part of the submission requirements in ISMIR research publications as well. This is useful for reproducibility and allows examining the energy cost that ISMIR research contributes to when developing AMGs and other models. However, in order to start such a discussion, it is essential to examine the current practices of reporting environmental impact-related information on AMG development at ISMIR.

3. METHOD

This study covers two corpora (total N = 113) of papers: 1) ISMIR papers published in the years 2017–2023 that introduce an AMG model, and 2) a complementary corpus of ISMIR papers from the years 2022–2023 that present analysis models and discuss processes that included heavy computations with GPUs. This will provide a perspective on training resource documentation in recent ISMIR conferences beyond AMG.

The **first corpus** was obtained by selecting papers that were specified as introducing an AMG model in the table compiled by Morreale et al. [1]. We extended this initial list by adding all papers from ISMIR 2023 that presented such a model in that year. This resulted in an overall list of 88 papers that present AMG models between 2013 and 2023. An analysis of the older papers revealed that the majority of papers published before 2017 did not involve DL models trained on GPUs, but rather shallow models (e.g., [38, 39]) or no training at all (e.g., [40, 41]). Therefore, we decided to exclude the 8 papers in the list by [1] published before 2017, resulting in 80 papers in this first corpus.

For each of these 80 papers, we documented whether there was information about the training time, whether the number of parameters was specified (search "param*" [2]), and whether the computational resources used for training were documented (search "GPU*", "CPU*", "TPU*"). We further searched the papers for discussions on energy consumption and environmental impact of the models, using terms "environment*", sustainab*, "ecolog*", "carbon", "energy" and "kWh". Finally, as an effort to connect these aspects to the wider perspectives of political ecology, we documented whether the paper indicated company connections in the author affiliations, whether funding information was included in the acknowledgments or elsewhere (search "fund*", "support*"), whether there were indications of full or partial corporate funding, as well as which countries were the author affiliations related to. The full information retrieved is available in a published data table [42]. Whereas our analysis mainly focuses on the documentation of energy consumption, the additional information included in our data collection was intended to facilitate further contextualization and future research investigations.

To account for the most recent work at ISMIR in our **second corpus**, we searched the proceedings documents of 2022 and 2023 for the keywords "GPU*" and "TPU*". We did not consider papers that discuss CPU usage in order to focus on DL models, and we excluded all papers that are already part of the first corpus. This way, we obtained a corpus of 33 papers that present models for analysis rather than generation, with some consideration of computational resources (15 papers from 2023, 18 papers from 2022). From the papers we obtained, we collected further information relating to training time, computational resources, and company connections. We focused on energy-related aspects in the second corpus in order to facilitate a comparison with AMG models.

For both corpora, the energy used in the training of a model was estimated for papers that provided information about the type and number of GPUs/TPUs used, along with the training time. We found that websites or GitHub sources did not add information for the vast majority of papers and, therefore, focused on information provided in the published papers. The Thermal Design Power (TDP) of each processor type was obtained from the datasheets of the manufacturer, and the energy used for a single training run was computed as the product of a number of processors, computing time in hours, and TDP. Using the TDP as a basis for energy consumption is a rather conservative estimate, as it ignores the energy consumption of the remaining computer hardware [19]. To take these factors into account, the use of tools (*e.g.* [3, 43]) to measure actual energy consumption during model development and the publication of this overall consumption would be required. We refrain from attempting to estimate the carbon emissions related to the computed energy consumption of individual papers because a reliable estimate would require detailed information about the energy sources used in the specific computation environment [43]. In our analysis, we do not consider the energy consumption related to model inference, but we will discuss insights related to the energy demands of inference.

Authors 1 and 2 collaborated on collecting data from both corpora by dividing the conference years between

[1] Interestingly, the first editions (2021, 2022) of this checklist included a recommendation to use a CO2 emissions tracker [37], but this aspect has been omitted from the latest version of the guidelines.

[2] The asterisk character (*) is used to find all spelling variations of a search term, *e.g.* parameter, parameters, parametric etc.

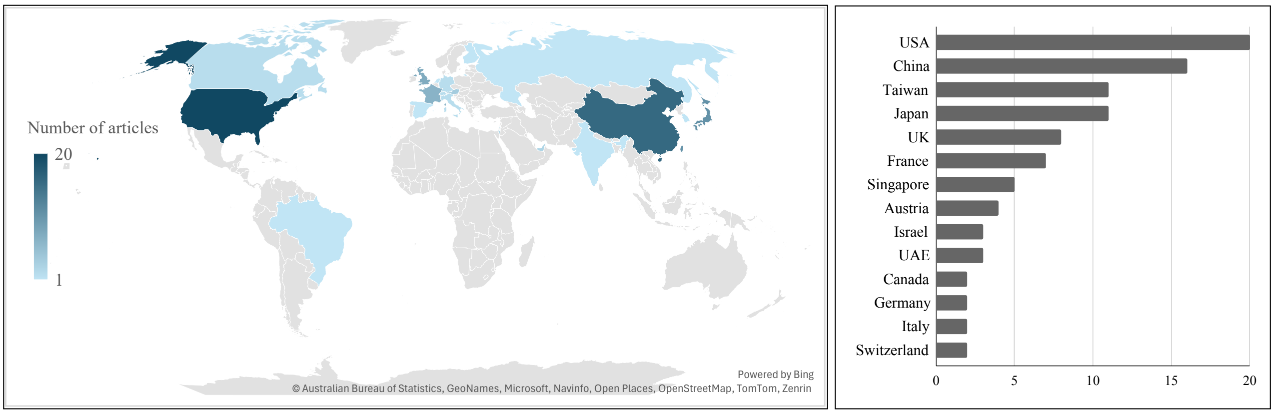

Figure 1. Geographical distribution of the authors in corpus 1 (N=80) at ISMIR 2017–2023, as indicated by the author affiliations. The block chart includes only countries from which at least two publications were found in the corpus.

them. The proceedings' PDF files were searched for the above-listed terms, and the identified occurrences were analyzed manually without the use of scripts. Papers that presented unclear aspects in the data collection were flagged and discussed between both co-authors. Author 1 conducted the estimates of energy use for models in both corpora.

4. RESULTS

4.1 Results from Corpora 1 and 2

Our analysis revealed that 60 of the 80 papers (75%) in corpus 1 do not provide any information about the time and hardware required to train the model proposed in the paper. Of the remaining 20 papers, seven only provide information about the type and numbers of GPU but do not specify the time required for training. The remaining 13 papers provide full information about GPUs as well as training time. For corpus 2 (33 papers), we identified 13 papers that disclose full information about the computational hardware and training time, another 15 papers that provide partial compute information, and 5 papers that do not include such information. Overall, only 23 % of the papers in corpora 1 and 2 are fully transparent about the hardware and the model training.

When investigating potential change over time in corpus 1, we see that between the years 2017 and 2021, one or two papers annually disclose the full training and hardware data. In 2022, an exceptional six papers provided the full information (30% of the submissions analyzed for that year), whereas, in 2023, only three papers were partially transparent with the information about the GPU budgets. While this may indicate a general trend of increasing transparency in reporting the computational hardware and training cost of the AMG models, it would be misleading to claim this as the current norm in the ISMIR community. The lack of general reflection around the issues of environmental impact is furthermore evident from how the keywords "environment*", "sustainab*", "ecolog*", "carbon', "energ*", and "kWh" were completely missing from the analyzed corpus (0 hits for proceedings of 2013–2023).

A few recent papers include reflections regarding increasing computational demands [44–46], but these reflections are motivated by the cost of computing and do not make a relation to environmental impact explicit.

It is also noteworthy how the increase in research engagement with AMG, as documented by Morreale at al. [1], coincides with corporate participation in these efforts. In the years 2020 until 2023, ca. 40 % of the papers in our corpus are co-authored by individuals with affiliations in private companies, compared to 27% in the years before that. This interestingly compares to the 27% of papers with industry-affiliated co-authors in our second corpus, *i.e.* papers that train models for non-generative purposes. Overall, these numbers suggest a certain focus of corporate interest on generative approaches.

Direct corporate funding of the research efforts is, however, rarely documented, with only four papers in the whole analyzed first corpus (N = 80) indicating either involvement of private funding or GPU support from NVIDIA. Overall, a slight majority of papers does not report any funding sources at all. The remainder refers mainly to public funding agencies, most likely as a response to the demands by the agencies for acknowledgment. This suggests an overall situation in which vested financial interests — by private and public stakeholders — are documented in a way that is not very transparent.

As shown in Figure 1, the majority of the AMG model development comes from researchers affiliated with institutions in the US (20 papers) or China (16), followed by Taiwan (11), Japan (11), and the UK (8). In total, these five countries account for over 60% of the ISMIR publications included in corpus 1. These numbers will gain significance in the context of carbon footprint estimates in the next section.

4.2 Energy Use Calculations

For the set of models from corpora 1 and 2 described above that reported the full details of the computational hardware and training time (N = 26), we conducted calculations on the estimated energy use based on the type and number of GPUs/TPUs, the reported training time, and the Thermal

Design Power (TDP) of each processor type, as provided by the manufacturers. The energy used for a single training run was consequently computed as the product of the number of processors, computing time in hours, and TDP. The results of this calculation for each model analyzed are shown in Table 1.

Based on our calculations, the mean/median amount required to train an ISMIR model (for either corpus 1 or 2) is about 224.8kWh (mean) and 18.46kWh (median). This amounts roughly to the energy demand of a single-person household for two months/three days in a Western country, such as Germany.[3] As is evident from Table 1, there is no clear distinction between the energy use of generative or analytic models, which implies that the pursued MIR task may not be an important factor. Instead, the distribution of values is strongly focused around smaller values, and only four outlier models require an amount of energy that lies above the average of 225kWh (hence, the large difference between mean and median). Out of these, the three most energy-demanding models in terms of training come from large IT corporations. The amount of energy required to train the models provided in these papers sums to 5.11MWh, which is about 87% of the total energy demand related to all 26 papers with full resource disclosure. In total, taking into account the full range of energy requirements, the papers with industry-affiliated authors demand about 89% of the total resources related to all 26 full-disclosure papers. In contrast, industry-affiliated authors are found only in 40 out of the 113 papers (35%) in our two corpora.

As mentioned in Section 3, an estimate of the actual carbon footprint requires – among other aspects – detailed information about the data centers at which the computation takes place and their energy sources. Nevertheless, we will approach a preliminary estimate of the carbon footprint related to model training at the most recent ISMIR conference. We carefully checked all papers in ISMIR 2023 and determined the number of papers that train a machine learning model, resulting in 62 out of 104 papers (59.6%). We accommodate for the fact that a small amount of these papers train "shallow" machine learning models and use an estimate of 50% of ISMIR papers that train deep learning models in recent years. Assuming the median as the representative statistic for the average energy consumption for training a model, we arrive at an energy consumption of 18.46kWh * 52 papers = 959.92kWh.

Starting from this number, two further obstacles impede a reliable estimate of the carbon footprint: 1) In each paper, the model has not been trained only once, but the total development of the presented model will have required more energy. Strubell et al. [14] have documented how the process of fine-tuning a specific model exceeded the energy demand of one training run by 24 times, and that a whole R&D cycle is three orders more expensive than a single training run. Lacking more precise numbers, it seems, therefore, fair to assume that the actual energy consumption related to a paper is at least that of fine-tuning an existing model. Hence, with a very conservative assumption of a factor of 20, we arrive at an estimate of $E_{est} = 19.20MWh$ for all model development related to a recent ISMIR conference.

[3] 5.77kWh per day for a one-person household in 2021 in Germany according to www.destatis.de.

Article	Corpus	Energy cost
Hawthorne et al 2022 [47] [4]	1	4 375 kWh
McCallum et al 2022 [44]	2	444 kWh
Toyama et al 2023 [48]	2	296 kWh
Sarkar et al 2022 [49]	2	240 kWh
Ma et al 2023 [50]	2	144 kWh
Alonso-Jiménez et al 2023 [51]	2	79 kWh
Perez et al 2023 [52]	2	36 kWh
Brunner 2018 [53]	1	33 kWh
Teng 2017 [54]	1	29 kWh
Di Giorgi et al 2022 [55]	2	24 kWh
Wu, Hsiao et al 2022 [56]	1	22 kWh
Zhao et al 2022b [57]	2	20 kWh
Donahue et al 2019 [58]	1	20 kWh
Donahue et al 2022 [59]	2	17 kWh
Yeh et al 2022 [60]	1	12 kWh
Wu, Chiu et al 2022 [61]	1	12 kWh
Singh et al 2022 [62]	2	10 kWh
Wei et al 2022 [46]	2	8 kWh
Wu & Yang 2020 [63]	1	6 kWh
Pasini & Schlüter 2022 [64] [5]	1	6 kWh
Zhao et al 2022a [65]	1	4 kWh
Zhang et al 2022 [66]	1	3 kWh
Srivatsan & Berg-Kirkpatrick 2022 [67]	2	3 kWh
Mittal et al 2021 [68]	1	3 kWh
Foscarin et al 2023 [69]	2	0,3 kWh
Peracha 2020 [70]	1	0,2 kWh

Table 1. Energy cost of model training in corpora 1 (N=13) and 2 (N=13).

The second obstacle is that the location of the data center at which computation took place is not documented. Therefore, we decided to use the countries of author affiliations as an indicator of where computation took place. In terms of carbon footprint, this has an impact as the USA and China are both on the high end of the carbon intensity spectrum [19]. We retrieved the average carbon intensity of the grids in 2022[6] for each country depicted in Figure 1, I_c (in gCO2eq/kWh) and computed the estimate for the total carbon footprint C_{total} of one conference as

$$C_{total} = (E_{est}/N_{total}) \cdot \sum_{c \in C} N_c \cdot I_c \qquad (1)$$

with N_c being the number of times co-authors were from

[4] For the four models in this paper, only the minimum and maximum training times were specified. We use the mean of these two values as an estimate.

[5] Full compute info for one of the included models only.

[6] https://ember-climate.org/data-catalogue/yearly-electricity-data/

a specific country out of the set C of all countries as depicted in Figure 1, $N_{total} = 96$ is the total count of the histogram. This results in an estimate of $C_{total} = 7.593$ tons of carbon dioxide from training processes related to one recent ISMIR conference.

Putting this number into context, according to the estimates by [43], the training of GPT-3 has caused energy consumption of about 189 MWh. With the carbon intensity of the USA in 2017 (higher than in 2022) of 449.06 gCO2eq/kWh, this has produced 85 tons of carbon dioxide, one order larger than our estimate for the whole of ISMIR.

5. DISCUSSION

While this paper focused on the training phase of music-AI models, more information is needed about the energy consumption along the full pipeline of model development, inference [7], and deployment. To this end, authors of ISMIR papers should – at the very least – clearly document the resources (compute time; type and number of processors) needed for training and inference, and – ideally – include more minute documentation of actual energy use during the whole development cycle. We encourage a discussion to adopt standards similar to NeurIPS within the ISMIR submission process.

A commonly used framework that can guide the direction towards considering the environmental impact of ISMIR in a broader sense can be found in the concept of planetary boundaries [71]. There are nine planetary boundaries that can help us to understand and analyze how our actions might influence the environmental systems. These include, for example, biodiversity loss and species extinction, stratospheric ozone depletion, ocean acidification, land-system change/deforestation, freshwater use, and atmospheric aerosol load. Taking the example of freshwater use, these dimensions can be directly applied to ISMIR research to examine the environmental impact in relation to the planetary boundaries. Efforts can be directed toward questions such as what is the level of water use for hardware cooling in computational tasks at ISMIR, and whether the life cycles of the used hardware are contributing to environmental processes such as ocean acidification or species extinction. Unfortunately, six of nine planetary boundaries are currently transgressed [72], and that places us on track for increased climate change and breakage of the prevailing ecosystems.

While energy estimates provide a baseline for understanding the scale of the specific issue of energy consumption and for comparing individual model types to one another, they are not in and of themselves a sufficient solution to the problem of environmental sustainability in model development at ISMIR or elsewhere. In order to address the complexity of the issues in all dimensions of the planetary boundaries, context-specific inquiries into the impact and effect of the ISMIR research and technologies developed and used by the community are needed. Furthermore, a broader cultural shift in thinking around AI development is necessary to bring environmental sustainability to the ISMIR research agenda. We argue that ISMIR can lead by a good example of more environmentally conscious model development, more mindful and minimalistic energy use, and reflective accounting for the environmental externalities and their political economies in current research and development practices.

We acknowledge that the calculations presented here are necessarily tentative by their nature. This is inherently a result of the lack of transparency in the ISMIR publications. While the information currently provided can provide us with indications of the scale of energy used in training the models, there are several details that may impact the exact values of these variables, which cannot be accounted for due to partial or lacking information. Such inaccuracies may skew the implied environmental impact, with undesirable consequences for social practices in the community. However, we argue that our estimate is very conservative on several points: First, the factor of 20 multiplied with the energy used for one training is below the estimates of [14], second, the use of TDP ignores all additional energy consumption by other hardware, and third, we use the median as a statistic. We would therefore like to point out that the likely underestimated energy costs could lull the research community into a false sense of security and encourage it to refrain from efforts that would be valuable for the environment. These estimates nevertheless provide an important basis upon which further inquiries into the complete environmental and ecological footprint of the conference can build.

Furthermore, we understand that the authors who contributed to our estimates were those who actively documented resource requirements. These papers may seem unfairly a focus of critique in our work, as many other authors who did not volunteer resource information at all were not cited in the paper. We believe it is instrumental to document the need for specifying the use of resources in the ISMIR community, and encourage further proactive efforts toward that goal.

6. CONCLUSION

In the era of acute climate crisis, the interest in resource-demanding music generation and analysis tasks shows signs of acceleration rather than slowing down. It is essential that research communities such as ISMIR apply critical self-reflection and acknowledge their role in promoting practices that may be excessively harmful to the environment. Increased transparency in documentation in ISMIR papers would serve better accounting for the current impacts of the research, steering the community norms and guidelines towards more sustainable practices, and providing a positive example for the wider industry. We encourage the ISMIR community to continue these critical discussions around the ethical impacts of MIR, including environmental sustainability and its political ecologies and beyond.

[7] Two models in the second corpus discuss the use of GPU resources for inference, but the included information does not allow conclusions about the energy consumption during the experiments.

7. ACKNOWLEDGMENTS

This paper is an outcome of a project funded by the Marianne and Marcus Wallenberg Foundation (MMW 2020.0102).

8. ETHICS STATEMENT

This work is based on secondary data that is publicly accessible in online sources. We acknowledge that our paper presents estimates with many uncertainties. Therefore, the numbers may imply an environmental impact larger or smaller than the actual one, in both cases to the detriment of the community. To mitigate this, we put great effort into clarifying the uncertainties in our method. We also see a risk that a paper focusing on quantitative aspects of environmental impact may fail to motivate larger-scale behavior change, which in the context of global crisis may be seen as an ethical shortcoming.

9. REFERENCES

[1] F. Morreale, M. Sharma, and I.-C. Wei, "Data collection in music generation training sets: A critical analysis," in *International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, Nov. 2023.

[2] A.-L. Ligozat, J. Lefèvre, A. Bugeau, and J. Combaz, "Unraveling the hidden environmental impacts of AI solutions for environment," *Sustainability*, vol. 14, no. 9, 2022.

[3] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres, "Quantifying the Carbon Emissions of Machine Learning," Nov. 2019. [Online]. Available: https://arxiv.org/abs/1910.09700

[4] P. Jääskeläinen, D. Pargman, and A. Holzapfel, "On the environmental sustainability of Ai art(s)," in *8th Workshop on Computing Within Limits*, Online, Jun. 2022.

[5] International Energy Agency, "Electricity 2024," Jan. 2024. [Online]. Available: https://www.iea.org/reports/electricity-2024

[6] E. Markey, "Press release: Markey, Heinrich, Eshoo, Beyer introduce legislation to investigate, measure environmental impacts of artificial intelligence," 2024. [Online]. Available: https://bit.ly/3PSwHzm

[7] J. Sevilla, L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn, and P. Villalobos, "Compute Trends Across Three Eras of Machine Learning," in *2022 International Joint Conference on Neural Networks (IJCNN)*, Padua, Italy, Jul. 2022, pp. 1–8.

[8] R. Masu, A. Melbye, J. Sullivan, and A. Jensenius, "NIME and the environment: Toward a more sustainable NIME practice," in *International Conference on New Interfaces for Musical Expression (NIME)*, Online and NYU Shanghai, Jun. 2021.

[9] A. C. on Neural Information Processing Systems (NeurIPS), "NeurIPS code of ethics," 2024. [Online]. Available: https://neurips.cc/public/EthicsGuidelines

[10] J. Barnett, "The Ethical Implications of Generative Audio Models: A Systematic Literature Review," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, Montréal, Canada, Aug. 2023, pp. 146–161.

[11] S. Oğul, "In tune with ethics: Responsible artificial intelligence and music industry," *Reflectif Journal of Social Sciences*, vol. 5, no. 1, pp. 139–149, 2024.

[12] F. Morreale, "Where does the buck stop? Ethical and political issues with AI in music creation," *Transactions of the International Society for Music Information Retrieval (ISMIR)*, vol. 4, no. 1, pp. 105–113, Jul. 2021.

[13] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Communications of the ACM*, vol. 63, pp. 54–63, Nov. 2020.

[14] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in NLP," in *Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019, pp. 3645–3650.

[15] P. Dhar, "The carbon impact of artificial intelligence," in *Nature Machine Intelligence*, vol. 2, Aug. 2020, pp. 423–425.

[16] P. Henderson, J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, and J. Pineau, "Towards the systematic reporting of the energy and carbon footprints of machine learning," *Journal of Machine Learning Research*, vol. 21, no. 248, pp. 1–43, 2020.

[17] A. S. Luccioni, S. Viguier, and A.-L. Ligozat, "Estimating the carbon footprint of BLOOM, a 176B parameter language model," *Journal of Machine Learning Research*, vol. 24, pp. 1–15, 2023.

[18] C.-J. Wu, R. Raghavendra, U. Gupta, B. Acun, N. Ardalani, K. Maeng, G. Chang, F. A. Behram, J. Huang, C. Bai, M. Gschwind, A. Gupta, M. Ott, A. Melnikov, S. Candido, D. Brooks, G. Chauhan, B. Lee, H.-H. S. Lee, B. Akyildiz, M. Balandat, J. Spisak, R. Jain, M. Rabbat, and K. Hazelwood, "Sustainable AI: Environmental implications, challenges and opportunities," 2022. [Online]. Available: https://arxiv.org/abs/2111.00364

[19] A. S. Luccioni and A. Hernandez-Garcia, "Counting carbon: A survey of factors influencing the emissions of machine learning," 2023. [Online]. Available: http://arxiv.org/abs/2302.08476

[20] C. Douwes, P. Esling, and J.-P. Briot, "Energy consumption of deep generative audio models," 2021. [Online]. Available: https://arxiv.org/abs/2107.02621

[21] C. Douwes, G. Bindi, A. Caillon, P. Esling, and J.-P. Briot, "Is quality enough f integrating energy consumption in a large-scale evaluation of neural audio synthesis models," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes, Greece, Jun. 2023, pp. 1–5.

[22] L. Zhuo, R. Yuan, J. Pan, Y. Ma, Y. Li, G. Zhang, S. Liu, R. Dannenberg, J. Fu, C. Lin *et al.*, "Lyricwhiz: Robust multilingual lyrics transcription by whispering to ChatGPT," in *International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, 2023.

[23] Y. Ding and A. Lerch, "Audio embeddings as teachers for music classification," in *International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, 2023.

[24] P. Jääskeläinen and A. Biorn-Hansen, "Critical questions for sustainability research in computational creativity," in *ICCC International Conference of Computational Creativity*. ICCC, 2024.

[25] P. Jääskeläinen, A. Holzapfel, and C. Åsberg, "Exploring more-than-human caring in Creative-Ai interactions," in *Proceedings of the 18th Nordic Conference on Human-Computer Interaction (NordiCHI)*, Aarhus, Denmark, Oct. 2022.

[26] Y. Strengers, "Smart energy in everyday life: are you designing for resource man?" *Interactions*, vol. 21, no. 4, pp. 24–31, 2014.

[27] P. Li, J. Yang, M. A. Islam, and S. Ren, "Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models," 2023. [Online]. Available: https://arxiv.org/abs/2304.03271

[28] K. Devine, *Decomposed: The political ecology of music*. Cambridge, Massachusetts: The MIT Press, 2019.

[29] M. Brennan, "The environmental sustainability of the music industries," in *Cultural Industries and the Environmental Crisis. New Approaches for Policy*, K. Oakley and M. Banks, Eds. Springer, 2020, pp. 37–49.

[30] L. Hilty, "Why energy efficiency is not sufficient - some remarks on "Green by IT"." Dessau, Germany: Shaker Verlag, 2012, pp. 13–20.

[31] A. Holzapfel, "Introducing political ecology of Creative-Ai," in *Critical Studies of Artificial Intelligence*, S. Lindgren, Ed. Glos: Edward Elgar Publishing, 2023, pp. 691–703.

[32] F. Ronchini and R. Serizel, "Performance and energy balance: a comprehensive study of state-of-the-art sound event detection systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, Korea, 2024.

[33] S. Keith, S. Collins, A. Renzo, and A. Mesker, "Slave to the ´Rithm. The AI Turn in the Music Industries," in *AI and the Future of Creative Work*. New York: Routledge, 2023, pp. 36–54.

[34] The International Conference on New Interfaces for Musical Expression (NIME), "NIME conference environmental statement," 2020. [Online]. Available: https://www.nime.org/environment/

[35] J. Sullivan, R. Masu, and A. P. Melbye, "Info and resources for sustainable NIME research," 2021. [Online]. Available: https://eco.nime.org/

[36] Annual Conference on Neural Information Processing Systems (NeurIPS), "NeurIPS Paper Checklist Guidelines," 2023. [Online]. Available: https://neurips.cc/public/guides/PaperChecklist

[37] ——, "NeurIPS paper checklist guidelines," 2021. [Online]. Available: https://neurips.cc/Conferences/2021/PaperInformation/PaperChecklist

[38] I.-T. Liu, Y.-T. Lin, and J.-L. Wu, "Music cut and paste: A personalized musical medley generating system," in *International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, Brazil, Nov. 2013.

[39] S. Cherla, T. Weyde, and A. S. d'Avila Garcez, "Multiple viewpiont melodic prediction with fixed-context neural networks," in *International Society for Music Information Retrieval Conference (ISMIR)*, 2014.

[40] G. Percival, S. Fukayama, and M. Goto, "Song2quartet: A system for generating string quartet cover songs from polyphonic audio of popular music," in *International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, Spain, Oct. 2015.

[41] H. Lim, S. Rhyu, and K. Lee, "Chord generation from symbolic melody using blstm networks," in *International Society for Music Information Retrieval Conference (ISMIR)*, 2017.

[42] A.-K. Kaila and A. Holzapfel, "Green MIR data table 2024," 2024. [Online]. Available: https://doi.org/10.5281/zenodo.12750436

[43] L. F. W. Anthony, B. Kanding, and R. Selvan, "Carbontracker: Tracking and predicting the carbon footprint of training deep learning models," in *ICML Workshop on "Challenges in Deploying and monitoring Machine Learning Systems" at the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

[44] M. McCallum, F. Korzeniowski, S. Oramas, F. Gouyon, and A. Ehmann, "Supervised and unsupervised learning of audio representations for music understanding," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[45] G. Plaja-Roglans, M. Miron, and X. Serra, "A diffusion-inspired training strategy for singing voice extraction in the waveform domain," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[46] W. Wei, P. C. Li, Y. Yu, and W. Li, "Hppnet: Modeling the harmonic structure and pitch invariance in piano transcription," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[47] C. Hawthorne, I. Simon, A. Roberts, N. Zeghidour, J. Gardner, E. Manilow, and J. Engel, "Multi-instrument music synthesis with spectrogram diffusion," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[48] K. Toyama, T. Akama, Y. Ikemiya, Y. Takida, W. Liao, and Y. Mitsufuji, "Automatic piano transcription with hierarchical frequency-time transformer," in *International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, Nov. 2023.

[49] S. Sarkar, E. Benetos, and M. Sandler, "EnsembleSet: a new high quality synthesised dataset for chamber ensemble separation," in *International Society for Music Information Retrieval Conference (ISMIR)*. Bengaluru, India: ISMIR, Nov. 2022.

[50] Y. Ma, R. Yuan, Y. Li, G. Zhang, C. Lin, X. Chen, A. Ragni, H. Yin, E. Benetos, N. Gyenge, R. Liu, G. Xia, R. B. Dannenberg, Y. Guo, and J. Fu, "On the Effectiveness of Speech Self-Supervised Learning for Music," in *International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, Nov. 2023.

[51] P. Alonso-Jiménez, X. Serra, and D. Bogdanov, "Efficient Supervised Training of Audio Transformers for Music Representation Learning," in *International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, Nov. 2023.

[52] M. Perez, H. Kirchhoff, and X. Serra, "Triad: Capturing harmonics with 3d convolutions," in *International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, Nov. 2023.

[53] G. Brunner, A. Konrad, Y. Wang, and R. Wattenhofer, "MIDI-VAE: Modeling dynamics and instrumentation of music with applications to style transfer," in *International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2018.

[54] Y. Teng, A. Zhao, and C. Goudeseune, "Generating nontrivial melodies for music as a service," in *International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China, Oct. 2017.

[55] B. D. Giorgi, M. Levy, and R. Sharp, "Mel spectrogram inversion with stable pitch," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[56] D.-Y. Wu, W.-Y. Hsiao, F.-R. Yang, O. D. Friedman, W. Jackson, S. Bruzenak, Y.-W. Liu, and Y.-H. Yang, "DDSP-based singing vocoders: A new subtractive-based synthesizer and a comprehensive evaluation," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[57] J. Zhao, G. Xia, and Y. Wang, "Beat Transformer: Demixed Beat and Downbeat Tracking with Dilated Self-Attention," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[58] C. Donahue, H. H. Mao, Y. Li, G. Cottrell, and J. McAuley, "Lakhnes: Improving multi-instrumental music generation with cross-domain pre-training," in *International Society for Music Information Retrieval Conference (ISMIR)*, Delft, Netherlands, Nov. 2019.

[59] C. Donahue, J. Thickstun, and P. Liang, "Melody transcription via generative pre-training," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[60] Y.-T. Yeh, B.-Y. Chen, and Y.-H. Yang, "Exploiting pre-trained feature networks for generative adversarial networks in audio-domain loop generation," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[61] Y.-K. Wu, C.-Y. Chiu, and Y.-H. Yang, "Jukedrummer: Conditional beat-aware audio-domain drum accompaniment generation via transformer VQ-VAE," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[62] A. Singh, K. Demuynck, and V. Arora, "Attention-based audio embeddings for query-by-example," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[63] S.-L. Wu and Y.-H. Yang, "The jazz transformer on the front line: Exploring the shortcomings of AI-composed music through quantitative measures," in *International Society for Music Information Retrieval Conference (ISMIR)*, Montréal, Canada, Oct. 2020.

[64] M. Pasini and J. Schlüter, "Musika! Fast infinite waveform music generation," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[65] J. Zhao, G. G. Xia, and Y. Wang, "Domain adversarial training on conditional variational auto-encoder for controllable music generation," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[66] D. Zhang, J.-C. Wang, K. Kosta, J. B. L. Smith, and S. Zhou, "Modeling the rhythm from lyrics for melody generation of pop songs," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[67] N. Srivatsan and T. Berg-Kirkpatrick, "Checklist models for improved output fluency in piano fingering prediction," in *International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, Dec. 2022.

[68] G. Mittal, J. Engel, C. Hawthorne, and I. Simon, "Symbolic music generation with diffusion models," in *International Society for Music Information Retrieval Conference (ISMIR)*, Online, 2021.

[69] F. Foscarin, D. Harasim, and G. Widmer, "Predicting music hierarchies with a graph-based neural decoder," in *International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, 2023.

[70] O. Peracha, "Improving polyphonic music models with feature-rich encoding," in *International Society for Music Information Retrieval Conference (ISMIR)*, Delft, Netherlands, Nov. 2019.

[71] J. Rockström, W. Steffen, K. Noone, F. S. C. Å. Persson, E. Lambin, and et al., "Planetary boundaries: exploring the safe operating space for humanity," *Ecology and society*, vol. 14, no. 2, 2009.

[72] K. Richardson, W. Steffen, W. Lucht, J. Bendtsen, S. E. Cornell, J. F. Donges, M. Drüke, I. Fetzer, G. Bala, W. von Bloh, G. Feulner, S. Fiedler, D. Gerten, T. Gleeson, M. Hofmann, W. Huiskamp, M. Kummu, C. Mohan, D. Nogués-Bravo, S. Petri, M. Porkka, S. Rahmstorf, S. Schaphoff, K. Thonicke, A. Tobian, V. Virkki, L. Wang-Erlandsson, L. Weber, and J. Rockström, "Earth beyond six of nine planetary boundaries," *Science Advances*, vol. 9, no. 37, Sep. 2023.