



Fairness Score for Platforms

D 5.3 Fairness Score for Platforms

WP 5, T 5.3

Authors: Daniel Bedoya (ULILLE); Antoine Henry (ULILLE)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Technical references

Project Acronym	Fair MusE
Project Title	Promoting Fairness of the Music Ecosystem in a Platform-Dominated and Post-Pandemic Europe
Project Coordinator	Giuseppe Mazziotti Universidade Católica Portuguesa (UCP) gmazziotti@ucp.pt
Project Duration	1 March 2023 – 28 February 2026 (36 months)

Deliverable No.	D5.3
Dissemination level*	PU
Work Package	WP 5 – Research Outcomes & Policy Solutions
Task	T5.3 – Creation of a Fairness Score



Fairness Score for Platforms

Lead beneficiary	4 (ULILLE)
Contributing beneficiaries	1 (UCP), 3 (AAL) 5 (ULIEGE)
Due date of deliverable	31/12/2025
Actual submission date	29/12/2025

* PU = Public – fully open (automatically posted online)

SEN = Sensitive – limited under the conditions of the Grant Agreement

v	Date	Beneficiary	Author
1.0	29/12/2025	ULILLE	Daniel Bedoya, Antoine Henry



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Table of Contents

1. Introduction	5
2. The scoring question	6
2.1 The Nutri-Score system	9
3. Fairness Score Guidelines	16
3.1 Motivation	16
3.2 Objectives	17
3.2.1 For end users	17
3.2.2 For MSSs	18
3.4 Calculation of the Fairness Score	18
3.4.1 What elements are scored?	19
3.4.1 Data practices	20
3.4.2 Power balance	24
3.4.3 Remuneration	27
3.4.4 Representation	30
3.4.5 Transparency	35
3.4.6 Score total	38
4. How to use the Fairness Score	40
4.1 Implementation of the Fairness Score	41
4.2 Generalization	44
4.3 Evaluation and iteration	44
5. Application of the Fairness-Score	46
Application example: Spotify	46
Individual results	47
Aggregated results	49
6. Limits of the Fairness-Score	50
7. Conclusion	52
Appendix	53
1. Template for GDPR export files structure	53
2. Template for Credits Metadata Completeness	53
3. Tuning the Fairness Score scale	55
4. Parametric Wireframe Example	57
References	59



1. Introduction

This deliverable is the third from the Fair MusE work package 5: “Research Outcomes & Policy Solutions”. The aim of the work package is to consolidate materials and outcomes from previous work packages to deliver the task 5.3 “Creation of a Fairness Score”. As fairness has been worked on by the project during the first two years, we have developed an approach to evaluate it in a comprehensible way for end users of streaming services.

Throughout this deliverable we develop and explain our approach to assess fairness within Music Streaming Services (MSSs). This document will present methodological choices and explanations about how the Fairness Score is calculated. This is a complete presentation of the Fair MusE’s Key Exploitable Results (KER 3), namely the “Fairness Score”. The online related materials will allow the implementation of the Fairness Score within an information system. Our implementation tries to overcome difficulties related to fairness by having an outcomes-based approach and by considering human subjectivity as much as possible. To achieve that, we work on the agency of users and a parametric approach to allow them to have an impact on the recommendation.

Based on the fairness indicators defined in our deliverable D4.3, we will provide a detailed explanation of how each indicator works and is calculated. We ground our development in the scoring literature, drawing on the French nutrition score (Nutri-Score) as an example.

The structure of this deliverable is as follows: First, we present general positions in the discussion of how to create a scoring system. Next, we present an in-depth description of the indicators in our Fairness Score and how they are calculated. Finally, we explain how to implement, deploy, and scale its use, exemplifying its application process using real MSSs data, while considering its caveats and limitations.



2. The scoring question

Recommender Systems (RSs) use mathematical and logical operations to provide suggestions for various decision-making processes, targeting items most likely to be of interest to a particular user (Ricci et al., 2022). In the case of music, in particular, these suggestions include the next song to listen to in a playlist, which artists to follow, and which genres to discover next. Within the deliverables D4.1 and D4.2 of the Fair MusE project, we have illustrated how technologies within RSs (e.g., big data, deep learning, and supervised learning) have strong incentives to utilize as much data as possible to perform calculations optimizing one outcome, maximizing the time users spend on their site/application, disregarding the end user's perceived satisfaction or even causing their dissatisfaction.

To counter this trend, research on algorithmic fairness has gained significant traction over the past decades to reduce inequality and better inform people about what they consume, focusing primarily on bias measurement and mitigation (see Deliverable D4.3). While drawing inspiration from these initiatives, we designed an outcome-oriented approach and a scoring system that includes features typically missing in the field of computer science, as outlined by the actors and techniques mentioned throughout this document.

By creating a scoring system, we aim to evaluate the fairness of recommender systems, specifically MSS, in their current state and encourage platforms to refocus and realign their incentives with users and creators, improving users' experiences while making the music ecosystem healthier.

What is a scoring system?

Evaluation is a complex, multifaceted process that aims to determine facts about the properties of real-world systems (Mitchell et al., 2021). It is often performed by creating a model and verifying its validity through measurement and testing (e.g., experiments, social research, statistics, and



Fairness Score for Platforms

benchmarking) across a sample of subjects (Zhan et al., 2024). Although different terms and techniques are used across disciplines, all incorporate data (collected, observed, synthesized) into a value-judgment function, using its results to certify, enforce, and make decisions affecting their fields of study (Mathison, 2004).

Raw or aggregated results may be arranged into a single score or a discrete scale-score using letters and numbers, often with only a few levels to make it more comprehensive. These scoring systems are used to evaluate a wide range of subjects, from the very concrete, using objective physical quantities, such as the ATMO air quality index¹, to the very abstract, using subjective and self-reported data such as the coffee value assessment². They have been popularized at a vast scale by various fields, for example:

- Computer science: Typical examples include classification algorithms that determine whether objects belong to a category. To do so, the algorithm calculates a score that predicts an outcome based on a specific set of features and a mathematical function, algorithm, or technique. Recommendation systems are an interesting example, as they rank, recommend, and aggregate content, producing a single output among millions of other possibilities that supposedly best fit the user's expectations (Hansen et al., 2020; Park et al., 2012; Zhao et al., 2019). The influence of these systems on users' behavior is noticeable, as they profoundly affect which items are recommended (Abul-Fottouh et al., 2020; Ferwerda et al., 2023; N. et al., 2018).
- Economy and finance: Credit scoring models are used to evaluate the trustworthiness of loan applicants based on their likelihood of defaulting on repayments, with improved automated techniques like evolutionary algorithms (Marqués et al., 2013). Similarly, AI-based portfolio

¹ <https://www.atmo-france.org/>.

² <https://sca.coffee/value-assessment>.



Fairness Score for Platforms

selection models can predict stock returns, thereby informing investors based on their risk profiles (Rather et al., 2017).

- Education: Automated essay scoring (AES) is an innovative research topic³ that draws on natural language processing (to evaluate structure, grammar, style, and discourse) and machine learning (e.g., word embedding models, convolutional neural networks) (Hussein et al., 2019). Rubrics are a more traditional but versatile scoring system that clearly communicates expectations and outcomes to students and teachers. They are presented in a table with evaluation criteria (rows), achievement levels (columns), and detailed performance descriptors (cell values) for validation (Boer et al., 2021). We use rubrics to describe the completion criteria for the Fairness Score calculation.
- Medicine and health: Scoring systems are also standard in the medical field, where scientists and medical doctors use them to rank the most effective treatment or diagnose the most probable disease (e.g., CHA2DS2-VASc predicts cardiac events, MELD predicts mortality from end-stage liver disease) (Yoo et al., 2023). Intensive care units (ICUs) apply scoring systems (e.g., APACHE, Glasgow coma scale) for managing patients, by evaluating their characteristics on admission (case mix) to estimate specific outcomes, i.e., death before discharge (Gunning & Rowan, 1999).

Muller (2018) presents an important caveat regarding measurement and evaluation: sometimes, what is measured has no relationship to what needs to be evaluated. In other words, focusing on measurement and scoring may create an abstraction that detaches the data from the primary goal, whether intended or not. Thus, the usefulness of scoring and metrics may be threatened by the overemphasis on measuring, replacing qualitative judgment (human experience and expertise) by quantitative evaluation (numeric one-size-fits-all simplifications of complex problems), as stated in

³ The ASAP competition awards cash prizes for the best AES algorithms: <https://www.kaggle.com/c/asap-aes>.



Fairness Score for Platforms

Goodhart's law: "When a measure becomes a target, it ceases to be a good measure" (Hoskin, 1996).

It is also important to consider who creates the score and what interests they serve. Most scoring systems are (1) produced and presented within an organization and (2) destined for experts who understand the technical nature of their methods and results. However, independent actors may develop their own scoring systems with results that are easy to understand and are publicly available. Notable examples are repairability scores⁴ and nutrient profiling models. Indeed, since the Nutri-Score nutrient profiling system exhibits many of the characteristics we want to incorporate into scoring fairness for MSSs, it will be discussed in more depth in the next section.

Successful scoring systems strike a balance between two extremes: On the one hand, proprietary and obscure scoring systems may game metrics in the name of accountability to obtain rewards (avoid punishment) while disregarding real improvement. On the other hand, third-party indicators often fall victim to managerialism, i.e., the pursuit of generic, measurable optimization at the cost of domain-specific needs. This objective is reflected in the design of our score and its implementation, as detailed in the Fairness Score Guidelines section.

2.1 The Nutri-Score system

In the health and nutrition fields, a widely used scoring system is the Nutri-Score (Julia & Hercberg, 2017; Merz et al., 2024). This score evaluates the nutritional quality of food. It is based on the assumption that healthy ingredients lead to healthy food, and that healthy food leads to healthy people. A simplified communication system makes it easier to understand whether a product is good (like fruits and vegetables) or harmful to consumers' health (like ultra-processed products high in sugar and saturated fat). The calculation was based first on the model "WXYfm" from the UK Food Standards Agency (Arambepola et al., 2008). Within the calculation, the algorithm (in its simplest

⁴ See https://www.ifixit.com/Wiki/Repairability_Scoring_Rubric_v2.0 or <https://www.indicereparabilite.fr/>.



Fairness Score for Platforms

form) assigns negative points to less desirable nutrients and positive points to beneficial nutrients, which are combined into a final score. In 2017, the Nutri-Score was first adopted in France, then other countries (Belgium, Switzerland, Germany, Luxembourg, and the Netherlands) also decided to use it on a voluntary basis.

The motivation for using the Nutri-Score is to rank food according to its nutritional value to promote consumers' health. To achieve that goal, objective measures from existing data (i.e., nutritional information of constituents and energy values of foods) are used to calculate a score. The score consists of a five-letter system and a multicolored visual identity to inform and ease consumer choice. The score is limited to the evaluation of the nutritional quality of food, so it complements other labels or standards (e.g., the Nutrition Facts label⁵).

2.1.1 Technical description of the Nutri-Score system

Quantities and units:

The Nutri-Score uses a table with specific values for each food nutrient, provided by producers through a publicly available, mandatory nutrition declaration⁶. The declaration includes: (1) a recommended serving size, i.e., a portion, or consumption unit, and (2) a standard serving size, expressed in relation to a 100 g or 100 ml portion, as the recommended serving size is often smaller than what is consumed. Using the standard serving size to calculate the score ensures coherence and consistency between different types of foods.

Score calculation:

⁵ FDA's Nutrition Facts Label: <https://www.fda.gov/food/nutrition-education-resources-materials/nutrition-facts-label>

⁶ The nutrition declaration provides food and beverages nutrient values of energy (in kilojoules kJ and kilocalories kcal) and mass (in grams g, milligrams mg or micrograms µg), according to Regulation (EU) No 1169/2011, available at the URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02011R1169-20180101>.



Fairness Score for Platforms

The score is calculated based on the following nutrients in foods and beverages: Energy (kJ), sugar (g), protein (g), salt (g), fiber (g), and saturated fatty acids (g). It uses a discrete, continuous scale from -15 (most healthy) to +40 (least healthy)⁷. For any given food item, the lower the score is, the better. A significant amount of energy, sugar, salt, or saturated fatty acids is assigned positive points (0-10), which will lead to a higher score, indicating that the food's nutritional dimension is not healthy. On the contrary, foods with fiber, protein, or fruits/vegetables and legumes are assigned negative points (0-5) and categorized as healthy foods (Julia & Hercberg, 2017). The total numerical score is associated with a letter and a color (see Table 1), which are displayed as a label on food packaging.

Visual	Letter	Color	Score
	A	Green	-15 to -2
	B	Light Green	-1 to +3
	C	Yellow	+4 to +11
	D	Orange	+12 to +16

⁷ An official Nutri-Score calculator is available at the URL: https://www.santepubliquefrance.fr/content/download/546762/file/Nutri-Score_calculation_tool_for_the_updated_algorithm2023.xlsx.



Fairness Score for Platforms

	E	Red	+17 to +40
---	---	-----	------------

Table 1: Attribution of colors and letters (visual identity) based on Nutri-Score points.

A scoring system mapping letters and colors to a discrete range to simplify recognition and adoption could be applied across various contexts and environments. It is noteworthy that the ranges defining each letter of the Nutri-Score system are unequally spaced. For instance, the lowest and highest levels (E and A) are the largest, while the ranges are significantly smaller at the middle levels (B, C, D).

Within Fair MusE, we decided to implement an approach analogous to the Nutri-Score system to evaluate the fairness of MSSs, with a similar non-linear distribution of levels (as explained in Section 3.4.6 and Appendix 3). As a prototype of the Fairness Score, this document focuses on evaluating playlists and algorithmically generated personal radios. The same principles will hold for future versions of the score, which will be applied to other categories, such as the content of a platform’s main page and the ranking of artist recommendations.

Scoring Fairness

Scoring fairness is evolving from a high-level description of fairness with various dimensions (e.g., disparate impact, equal opportunity, equalized odds) (Beigang, 2023; Bellamy et al., 2019; Hellman, 2020) to a specific focus where fairness dimensions are calculated considering the subjective aspects of fairness as much as possible. We discussed and illustrated the difficulties of defining a strict perimeter for what constitutes fairness within D4.1, D4.2, and, more precisely, D4.3. Indeed, this concept is deeply rooted within human subjectivity. The difficulty here is not in creating measurement instruments and techniques; the debate centers on deciding what should be measured. In D4.3, we argued that fairness indicators should address consumer and creator



Fairness Score for Platforms

concerns (e.g., power imbalances, opaque algorithms and practices, unfair remuneration) and they should emphasize and reinforce end users' agency. Once implemented, our scoring system should also contribute to:

- Strengthening digital literacy among MSS users by making them aware of how recommendation systems work.
- Giving users direct feedback on how the tracks they are recommended relate to various fairness dimensions.

Our approach goes beyond traditional algorithmic fairness metrics, which focus primarily on optimizing one or two features (as mentioned above) or on protected attributes (e.g., race, gender, and religion). We highlight the importance of outcomes over processes. While process-oriented approaches prioritize the actions of defining, measuring, and unbiasing, outcome-oriented approaches, in contrast, consider the consequences of the process in a broader scope and context (e.g., social impact, cultural relevance, environmental sustainability). The shift in perspective results in a compass that views fairness as a complex, nuanced topic.

The Fair MusE project defined nine criteria to address algorithmic fairness in the music ecosystem (see D3.2), which, in turn, define the type of data needed to calculate five fairness features (see Figure 1). These features constitute the basis of our Fairness Score. Ultimately, they should be implemented within the industrial pipeline of a "fairness by design" music recommender system, thereby improving its outcomes.



Fairness Score for Platforms

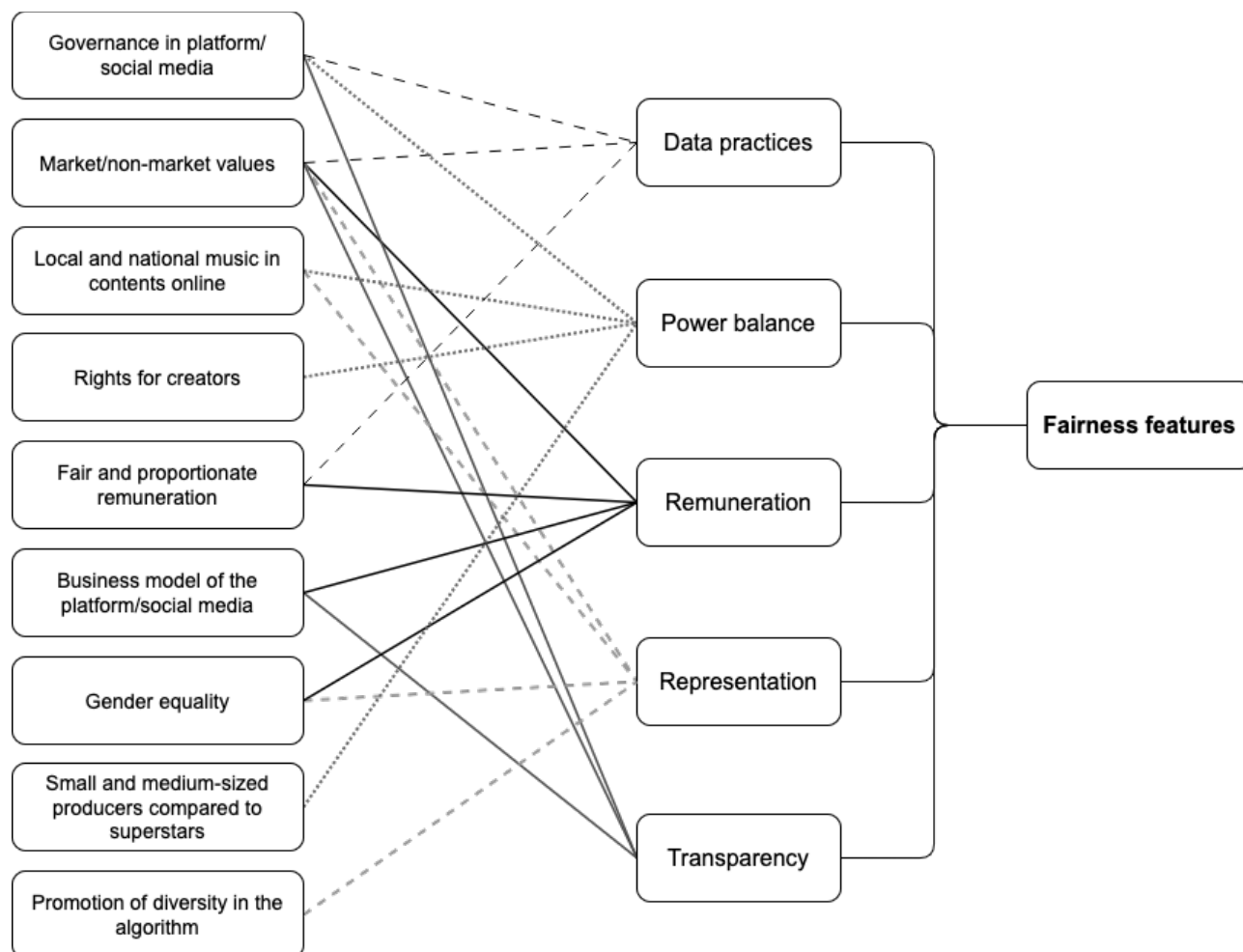


Figure 1. Correspondence between fairness criteria and fairness indicator dimensions defined in D4.3. Lines connecting to each dimension use a consistent type and thickness for visual clarity only, i.e., there is no hierarchy difference between line types or widths.

Through our Fairness Score, we try to provide a systematic approach to classify and calculate dimensions of fairness. This score will be used to evaluate content automatically generated through MSS recommendation systems, such as:

- Algorithmic stations; playlists based on a single artist or track.
- Algorithmic playlists; created using user actions on the platform (i.e., the ranking on a playlist may be influenced by skipping, sharing, or saving), similar track features (e.g., mood, genre,



Fairness Score for Platforms

or release date), or collaborative filtering (i.e., tracks are recommended because similar users have listened to them).

- Algo-torial playlists; i.e., combining editorial and algorithmic content.



3. Fairness Score Guidelines

The Fairness Score provides a concrete answer to Fair MusE's purpose of fostering fairness within the music ecosystem. This section presents the description, the methodology and the technological implementation of the score⁸.

3.1 Motivation

An unfair music ecosystem leads to issues such as increased inequalities between artists, unfair remuneration, stream manipulation, and recommendations that prioritize increasing the MSS's bottom line. To tackle these concerns, institutions (e.g., the European Parliament⁹), Collective Management Organisations (CMOs), independent labels (e.g., Impala¹⁰), artists, consumers, and policymakers undertake a collective effort to promote a fairer music ecosystem, yet they face countless difficulties (described throughout the Fair MusE project's deliverables) because MSSs have been reluctant to cooperate (Eriksson et al., 2019). To help coordinate these initiatives, promote a better understanding of the content MSSs are pushing (through their recommendation systems or curated playlists), and enable stakeholders to influence the recommended content, we have developed a prototype Fairness Score dedicated to music streaming.

The Fairness Score Guidelines provide a detailed description of each criterion, how they are assessed, and how the global score produces an aggregated evaluation of fairness in MSSs. As explained in our deliverable D4.1, the initial dataset used to produce and test the Fairness Score is derived from a database constructed by triangulating open-source information from MusicBrainz,

⁸ A prototype of the Fairness Score package is to be accessible for download through GitLab at the URL: https://gitlab.com/fairmuse1/fairness_score/.

⁹ https://www.europarl.europa.eu/doceo/document/E-10-2025-003362_EN.html.

¹⁰ IMPALA - Independent Music Companies Association: <https://impalamusic.org/>.



Fairness Score for Platforms

data provided by end users, and data from a data broker (Soundcharts). This approach aims to be independent of record labels and streaming platforms. Hacoen (2023) as well as Vombatkere et al. (2024) acknowledge the difficulty of motivating users to donate data. Within Fair MusE's deliverables D4.1, D4.2, and D4.3, we illustrated the rationale, implementation, and limitations of using data donations to study fairness in MSSs. The recruitment and data collection campaign remains active until the end of the project. We attained 212 exploitable donations in the first phase of the project, which were used for experimentation, preliminary analyses, and score prototyping. We are open to adopting larger datasets (e.g., ListenBrainz) to scale up our pipeline.

The Fairness Score is based on a core assumption: Better incentives lead to fairer recommendation algorithms, and fair recommendations lead to a healthy music ecosystem. The fairness guidelines will explore how to define these incentives and how to hold MSSs accountable.

The score must also be simple to understand and visually salient. First, by aggregating fairness features relevant to stakeholders (end users and creators), we will reinforce agency and promote understanding of the underlying recommendation systems. Second, by providing clear, identifiable visual information, we will help raise stakeholders' awareness of how the tracks recommended have been selected.

3.2 Objectives

3.2.1 For end users

- Simplify the choice of fair-listening: The algorithm will aggregate multiple dimensions of fairness into an easy-to-understand single-letter score.
- Develop a clear visual identity: The score visuals should be recognizable and easy to interpret at a glance.



Fairness Score for Platforms

- Foster fairness literacy: The algorithm and its detailed documentation will be publicly available.
- Reinforce user and creator agency: The Fairness Score is intended to inform, not to coerce listening habits, so that people keep joy in music listening.

3.2.2 For MSSs

- Encourage adoption among MSSs: The guidelines, the demonstrator, and the visual identity of the score are designed to be straightforward and easily implemented by platforms.
- Cultivate consistency and comparability: The evaluated features are applicable for any platform regardless of their specificities, leveling the playing field and comparing what is comparable.
- Promote explainability and transparency: The Fairness Score will encourage more open and explicit descriptions of the reasons behind rankings and recommendations in MSSs.
- Support scalability: The prototype presented in this document is compatible with a large-scale implementation by direct use within MSS servers and important independent actors in the open-source community.

3.4 Calculation of the Fairness Score

The fairness features described in this section were defined in Fair MusE’s D4.3 deliverable and in a preprint manuscript currently under review for publication (Bedoya et al., 2025). Here, we expand on their definitions and calculations for their use in Fair MusE’s scoring system. Each feature is depicted by: a table presenting its description, what data it uses, and its source; a rationale for its inclusion with recommended actions for MSSs; and a scoring rubric with explicit criteria for its evaluation within the score, designed similarly to a checklist fairness evaluation tool (Richardson & Gilbert, 2021). Individual features can have a smaller or larger impact on the total score according to their weights, as shown in Section 3.4.6.



3.4.1 What elements are scored?

As previously mentioned, nutrition profiling systems work by scoring food and beverages based on a precise nutrient table. However, since there is no standard list of ingredients composing MSSs, the Fairness Score uses the following type of data, which can be compared across MSSs, with obvious caveats that will be addressed further:

- Personal data files: Files exported by exercising rights under Article 15 of the General Data Protection Regulation¹¹ (GDPR) will serve to ensure that MSSs user data exports are thorough, providing sufficient descriptive data about the tracks (e.g., name, credit, and album) as well as cross-platform unique identifiers (such as the ISRC¹² or the ISNI¹³). The use of machine-readable (intended for automatic analysis) export formats (e.g., JSON or CSV) is also incorporated.
- Platform practices: Under this umbrella, the score evaluates issues such as remuneration, transparency, and policies on copyright and representation.
- Tracks in playlists: Examining playlists aims to help identify dark patterns on MSSs. This prototype of the Fairness Score uses a sample of playlists (editorial, algorithmic, and algorithmic) as a proxy for fairness practices within the platform. A playlist track set with a lower limit of 10 tracks, an upper limit of 100 tracks, and a minimum duration of 10 minutes was evaluated in this prototype where pertinent.
- Real-time recommendations: The current prototype does not incorporate real-time recommendations when calculating a platform's Fairness Score. This feature would require the following three steps, deferred for future work: First, access to a much larger catalogue (used as a training dataset) than our dataset. Second, the creation of an impartial

¹¹ Art. 15 GDPR "Right of access by the data subject": <https://gdpr-info.eu/art-15-gdpr/>

¹² International Standard Recording Code (ISRC): <https://www.iso.org/standard/64817.html>

¹³ International Standard Name Identifier (ISNI): <https://isni.org/>



recommendation algorithm that applies the philosophy and parameters highlighted in this document. Third, a pre-processing or in-processing implementation of this algorithm.

3.4.1 Data practices

To evaluate data practices from MSSs, we rooted our approach within the FAIR principles: findability, accessibility, interoperability, and reusability. Those principles have been shown to contribute to better transparency and more metadata control (Radha et al., 2025). The score evaluates GDPR-compliant export files from MSSs against the FAIR principles with a specific focus on improved interoperability (e.g., by using controlled vocabularies, unique identifiers, and metadata standards). By applying these principles, spurious metadata entries would be easier to detect, and platforms would be discouraged from being lenient on bad-faith actors.

Data access

Description	Provide standardized data/metadata in human and machine-readable formats through a simple process
Associated data	GDPR export from user
Data acquisition / Data location	Experimentation

Streaming platforms should provide users with a simple, standardised process to access and download their data in human and machine-readable formats. Data listening history should include standardised track and creator identifiers (e.g., ISRC), timestamps with standardised or otherwise explicitly stated units (e.g., ms), and user actions (e.g., shuffle mode, playlist, skip, etc.) that could help rebuild a user's recommendation profile in a new platform. User data files should also include a user's created and saved playlists, as well as followed artists and tracks. Data formats should include both human-readable (e.g., TXT files) and machine-readable (e.g., JSON files), giving the users the choice of which one to use.



Fairness Score for Platforms

GDPR Article 20¹⁴ and Recital 68¹⁵ guarantee the right to data portability and interoperability by requiring that data be structured, machine-readable, and free of charge. MSSs comply with GDPR while providing a highly inconsistent level of detail to end users. While MSSs must use well-defined database schemas and unique identifiers internally to function correctly, the information end users receive may lack the structure needed to reconstruct the connections between their data (see Appendix 1). For instance, track names that are homonyms in a user’s listening history file cannot be paired one-to-one with a unique position within a playlist. End users are thus obliged to use a paid third-party intermediary or use the migration tools to import their data into a competing platform.

Similarly, the fair principles may be applied to assess sharing and downloading options in personal playlists. For example, while most platforms provide the lowest level of access, i.e., creating a sharing link, practices aligned with the FAIR principles would incorporate an option to download a JSON file (or plain text) with the playlist data, which could be used to be imported into another MSS.

Calculation

The FAIR (findable, accessible, interoperable, reusable) principles and the tool from the Australian Research Data Commons¹⁶ are used to evaluate data requested through GDPR article 15 with this indicator.

Points	Criteria
-1	The dataset does not comply with any of the FAIR principles. Data does not contain identifiers, it is not accessible or available. Data is unstructured, mostly in a proprietary format, or lacks interoperability (proper relationships between resources).
0	The dataset complies with some of the FAIR principles. Metadata is human-readable and structured, allowing the reconstruction of connections between different types of

¹⁴ GDPR Article 20 “Right to Data Portability”: <https://gdpr-info.eu/art-20-gdpr/>

¹⁵ GDPR Recital 68 “Right of Data Portability”: <https://gdpr-info.eu/recitals/no-68/>

¹⁶ <https://ardc.edu.au/resource/fair-data-self-assessment-tool/>



Fairness Score for Platforms

	data. However, it does not contain proper relationships between resources or meaningful links to other data resources.
+1	The dataset complies with all the FAIR principles. It is structured and contains rich metadata, which is accessible. The dataset contains controlled vocabularies and complete, qualified links to other resources in a machine-readable format.

Metadata standardisation

Description	Follow and enforce standard metadata collection/provision procedures
Associated data	GDPR export from user
Data acquisition / Data location	Experimentation

Music distribution services often require metadata standardisation methods such as an ISRC. However, platforms only require a limited number of fields to be provided. For example, the music genre label on a given track may differ from one streaming platform to another. Platforms should be transparent about which metadata standardisation methods they use, what their adoption ratio among their artist database is, and which fields are kept or replaced by the platform's own. Additionally, incomplete metadata fails to contribute to a clear and transparent compensation system for rightholders.

Calculation

We use Fair MusE's Music Copyright Infrastructure (see deliverable D5.1) as a reference to assess whether MSSs abide by a comprehensive metadata standardization policy, such as the DDEX standard¹⁷. This indicator focuses primarily on the use of identifiers for musical compositions (e.g., ISWC¹⁸), recordings (e.g., ISRC), and performers (e.g., IPN¹⁹, ISNI).

¹⁷ Digital Data Exchange (DDEX): <https://ddex.net/standards/>

¹⁸ International Standard Musical Work Code (ISWC): <https://www.iswc.org/>

¹⁹ International Performer Number (IPN): <https://www.scapr.org/services/>



Fairness Score for Platforms

Points	Criteria
-1	The MSS does not apply any metadata standard, does not provide interoperable unique identifiers for any of the rightsholders as a part of GDPR export files to end users, or provides only proprietary identifiers.
0	The MSS partially applies metadata standards, i.e., not using unique identifiers for all rightsholders, or not providing them (in a well-structured manner) as a part of GDPR export files.
+1	The MSS abides by industry metadata standards and provides end users well-structured GDPR export files that allow for a complete reconstruction of relationships between data fields.

Metadata completeness

Description	Ensure MSSs provide a publicly accessible comprehensive set of metadata
Associated data	GDPR export from user
Data acquisition / Data location	Experimentation

Streaming platforms should include extended metadata about recordings and performances that were previously common in physical media. For example, a CD booklet would traditionally include credits for the production team (recording, mixing, mastering), performers (orchestras, studio musicians), and places (recording studios) that are not present in major MSSs. While metadata standards such as the DDEX enable the inclusion of all these fields into a musical work's metadata, they are typically left out of what is communicated to Digital Service Providers (DSPs). Having a complete set of metadata not only recognizes the work of the people involved in creating the tracks but also supports a clear, transparent compensation system.

Calculation

The score compares the metadata provided through GDPR export files and track credits available on MSSs against the 25 metadata fields included in the credits metadata template in Appendix 2. Metadata from the following four categories is considered: Distribution (required for ingestion by



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Fairness Score for Platforms

DSPs), composition (abstract creation, authorship), production (process of creating and finalizing a track), and performance (people who actualized the music).

Points	Criteria
-1	The MSS only complies with distribution-critical metadata. The MSS does not provide any means to access any other category through the platform, its API, or the GDPR export files.
0	The MSS provides metadata from all categories, partially (<70% of the suggested fields). If the fields are not present directly on GDPR export files or displayed on the platform, they are accessible through an API or other resource by using a standardized unique identifier.
+1	The MSS provides metadata from all categories, mostly (≥70% of the suggested fields). If the fields are not present directly on GDPR export files or displayed on the platform, they are accessible through an API or other resource by using a standardized unique identifier.

3.4.2 Power balance

This dimension demonstrates the importance of our parametric approach to change the power balance toward users and creators. Table A2 in Appendix 4 illustrates how more control could be given to these stakeholders within the platform's interfaces.

Creator control

Description	Giving artists control over (1) the extent to which their content is used for training generative AI music algorithms via an opt-in method and (2) how their content is presented and used in recommendation algorithms.
Associated data	Case study
Data acquisition / Data location	Terms and conditions Artist's contracts

Giving artists control over (1) their content being used to train generative AI music algorithms via an opt-in method and (2) how their content is presented and used in recommendation algorithms.



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Fairness Score for Platforms

For example, allowing creators to promote their latest releases over more popular content presents additional context for their music.

Calculation

Points	Criteria
-1	There is no control given to artists of their content being used to train AI or control over how their content is presented through the recommendation systems
0	At least one of the criteria is met (AI consent or control over recommendation systems).
+1	Comprehensive creator control parameters are implemented by MSSs. Artists have access to AI consent and control over recommendation systems.

User data consent

Description	Giving end user control over (1) the use of their data for the recommendation systems or access to the catalogue without recommendations and (2) the use of users' data for training algorithms.
Associated data	Case study
Data acquisition / Data location	Terms and conditions

Users should be able to choose via an opt-in method whether they want to use algorithmic recommendations at all or only have access to the platform catalogue. In parallel, the choice to have their data used for training the recommendation algorithms of other users (e.g., via collaborative filtering) should also be optative.

Calculation

Points	Criteria
-1	There is no control given to the end user over the possibility of using or not recommending systems and regarding the use of its data (i.e., for training algorithms).



Fairness Score for Platforms

0	One of the criteria is met, ensuring better control of end users regarding their data.
+1	Both of them are met, ensuring full control of end users regarding their data.

User control

Description	User-centric parametric control over recommendation systems
Associated data	Case study
Data acquisition / Data location	Terms and conditions

Since fairness from the user's perspective is also important, platform users should have control of how recommendation algorithms generate their recommendations. Since balancing fairness metrics can be challenging, platforms should not make decisions on behalf of their whole user base. For instance, users should be able to choose a balance between accuracy of recommendations and algorithmic fairness or how much new music vs. known music they would like in a search/algorithmic playlist. Users should also have more control over how music is filtered in a search²⁰. For example, users may favour tracks based on their metadata (audio features, instruments, or music producers). The AI hype of LLMs is already being used as a marketing strategy that allows users to create personalised playlists using these hidden features.²¹ However, this should not substitute for a comprehensive user interface presenting these options.

Calculation

²⁰ In May 2025, Spotify added new controls (e.g., genre filtering, new sorting options, and temporarily removing tracks from recommendations) to improve music discovery, underscoring the high demand for a customizable listening experience:

<https://newsroom.spotify.com/2025-05-07/experience-a-new-dimension-of-music-discovery-with-more-controls-and-enhanced-tools/>

²¹ Spotify's AI-playlist feature is still on beta at the time of writing this document:

<https://newsroom.spotify.com/2025-04-24/spotify-expands-ai-playlist-in-beta-to-premium-listeners-in-40-new-markets/>



Fairness Score for Platforms

Points	Criteria
-1	The end user doesn't have access to an interface to choose between the accuracy of recommendations or fairness, new/known music, and has access to advanced research filters used by the recommender system.
0	One of the criteria is met, allowing a more advanced, yet still limited, control by the end user.
+1	User control criteria are met, allowing the end user control over the recommendations and how their personal data is used by the MSS recommender system.

3.4.3 Remuneration

Compensation transparency

Description	Provide data for revenue distribution and creator remuneration and a comparable conversion in widely used royalty calculation formats
Associated data	Case study
Data acquisition / Data location	Contracts

Royalty distribution on music streaming platforms is often vague. Using the example of Spotify, which has market dominance, creator royalties arise from dividing the platform's net revenue based on stream share (number of streams per rights holder divided by the total number of streams). However, they exclude some revenue sources (downside protection deals, data-related fees) from the total net revenue (Castle & Feijóo, 2021). The more straightforward metric of stream rates (net amount of revenue to distribute divided by the total number of streams on a platform) is also misleading. Though per-stream rates are widely used by proxy creators and users (e.g., most online royalty calculators websites require only a stream count), they do not fully reflect how major platforms (Spotify, Apple Music, YouTube Music) calculate the royalties they pay. Improved stream share distributions from streaming platforms must include revenue from stock sales, data collection,



Fairness Score for Platforms

and other deals. Alternative methods to calculate royalties are being proposed but it may take time for significant changes to happen (Jensen, 2024). In any case, detailed data of royalties for (anonymised) artist revenue tiers should be publicly available instead of curated marketing statements showcasing cherry-picked examples.

Calculation

An anonymized, tiered sample datasets of artists with their respective compensation data are needed.

Points	Criteria
-1	The MSS does not disclose all the details constituting their revenue, uses royalty distribution systems that majoritarially benefit the platform and its most prominent creators, or unilaterally introduces barriers that affect independent and small creators.
0	The MSS uses royalty distribution systems that majoritarially benefit the platform and its most prominent creators.
+1	The MSS is transparent about all the details that constitute their revenue, and it uses royalty distribution systems that equitably distribute revenue among creators based on actual listening data, i.e., user-centric systems.

Profit reinvestment ratio

Description	Indicate ratio of profit reinvested on creators over shareholder returns
Associated data	Case study
Data acquisition / Data location	Contracts

Major streaming platforms have the power and resources to improve the music ecosystem. They should provide transparent reports indicating their shareholder returns compared to community reinvestment efforts for artistic creation and improving the music industry. For example, incentives



Fairness Score for Platforms

and benefits targeted at emerging artists, donations to music education non-profits, or funding public use of music recording studios.

Calculation

$$PIR = \frac{SR}{CR}$$

Where PIR is the Profit Reinvestment Ratio (from 0 to 1), SR is the shareholder return, and CR is the Community Reinvestment.

Points	Criteria
-1	The MSS does not disclose the amount they reinvest to support artistic creation and improving the music industry and/or their profits. The ratio is not calculable.
0	The MSS discloses the ratio of reinvestment for supporting the music ecosystem to profit. $0 \leq PIR < 0.5$
+1	The MSS discloses the ratio of profit to reinvestment for supporting the music ecosystem. $0.5 \leq PIR \leq 1$

Copyright control

Description	Systematic provision of state-of-the-art copyright verification tools
Associated data	Case study
Data acquisition / Data location	Terms and conditions Contracts

Audio fingerprinting technologies such as Shazam's song recognition or YouTube's content ID allow for identifying specific content and avoiding the unauthorised use of audio on streaming platforms. These technologies are more important in the current landscape where AI-generated music foreshadows important challenges in upholding creators' copyrights (Afchar et al., 2025).



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Fairness Score for Platforms

Recognition and fair compensation for creators without a financial burden on their part require platforms to offer copyright verification options systematically, especially to emergent artists.

Calculation

Points	Criteria
-1	The MSS does not provide any copyright verification system.
0	A simple manual copyright verification system exists, accessible on demand by the creators.
+1	The MSS provides an automatic copyright verification system reinforcing the rights of artists.

3.4.4 Representation

Evaluating representation is a difficult task for MSSs, as multiple attributes can be highly over- or underrepresented depending on the context. For example, a workout playlist would underrepresent music genres such as blues or folk, and a top K-pop artists playlist should not be equally composed of artists from all nationalities.

The calculation of indicators in this category should differ when a playlist is centered on a single artist/group, genre, gender, or country. Although a general theme can be inferred by analyzing its title and description, playlist metadata does not include fields explicitly stating a focus on the attributes mentioned above.

While community-based annotations could enrich an existing database of playlists or generate playlists in real time, this method still poses technical challenges for scalability and adoption. For the current prototype, we focus on exemplifying the evaluation of fairness in playlists that do not require any special treatment. Any adjustments or particularities are described below.



Fairness Score for Platforms

Track diversity

Description	Display diversity and homogeneity metrics on editorial and major playlists.
Associated data	GDPR export from user
Data acquisition / Data location	User data

Streaming platforms fully control which tracks are added to their editorial playlists. Yet, many playlists contain many of the same tracks regardless of the differences their titles or themes may suggest²². A diversity ratio computed between tracks on editorial playlists would make it evident to users whether there is an actual breadth or only a perceived abundance of music.

Calculation

The value of d in the table below represents the indicator used to evaluate the selected track diversity aspect, which can be an individual metric or a combination of them. In any case, values of d must be expressed as a continuous range from 0 (less diverse) to 1 (more diverse). Examples of possible indicators are: the Gini index (the distribution of tracks in a category compared to all available elements in that category), a cosine distance metric based on the content of the music (audio-based descriptors like loudness, instrumentality, or valence), a clustering metric such as the clustering coefficient based on a network representation of tracks, or other metrics to explore.

Points	Criteria
-1	$0.0 \leq d < 0.4$: According to the chosen metric, there is little diversity between the tracks on the playlist.
0	$0.4 \leq d < 0.7$: According to the chosen metric, there is some diversity between the tracks on the playlist.

²² In 2024, multiple sources documented the seemingly unnatural way the song “Espresso” by Sabrina Carpenter was being pushed by Spotify’s algorithm in playlists where it wouldn’t normally belong. See Amy Davidson’s article, available at <https://www.nme.com/features/music-features/espresso-sabrina-carpenter-spotify-streaming-3757741> URL:



Fairness Score for Platforms

+1	$0.7 \leq d \leq 1$: According to the chosen metric, there is a reasonable amount of diversity between the tracks on the playlist.
----	---

Gender diversity

Description	Display gender disparity metrics on artists from algorithmic playlists and recommended content
Associated data	GDPR export from user Artist gender metadata (Soundcharts, MusicBrainz)
Data acquisition / Data location	User data Data brokers

Women have been historically underrepresented in the music industry, as reflected in MSSs, though they are more often overrepresented in the superstar category (Epps-Darling et al., 2020). Platforms should disclose the ratio of female to male artists, as well as the proportion of artists with other gender identities, to continuously examine past and current trends and take corrective measures against structural gender biases.

Calculation

Artist's gender diversity in tracks using disparate impact, calculated using the following formula:

$$DIR = \frac{PU}{PP}$$

Where DIR is the Disparate Impact Ratio (from 0 to 1), PU is the probability of positive outcomes given the unprivileged label (e.g., female), PP is the probability of positive outcomes given the privileged label (e.g., male). In the absence of predicted values, the base rate (disparate impact of the provided dataset) is calculated.

A DIR of 0 means the unprivileged group is completely excluded from the outcome, while a DIR of 1 indicates equal likelihood of favorable outcomes for both groups. Usually, $DIR < 0.8$ is considered biased against the unprivileged group, though these empiric rules have caveats (Watkins & Chen,



Fairness Score for Platforms

2024).

Here the calculation of diversity implies gender diversity within tracks inside a non-gendered playlist, i.e., in a context where gender should not matter for the output of the recommendation, as it would not be appropriate to apply this metric to a playlist designed around “Boy bands in the 2000s”.

Points	Criteria
-1	$0.0 \leq \text{DIR} < 0.4$
0	$0.4 \leq \text{DIR} < 0.8$
+1	$0.8 \leq \text{DIR} \leq 1$

Auto-play diversity

Description	Consider how automatically suggested tracks are musically or thematically relevant to the current playback cue and that members of all groups are represented equally.
Associated data	GDPR export from user, MusicBrainz catalog
Data acquisition / Data location	User data

Algorithmic playlists generate a list of tracks based on a song, music genre, or artist. However, users may find a cluster of popular songs queued repeatedly regardless of their actions on the platform (Snickars, 2017). Recommendation systems should introduce diversity in the appropriate attributes (e.g., location, gender) so that the same tracks are rarely repeated through the techniques mentioned in D4.3. The resulting diversity index should then be communicated.

Calculation

Equalized odds or Equal opportunity may be used to evaluate this feature, represented by the probability p in the table below. This feature focuses on the diversity of tracks when users activate the Autoplay option in MSSs. Thus, recommendation diversity is measured on the outcomes of their exported files and a comparison using open catalogues (e.g., MusicBrainz) as a proxy for calculating



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Fairness Score for Platforms

the probabilities and ground truth data of negative outcomes (Hardt et al., 2016). However, to improve accuracy of the metric, prediction should be contrasted with consumption and tested in a sandbox using a fair recommender system.

Points	Criteria
-1	$0.0 \leq p < 0.4$
0	$0.4 \leq p < 0.8$
+1	$0.8 \leq p \leq 1$

Domestic ratio

Description	Display the proportion of local tracks in editorial and algorithmic recommendations
Associated data	GDPR export from users Artist country metadata (Soundcharts, MusicBrainz)
Data acquisition / Data location	User data Music Streaming Platforms, Data brokers

The predominance of international over national content outside of dominant markets (e.g., the United States, the United Kingdom) hinders local artists from developing a career in music except for cases where entire mainstream genres are dominated by a linguistic background (e.g., rap in France, reggaeton for Spanish). A domestic ratio, i.e., how much national over international music is consumed in the context of algorithmic consumption, is needed. Platforms should provide this data because computing it from open datasets may be unreliable (Matrosova et al., 2024). Additionally, local quotas should be enforced for contents in algorithmic recommendations following legally adopted policies (e.g., music quotas for radio stations in France).²³

²³ [Deliberation no. 2021-103 adopted in December 8, 2021](#) relating to the commitments of radio services, Law on Freedom of Communication



Fairness Score for Platforms

Calculation

The proportion of local tracks to international tracks is calculated using the formula:

$$DR = \frac{DC}{DC + IC}$$

Where DR is the Domestic Ratio (from 0 to 1), DC is the Domestic Count (number of local tracks) and IC is the International Count (number of foreign tracks).

Each credited artist's country in a track is counted once, unless they all share the same country.

Points	Criteria
-1	$0.0 \leq DR < 0.3$
0	$0.3 \leq DR < 0.6$
+1	$0.6 \leq DR \leq 1$

3.4.5 Transparency

Transparency is difficult to measure and evaluate because it needs a dynamic system verifying data provided by platforms. Each element should be meticulously validated by independent sources.

Verified artists ratio

Description	Enforce creator verification and provide the proportion of verified creators
Associated data	Case study
Data acquisition / Data location	Terms and conditions

The proliferation of fake tracks presents a skewed view of the music catalogue on streaming platforms. For example, hours of content can be generated almost instantly with minimal effort by



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Fairness Score for Platforms

filtering white, pink, or brown noise and be uploaded to relaxation or focus playlists, competing on the same level as tracks produced with real artistic intent. Streaming platforms and regulation agencies should enforce policies to avoid these cases (e.g., multiple fake developers and games exist in the Steam video game store, but their parent company, Valve, has proven slow to react). A standardised method to separate real from fake content creators and a publicly available rate of verified artists would provide independent organisations and labels with an overview of the reality of the current music landscape.

Calculation

Points	Criteria
-1	The MSS does not provide any method to verify artists and separate real vs. fake creators.
0	The MSS is using a non-disclosed approach to verify artists. The percentage of verified artists is not disclosed and thus outcomes cannot be compared.
+1	The MSS is using an open, standardized approach, allowing other stakeholders to have a clear overview of the percentage of human to non-human creators so that end users can inform their listening experience.

Explainability

Description	Provide comprehensive explanations behind recommendation generation
Associated data	GDPR export from user + Terms and conditions
Data acquisition / Data location	User data + case study

Recommender systems typically make predictions on the streaming platform side based on the available data and model parameters. Users are served with the results of the whole system, sometimes without any context (e.g., algorithmic playlists). Comprehensive explanations of the reasons behind the generation of recommendations can help identify unfair algorithmic behaviours and improve user control and satisfaction. Platforms must include more straightforward



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Fairness Score for Platforms

explanations for users. Without revealing proprietary information about their recommendation algorithms, MSS may create scalable, informative textual explanations to improve users' experience. Examples of such systems are feature-based, example-based, or knowledge graph-based explanations (Afchar et al., 2022).

Calculation

The question to answer here is: Are the recommendations explained in plain language or are they presented through a black box model?

Points	Criteria
-1	The MSS does not include any explanation for any of their recommendations, including algorithmically generated playlists, recommended artists, or recommended genres.
0	The MSS uses vague, generic, or obtuse language to refer to their recommendations on algorithmic content. The platform uses terms such as "For you", "Your next listen", "Tailored to your taste", or similar without any additional clarification.
+1	The MSS uses specific, straightforward, instructive language to explain the reasons behind recommendations. If the explanations are not displayed directly, they are accessible through an optional, unintrusive contextual menu integrated into the user interface.

AI-created music ratio

Description	The rise of AI-generated music is threatening artists, while, at the same time, their remuneration is decreasing. Having an understanding of how an MSS is detecting and handling AI generated music is an important dimension regarding fairness for the artist.
Associated data	Case study
Data acquisition / Data location	User data + case study

Music creation is threatened by the advent of AI-generated music, which already has an acceptable quality and can be produced (Danesi, 2024) in a fraction of the time of real music. Exploiters can



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Fairness Score for Platforms

generate royalties with virtually no effort, pulling the rug out from under real emerging artists who have a small catalogue. Moreover, streaming platforms could also inflate their catalogue by including AI-generated music of their own. Methods for AI-generated music detection are in the vanguard of research, not without many challenges (Afchar et al., 2025). Streaming platforms should use AI-music detection methods and publicly disclose the ratio of AI to human music. In addition, when streaming platforms create AI-generated content, they must adequately label it as such.

Calculation

Points	Criteria
-1	The MSS does not use an AI-generated music detection system or does not disclose any information regarding the ratio of AI to human content inside the catalogue or the recommendations.
0	The MSS uses an AI-generated music detection system. AI-generated music is included by default in recommendations and it constitutes more than 40% of the total tracks recommended.
+1	The MSS discloses the ratio of AI-generated music to human music. The use of AI-generated music for recommendations is clearly labeled and is opt-in. End users can control the proportion of AI-generated content that is included in their recommendations.

3.4.6 Score total

Once individual indicators and dimensions are evaluated and scored, the overall score is obtained using a weighted sum:

$$s = \sum_{i=1}^n w_i x_i$$



Fairness Score for Platforms

Where s is the total score, w represents the individual weights, and x is the value of each fairness feature. Since this prototype of the score uses unitary weights ($w = 1$) the total score is simply a sum. Because there are 16 features and each one can take a value of -1, 0, or 1, the final Fairness Score ranges from -16 to +16.

After the features have been aggregated, the numerical score s is mapped to a five-level scale using closed intervals (see Table 2). The reasoning behind the definition of each interval in Table 2 is detailed in Appendix 3.

Letter	Score Interval	Meaning	How to improve
A	[12.62, 16]	Fully aligned	Comply with the highest level of all criteria (+1 point).
B	[8.79, 12.62)	Substantially aligned	Comply with the highest level of most criteria (+1 point).
C	[4.24, 8.79)	Partially aligned	Increase criteria from neutral to high (0 to +1 point).
D	[-1.69, 4.24)	Minimally aligned	Improve criteria with the lowest levels (-1 point).
E	[-16, -1.69)	Unaligned	Comply with the neutral level of most criteria (0 point).

Table 2: Fairness Score five-letter scale correspondence with numerical score intervals. Each level contains a short explanation of its meaning and how to potentially improve.



4. How to Use the Fairness Score

The purpose of the Fairness Score is to contribute to knowledge and understanding of consumers. The Fairness Score directly addresses the issue of accountability. It's not intended to dictate musical taste, but to provide a contextual understanding of the playlist's composition according to a predefined set of criteria. This shift from a process-oriented approach – where the focus is on the algorithm's inner workings – to an outcome-oriented approach, prioritizing the user experience, is central to our work.

The introduction of the Fairness Score allows for a reintroduction of the end user/consumer into the loop, acknowledging their subjectivity and empowering them to actively participate in shaping their musical consumption habits. This approach contributes to the reinforcement of freedom of choice, which has been tackled by algorithmic systems (N. et al., 2018). From this perspective, the Fairness Score suggests that a specific playlist or (personalize) radio is or isn't fair according to our calculation. The consumer is then able to keep listening or to choose something else. It encourages users to move from being recipients of algorithmic recommendations to active evaluators, capable of critically assessing the playlists presented to them. This aligns with broader theoretical frameworks that emphasize the importance of user agency and the potential for technology to be a tool for empowerment rather than a source of control. The success of the Fairness Score will depend not only on its technical implementation but also on its ability to resonate with users and inspire them to engage with the system.

For MSS, an important change is this shift from a process-oriented approach to a more outcome-oriented approach. This shift allows to reintroduce the end user/consumer in the loop and its subjectivity regarding what he wants to promote. Low scores from the Fairness Score aren't equal to "bad" music or "bad" quality of the playlist; it's more about having an overview of the playlist regarding our criteria.



With the Fairness Score, we would like to implement a parametric approach, allowing end users to reinforce their agency by having a direct possibility to influence recommendations. Indeed, the effectiveness of the Fairness Score hinges on the clarity and transparency of the underlying criteria used to evaluate playlists. Without a clear understanding of what constitutes “fairness” in this context, the score risks being perceived as arbitrary or even manipulative. The criteria themselves are not neutral; they reflect specific values and priorities that could be contested or challenged. By developing a UX interface link to the criteria, we would be able to give broader control to consumers over those criteria.

4.1 Implementation of the Fairness Score

Three complementary pieces support the implementation of the system depicted by this deliverable:

1. Guidelines and code demonstrator: The Fairness Score may be implemented by following the latest version of guidelines that have been outlined in this document or by installing the upcoming Python package directly. The package will include documentation on its use, which algorithms are applied, and how to customize it for a specific need.
2. Visual identity: As previously noted, the Fairness Score has been designed to be easy to understand and visually salient. For this reason, a set of distinct labels has been designed in collaboration with partners of the Fair MusE project to represent each level of the Fairness



Fairness Score for Platforms

Score scale, as illustrated in Figures 2, 3, 4, and 5.

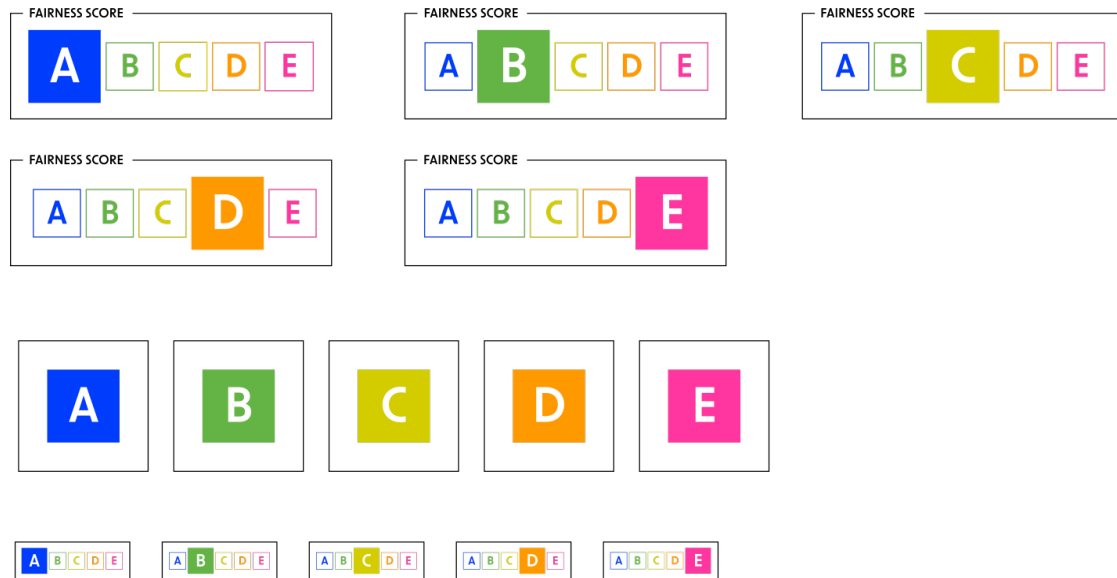


Figure 2. Visual identity of the Fairness Score (RGB version)

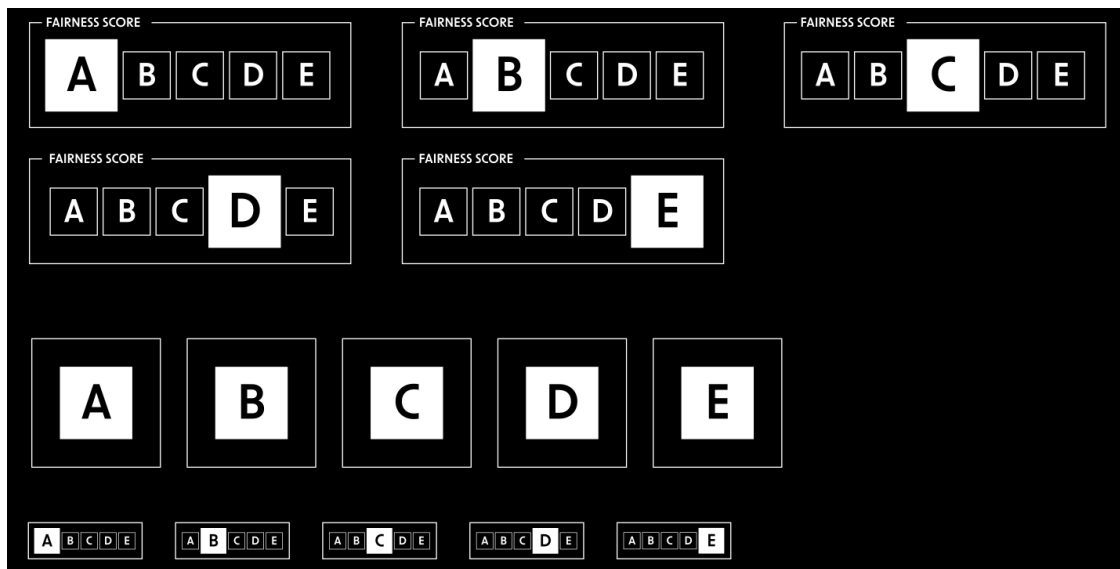


Figure 3. Visual identity of the Fairness Score (white over black version)



Fairness Score for Platforms

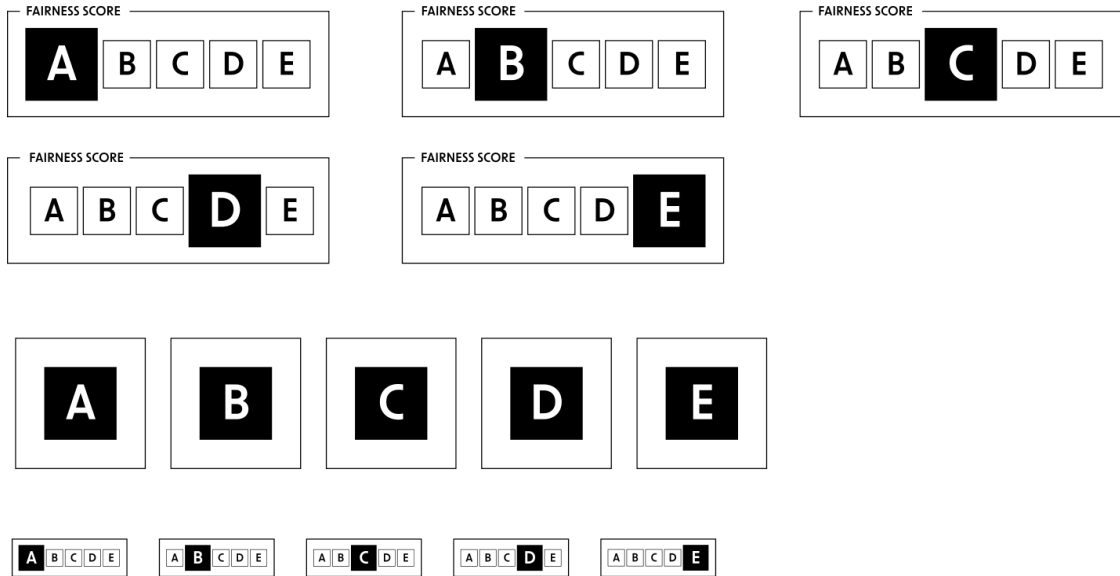


Figure 4. Visual identity of the Fairness Score (black over white version)

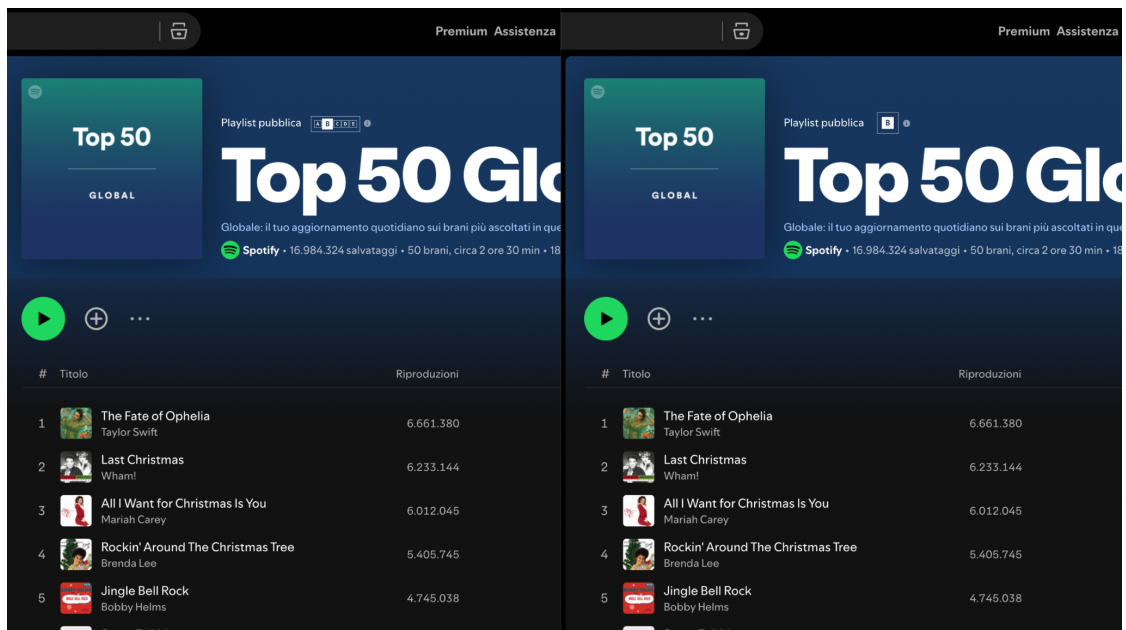


Figure 5. Visual identity of the Fairness Score (add-on element example)

3. Parametric wireframe: We propose a first wireframe for a possible implementation of the interface that allows end users to influence the MSSs' recommendations (see Appendix 4).



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

This parametric approach should reinforce human agency, promoting a more thoughtful and deliberate approach to music listening.

4.2 Generalization

This document and the first release of the Python package are scheduled as follows:

1. First phase: internal experimentation and prototyping. Currently, the Fairness Score is being prototyped and evaluated using data given by European consumers. Within this phase, we used mostly data from donors to ensure our independence and to have quality data.
2. Second phase: the score will be presented to open-source partners for use at a larger scale (e.g., ListenBrainz). This phase will allow us to collect feedback and make corrections to improve how the package works. We will benefit from their data/community to correct/improve our prototype and analyze the accuracy of our approach.
3. Third phase: the next score prototype will be presented to streaming platforms for implementation at a larger scale, but still as a pilot project. At that time, we should benefit from MSSs' data on a very large scale.

4.3 Evaluation and iteration

- Through our approach, algorithmic fairness can be evaluated directly by end users, further informing their perceptions of a platform's overall fairness policies, particularly its algorithmic content, such as playlists. We strive for a feedback process in which the efficacy of the Fairness Score is studied through controlled experiments involving MSSs or organizations that share Fair Muse's ethos, such as ListenBrainz, including qualitative and quantitative comparisons with other fairness toolboxes (see D4.3). This material may be used to create a survey of music industry professionals (to be defined), similar to Scarborough et al. (2007).



Fairness Score for Platforms

- To scale up and expedite the development of the Fairness Score, we recommend the creation of a scientific committee tasked with monitoring the impact of the score and the technical, economic, political, and legal developments within the field. This committee would be responsible for publishing new versions of the Fairness Score and its associated tools to maintain their effectiveness (Sarda et al., 2024).



5. Application of the Fairness-Score

Using outcomes from WP3, and WP4, the focus of this first version of the Fairness Score is on six Music Streaming Services:

- Amazon Music
- Apple Music
- SoundCloud
- Spotify
- TikTok
- YouTube Music

During the experimentation phase that follows the creation of this prototype, we will create detailed examples applying the Fairness Score for each of the six platforms mentioned above. The following section contains an example of the calculation of the Fairness Score applied to Spotify's platform options, GDPR export files, and one algorithmic sample playlist.

Application example: Spotify

The resulting values after applying the evaluation criteria described in Section 3 are detailed in Table 3:

Dimensions	Indicator	Points
Data practices	Data access	0
	Metadata standardization	-1
	Metadata completeness	-1
Power balance	Creator control	-1
	User data consent	-1



Fairness Score for Platforms

Dimensions	Indicator	Points
	User control	-1
Remuneration	Compensation transparency	-1
	Profit reinvestment	-1
	Copyright control	0
Representation	Track diversity	-1
	Gender diversity	-1
	Domestic ratio	0
	Auto-play diversity	N/A
Transparency	Verified artists ratio	0
	Explainability	-1
	AI-created music ratio	-1

Table 3. Example calculation of the Fairness Score on Spotify data.

Individual results

The example shown in Table 3 provides an explanatory summary of the nuances behind each indicator and its numerical value:

1. Data practices.

Files obtained under GDPR Article 15 contain structured data; they are linked through unique identifiers, although these are not widely used standards such as ISRC or ISNI, but rather Spotify's proprietary URIs. These identifiers are only accessible through and dependent on Spotify's own services. Metadata included in the platform and provided through data export files contains the bare minimum fields to identify a track and its rightsholders. However, it lacks complementary data on its composition, production, and performance.

2. Power balance.



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Fairness Score for Platforms

There are no explicit options enabling creators or user control over their data for algorithmic recommendations or training of AI-generated content.

3. Remuneration.

The platform does not explicitly disclose all the details of how its revenue is calculated; it uses a stream-share model and has introduced arbitrary barriers to revenue, such as the 1000-stream quota.

4. Representation.

Although this example uses a single algorithmic playlist, the same process may be generalized for all algorithmic and editorial playlists in MSSs. The details on the calculations are described below:

- Track diversity was calculated using an inverse Gini index on music genres. Out of 27 main genres, there were only 11 being represented, which gives a total of 0.14. To improve the relevance of this metric, in a further version, users may limit this calculation to the genres that are present on a playlist or exclude genres they do not listen to.
- Gender diversity was calculated using the base rate because we lack the full catalogue used for the recommendations. The disparate impact (ratio of female over male artists on the playlist) is 0.21. In the future, users may reassign the unprivileged/privileged groups.
- As before, auto-play diversity was not calculated because we lack the full catalogue that would create this playlist. In this prototype, N/A values default to a neutral level with 0 points.
- Since we do not have data linking the algorithmic playlist to a specific user, we calculated the domestic ratio for all countries on the playlist. We designated the



Fairness Score for Platforms

user's country of origin as the one with the highest domestic ratio, namely the United Kingdom. The domestic ratio is 0.59.

5. Transparency.

The platform does not disclose data regarding the amount of verified creators present on their database and lacks clear explanations on how most algorithmic playlists are created. There is no official data addressing the amount of AI-created music in the catalogue or included in recommendations.

Aggregated results

The resulting score according to the sum of all individual points is:

Total Score	-11
Letter	E
Meaning	Unaligned

According to the criteria evaluated by the Fairness Score, the platform is not aligned with the goal of creating a fair music ecosystem. The reason for this score is the amount of negative points, assigned when the conditions for the neutral level (0 points) are not attained. The clear path to increase this score is to work toward fairer protocols and practices across all evaluated dimensions.



6. Limits of the Fairness-Score

The deliberate adoption of the Nutri-Score as an analogous framework is insightful. The Nutri-Score's success (and criticisms) offers a useful lens through which to examine the challenges of operationalizing a complex ethical concept. The Nutri-Score aims to simplify nutritional information, translating complex scientific data into a readily understandable label. However, it is important to acknowledge that the Nutri-Score has not been without controversy. Criticisms have centered around its reductionist approach (over-simplifying complex nutritional profiles), its potential to be manipulated by food manufacturers, and its underlying assumptions about what constitutes a “healthy” diet (which are themselves culturally and socially contingent).

The same risks apply to the Fairness Score. The project's assumptions mirror a similar simplification process. Applying this logic to music streaming, it implies that certain attributes (e.g., artist compensation, genre diversity) inherently contribute to a “fair” music ecosystem. However, these attributes are themselves interconnected and influenced by a multitude of factors, including market forces, artist agency, listener preferences, and platform business models. Reducing this complexity to a single score risks obscuring the nuances of the system and potentially reinforcing existing biases.

In addition, by promoting an approach that reinforces subjectivity, we align with works in the social sciences (Tajouri & Tsoukiàs, 2024). However, we might be criticized by researchers who argue that predominantly objective metrics should be used to assess fairness. Currently, research on algorithmic fairness is focused on measuring and reducing bias through indicators (see D4.3) that model fairness as a binary classification problem or that optimize single attributes in isolation, although newer approaches are being explored (Ekstrand et al., 2022).



Fairness Score for Platforms

Likewise, we are aware of the risk of fair washing (Aivodji et al., 2019; Baule et al., 2023; Shahin Shamsabadi et al., 2022). This phenomenon may occur if our approach is misused to falsely portray fairness. Actors may optimize specific features to appear fair, even if they are not.

We have to acknowledge that this version of the Fairness Score is a first proposal and thus it is subject to changes and corrections originating from user and expert feedback, as well as trials within MSSs. We suggest the creation of an international scientific committee to decide needed changes or updates. Indeed, one difficulty with the evolution of the Fairness Score is technological change. This committee should be able to follow the rise of new technologies or approaches to be able to handle them within the Fairness Score.

Also, we need to acknowledge that the first version was built on a specific dataset, with intrinsic limitations (i.e., the number of donors) and it needs to scale up at the level of a global MSS. On the same level, the question of algorithmic performance will be an important aspect to ensure the implementation of the score by MSSs.



7. Conclusion

This deliverable describes the creation of a Fairness Score, an algorithm designed to evaluate the fairness of MSSs and their recommendations. Without the ability to access MSSs' data infrastructure directly, we collected data that would guide us toward a comprehensive understanding of how they work. Through our deliverables, we built an approach to model fairness and to evaluate it. As explained, the most challenging part is deciding what to measure to obtain results that are meaningful and easily reproducible across MSS.

This score uses nine fairness criteria across five dimensions, along with non-sensitive data (from the perspective of MSSs), to model it. Each element is described in detail, including specific instructions for calculating it. This document also presents guidelines for implementing, using, and applying the score.

Our approach is focused on a more subjective perspective on fairness to give deeper control to the end user/consumer. This allows us to consider multiple perspectives of fairness and thanks to a parametric approach, to give consumers bigger control and impact over recommendation systems.

In addition, to facilitate its assimilation and implementation, the first version of a Python package for the Fairness Score is included with this deliverable. It is under active development and will be proposed to MSSs for experimentation and eventual adoption.

From a policymaker's perspective, this score demonstrates what can be done. By using a technology building block, policymakers could ask actors like MSSs to implement it across the European Market, thereby embedding law and compliance directly into Information Systems. These actions aim to promote a Fair By Design technical infrastructure, where our package helps ensure a feedback loop on algorithmic fairness.



Appendix

1. Template for GDPR export file structure

Figure A1 presents a simplified structure of the listening history data for end users. Following the FAIR data principles, the data should be in a machine-readable format, well-structured, and linked using standardized, unique identifiers (see 3.4.1) that allow users to reconstruct relationships between fields.

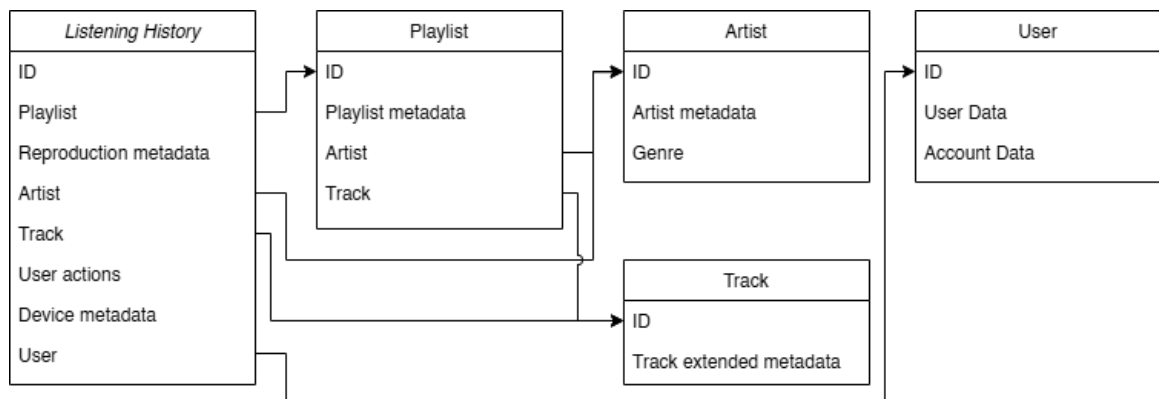


Figure A1. Suggested GDPR export file structure for MSSs.

2. Template for Credits Metadata Completeness

Table A1 was adapted from the Recording Academy’s “Minimum Data Requirements/Basic Credits List”²⁴ and the following DDEX standards: ERN (Electronic Release Notification), RIN (Recording Information Notification), and RDR (Rights Data Exchange).

²⁴ Available at the URL: https://naras.a.bigcontent.io/v1/static/credits_and_recording_metadata_09_27_2021



Fairness Score for Platforms

Credits list fields	DDEX Standard	Possible DDEX Element(s)
Artist Name	ERN	DisplayArtist/DisplayArtistName
Song Title	ERN / RIN	SoundRecording: DisplayTitle / MusicalWork: Title
Album Title	ERN	ReleaseTitle
Songwriters (Composer / Lyricist)	RDR	MusicalWorkContributor + ContributorRole: Composer / Lyricist
Producers	RIN	Contributor + ContributorRole: Producer
Recording Engineer	RIN	Contributor + ContributorRole: RecordingEngineer
Mixing Engineer	RIN	Contributor + ContributorRole: MixingEngineer
Mastering Engineer	RIN	Contributor + ContributorRole: MasteringEngineer
Performing Musicians	RIN	Contributor + ContributorRole: Performer
Other Contributors (arranger, conduction, art direction, etc.)	RIN	Contributor + appropriate ContributorRole or DisplayCredits
Presence of samples in the recording (True/False)	ERN / RIN	ContainsSamples
Recording Location (City / Country)	RIN	VenueAddress/TerritoryCode
Date of Recording	RIN	CreationDate / SessionDate
Original Recording Format	ERN	TechnicalDetails: SamplingRate, BitsPerSample, AnalogToDigital
Mix Recording Format	ERN	TechnicalDetails: AudioCodecType
Label / Media Company / Content Owner	ERN	LabelName / ReleaseParty / RightsController
Song Publisher	RDR	RightsController: Publisher
Instrument Played	RIN	InstrumentType (linked to Contributor)
Arranger / Conductor	RIN	ContributorRole: Arranger / Conductor
Recording Studio Name	RIN	VenueName
Song Publisher / Publishing Administrator	RDR	RightsAdministrator
Recording identifier (ISRC)	ERN	SoundRecordingId / ISRC
Songwriters' Performance Rights Organization	RDR	RightsHolder
Contributors Identifier (ISNI)	ERN / RIN / RDR	PartyId / ISNI



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088

Credits list fields	DDEX Standard	Possible DDEX Element(s)
Composition Identifier (ISWC)	RDR	MusicalWorkId / ISWC

Table A1. List of possible extended metadata fields within the DDEX standard.

3. Tuning the Fairness Score scale

Each interval in the final Fairness Score (Section 3.4.6) is obtained by using the function $F(s)$, which applies a min-max scaling on s followed by a power transform (using an arbitrary constant α), as expressed below:

$$F(s) = \left(\frac{s - s_{\min}}{s_{\max} - s_{\min}} \right)^{\alpha}$$

Where $F(s) \in [0,1]$ and $\alpha > 0$. This function introduces a non-linear component to the intervals defining each letter of the score while allowing equally spaced intervals for each level of the five-letter score (i.e., 0.0, 0.2, 0.4, 0.6, 0.8, 1.0). Figure A2 illustrates how, after scaling, the values of $\alpha < 1$ produce larger intervals for the highest levels, while $\alpha > 1$ produce larger intervals for the lowest levels.



Fairness Score for Platforms

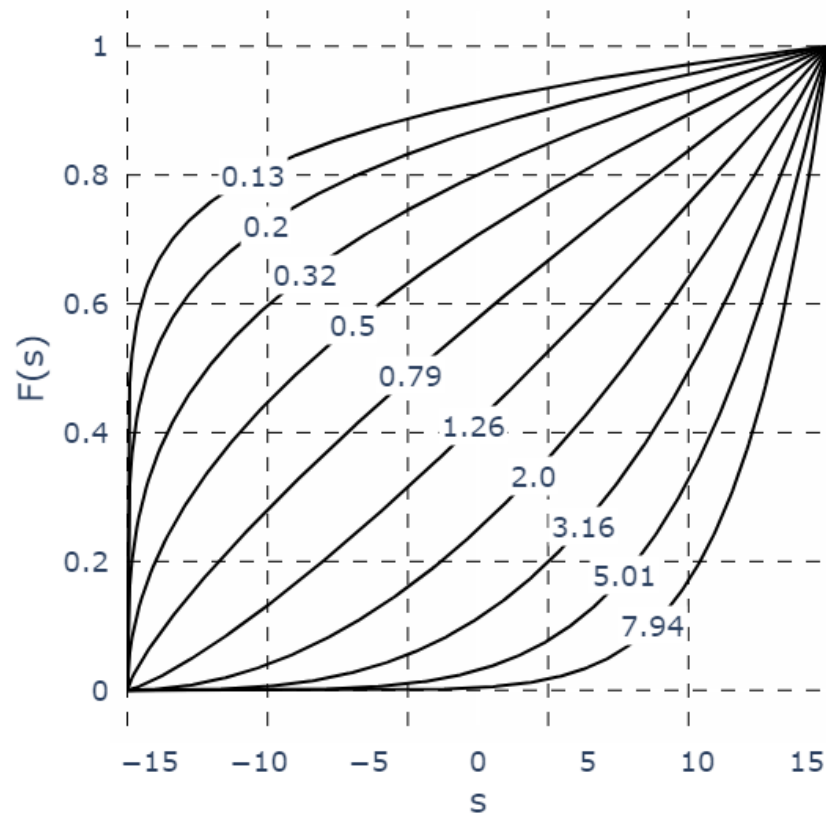


Figure A2. Normalized score $F(s)$ as a function of the Fairness Score s . Solid lines indicate changes to the five-letter Fairness Score scale for different values of α .

In the current prototype of the score, the value of $\alpha = 2$ was chosen because it raises the score needed to attain higher levels, encouraging MSSs to have a higher numerical score by complying with fairer practices for the music ecosystem. The resulting levels are shown in Figure A3.



Fairness Score for Platforms

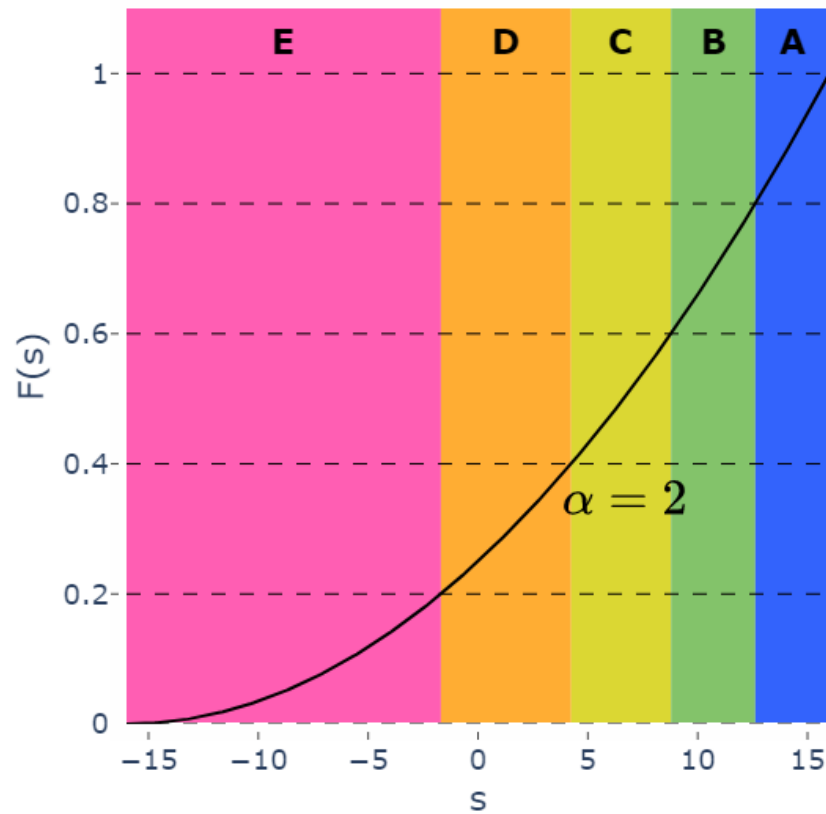


Figure A3. Chosen intervals for the Fairness Score. The solid line shows the scaling function $F(s)$ when $\alpha = 2$. Each of the five levels (shown as colored areas) of the Fairness Score is thus delimited by the intersection of the solid line and the dashed lines on the y-axis.

4. Parametric Wireframe Example

The following table provides a non-exhaustive list of parametric approaches that MSSs should adopt to broaden users' and creators' options and improve their overall experience. Average users would not need to interact with these parameters, but they would be available for those who want to regain control over their listening experience. Arbitrary values are given for numerical options; they do not represent suggested values.



Fairness Score for Platforms

Field	Description	Input type	Options
Organic listening	Use recommendation system features (suggested artists, playlists, AutoPlay, etc.). If turned off, access to the catalogue is maintained.	Toggle	On (default), Off
Collaborative recommendations	Use personal data (listening activity/collaborative filtering, actions) used to tune recommendation algorithms for you and other users.	Toggle	On (default), Off
AI features	Turn on generative AI features (AI playlists, prompts, training on personal data, etc.)	Toggle	On, Off (default)
Optimization parameter	Choose which parameter should be optimized when creating a recommended playlist or the next item in a queue.	Dropdown	Automatic (default), Accuracy, Novelty, Diversity, Custom (user defined)
Local content	Select a balance between recommendations from countries around the world and from your country of residence.	Slider	0 - 100 (Default: 50)
Diversity	Select a balance between recommendations from traditionally privileged groups and unprivileged groups.	Slider	0 - 100 (Default: 50)
Popularity	Select a balance between recommendations from independent, niche creators and well-known creators with an established audience.	Slider	0 - 100 (Default: 50)
Preferred languages	Select your preferred spoken language(s) for recommendations. If None is selected, the platform will infer this information.	Dropdown multiselect	None (default), English, French, Spanish, Arabic, etc.

Table A2. Wireframe of possible user-controlled recommendation algorithm parameters.



References

- Abul-Fottouh, D., Song, M. Y., & Gruz, A. (2020). Examining algorithmic biases in YouTube's recommendations of vaccine videos. *International Journal of Medical Informatics*, 140, 104175. <https://doi.org/10.1016/j.ijmedinf.2020.104175>
- Afchar, D., Melchiorre, A., Schedl, M., Hennequin, R., Epure, E., & Moussallam, M. (2022). Explainability in Music Recommender Systems. *AI Magazine*, 43(2), Article 2. <https://doi.org/10.1002/aaai.12056>
- Aivodji, U., Arai, H., Fortineau, O., Gambs, S., Hara, S., & Tapp, A. (2019). Fairwashing: The risk of rationalization. In K. Chaudhuri & R. Salakhutdinov (Eds), *Proceedings of the 36th international conference on machine learning* (Vol. 97, pp. 161–170). PMLR. <https://proceedings.mlr.press/v97/aivodji19a.html>
- Arambepola, C., Scarborough, P., & Rayner, M. (2008). Validating a nutrient profile model. *Public Health Nutrition*, 11(4), 371–378. <https://doi.org/10.1017/S1368980007000377>
- Baule, R., Münchhalfen, P., Shkel, D., & Tallau, C. (2023). Fair-washing in the market for structured retail products? Voluntary self-regulation versus government regulation. *Journal of Banking & Finance*, 148, 106749. <https://doi.org/10.1016/j.jbankfin.2022.106749>



Fairness Score for Platforms

Bedoya, D., Henry, A., Sørensen, J. K., Holm, R., & Hernandes, S. A. (2025). *Fairness in Music Information Systems: From a Process-Oriented Perspective to an Outcome-Oriented Approach* (SSRN Scholarly Paper No. 5290737). Social Science Research Network. <https://doi.org/10.2139/ssrn.5290737>

Beigang, F. (2023). Yet Another Impossibility Theorem in Algorithmic Fairness. *Minds and Machines*, 33(4), 715–735. <https://doi.org/10.1007/s11023-023-09645-x>

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilović, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, 63(4/5), 4:1-4:15. IBM Journal of Research and Development. <https://doi.org/10.1147/JRD.2019.2942287>

Boer, I. de, Vegt, F. de, Pluk, H., & Latijnhouwers, M. (2021). *Rubrics – a tool for feedback and assessment viewed from different perspectives: Enhancing learning and assessment quality*. Springer Nature.

Ekstrand, M. D., Das, A., Burke, R., & Diaz, F. (2022). Fairness in Recommender Systems. In F. Ricci, L. Rokach, & B. Shapira (Eds), *Recommender Systems Handbook* (pp. 679–707). Springer US. https://doi.org/10.1007/978-1-0716-2197-4_18



Fairness Score for Platforms

Epps-Darling, A., Cramer, H., & Bouyer, R. T. (2020). *Artist gender representation in music streaming*. Proceedings of the 21th International Society for Music Information Retrieval Conference, ISMIR 2020, Montreal, Canada, October 11-16, 2020. <http://archives.ismir.net/ismir2020/paper/000148.pdf>

Eriksson, M., Fleischer, R., Johansson, A., Snickars, P., & Vonderau, P. (2019). *Spotify Teardown: Inside the Black Box of Streaming Music*. The MIT Press. <https://doi.org/10.7551/mitpress/10932.001.0001>

Ferwerda, B., Ingesson, E., Berndl, M., & Schedl, M. (2023). I Don't Care How Popular You Are! Investigating Popularity Bias in Music Recommendations from a User's Perspective. *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, 357–361. <https://doi.org/10.1145/3576840.3578287>

Gunning, K., & Rowan, K. (1999). ABC of intensive care: Outcome data and scoring systems. *BMJ*, 319(7204), 241–244. <https://doi.org/10.1136/bmj.319.7204.241>

Hansen, C., Hansen, C., Maystre, L., Mehrotra, R., Brost, B., Tomasi, F., & Lalmas, M. (2020). Contextual and Sequential User Embeddings for Large-Scale Music Recommendation. *Fourteenth ACM Conference on Recommender Systems*, 53–62. <https://doi.org/10.1145/3383313.3412248>



Hardt, M., Price, E., & Srebro, N. (2016). Equality of Opportunity in Supervised Learning. *Advances in Neural Information Processing Systems*, 29. https://proceedings.neurips.cc/paper_files/paper/2016/hash/6a9659feb1216f14f7384ba499518b38-Abstract.html

Hellman, D. (2020). Measuring Algorithmic Fairness. *Virginia Law Review*, 106, 811–866.

Hoskin, K. (1996). The ‘awful idea of accountability’: Inscribing people into the measurement of objects. In R. Munro & J. Mouritsen (Eds), *Accountability: Power, ethos and the technologies of managing*. Cengage Learning Emea. <https://www.semanticscholar.org/paper/The-%E2%80%98awful-idea-of-accountability-%E2%80%99-inscribing-Hoskin-Vos/e0d35b72c3bc26ec3f01019678ea18ced037c805>

Hussein, M. A., Hassan, H., & Nassef, M. (2019). Automated language essay scoring systems: A literature review. *PeerJ. Computer Science*, 5, e208. <https://doi.org/10.7717/peerj-cs.208>

Julia, C., & Hercberg, S. (2017). Development of a new front-of-pack nutrition label in France: The five-colour Nutri-Score. *Public Health Panorama*, 3(4), 712–725.

Marqués, A. I., García, V., & Sánchez, J. S. (2013). A literature review on the application of evolutionary computing to credit scoring. *Journal of the Operational Research Society*, 64(9), 1384–1399. <https://doi.org/10.1057/jors.2012.145>

Mathison, S. (2004). *Encyclopedia of Evaluation* (1st edn). SAGE Publications.



Fairness Score for Platforms

Matrosova, K., Marey, L., Salha-Galvan, G., Louail, T., Bodini, O., & Moussallam, M. (2024). Do Recommender Systems Promote Local Music? A Reproducibility Study Using Music Streaming Data. *18th ACM Conference on Recommender Systems*, 148–157. <https://doi.org/10.1145/3640457.3688065>

Merz, B., Temme, E., Alexiou, H., Beulens, J. W. J., Buyken, A. E., Bohn, T., Ducrot, P., Falquet, M.-N., Solano, M. G., Haidar, H., Infanger, E., Kühnelt, C., Rodríguez-Artalejo, F., Sarda, B., Steenbergen, E., Vandevijvere, S., & Julia, C. (2024). Nutri-Score 2023 update. *Nature Food*, 5(2), 102–110. <https://doi.org/10.1038/s43016-024-00920-3>

Mitchell, S., Potash, E., Barocas, S., D’Amour, A., & Lum, K. (2021). Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application*, 8(1), 141–163. <https://doi.org/10.1146/annurev-statistics-042720-125902>

Muller, J. Z. (2018). *The Tyranny of Metrics*. Princeton University Press. <https://doi.org/10.23943/9781400889433>

N., K. A. R. A. K. A. Y. A. L. I., B., K. O. S. T. E. M., & I, G. (2018). Recommendation systems as technologies of the self. Algorithmic control and the formation of music taste ». *Theory, Culture and Society*, 35(2), 3–24. <https://doi.org/10.1177/0263276417722391>



Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39(11), 10059–10072. <https://doi.org/10.1016/j.eswa.2012.02.038>

Radha, S. K., Kuehlkamp, A., & Nabrzyski, J. (2025). Advancing Transparency and Responsibility in Machine Learning: The Critical Role of FAIR Principles - A Comprehensive Review. *ACM J. Responsib. Comput.*, 2(4), 20:1-20:38. <https://doi.org/10.1145/3768151>

Rather, A. M., Sastry, V. N., & Agarwal, A. (2017). Stock market prediction and Portfolio selection models: A survey. *OPSEARCH*, 54(3), 558–579. <https://doi.org/10.1007/s12597-016-0289-y>

Ricci, F., Rokach, L., & Shapira, B. (2022). Recommender Systems: Techniques, Applications, and Challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds), *Recommender Systems Handbook* (pp. 1–35). Springer US. https://doi.org/10.1007/978-1-0716-2197-4_1

Richardson, B., & Gilbert, J. E. (2021). *A Framework for Fairness: A Systematic Review of Existing Fair AI Solutions* (No. arXiv:2112.05700). arXiv. <https://doi.org/10.48550/arXiv.2112.05700>

Sarda, B., Kesse-Guyot, E., Deschamps, V., Ducrot, P., Galan, P., Hercberg, S., Deschasaux-Tanguy, M., Srour, B., Fezeu, L. K., Touvier, M., & Julia, C. (2024). Consistency of the Initial and Updated Version of the Nutri-Score with Food-Based Dietary Guidelines: A French



Fairness Score for Platforms

Perspective. *The Journal of Nutrition*, 154(3), 1027–1038.

<https://doi.org/10.1016/j.tjnut.2024.01.029>

Scarborough, P., Boxer, A., Rayner, M., & Stockley, L. (2007). Testing nutrient profile models using data from a survey of nutrition professionals. *Public Health Nutrition*, 10(4), 337–345.

<https://doi.org/10.1017/S1368980007666671>

Shahin Shamsabadi, A., Yaghini, M., Dullerud, N., Wyllie, S., Aïvodji, U., Alaagib, A., Gambis, S., & Papernot, N. (2022). Washing the unwashable: On the (im) possibility of fairwashing detection. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds), *Advances in neural information processing systems* (Vol. 35, pp. 14170–14182). Curran Associates, Inc.

https://proceedings.neurips.cc/paper_files/paper/2022/file/5b84864ff8474fd742c66f219b2eaac1-Paper-Conference.pdf

Snickars, P. (2017). More of the same – on Spotify radio. *Culture Unbound: Journal of Current Cultural Research*, 9(2), 184–211. <https://doi.org/10.3384/cu.2000.1525.1792184>

Watkins, E. A., & Chen, J. (2024). The four-fifths rule is not disparate impact: A woeful tale of epistemic trespassing in algorithmic fairness. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 764–775.

<https://doi.org/10.1145/3630106.3658938>



Fairness Score for Platforms

Yoo, R. M., Dash, D., Lu, J. H., Jenkins, J. Z., Rabbani, N., Fries, J. A., & Shah, N. H. (2023). Investigating Real-world Consequences of Biases in Commonly Used Clinical Calculators. *American Journal of Managed Care*, 29(1), e1. <https://doi.org/10.37765/ajmc.2023.89306>

Zhan, J., Wang, L., Gao, W., Li, H., Wang, C., Huang, Y., Li, Y., Yang, Z., Kang, G., Luo, C., Ye, H., Dai, S., & Zhang, Z. (2024). Evaluatology: The science and engineering of evaluation. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 4(1), 100162. <https://doi.org/10.1016/j.tbench.2024.100162>

Zhao, Z., Hong, L., Wei, L., Chen, J., Nath, A., Andrews, S., Kumthekar, A., Sathiamoorthy, M., Yi, X., & Chi, E. (2019). Recommending what video to watch next: A multitask ranking system. *Proceedings of the 13th ACM Conference on Recommender Systems*, 43–51. <https://doi.org/10.1145/3298689.3346997>

