

## OVERVIEW ARTICLE

# Ethical Dimensions of Music Information Retrieval Technology

Andre Holzapfel\*, Bob L. Sturm† and Mark Coeckelbergh‡

This article examines ethical dimensions of Music Information Retrieval (MIR) technology. It uses practical ethics (especially computer ethics and engineering ethics) and socio-technical approaches to provide a theoretical basis that can inform discussions of ethics in MIR. To help ground the discussion, the article engages with concrete examples and discourse drawn from the MIR field. This article argues that MIR technology is not value-neutral but is influenced by design choices, and so has unintended and ethically relevant implications. These can be invisible unless one considers how the technology relates to wider society. The article points to the blurring of boundaries between music and technology, and frames music as “informationally enriched” and as a “total social fact.” The article calls attention to biases that are introduced by algorithms and data used for MIR technology, cultural issues related to copyright, and ethical problems in MIR as a scientific practice. The article concludes with tentative ethical guidelines for MIR developers, and calls for addressing key ethical problems with MIR technology and practice, especially those related to forms of bias and the remoteness of the technology development from end users.

**Keywords:** Ethics; Music Information Retrieval; Bias; Socio-Technical Approach

## 1. Introduction

This article<sup>1</sup> examines some of the ethical dimensions of research and technologies in a specific field of computer science, Music Information Retrieval (MIR). The goal is not to present a complete rulebook for ethical conduct or a comprehensive list of ethical issues in MIR, but to initiate a discussion of ethics in MIR. Computing as a profession is in need of a sense of responsibility that goes beyond a view of computing as a problem solving exercise (Gotterbarn, 2004). The design of computer systems is often driven and constrained by system design issues and trade-offs in design and performance, often without systematic considerations of potential negative impacts on society or interactions with other systems in real use cases (Huff, 2003; IEEE GIEAIS, 2017). The design paths are influenced by personal choices of developers, funding politics, and other aspects beyond efficiency and productivity (Winner, 1980), so that information technology – and technology in general – is not value-neutral (Friedman, 1996).

The field of MIR has been defined in various ways, for instance by Serra et al. (2013); Casey et al. (2008); Downie

(2003); CCRMA (2016). The importance of content-based processing is emphasized by Casey et al. (2008), where music, usually in the form of an audio recording or notation, is considered to contain information that can be extracted using computational methods. On the other hand, the management of information and the interactions between users and music content play an important role in MIR as well (Orio, 2006; Downie, 2003).

In accordance to these definitions of MIR, the call for participation at the ISMIR conference<sup>2</sup> lists in great detail task-oriented algorithmic development that aims to extract particular kinds of information from music signals. For instance, the list mentions “extracting musical features and properties”, such as genre, “estimating music metadata”, such as identifying the primary source of the piece underlying a musical performance, or manipulating musical sequences, such as synthesizing new melodies in a certain compositional style. Other topics, such as methodology and user studies, are found at the end of the call for participation, interestingly with the appendix “concerns” in both cases. In particular, user studies – arguably crucial to understand the interaction between society and music content – represent a very small part of ISMIR papers over the years (Lee and Cunningham, 2013).

The extent of discussion of ethical issues within the MIR community has been limited. The only ethical issue that Serra et al. (2013) identify in their roadmap for future directions in MIR is “addressing legal and ethical issues concerning data”, such as the consideration of what data “we should have” taking into account privacy issues.

\* Department of Media Technology and Interaction Design, KTH Royal Institute of Technology, Stockholm, SE

† Speech, Music and Hearing Division, KTH Royal Institute of Technology, Stockholm, SE

‡ Department of Philosophy, University of Vienna, Vienna, AT

Corresponding author: Andre Holzapfel ([holzap@kth.se](mailto:holzap@kth.se))

To the best of our knowledge, the only paper that has focused on ethical issues during the ISMIR conference was the 2003 keynote talk by the ethnomusicologist Anthony Seeger (Seeger, 2003). His keynote pointed out various reasons beyond copyright for music being restricted in access, such as the privacy or other cultural interests of the involved performers. Seeger observed that a distinct bias towards “US popular music or a restricted part of European ‘Classical concert music’” exists in MIR research, and demanded that retrieval systems should be designed to “include all kinds of music”. Eleven years later, a central point of an ISMIR tutorial on ethics (Holzapfel and Tzanetakis, 2014) was that current MIR algorithms are still likely to be unable to retrieve accurate information from arbitrary music signals. One of the identified ethical implications of this limitation is the potentially unfair treatment of biased recommendation algorithms that might discriminate against musicians of certain styles.

We argue that research and development in MIR already includes ethical motivations, but is insufficiently informed by more general practical ethics and ethics of technology, and is too limited in scope. It is currently guided by value judgements concerning system design constraints, whereas considerations of the interactions of developed technology with other systems and the wider society within MIR are yet to be performed. This latter consideration is being recognised more in general machine learning research. Some work in the growing domain of “interpretable machine learning” (Molnar, 2018) seeks to address problems that arise from models that inadvertently learn social prejudices from data (Caliskan et al., 2017), and that effectively propagate discrimination when applied in the real-world (Angwin et al., 2016). This is motivating the establishment of ethical standards for research and development in artificial intelligence (Bryson and Winfield, 2017), and the consideration of human rights, metrics of well-being, accountability of engineers, transparency of technology, and risk mitigation by *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems* (IEEE GIEAIS, 2017).

The following subsection grounds our discussion about ethics in MIR with examples of ethical dilemmas with impacts on various stakeholders: researchers, peer reviewers, publishers, users, and the general public. Section 2 provides the theoretical basis and the motivation for our discussion of ethics in MIR. In Section 3, we first provide a concise analysis of common practices in MIR, and then return to the motivating examples of Sec. 1.1 to highlight some ethical problems for research in MIR. Section 4 proposes initial strategies and tentative ethical guidelines to stimulate a discussion of ethics in the MIR community.

### 1.1. Grounding examples

We now present five examples to ground our discussion of ethics. Names in the examples refer to “fictional” groups of researchers and companies, but the examples we constructed represent situations based on real events. The examples do not cover all potential issues related to MIR, but instead address a variety of issues: the concept

of copyright (Examples 1 and 3); the ways in which MIR may change music (Example 1); the effects of using MIR approaches in conditions for which they were not tested (Example 2 and 3); the impact of biases encountered in MIR (Examples 2 and 4); and common practices related to datasets and evaluation measures (Example 5).

**Example 1: Traditional music modelling and generation.** Related to their work supported by a public research grant, Adaetal wonder how well a particular machine learning method can be used for music modelling and generation in a specific style, given machine-readable notations. Adaetal know about the website <http://pabcmusic.com>, which has tens of thousands of transcriptions of traditional music contributed by hundreds of people who play the music. Adaetal download all of the transcriptions and use the machine learning method to build a model of the collection.

Adaetal wonder how well the resulting model has captured the stylistic conventions of the music, and so they use the model to generate thousands of new transcriptions, synthesise them using a variety of stylistically appropriate MIDI-controlled instruments, and post them online for anyone to hear. Adaetal return to <http://pabcmusic.com> and ask its users in a discussion forum to listen to the results and say what they think. Some users are amused, some are not impressed, and some are offended.

To try to encourage a more substantial discussion about the quality of the computer-generated transcriptions, Adaetal selects 3,000 of them at random to create a volume of transcriptions. Adaetal then hires music experts to look over the volume, and perform some at concerts. One of the musicians Adaetal are working with suggests recording and releasing a CD of the computer-generated music, but passed off as composed by a real, but unknown composer in that tradition. The CD could be reviewed by a specialist, and then revealed as being generated by a computer. Adaetal write a grant proposal to support such a project, but they have second thoughts when the idea invokes a strong rebuke from another expert in the tradition. Is it ethically acceptable to deceive an audience about the origin of the music, even momentarily?<sup>3</sup>

**Example 2: Digital audio workstation.** Drumetal conduct research in computational methods that emulate the human perception of music similarity, for instance in terms of rhythmic content. They derive a method for rhythmic similarity that makes use of an estimation of the beat in a piece of music. They focus their evaluations on Fourland popular music, because of its fairly stable tempo characteristics and even meter in 4/4, and because a dataset that is organized into rhythm-related classes is available in these styles. They show the system performs well on this particular dataset.

Abletal – a company developing Digital Audio Workstation (DAW) software – become aware of the technology developed by Drumetal, and approach them for a collaboration in order to incorporate their similarity method into a DAW. The DAW is popular all over the world, and producers in Fiveland, where all local dance music is in 5/4, discover that the new tools in the DAW do not really work for them. They decide to adapt their

productions to the rhythmic style in Fourland because they want the full advantages of the DAW. Is it just that Fiveland musicians and producers have to adapt their productions and style to Fourland rhythmic styles because of the cultural dominance of the latter, thereby neglecting their own musical tradition?

**Example 3: Free music.** Systems for music detection and recognition can be applied to count the number of hours that radio stations broadcast music, and/or to identify the recordings played. The government of Lalaland decides to apply a cover song detection system to all radio broadcasts to prevent the illegal reproduction of copyright-protected music. Radio broadcasters are obliged to pay amounts according to the time measured by the installed systems, but traditional Lalalandian tunes that do not underlie any copyright restrictions get a large part of the airplay in Lalaland. However, the system incorrectly attributes most of the indigenous music as cover versions of Beatles songs. Broadcasters thus reduce the amount of time they broadcast local (and other) music, in order to reduce their costs. Is this fair towards local musicians? Does the technology unjustly discriminate against them?

**Example 4: The long tail.**<sup>4</sup> Spotetal are a large music streaming company using a recommendation system that satisfies many users. However, a large number of artists is never recommended by their systems, due to a lack of user data or other artefacts that are not completely understood. The recommendation system incorporates state-of-the-art machine learning, and takes into account user data as well as the audio data of the recordings. Some Fiveland musicians complain about the fact that Spotetal never recommends them, and they claim that they lose larger amounts of money due to this situation. They demand an explanation for why they are discriminated against by Spotetal. Is this a case of discrimination, and if so, can it be avoided by re-designing the technology?

**Example 5: Flawed dataset/experiment.** Jimetal have developed an algorithm that extracts features from audio recordings, and wonder how well it works for the problem of music genre recognition, *i.e.*, the classification of audio recordings to a defined set of music genre labels. Jimetal know a dataset that has been widely applied for the evaluation of such systems. They use a standard experimental design with this benchmark dataset, and find their new feature to be very successful. Jimetal refine their feature extraction method, and successfully publish two conference papers and a journal article detailing their results.

Sametal are working on the same problem, and wonder why the feature of Jimetal works so well. They attempt to reproduce the results of Jimetal, but have considerable trouble. Sametal communicate these troubles with Jimetal. After some correspondence, it becomes apparent to Jimetal that their original experiments had an error that inflated their results, and that the benchmark dataset has a variety of flaws. This invalidates the results in their three publications, and calls into question the outcomes of experiments in hundreds of publications that use the same dataset. How should the researchers deal with this? What do research ethics require from Jimetal?

## 2. Theoretical basis and motivation for ethics of MIR

### 2.1. A socio-technical approach as a theoretical basis for ethics of MIR

Practical ethics establishes norms for acceptable conduct, and provides a framework for analysis that informs decision and action (Resnik, 2015). An example of practical ethics is engineering ethics, which amounts to the set of ethical principles of obligation, rights, and ideals that ought to be endorsed by those engaged in engineering (Martin and Schinzinger, 1996, p.4). Various engineering societies have codes of ethics, such as the Institute of Electrical and Electronics Engineers (IEEE), and the Association for Computing Machinery (ACM).

Establishing codes of ethics for software development has been informed by the discourse in computer ethics, which deals with issues that occur due to the employment of digital technologies (Moor, 1985). Sociotechnical computer ethics (Johnson, 2009) acknowledges that a piece of software is not an isolated object, but a combination of human arrangement, technical artefacts, and social practices into a sociotechnical system (Hughes, 1994). A sociotechnical system involves an interaction between technology and society, in which both shape each other. The concept thus denies the idea that development in technology can be considered solely from a “within” perspective targeting progress towards objectively improved technological artefacts. Motivated by such a wider perspective, non-neutrality and the importance of transparency of algorithmic decision making are now being discussed more widely (Algorithm Watch, 2017; Mittelstadt et al., 2016; IEEE GIEAIS, 2017; Molnar, 2018).

Huff (2003) conceptualizes the interactions between technology and society as four levels of constraints on system design, summarised by **Table 1**. At the lowest level, one designs a computer system without interaction with any environment, and constrained only by design issues, and trade-offs in design and performance. At the second level, design is constrained by logistical issues, *e.g.*, company policies, budgets, and time-lines. At the third level, system designers anticipate how the system will interact with other technologies. At the highest level, system designers consider larger “impact on society” issues, *e.g.*, privacy, property, power, and equity. Huff’s levels shows how interactions that constrain system

**Table 1:** Huff (2003) conceptualizes the constraints on system design along four levels. Higher levels relate to increasingly wider circles of society.

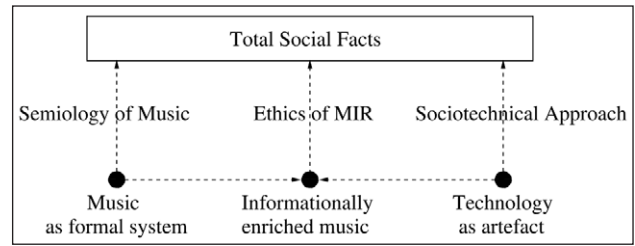
Level 4	Larger “impact on society” issues ( <i>e.g.</i> privacy, property, power, equity)
Level 3	Anticipated uses and effects: interactions with other technologies and systems
Level 2	Company policies, specifications, budgets, project time-lines
Level 1	System design issues, trade-offs in design and performance

design relate to increasingly wider circles of society, and so it can be considered a taxonomy of design regarded from a sociotechnical perspective.

Our proposal is guided by the sociotechnical approach, and aims to establish an ethics of MIR motivated by two concepts: *informational enrichment* and *total social fact*. First, an increasing interaction between music and technology leads to what Moor (2003) would refer to as “informationally enriched music”. The concept of informational enrichment was suggested by Moor (2003) in the context of money, which through a process of digitization increasingly lost its original physical interpretation and became informationally enriched by means of internet transactions. Similarly, the lines between music and technology become hard or impossible to draw in environments where digital devices and algorithms take part in music creation, distribution, and listening.

Second, the relevance of considering MIR (or more generally music technology) using the sociotechnical approach resonates with the proposition of Molino et al. (1990) to take into account the neutral level of the music datum (the music signal) as well as the creative/performative and receptive aspects of music in music analysis. He goes beyond the sender-receiver model of music communication towards a semiology of music, which regards music as symbolic systems with many possible interpretations. Nattiez (1990) takes the same approach. Music within such a system represents a total social fact, *i.e.*, a human activity with sociological, historical, and physio-psychological dimensions, which has the potential to set in motion individuals, groups, and even the whole society. For instance, the importance of music to developing individual and group identities has been well recognized within social sciences (DeNora, 1999; North and Hargreaves, 2008). The related phenomena are at the same time legal, economic, religious, aesthetic, and morphological. Whereas the phenomena exist at the social level, they can only be perceived in concrete data, which in the case of music can be the musical text or sound, and also the behavior of performers, listeners, or dancers. This concept of a total social fact has strong relations to the sociotechnical approach, in which technology is not considered as an isolated artifact but as a system interacting with all parts of society.

**Figure 1** illustrates the relationship between music semiology and the sociotechnical approach. In music semiology, music as a formal system (described by a set of rules) forms the neutral basis. In the case of the sociotechnical approach this basis is technology as an artefact with no interaction with society. Both music semiology and the sociotechnical approach argue for an interpretation of either music or technology in terms of a total social fact. The process of informational enrichment leads to an increasing relatedness of music semiology and the sociotechnical approach. This makes an analogous ethical framework for music information retrieval necessary, which examines the ethical implications of technology development for enriched music as a total social fact, and not as isolated datum or artefact. Within this framework, the four levels of constraints (**Table 1**) are



**Figure 1:** The interpretation of technology as total social fact and not as neutral artefact is established by the sociotechnical approach, whereas the semiology of music makes the analogous connection in the context of music. With information enrichment as a process that merges music and technology – symbolized by the two horizontal arrows – Ethics of MIR will interpret informationally enriched music as total social fact (*i.e.*, establish the middle vertical arrow). We suggest that the ethics of MIR should be inspired by both the semiology of music and the sociotechnical approach.

instrumental to guide research and system design within MIR. Level 1 represents design constraints that solely take into account technology as an isolated artefact, *i.e.*, the basis of **Figure 1**. Considering constraints up to Level 4 may avoid ethical conflicts, because the related constraints account for music as a total social fact. For instance, in Example 2 developers did not take into account the potential interaction of their similarity method when applied in a digital audio workstation (Level 3). Example 4 illustrates how a recommendation system may affect the equity of musicians in accessing the market (Level 4).

**2.2. Motivations for an ethics of MIR**

We see three wider contexts within which MIR operates, aspects that strongly motivate a more thorough discussion of ethics in MIR. First, recent developments in the music industry indicate a growing importance of MIR technology in music distribution. Part of the growth of the music industry relies on the success of automatic recommendation in countries where the music styles that dominate the market are significantly different from the music in US or European charts. Motivation for a discussion of ethics in MIR arises from the discrepancy between the variety of music in terms of style and market conditions on the one hand, and the unknown consequences of applying MIR methods to music and distribution situations for which they have not been evaluated.

Second, the European Union established a new set of regulations that restricts automated decision-making on user-level predictors, and grants users the right to an explanation of algorithmic decisions (Goodman and Flaxman, 2016). An important question then is if machine learning algorithms trained on biased music data should be taken into account in the context of the new legislation. Even if MIR research – for instance, music recommendation (Example 4) – is not affected by these laws, the question remains if recommendation algorithms that favour artists who are well-represented in the training data do not discriminate against less-represented artists

in terms of their market share. The unfair treatment of individuals or groups caused by biased training data, and the lack of human interpretability and transparency in machine learning, have been identified as core ethical concerns in algorithmic decision making (Mittelstadt et al., 2016; IEEE GIEAIS, 2017). The MIR community can contribute to such a discussion about the novel aspect of machine learning approaches applied to music.

Third, an engineering field should go through a process of establishing and discussing ethical norms in order to maintain a reputation as a responsible and mature discipline. The urge to publish and develop at a high pace resembles a “game player” (Maccoby, 1976): Engineers engage in problem solving games, enjoy technological work, and move ahead in a competitive world. Bibliometrics become more important than a substantial and critical reflection of how the wider public interacts with our work (Example 1). Such competition encourages engineers to focus on Levels 1 and 2 of system design (Table 1). However, in order to gain respect as a profession and as a scientific field, reflections on ethics are needed to move towards a discipline that actively avoids potential harm that emerges from interactions with other technologies and systems (Level 3), and from the impact the technology may have on society in general (Level 4).

Technology developed by MIR is not an isolated artefact, but embedded into a sociotechnical system. Music is not an isolated item in a database, but rather a total social fact with many facets that continuously interact with human practices, societies, and cultures. MIR, with its technologies for analysis and synthesis of music data, actively contributes to the permanent reshaping of music and our interactions with music in the digital world. Critical reflections of this process will improve the understanding of the underlying rules that direct MIR research and shape the use and meaning of the technology (Coeckelbergh, 2017).

### 3. Identifying ethical problems in MIR

MIR studies that pursue task-oriented algorithm development often make decisions without explicit specification of a use case or the design of the evaluation (Sturm et al., 2014; Sturm, 2016). As a formal concept, the use case of a system includes the music universe in which it operates, the choice and format of music recordings, a vocabulary and semantic rules for descriptors applied to the music, and a specification of its criteria for success. The evaluation design includes decisions regarding the specific testing corpus (usually a collection of music recording material intended to be a proxy for the task addressed by a system), a quantitative measure intended to be relevant to the success criteria, and the specification of other algorithms for making comparisons. Hence, the use case and evaluation involve making decisions on the levels of algorithms, evaluation methods, and datasets.

These decisions are usually motivated not exclusively by the desire for scientific rigor. Researchers may choose algorithms based on their own or their supervisor’s background, the availability of implementations, as well

as considerations of which ones are likely to be accepted in high-ranking peer-reviewed publications. A researcher might decide which evaluation measure to use by convenience and popularity. Even when shortcomings of evaluation methods and measures have been documented, they are still applied because their improvement is an additional effort that is considered not to be part of the presented study.

The compilation of datasets is very time consuming as well, so that datasets might still be used even when their problems have been documented. Furthermore, corpora are limited in size and diversity since the annotation of music is time-consuming, and the choice of styles is restricted to what can be annotated with the information that is the target of the algorithm. This results in the following factors that limit size and diversity. First, for complex annotation tasks the annotators are restricted to musical idioms that they are familiar with. Second, only annotations can be performed that make sense in the context of the music. For instance, there may be little point in annotating functional harmony in modal musics of Turkey or India. Finally, the music material needs to be available to the annotator. For these reasons, the most common music in datasets is popular music from the cultures of the MIR developers. Apart from the various kinds of decisions that shape MIR research outcomes, the software developed for a publication is made available online only in some cases, and the results of a publication are often not immediately reproducible because datasets are shared only on request, if at all.

Many of the decisions identified above are of methodological nature, but can be based on value judgements. Some of these issues are discussed intensively, some to a lesser degree. The important question, in the context of this paper, is what ethical dimensions exist in MIR researchers’ decisions and the actions that follow them. Is it conceivable that an MIR practice discriminates against persons or groups through its implementation of algorithms in distributable software? Could any person or group be disadvantaged? While it is likely that nobody will die from an MIR algorithm,<sup>5</sup> the impact of the technology still deserves reflection.

We propose a set of thematic units within the realm of MIR research that, in our opinion, identify some of the most important ethical problems within MIR. Legal and ethical problems of exchanging music recordings for research purposes have been discussed by Seeger (2003) and Serra et al. (2013). We go beyond this aspect and discuss the transformation of music in the digital domain (Section 3.1), the unintended use of software and the impact of various biases in MIR (Section 3.2), the focus on Anglo-American copyright concepts (Section 3.3), and the scientific practices in the field of MIR in Section 3.4.

#### 3.1. Music: an informationally enriched total social fact

In accordance with the theoretical basis established in Section 2.1, we need to understand that music is not merely sound or notation, but includes the information aspects that are continually added to it by the development of

technology. This means that the invention of algorithmic structures to analyse or generate music alters its status as information entity. A central aspect of information ethics according to Floridi (2008, p.12) is that an information entity – such as music – can be considered to have certain rights to persist and to be respected in its integrity in the interaction with agents. When music is not solely considered as isolated information entity, but in relation with the many ways it interacts with society as total social fact, these rights gain even more weight by humans creating and listening to music.

From our point of view, the field of MIR has an opportunity to lead a discussion about how artistic work should be treated in informationally enriched environments. The re-processing of artistic work by means of technology, for instance in the form of remixes and mash-ups (Sinnreich et al., 2009), have been discussed from ethical and legal perspectives (Gunkel, 2016; Sturm, 2006). However, the ethical implications of a digital work of art being reshaped by MIR technology has not been discussed. Such reshaping can happen, for instance, by means of transforming the structure of the piece of music<sup>6</sup> by automatic mashups (Davies et al., 2014), by adding various layers of information to a work (for instance by annotating it with structural information), or by generating new pieces of music within a specific idiom by using the outputs of generative networks trained on music corpora (see Example 1).

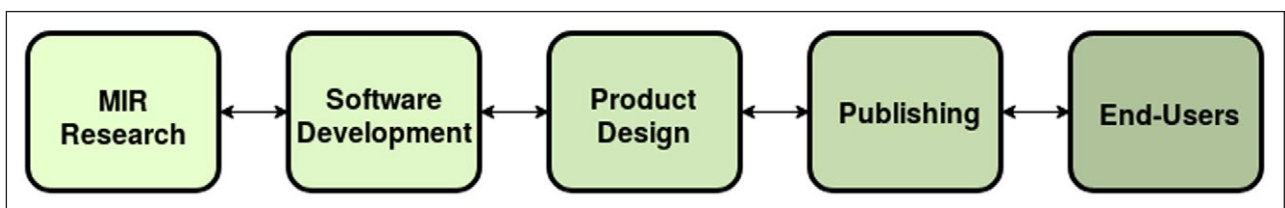
In Example 1, a corpus of music is used as material for a machine learning algorithm, which then generates “new” examples following inferred characteristics. This reshapes a musical idiom, and potentially shifts its borders by the process of algorithmic composition. Such an activity might inspire human composers to create work that may not have emerged otherwise. The attempted addition of new material into traditional repertoire based on computer output may offend people involved in the specific music community, and could deprive some musicians and composers of their means of existence. These few examples of social interaction with the digital artefact demonstrate that methods reshaping digital artwork should not only be considered under their performance and system design aspects, but also regarding their anticipated uses and their impact on society, both positive and negative, in order to meet the ethical concerns raised by the scientific endeavour.

### 3.2. Unintentional power and bias

The conditions under which a software system will be used are in many cases hard to predict for the developer. From our point of view, this is particularly important in the case of MIR. **Figure 2** depicts a simplified value chain (Kaplinsky and Morris, 2001) for the process, in which MIR research outputs make their way to end users. In many cases, basic concepts and algorithms are developed in academic institutes (first block in **Figure 2**), and from this pool of ideas choices are made by software developers in companies that want to incorporate specific functionalities into their projects (second block).

A central problem is that only limited communication is established between MIR research and other parts of the value chain. This way, usually no feedback from users can be obtained regarding software that employs specific research ideas. It has been shown that even software designers – one element closer to end users in the MIR value chain – are often too remote from the situations in which the power of their products has its effects (Huff, 2003). Such a remoteness from users has been documented for the developers of financial technologies by Coeckelbergh (2015). With the informational enrichment of music paralleling that of money (see Section 2.1), a similar remoteness between developers and users may be expected. An example can be conceived of in relation to rhythm, where most MIR tools focus on common time signatures, which finds its continuation in tools within digital audio workstations (Example 2). Similarly, developers of MIR cover song detection algorithms might not have been aware of the consequences that an application of their ideas in cultural environments has, in which the concept of copyright is different from the Anglo-American (see Example 3).

Both Examples 2 and 3 illustrate the emergence of unintentional power, which might at least be partially a result of bias in algorithmic, music corpus, and/or evaluation measure decisions of the developer. Computer systems are often biased, *i.e.*, they can systematically and unfairly discriminate against certain individuals or groups (Friedman and Nissenbaum, 1996; Angwin et al., 2016; Bryson and Winfield, 2017; IEEE GIEAIS, 2017; Molnar, 2018). Such a bias was documented for recommendation systems by Bozdag (2013), who remarks that these systems are not mere algorithms, but influenced by



**Figure 2:** MIR value chain: MIR research outcomes represent a pool of ideas. Some of the ideas are chosen to be incorporated into software solutions that aim at a specific task. In the steps that follow, the software is shaped into a product that targets specific user groups and purposes (*Product Design*), and which then needs to be marketed and distributed to the users (*Publishing*). MIR research, as most engineering research, is often not immediately involved in the following steps through the value chain. This leads to a barrier between MIR research and the higher levels of constraints on system design (Table 1).

human behaviour and concepts. By including such bias, information intermediaries become the emergent gate keepers of our society, and within MIR such a bias arguably plays a role in music recommendation (Example 4).

Bozdog (2013) differentiates between three forms of bias, which we illustrate with adaptations to the case of the MIR field:

1. *Pre-existing*: The MIR community, as many engineering research communities, is not characterized by a particularly rich socio-cultural background. MIR researchers are typically WEIRD (white, educated, industrialized, rich, operating within democracies) (Henrich et al., 2010), from a limited set of geographical origins, and a majority is male despite efforts of the “Women in MIR” initiative.<sup>7</sup>
2. *Technical*: Datasets are biased towards Eurogenetic forms of music, and consequently MIR tasks are biased towards challenges that are meaningful in these idioms, such as the transcription of music using a piano-roll representation. Technical bias can also be identified in evaluation measures, such as for the task of beat tracking, in which many measures assume the existence of an isochronous beat, a typical trait for Eurogenetic meter but not for many other musical idioms.
3. *Emergent*: Local music industries differ widely in their organization and regarding musical style across the globe (Slobin, 1992). This implies that the application scenario encountered by algorithms might be fundamentally different from the one anticipated during the implementation of the software and the development of the algorithm. Such a change of users and stakeholders who interact with a software is one of the main sources of an emergent bias (Friedman and Nissenbaum, 1996). Another source of bias is the advance in knowledge, as for instance by the discovery of flaws in a dataset used in training of a machine learning algorithm (see Example 5). Even if such knowledge is documented, the process of its integration in existing systems might not be straight-forward due to the large number of individuals and organisations involved (Huff, 2003).

We argue that the value chain and these three forms of bias can result in the unfair treatment of musicians as market participants. MIR methods, in the widest sense, conduct semantic interpretations of music, connecting measurable quantities of digital music to higher level concepts, *e.g.*, beat, artist, genre or style. Music that is under-represented in MIR datasets, or that does not fit MIR tasks and evaluation measures, is unlikely to be interpreted in a semantically correct way by methods that emerge from the biased MIR community. This is, for instance, very likely to result in situations where some artists are not recommended by content-based systems, and therefore receive less compensation by streaming content providers. This illustrates how under-representation in terms of a dataset may affect people related to such

under-represented styles, an effect that emphasizes the importance to consider music as total social fact.

### 3.3. Cultural relativity of copyright

Legislation in many countries grants creator(s) of music – or the publishing company – a set of exclusive rights, for instance, regarding reproduction, public performance, and distribution. Situations have emerged in which artists active within the Anglo-American copyright framework use creations from artists from developing countries without their agreement or without providing them financial compensation (Feld, 1996; Wallis and Malm, 1984). Whereas such conduct might have been legal in these cases, it could be considered as unfair treatment and as such unethical, especially since the opposite case – stealing from music that is protected by Anglo-American copyright – may result in legal consequences.

An example of ethical but illegal practice is sharing a music dataset that has been used to produce research results. Since this increases reproducibility, it can be considered desirable from an ethical point of view, but might be illegal since it does not conform to copyright restrictions. Legal/illegal and ethical/unethical can be regarded as axes of a two-dimensional space, with a culturally dependent mapping of cases and situations. Finding a common, universal configuration that promotes creative use and fair treatment is clearly beyond the focus of this paper. Protection of minority populations and maintaining information flow would need to be balanced out (Brown, 1998), and abandoning the dichotomy of “original” and “copy” might be a necessary step (Gunkel, 2016). Instead, we would like to point out how automation of intellectual property right (IPR) management by using MIR technology may lead to ethical problems (see Example 3).

Research in MIR provides many tools that support the automatic processing of audio for IPR management, which form an important basis in online music distribution in the so-called West (for instance, the *content ID* mechanism of YouTube<sup>8</sup>). The automatic detection of the presence of music in radio broadcasts (Schlüter and Sonnleitner, 2012) is an effective solution for billing radio stations according to national licensing agreements. In automatic music detection, the individual music sources are not identified, but the overall duration of detected music is used as a basis, ignoring the possibility that some of the music might be copyright-free in the specific legal context.

One source of ethical problems is that the notion of copyright that informs MIR systems, such as cover song detection, is derived from Anglo-American copyright laws. This leads to problems, for instance, if a melody is considered traditional in one national context, but is protected by copyright in another (Wallis and Malm, 1984, Chapter 6). MIR technology could amplify this existing power relation of unfair treatment by protecting intellectual property of individuals or corporations from specific cultures. Furthermore, algorithms for audio fingerprinting and cover-song detection are unable to consider “fair use”, *i.e.*, they cannot determine whether the use constitutes theft or is acceptable in a specific context,

for instance in education or parody. Who is responsible for decisions that are delegated to an automated, possibly machine-learning based IPR management system is a completely open question (Mittelstadt et al., 2016). MIR could contribute specific cases and viewpoints to an interdisciplinary dialogue about algorithmic decision making and fair inter-cultural copyright. Such a contribution would help to analyze the interactions of IPR management systems with intellectual property concepts in diverse socio-cultural environments.

### 3.4. MIR scientific practices

Scientific practices typically seen as objective can in fact be based on value judgements, which are widely agreed upon within a community but that include a variety of subjective elements (Longino, 1990). Whereas such value judgments affect, for instance, the way a scientific community handles publication practices and review processes, we will focus on two aspects that are more specifically related to music.

**Datasets** used for evaluation in MIR can present several problems. First, limited availability of data affects research transparency and reproducibility. Due to copyright restrictions, evaluation datasets cannot be publicly shared. This usually results in evaluation data being available only upon request from the authors, if at all. Second, the compilation and annotation of datasets is time-consuming, and therefore datasets are often of limited size. The annotations stem in most cases from one particular source, which is in many cases one of the authors of the publication first using the dataset. This combination of limited availability and size of annotated datasets may be the source of a propagation of problems that are incorporated in the data (see Example 5). A recent attempt to alleviate this dilemma<sup>9</sup> circumvents copyright restrictions by collecting large amounts of computed features, without actually sharing the music recordings. However, a specific set of features is pre-defined, which limits flexibility, and the emphasis on data quantity is not likely to reduce the bias towards certain types of music.

Several **evaluation measures** have been proposed to assess the performance of MIR algorithms on datasets for certain tasks. The consideration of statistically significant improvement regarding such measures is considered to be an “objective” indicator of progress. The problems of datasets and the evaluation measures are widely neglected in publications, even though their problems have been subject to discussion in the community. Collaborations with music archives could pave a way to improve availability, size, and quality of data, and facilitate different forms of evaluation, but apart from first initiatives the full potential of such collaborations remains to be explored (de Valk et al., 2017).

We argue that when we are not able (or willing) to reflect on our research practices – as for instance in relation to datasets and evaluation – then we ignore some of the dimensions of our work and obscure their visibility. This way we create latent value judgments that influence the outcome of our research but that are not

clearly documented or apparent to the wider public. We might present something as desirable progress to readers outside of MIR that is rather an artefact of our framework of value judgments. The increasing demand for research publications to be reproducible and transparent, for instance in the machine learning community, resonates with these issues. The ethical implications emerge from the fact that errors in research can be more easily understood when value judgments are documented, whereas their concealment leads to a blurry situation in which responsibilities are unclear. Therefore, discussing and revealing value judgments in MIR may avoid undesired interactions with other systems and help future users to identify reasons for malfunctions.

## 4. Discussion and Conclusion

We conclude with a critical perspective on one guiding example from Section 1.1, chosen because one of the authors has personal experience with it. Since ethical considerations are reflections of an autonomous human being on the consequences of his/her practices, this perspective is necessarily personal. We then synthesize some potential guidelines for ethics in MIR, which we hope will motivate more researchers to reflect on the ethical dimensions and impacts of their research. These guidelines are only meant as points of departure, and not to be considered exhaustive.

### 4.1. Critical perspective on traditional music modelling and generation

The research question of Example 1 seems harmless enough: how well can machine learning model a specific style of folk music? However, the specific style Adaetal chose – initially by reasons of convenience and availability of data (value judgements that Adaetal should make explicit) – is of a “living” tradition, with modern-day practitioners, and some who see themselves as “gate keepers”. The fact that the crowd-sourced collection of music transcriptions is available publicly does not necessarily justify *any* use of the material. The contributors to that collection likely did not foresee the use of the data far outside the preservation of their practice. Hence, it is not surprising that some of the practitioners are offended by what they perceive as a trivialisation of their tradition. That someone is offended does not necessarily mean something is unethical, but it does point to issues that deserve careful reflection.

“Folk music” might be seen as owned by no one in particular, and so issues of ownership and rights could be irrelevant. However, the research of Adaetal is supported by taxpayer money (public research grants). Adaetal is profiting from research papers, invited presentations, media appearances, and job offers. This leaves Adaetal with serious questions: Is this research only benefiting Adaetal, to the detriment of the tradition they are using? How is the work of Adaetal contributing to this tradition? Implications go beyond the music tradition. Adaetal chooses particular examples generated by the models with titles they find humorous, e.g., “The Drunken Pint.” Given that much of the training material includes Irish traditional music, a focus on examples having to do with alcohol perpetuates



a harmful stereotype of Irish people as alcoholics. Weak responses to these criticisms include claiming their research is attracting attention to a living tradition; or, their computer models provide new or different ways of understanding the music. A stronger response comes if Adaetal empower the tradition's practitioners by hiring them to play traditional and computer-generated music, to give expert feedback for algorithm development, and if Adaetal maintains an honest, respectful and open dialogue with the practitioners. This also serves to decrease the separation between the research and the music practitioners, and can highlight ways to improve the models (Wagstaff, 2012).

#### 4.2. Potential guidelines

First, if the development of a technology relies on exploiting data, then MIR developers should carefully consider the relevance and quality of that data with respect to the problem they are trying to solve. This covers the presence of distortion, incomplete metadata, mislabelings, repetitions, and file corruption, as well as how the dataset is connected to the success criteria for addressing the defined problem. In addition, listening to items of a dataset might reveal specific properties of the music, a process in which the consultation of musicologists, musicians, and expert listeners may be helpful. For instance, Drumetal in Example 2 might have discovered the bias towards simple rhythmic characteristics in their evaluation data, and might either have documented this bias or extended their data by samples of Fiveland music. Accounting for bias in datasets and restricting machine learning methods from operating upon protected attributes of personal data are current areas of research in machine learning (Bryson and Winfield, 2017; Caliskan et al., 2017; Molnar, 2018).

Second, we suggest that the diverse **cultural biases** that MIR research necessarily produces must become more explicit and reflected upon in the light of the value of cultural diversity. We should be aware of the diverse markets and cultural institutions that exist throughout the world, with various music, customs, and concepts of intellectual property. Taking into account the diversity of music, both in strictly acoustic but also in wider cultural terms, the adaptation of MIR developments to different conditions must be either possible without larger engineering expertise, or the fact that a tool cannot be adapted and is constrained to specific conditions must be clearly stated. One key point here is the documentation of the data that was used to train machine learning algorithms, which can facilitate the investigation of bias, and improve the transparency of music recommendation systems. Example 4 illustrates how a bias in a dataset that favors certain data over others may be strongly connected to a bias in recommendations, which negatively affects cultural groups related to under-represented data. In order to avoid such consequences, the diversity of datasets needs to be increased, and collaborations with music archives may provide both access to data and to knowledge of its related cultural context.

Third, since many methodological choices are based on **value judgements** that dominate in our field, we are

in need of documenting and questioning pre-dominant value judgements. Once we acknowledge that (MIR) technology is not value-neutral, this facilitates making values – such as a freedom of bias and user autonomy as suggested by Friedman (1996) – explicit in the design process as it proceeds through the MIR value chain. We need to consider that widely used music collections are not necessarily good datasets. If we become aware of problems, we need to document them, and use the affected collection only if the problems can be mitigated, or if the limitations of conclusions drawn from that dataset are made explicit. This way, error propagation caused by flawed datasets (see Example 5) can be mitigated. Importantly, we need to initiate a discussion of whether evaluation that is standard in machine learning or information retrieval might be inappropriate for some MIR problems (Sturm, 2014, 2016). This hinges upon the explicit definition of a problem and the success criteria of its solution (Sturm et al., 2014). Sometimes many of the aspects that we aim to analyse do not possess an “objective” ground-truth. This is widely recognized but often approached as a methodological problem, and not as a problem that is inherent in the phenomenon of music (see, for instance, McFee et al. (2015)). This follows from a consideration of music not in terms of a simplistic sender-receiver communication, but by taking into account the wide variety of creative and interpretative perspectives that any human subject may have towards music (Molino et al., 1990; Nattiez, 1990). The resulting ambiguity, or rather richness of possible interpretations, especially affects problems that involve a high degree of subjectivity. The existence of a cleanly or expertly labeled dataset does not mean that the problem is well-defined (Sturm, 2016).

Finally, we suggest that the **remoteness** of MIR from the actual music and the related people, practices and culture should be minimised. This is in line with a recommendation of Wagstaff (2012) to include practitioners of the originating problem domain in the development and assessment of a technology. A more collaborative approach that involves a wider basis of musicians and listeners in choosing successful algorithmic compositions in Example 1 could reveal the positive creative aspects of the work, as well as its limitations. Research problems should be defined as much as possible by use cases and formalism (Sturm et al., 2014). Projects should include implementation parts, in which prototypes are developed and tested in the planned environment. Such a planned environment for an MIR algorithm must be clearly specified in the documentation of the algorithm. With such a documentation, the Lalaland administration could have anticipated undesired effects of cover song detection algorithms in their cultural environment (Example 3).

#### 4.3. Conclusion

A clear step beyond the various ethical problems regarding remoteness, cultural biases, and value judgements could be addressed by moving from dataset-based evaluation towards user-based evaluation (Wagstaff, 2012). More inclusive research projects can incorporate systematic user

studies. Examples are collaborations with music archives that provide access for streaming users, exploration of the use of music to achieve certain emotional states (Demetriou et al., 2016), or the application in therapeutic contexts (Li et al., 2010). Including users into a research process may help avoid conflicts with communities (see Example 1) by allowing them to participate as catalysts for research that matters (Wagstaff, 2012). This way the developed technology will have a stronger connection to social practices in the aimed user groups. Furthermore, there are not only the users of the technology but also other stakeholders including musicians, companies, etc. If possible they should also be involved in the process. This, however, is accompanied by other ethical issues of using human subjects, not to mention an increase in the cost of an evaluation.

Most of our propositions demand a long-term engagement with ethics and require more research, dialogue and discussion. We believe that such a process will increase the reputation of MIR as a mature scientific field, will lead to a more responsible treatment of the people who have a stake in MIR, and will be more respectful of the total social fact of music.

## Notes

- <sup>1</sup> Author contributions: The topic of this article grew from several discussions between AH and BLS. AH did most of the writing. BLS added grounding examples, and performed extensive editing. MC provided guidance in general aspects of ethics of technology, helped with structuring and editing, and wrote the abstract.
- <sup>2</sup> During the process of writing, the CFP from 2017 was considered (<https://ismir2017.smcnus.org/call-for-participations>). Changes to the CFP in 2018 seem to indicate an increasing awareness of the strong focus on task-oriented development, and the ambition to alleviate this focus.
- <sup>3</sup> The debate on music generation is not restricted to traditional music, as demonstrated by the algorithm-aided composition of a Broadway musical (Colton et al., 2016).
- <sup>4</sup> The term *long tail* has been applied to music recommendation by Celma (2010), referring to the problem of few music items being very popular, and a large amount hardly ever being accessed by listeners.
- <sup>5</sup> One possibly fatal consequence could be a personalised exercise playlist that is too fast for the listener's heart. Another possibility is using a music similarity algorithm to generate playlists for use in torture.
- <sup>6</sup> <https://musicmachinery.com/2012/11/12/the-infinite-jukebox/>.
- <sup>7</sup> <http://www.ismir.net/wimir.html>.
- <sup>8</sup> <https://www.sourceaudio.com/youtube/>.
- <sup>9</sup> <http://acousticbrainz.org>.

## Acknowledgements

The authors would like to thank Per-Anders Forstorp for comments and suggestions on an earlier version of this manuscript.

## Competing Interests

The authors have no competing interests to declare.

## References

- Algorithm Watch.** (2017). The ADM manifesto. Online, last accessed 20 June 2017, <https://algorithmwatch.org/en/the-adm-manifesto/>.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L.** (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. Online, last accessed 27 June 2017, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Bozdag, E.** (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3), 209–227. DOI: <https://doi.org/10.1007/s10676-013-9321-6>
- Brown, M. F.** (1998). Can culture be copyrighted? *Current Anthropology*, 39(2), 193–222. DOI: <https://doi.org/10.1086/204721>
- Bryson, J., & Winfield, A.** (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, 50(5), 116–119. DOI: <https://doi.org/10.1109/MC.2017.154>
- Caliskan, A., Bryson, J. J., & Narayanan, A.** (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186. DOI: <https://doi.org/10.1126/science.aal4230>
- Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M.** (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4), 668–696. DOI: <https://doi.org/10.1109/JPROC.2008.916370>
- CCRMA.** (2016). Music information retrieval. Online, last accessed 27 June 2017, <https://ccrma.stanford.edu/workshops/music-information-retrieval-2016>.
- Celma, O.** (2010). Music recommendation. In: *Music Recommendation and Discovery*, 43–85. Springer. DOI: [https://doi.org/10.1007/978-3-642-13287-2\\_3](https://doi.org/10.1007/978-3-642-13287-2_3)
- Coeckelbergh, M.** (2015). *Money Machines: Electronic Financial Technologies, Distancing, and Responsibility in Global Finance*. Routledge.
- Coeckelbergh, M.** (2017). Technology games: Using Wittgenstein for understanding and evaluating technology. *Science and Engineering Ethics*, 1–17. DOI: <https://doi.org/10.1007/s11948-017-9953-8>
- Colton, S., Llano, M. T., Hepworth, R., Charnley, J., Gale, C. V., Baron, A., Pachet, F., Roy, P., Gervas, P., Collins, N., Sturm, B., Weyde, T., Wolff, D., & Lloyd, J. R.** (2016). The “Beyond the Fence” musical and “Computer Says Show” documentary. In: *Seventh International Conference on Computational Creativity*.
- Davies, M. E., Hamel, P., Yoshii, K., & Goto, M.** (2014). Automashupper: Automatic creation of multi-song music mashups. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 22(12), 1726–1737.

- Demetriou, A., Larson, M., & Liem, C.** (2016). Go with the flow: When listeners use music as technology. In: *International Society for Music Information Retrieval Conference*, 292–298.
- DeNora, T.** (1999). Music as a technology of the self. *Poetics*, 27(1), 31–56. DOI: [https://doi.org/10.1016/S0304-422X\(99\)00017-0](https://doi.org/10.1016/S0304-422X(99)00017-0)
- de Valk, R., Volk, A., Holzapfel, A., Pikrakis, A., Kroher, N., & Six, J.** (2017). MIRchiving: Challenges and opportunities of connecting MIR research and digital music archives. In: *4th International Workshop on Digital Libraries for Musicology*, 25–28. ACM. DOI: <https://doi.org/10.1145/3144749.3144755>
- Downie, J. S.** (2003). Music information retrieval. *Annual Review of Information Science and Technology*, 37(1), 295–340. DOI: <https://doi.org/10.1002/aris.1440370108>
- Feld, S.** (1996). Pygmy POP. A genealogy of schizophonic mimesis. *Yearbook for Traditional Music*, 28, 1–35. DOI: <https://doi.org/10.2307/767805>
- Floridi, L.** (2008). Foundations of information ethics. In: *The Handbook of Information and Computer Ethics*, chapter 1, 38–59. Wiley. DOI: <https://doi.org/10.1002/9780470281819.ch1>
- Friedman, B.** (1996). Value-sensitive design. *Interactions*, 3(6), 16–23. DOI: <https://doi.org/10.1145/242485.242493>
- Friedman, B., & Nissenbaum, H.** (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330–347. DOI: <https://doi.org/10.1145/230538.230561>
- Goodman, B., & Flaxman, S.** (2016). European union regulations on algorithmic decision-making and a ‘right to explanation’. *arXiv preprint arXiv:1606.08813*.
- Gotterbarn, D.** (2004). Informatics and professional responsibility. In: Bynum, T. W., & Rogerson, S. (Eds.), *Computer Ethics and Professional Responsibility*, 107–118. Blackwell.
- Gunkel, D. J.** (2016). *Of Remixology: Ethics and Aesthetics after Remix*. MIT Press.
- Henrich, J., Heine, S. J., & Norenzayan, A.** (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. DOI: <https://doi.org/10.1017/S0140525X0999152X>
- Holzapfel, A., & Tzanetakis, G.** (2014). Why is Greek music interesting? Towards an ethics of MIR. *Tutorial presented at the International Society for Music Information Retrieval Conference*.
- Huff, C.** (2003). Unintentional power in the design of computing systems. In: *Computer Ethics and Professional Responsibility*, chapter 4, 98–106. Blackwell.
- Hughes, T. P.** (1994). Technological momentum. In: Smith, M. R. (Ed.), *Does Technology Drive History? The Dilemma of Technological Determinism*, 101–113. MIT Press, Cambridge, MA.
- IEEE GIEAIS.** (2017). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems. Online, last accessed 20 June 2017, <https://ethicsinaction.ieee.org/>. *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*.
- Johnson, D. G.** (2009). *Computer Ethics*. Pearson Education.
- Kaplinsky, R., & Morris, M.** (2001). *A Handbook for Value Chain Research*. International Development Research Centre, Ottawa.
- Lee, J. H., & Cunningham, S. J.** (2013). Toward an understanding of the history and impact of user studies in music information retrieval. *Journal of Intelligent Information Systems*, 41(3), 499–521. DOI: <https://doi.org/10.1007/s10844-013-0259-2>
- Li, Z., Xiang, Q., Hockman, J., Yang, J., Yi, Y., Fujinaga, I., & Wang, Y.** (2010). A music search engine for therapeutic gait training. In: *Proceedings of the 18th ACM International Conference on Multimedia*, 627–630. ACM. DOI: <https://doi.org/10.1145/1873951.1874037>
- Longino, H. E.** (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- Maccoby, M.** (1976). *The Gamesman: The New Corporate Leader*. Simon and Schuster.
- Martin, M. W., & Schinzinger, R.** (1996). *Ethics in Engineering*. McGraw-Hill.
- McFee, B., Nieto, O., & Bello, J. P.** (2015). Hierarchical evaluation of segment boundary detection. In: *International Society for Music Information Retrieval Conference*, 406–412.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L.** (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). DOI: <https://doi.org/10.1177/2053951716679679>
- Molino, J., Underwood, J., & Ayrey, C.** (1990). Musical fact and the semiology of music. *Music Analysis*, 9(2), 105–156. DOI: <https://doi.org/10.2307/854225>
- Molnar, C.** (2018). Interpretable machine learning: A guide for making black box models explainable. Online, last accessed 27 June 2017, <https://christophm.github.io/interpretable-ml-book/>.
- Moor, J. H.** (1985). What is computer ethics? *Metaphilosophy*, 16(4), 266–275. DOI: <https://doi.org/10.1111/j.1467-9973.1985.tb00173.x>
- Moor, J. H.** (2003). Reason, relativity, and responsibility in computer ethics. In: *Computer Ethics and Professional Responsibility*, chapter 1, 21–38. Blackwell.
- Nattiez, J.-J.** (1990). *Music and Discourse: Toward a Semiology of Music*. Princeton University Press.
- North, A., & Hargreaves, D.** (2008). *The Social and Applied Psychology of Music*. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780198567424.001.0001>
- Orio, N.** (2006). Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval*, 1(1), 1–90. DOI: <https://doi.org/10.1561/1500000002>
- Resnik, D. B.** (2015). What is ethics in research & why is it important? Online, last accessed 20 June 2017,

- <https://www.niehs.nih.gov/research/resources/bioethics/whatis/>.
- Schlüter, J., & Sonleitner, R.** (2012). Unsupervised feature learning for speech and music detection in radio broadcasts. In: *Proceedings of the 15th International Conference on Digital Audio Effects*.
- Seeger, A.** (2003). I found it, how can i use it? Dealing with the ethical and legal constraints of information access. In: *International Society for Music Information Retrieval Conference*.
- Serra, X., Magas, M., Benetos, E., Chudy, M., Dixon, S., Flexer, A., Gómez, E., Gouyon, F., Herrera, P., Jorda, S., Paytuvi, O., Peeters, G., Schlüter, J., Vinet, H., & Widmer, G.** (2013). Roadmap for music information research. Online, last accessed 5 September 2018, [http://mires.eecs.qmul.ac.uk/files/MIRES\\_Roadmap\\_ver\\_1.0.0.pdf](http://mires.eecs.qmul.ac.uk/files/MIRES_Roadmap_ver_1.0.0.pdf).
- Sinnreich, A., Latonero, M., & Gluck, M.** (2009). Ethics reconfigured: How today's media consumers evaluate the role of creative reappropriation. *Information, Communication & Society*, 12(8), 1242–1260. DOI: <https://doi.org/10.1080/13691180902890117>
- Slobin, M.** (1992). Micromusics of the west: A comparative approach. *Ethnomusicology*, 36(1), 1–87. DOI: <https://doi.org/10.2307/852085>
- Sturm, B. L.** (2006). Concatenative sound synthesis and intellectual property: An analysis of the legal issues surrounding the synthesis of novel sounds from copyright-protected work. *Journal of New Music Research*, 35(1), 23–33. DOI: <https://doi.org/10.1080/09298210600696691>
- Sturm, B. L.** (2014). The state of the art ten years after a state of the art: Future research in music information retrieval. *Journal of New Music Research*, 43(2), 147–172. DOI: <https://doi.org/10.1080/09298215.2014.894533>
- Sturm, B. L.** (2016). Revisiting priorities: Improving MIR evaluation practices. In: *International Society for Music Information Retrieval Conference*, 488–494.
- Sturm, B. L., Bardeli, R., Langlois, T., & Emiya, V.** (2014). Formalizing the problem of music description. In: *International Society for Music Information Retrieval Conference*.
- Wagstaff, K. L.** (2012). Machine learning that matters. In: *International Conference for Machine Learning*, 529–536.
- Wallis, R., & Malm, K.** (1984). *Big Sounds from Small Peoples: The Music Industry in Small Countries*, volume 2. Pendragon Press.
- Winner, L.** (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136.

**How to cite this article:** Holzapfel, A., Sturm, B. L., & Coeckelbergh, M. (2018). Ethical Dimensions of Music Information Retrieval Technology. *Transactions of the International Society for Music Information Retrieval*, 1(1), pp. 44–55. DOI: <https://doi.org/10.5334/tismir.13>

**Submitted:** 26 January 2018

**Accepted:** 13 July 2018

**Published:** 21 September 2018

**Copyright:** © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

**] *Transactions of the International Society for Music Information Retrieval* is a peer-reviewed open access journal published by Ubiquity Press.**

**OPEN ACCESS** 