

Revising research practices for singing data collection

Downloaded from: https://research.chalmers.se, 2026-11-10 03:06 UTC

Citation for the original published paper (version of record):
Cotton, K., Holzapfel, A., de Vries, K. et al (2026). Revising research practices for singing data collection. AI and Society, 41(10): 9617-9644. http://dx.doi.org/10.1007/s00146-026-03205-4

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

AI & SOCIETY (2026) 41:9617–9644
https://doi.org/10.1007/s00146-026-03205-4

OPEN FORUM

Revising research practices for singing data collection

Kelsey Cotton[1,2] · **André Holzapfel**[3] · **Katja de Vries**[4] · **Karl Berglund**[5] · **Kıvanç Tatar**[1,2]

Received: 8 December 2025 / Accepted: 21 June 2026 / Published online: 12 July 2026
© The Author(s) 2026

Abstract

As AI voice synthesis enables increasingly sophisticated vocal deepfakes and non-consensual voice cloning, the governance, licensing and access of singing datasets has become an urgent concern for data-contributors, who face significant harms from downstream and non-consensual usage of their singing data. Singing datasets are foundational to the development of high fidelity voice AI synthesis, yet current data collection practices pose challenges: data-contributors have an event-centric contribution to datasets which constrains the scope of the power they have in influencing dataset licensing and access decisions; and they face greater potential harms from non-consensual downstream use of their data. This is in contrast to data-collectors, whose involvement with the dataset directly concerns licensing and access decisions, and who do not risk the same harms as data-contributors. Current singing data collection practices suggest an imbalance in the power-to-interest stakes held by data-contributors relative to data-collectors. To investigate the power-to-interest differences between data-contributors and data-collectors, we apply the Ethically Aligned Stakeholder Elicitation (EASE) framework to three singing datasets, mapping the power-to-interest stakes of data-contributors and considering these stakes against their exposure to potential harms from downstream data usage. Our analysis consistently demonstrates that data-contributors occupy lower-power positions compared with data-collectors; they hold minimal decision-making authority over how their data may be accessed; and they face greater risk of harm from non-consensual downstream use of their data. A persistent temporal symmetry emerges: while data-contributors' involvement in datasets is event-centric, their vulnerability to harm is temporally unbounded, extending beyond the moment of contribution as new AI capabilities develop. Drawing from a cross-section of data ethics, decolonial data perspectives and artificial intelligence legislation, we propose revised research guidelines which reposition data-contributors more centrally in dataset governance and licensing decisions. We contribute with three consideration factors to better consider data-contributors' interests: stakeholder authority in dataset licensing, use and access decisions; the temporal scope of stakeholder involvement with the dataset; and vulnerability to harm from downstream data usage. Our findings highlight an urgent need for researchers to consider what licensing-related steps could be taken to protect the personality or identity rights of human participants within an AI climate in which singing data is increasingly treated as a freely available commodity for downstream use.

Keywords singing data · downstream risks · music information retrieval · voice AI · data

✉ Kelsey Cotton
kelsey@chalmers.se

André Holzapfel
holzap@kth.se

Katja de Vries
katja.devries@jur.uu.se

Karl Berglund
karl.berglund@littvet.uu.se

Kıvanç Tatar
tatar@chalmers.se

[1] Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden

[2] University of Gothenburg, Gothenburg, Sweden

[3] Division of Media Technology and Interaction Design, KTH Royal Institute of Technology, Stockholm, Sweden

[4] Department of Law, Professors, Instructors, Researchers, Uppsala University, Uppsala, Sweden

[5] Department of Literature and Rhetoric, Uppsala University, Uppsala, Sweden

1 Introduction

Ongoing discourse in academia (Adam 1995; Johnson and Verdicchio 2017; 2025; Hao 2021) and popular media (Johnson 2019; Gray 2022) has been increasingly concerned with the ethics of how we clone and generate human vocal likeness using AI models (Arik et al. 2017; 2018; Naik et al. 2022; Napolitano 2022). This is especially relevant to researchers engaged within the collection and dissemination of singing data, as these datasets have demonstrated a high propensity for downstream utilisation in the creation of deepfakes. Recent cases within the music industry include the unauthorised cloning of artists Drake and The Weekend in a fabricated collaboration made headlines in 2023 (Willman 2023; Coscarelli 2023; McEvoy 2023; Shanfeld 2023)- illustrating how such downstream utilisation is already occurring. YouTubers are also increasingly noting how their voice data is being scraped from their videos and used–non-consensually–for commercial purposes (Geerling 2024a, b; Norms 2024, 52 Card Laboratory 2024).

Aside from the commercial implications of vocal deepfakes, the formation of high-quality singingdatasets also has significant implications for human identity and personality rights. These datasets can be used to create deepfakes of data-contributors, which have been used in various settings that the data-contributors have not explicitly consented to (Altulahi et al. 2022; Hutiri et al. 2024). Choices made in the creation of datasets have real-world implications (Birhane et al. 2021; Mannie 2023; Cross 2022; Jeong and Lim 2019; Patel 2023; Rothman 2018) for data-contributors, whose voices have been subsequently used to train AI voice models for non-research-related usage such as advertising (Midler v. Ford Motor Co. 1988; McPherson 2003) and digital replicas (Olsen 2023; Tobin 2024).

These developments pose challenges for researchers collecting singing data, whereby current data collection practices may inadvertently enable such downstream usage. Our core research question examines how existing research practices around data collection and dissemination inadvertently enable downstream usage. This is especially pressing, given how Music Information Retrieval (MIR) systems depend upon, engage with and implicate "human practices, societies, and cultures" (Holzapfel et al. 2018). In relation to singing data, this is especially relevant as data is often from particular musical (Cuesta et al. 2019; 2020; Wilkins et al. 2018; Srinivasamurthy et al. 2021; Weiß et al. 2020; Rosenzweig et al. 2020), religious or sacred contexts (Proutskova et al. 2023). The increasing capacity to construct a synthetic likeness of a human voice necessitates revising the current research practices for how we collect data, licence and publish singing datasets.

Despite growing awareness of the data ethics challenges in MIR research (Holzapfel et al. 2018; Morreale et al. 2023; Lee et al. 2023), the power dynamics between data-contributors and data-collectors in singing datasets remain underexamined. Our motivations are to establish norms beneficial to research whilst avoiding negative impacts for data-contributors. We do not discount the critical importance of singing data for MIR research, rather we argue that these needs are served by a stance on data collection which positions data-contributors explicitly in the centre.

In many research-specific contexts, data-contributors face potential harms from downstream usage of their data–including non-consensual cloning and commercial applications–and where existing licensing and legislation frameworks do not yet provide adequate protections (Hutiri et al. 2024; Sharma et al. 2025; Pasquale and Sun 2024). In such contexts, data-contributors are disadvantaged because they have a lower power stake in "influenc[ing] the direction and decisions for a project" (Newcombe 2003). Central to our argument is the question of what licensing-related steps researchers can take to better protect the personality and identity rights of data-contributors, whose singing data risks being treated as a freely available commodity for downstream and non-consensual AI training (Cotton et al. 2024; Tatar et al. 2023).

To investigate how the power-to-interest stake of data-contributors correlates with their exposure to potential harms from downstream usage, we use the Ethically Aligned Stakeholder Elicitation (EASE) framework proposed by Kaila et al (2023). We selected this framework over alternatives because its stakeholder-centred approach directly addresses our concern with power and vulnerability, in contrast to existing frameworks such as FAIR (Wilkinson et al. 2016) which prioritise the interests of data users. We apply EASE to three singing datasets: VocalNotes, Saraga and Rembetiko–which span different historical approaches to data stewardship, licensing, and access. Across these three singing datasets we ask: which stakeholders hold decision-making authority over datasets, and which do not? Is their authority correlated with the potential harms they face from downstream usage of the voice dataset? How do temporal asymmetries—where data contribution is event-centric but vulnerability to harm is unbounded—affect different stakeholders' capacity to protect their interests? What wider considerations and changes need to be implemented to current practices to support continued research developments, whilst ensuring that communities and cultures are not inadvertently exploited? Applying EASE across these three datasets, we consistently found that data-contributors occupy lower power-to-interest positions relative to data-collectors regardless of the licensing and access mechanisms of the dataset (Longpre et al. 2024; Sharma et al 2025; Cotton et al. 2024; Hutiri et al. 2024). A key factor is a temporal asymmetry:

data-contributors' involvement with a dataset is event-centric yet their vulnerability to harm can extend far beyond the moment of their data collection as voice AI technology continues to develop.

Our contributions are as follows. We first present an adaptation of EASE for retrospective analysis of existing and published singing datasets. This adaptation further incorporates three consideration factors: legal authority over licensing and access decisions; the event-centric versus longitudinal nature of stakeholder involvement; and the vulnerability to potential downstream harms. When integrated with EASE, these consideration factors form a method for assessing power-to-interest dynamics in singing dataset contexts. Second, our application of the adapted EASE in three case studies indicates a pattern of data-contributors occupying lower power-to-interest positions than data-collectors. Third, we draw from data ethics, decolonial perspectives, and AI legislation to propose revised research guidelines for data-collectors which reposition data-contributors and their interests more prominently in dataset governance, licensing and access decisions. These revised guidelines are intended to maintain support for the continued development of MIR research.

Section 2 establishes contextual background and outlines pertinent related work. Section 3 presents our EASE conducted on three singing dataset case studies, with results reported in Section 4.followed by our proposed research guidelines in Section 5 proposes revised research guidelines, followed by broader discussion in Section 6.

2 Background and related work

In the following we establish the contextual foundation and define key terminology. We outline core data ethics considerations in Section 2.2. In Section 2.3 we discuss current legislation and data protections relevant to singing datasets. Section 2.4 outlines some approaches to voice licensing, from open-source approaches to commercial sector practices.

2.1 Key definitions

Within this paper, we use 'singing data' to refer to the recorded and digitised outputs of human-produced vocal material. In line with this paper's focus, we include phonatory gestures (Zhang 2016) and singing (Kelsey 1949) as types of vocal material we classify 'singing data'. As our paper is concerned with singing data collection, we intentionally constrain our discussion around data to refer explicitly to voiced outputs made by a human. Building upon Cotton et al. (2024), we distinguish between human created voice as an embodied vocal act occurring in physical space (Barthes 1978; Eidsheim 2015; Weidman 2014); the machine recorded voice that is captured through some form of recording (analog, digital, and with/without metadata) (Sterne 2003, 2022); and the machine synthesised voice. The singing datasets we examine in Section 3 are examples of machine recorded voice. The downstream harms that are central to this paper concern the transformation of machine recorded voice into machine synthesised voice, without the knowledge or consent of the data-contributor.

'Singing' refers to the act of producing phonated material–the wide range of sounds that can be made with the voice (Sundberg 1989, 1996)–within a system of organisation such as a cultural aesthetic framework which may inform pitched, rhythmic, harmonic and/or tonal structures (Kelsey 1949). As defined by Cotton and Tatar (2024), we further understand singing as encompassing both conventional singing across musical traditions (Blacking 1976; Turino 2008) and non-normative practices positioned within artistic and musical contexts. That is, phonated material such as vocal fry, keening or ululation are considered 'singing' where these sounds occur within a musical or cultural performance framework. Our use of 'singing' throughout this paper therefore reflects the vocal performances from our three case studies, which are from varying cultural and musical traditions. We further note that where the term 'voice' appears (in Section 2.3 and Section 2.4) it follows the terminology of the licensing and legislative literature, which engages with vocal data that extend beyond singing.

Throughout this paper, we refer to data-contributors and data-collectors as person(s), communities or institutions involved in dataset collection, curation and disseminations. We define data-contributors as individuals or communities who contribute their singing data for inclusion in singing datasets. Our usage of the term 'data-collectors' refers to researchers and institutions that are engaged with activities concerned with the recording, processing, annotation and management of singing datasets. We further establish our definition of a singing dataset as a corpus of singing recordings- encompassing audio recordings; visual representations (in the form of spectrograms and/or waveforms); textual transcriptions and annotations; and associated metadata (i.e. speaker demographics, recording conditions) organised in some form of structure for the purposes of research, technology development or archiving (Hillmann et al. 2008; Edwards et al. 2011).

We make reference in our paper to 'downstream usage' of data. In line with the scope of this paper, this refers to applications of singing data which diverge beyond the intended research purpose under which the data was originally collected (Longpre et al. 2024; Sharma et al. 2025). Some examples of downstream usage include commercial deployment of tools or software dependent on singing datasets; the training of AI voice models for singing

voice synthesis and cloning (Lee et al. 2023) or utilisation of singing data occurring without the knowledge or consent of the data contributor(s) of the dataset (Hutiri et al. 2024; Sharma et al. 2025; Pasquale and Sun 2024). We emphasise here the ethical challenges posed by both the temporal and contextual distance (Cotton et al. 2024) between the initial data collection activities, and instances where downstream usage occurs.

Although this paper's analysis is deployed on singing datasets, our findings (presented in Sect. 4) carry implications for more broad voice data collection (i.e., speech data sets for emotion recognition and synthesis (Baird and Parada-Cabaleiro 2019; Cowen et al. 2019), speaker diarization and recognition (Przybocki and Alvin 2001; Canavan et al. 1997). As singing data has the same biometric characteristics as other forms of voice data (Gonzalez-Rodriguez et al. 2008, Beranekt 2013 and Rusti et al. 2023), data-contributors for both singing and broader voice datasets are subject to similar downstream risks (Sharma et al. 2025; Gu et al. 2023; Bäckström 2025).

2.2 Data ethics

Amidst the surge of AI model development and deployment, there has been a similarly surging discourse concerning the ethics surrounding the data that is scraped, collected and used to train these AI models (King 2023; Leavy et al. 2021; Peña and Varon 2019; Ricaurte 2019). Notable papers are concerned with equitable access to data (Carroll et al. 2020); the visibility or invisibility (Hampton 2021; Newlands 2021) of stakeholders compromised by treatment of their own data (Cotton et al. 2024) and proposals concerning transparency measures for datasets (Gebru et al. 2021).

Carroll et al (2020) formulate a care-centric practice of data, building upon concerns related to data sovereignty and self-management within Indigenous communities across Oceania, the United States and Canada. Their care-centric principles invite an examination of collective benefit; authority to control; responsibility and ethics. In a similar vein, Gray and Witt (2021) address how feminist perspectives and feminist data ethics may be more assertively integrated into machine learning pipelines. They propose five interventions into the economy of machine learning, encompassing: 1) diversity with regards to representation and participation in the machine learning economy; 2) critique of positionality; 3) foregrounding human(s) throughout a machine learning pipeline; 4) ensuring the implementation of accountability and transparency measures; and 5) equitable distribution of responsibility.

A contrasting approach is offered by Gebru et al (2021), who propose the addition of technical documentation for datasets–"datasheets"– akin to existing protocols for electronics hardware. They argue that datasets would benefit from the application of technical datasheets to provide contextual information about how the data was collected; what protocols for securing the giving (and withdrawal) of consent were undertaken; as well as whether any analysis on the "the potential impact of the dataset and its use on data subjects" was conducted. They argue that posing a series of questions to dataset creators with the aim "to promote transparency" and to "flag potential issues" would improve the transparency and reliability of how datasets have been constructed.

Wilkinson et al (2016) have proposed the FAIR Data Principles, which aim to support data stakeholders to ensure the re-use of data. They posit that the data collection practices should extend beyond "collection, annotation and archiv[ing]" of digital assets, and further prioritise later discovery and re-use of data for "downstream investigations" (Wilkinson et al. 2016). They propose 4 core principles: Findable; Accessible; Interoperable and Reusable, and discuss how these principles are evident in some examples of research repositories (Dataverse, ISA, Open PHACTS and UniPort). Wilkinson et al further emphasise the importance of FAIR as it pertains to "computational stakeholders"—the computational agents and systems utilised by humans to "undertake discovery and integration tasks". They posit that ensuring "FAIRness" of digital objects "improve[s] knowledge discovery through assisting both humans, and their computational agents, in the discovery of, access to, and integration and analysis of, task-appropriate scientific data and other scholarly digital objects" (Wilkinson et al. 2016).

A further relevant framework is the Ethically Aligned Stakeholder Elicitation (EASE) developed by Kaila et al (2023). EASE was originally designed for musical AI applications, and provides a two-part process for first identifying stakeholders and then mapping their power-to-interest stakes. Unlike the other discussed frameworks in this section, EASE centralises the relationships between stakeholders and the power asymmetries between them. Table 1 is a summary of the relevant frameworks discussed in this paper. It should not be taken as an exhaustive list. Our full rationale for selecting EASE, and the adaptations we made in applying EASE to our case studies is discussed in full detail in Section 3.1.

2.2.1 Ethics of musical and singing-specific data collection

The broader data ethics frameworks discussed in Section 2.2 provide valuable contact for understanding extant ethics discussions within music research communities. In this regard, discussion of ethics within research communities engaging with musical data collection has seen some progress (Seeger 2003; Serra et al. 2013; Holzapfel et al. 2018; Morreale et al. 2023; Gotham et al. 2025; Ma et al. 2023; Cuthrell 2024; Birhane et al. 2021; Christodoulou and Jensenius 2024). The

Table 1 A summary of the frameworks discussed in this section with respects to the central concerns of this paper: power asymmetries between data, contributors and data-collectors; and data-contributors vulnerability to downstream harms

Framework	Primary purpose	Stakeholder focus	Addresses power asymmetries between data-contributors and data-collectors?	Addresses vulnerability to downstream harm?
FAIR (Wilkinson et al. 2016)	Data management: findability; accessibility; interoperability; reusability	Data users and computational agents	No	No
CARE (Carroll et al. 2020)	Indigenous data sovereignty; data self-determination	Indigenous communities and data-contributors	Partially: through authority to controller	Partially: within Indigenous contexts
Datasheets for Datasets (Gebru et al. 2021)	Dataset documentation and transparency	Dataset creators	No	Partially: Gebru et al 2021 refers to potential harms to data subjects
EASE (Kaila et al. 2023)	Stakeholder power-interest mapping;	All stakeholders, with particular focus on power distribution	Yes	Partially: our application of EASE extends this, to directly address vulnerability to downstream voice data usage
ISMIR Ethics Guidelines (Bauer et al 2020)	MIR specific ethical research practices	MIR researchers	No	Partially: addresses privacy and usage considerations but does not discuss formal protections
Holzapfel et al 2018	Ethical dimensions of MIR; proposes guidelines for engaging with musical communities	MIR researchers and musical communities	Partially: calls for decreasing the distance between researchers and practitioners	No: focused on research practices, not downstream usage of data
Morreale et al 2023	Critical analysis or Automatic Music Generation (AMG) data collection; consent and documentation guidelines	Data.collectors and data-contributors	Partially: identifies ethical engagement concerns within the field	Partially: emphasises consent and documentation, but does not address downstream harms

academic discourse has established valuable foundational guidelines for data-collectors engaged with musical data collection. However, the prevailing age of rapid and continually evolving generative AI development is such that we face new challenges in how we may approach the collection of musical data. This is especially relevant for singing dataset collection, as singing data carries biometric characteristics-which means that it can be used to identify individuals (Rusti et al. 2023; ; Nnabueze Ajimah and Iloanusi 2024; Örpek et al. 2024; Pathak et al. 2025; Leix Palumbo and Prey 2024; González-Rodríguez et al. 2008; Beranekt 2013).

Holzapfel et al (2018) identify the lack of lively discourse concerning a sustained engagement with data ethics, contending that the research conducted within MIR communities are "embedded into a sociotechnical system...continuously interact[ing] with human practices, societies, and cultures" and thus have a scientific and community-minded obligation to consider and actively develop "a long-term engagement with ethics...dialogue and discussion". They analyse common practices in MIR related to task-oriented development of algorithms, encompassing the choice of algorithm; evaluation methods; and choice of data corpus. They outline a number of ethical considerations related to case examples from MIR, proposing research guidelines and reflective questions researchers should ask themselves when conducting research within a musical tradition or practice.

Further progress with respect to data ethics discussions in MIR research communities is explored by Chen et al (2019) and Lee et al (2022). Their work examines the tensions between open data practices and ethical data governance within music research communities. Chen et al. address the evolution—and musical diversity—of dataset publication within the MIR community, focusing on diversity as it pertains to the breadth of musical genres reflected in datasets published within the MIR community. They acknowledge a clear bias towards Western popular music within their surveyed datasets, attributing this "due to lack of resources" (Chen et al. 2019). They propose the development of a distributed Music Information Retrieval Exchange (MIREX); new licencing models from companies with existing developer-centric licences to enable MIR practices; and greater involvement from the MIR community to engage with government bodies which "create the infrastructures that MIR researchers work under" (Chen et al. 2019).

Lee et al's (2022) work explores community and developer responses to Singing Voice Synthesis (SVS) systems. They investigate how the general public reacts to AI-generated singing voices; the implications of general public reactions for SVS developers; and the ethical issues of concern to the general public and SVS developers. They perform content analysis of comments on videos in which SVS is used or its applications are demonstrated (largely YouTube videos of singing voice systems demonstrated on Korean TV shows). Lee et al then conducted interviews with researchers and developers working with developing SVS technology. The interviews and viewer comments were then coded using an inductive thematic analysis to distill: 1) the core use scenarios of SVS; 2) attitudes towards SVS; 3) different understandings of "creativity" across researcher/developers and viewers; and 4) attitudes towards notions of ownership, consent and voice data rights. From their thematic analysis, they identified discrepancies between viewers' and developers' expected SVS-usage scenarios as well as similarities in how developers and viewers consider what constitutes "creativity" in the age of AI.

The International Society for Music Information Retrieval (ISMIR) has previously proposed "Ethics Guidelines in Music Information Retrieval" (Bauer et al. 2020), which are oriented towards defining ethical practices for researchers within the MIR field. The guidelines concentrate on publication practices; societal and environmental wellbeing; and transparent research methods. Although perspectives on data privacy, documentation and governance are addressed, these are formulated to prioritise technical development and scientific inquiry. Within the context of data collection, the guidelines do give suggestions regarding privacy and usage constraints. However, these do not engage with formal measures that actively ensure legal protections from non-permitted usage (such as those provided by a data licence).

Morreale et al (2023) have similarly conducted a critical analysis of data collection practices for Automatic Music Generation (AMG) tasks published in the ISMIR conference across a 10-year period (2013-2022). They examined 121 papers, identifying how the datasets were constructed; the content of the data itself (performances, arrangements); and the nature of involvement of the musicians featured in the dataset and the level of knowledge the musicians appear to have had regarding the purposes of the collection of data. Their analysis indicated a "lack of engagement with ethical issues", noting "Only four datasets included any ethical considerations, and only two of them contained an explicit ethics statement" (Morreale et al. 2023). Their work, and findings regarding prevailing attitudes and practices towards voice data collection are directly relevant to this paper's concern: *how are data-contributors' interests weighted against the technical development priorities of data-collectors?* Our recommendations extend upon Morreale et al's work to address singing-specific challenges, including consideration for personality rights and collective singing traditions.

2.2.2 Decolonical perspectives on data

Whilst the frameworks discussed in the previous section emphasise the epistemological importance of data being accessible and reusable, decolonial perspectives challenge this idea. This opposition is especially relevant for singing

datasets, as such frameworks may inadvertently rely on extractive processes wherein singing data from vulnerable, marginalised or underrepresented communities is collected, shared and published. Decolonial scholars such as Ricaurte discuss what they describe as a "regime of data colonialism", which in turn begets "new forms of colonization through data, grounded in material infrastructures and symbolic constructions that reinforce these practices" (Ricaurte 2019, Ricaurte argues that data colonialism reflects a data-centric epistemology wherein "1) data reflects reality, (2) data analysis generates the most valuable and accurate knowledge, and (3) the results of data processing can be used to make better decisions about the world." Such epistemology is considered by Indigenous scholars as reflective of Western and European epistemological dominance (Ricaurte 2019; Varon and Peña 2021; Muldoon and Wu 2023), where data are framed as property-ownable, transferable, and commodifiable (Duarte 2020).

This establishes an important counterpoint that is relevant for this paper's focus: data-contributors and data-collectors may have legitimate reasons to not make singing data findable, accessible, interoperable or reusable. Singing data from particular cultures may embody collective knowledges, sacred practices, or community-specific vocal techniques. Ethnomusicological scholarship has historically noted how singing practices are integral to and embedded within cultural knowledge systems (Feld 2012; Turino 2008). Further, as singing data carries biometric characteristics that may render the data-contributors identifiable, this poses significant concerns about the potential harms of downstream usage of this data (Stoever 2016; Eidsheim et al. 2019).

Although frameworks such as FAIR prioritise engineering and research development, these may not prove *fair* to data-contributors who bear the brunt of the potential downstream harms (Benjamin 2019; Noble 2018). Decolonial perspectives remind us that some communities may rightfully choose to restrict access to their data as an act of sovereignty, cultural preservation, and protection against extractive practices (Christen et al. 2018; Carroll et al. 2020). The tension between these perspectives underscores the need for dataset collection frameworks that respect both the potential value of data sharing and the potential harms to communities whose data is discoverable, accessible, and reusable in downstream contexts.

2.3 AI laws relevant to singing data

Voice databases and the resulting trained voice models are legally protected in a large variety of manners. As these may differ between different jurisdictions, we focus on European law. This is due to the European context being where our legal expertise is in, and the region within which we as authors are based. We note that future work outside of this context would prove immensely valuable, and especially where legal approaches differ.

Voice data and models can trigger the protection of different legal instruments that differ in their subject matter (what is the general concern of a law?) and material scope (which facts trigger applicability of the law?). On an EU level the three legislative instruments that are most likely to be of relevance are the General Data Protection Regulation (EU) (2016) (GDPR 2016/679); the Directive on Copyright in the Digital Single Market (CDSM 2019/790), notably Articles 3 and 4 on the Text and Data Mining (TDM) exception which in some cases might apply to AI training models; and the product safety rules with regard to AI systems found in the European AI Act (AI Act 2024/1689).

Regarding GDPR, if voice data can be tied to an identified or identifiable person they qualify as personal data, and a large set of data protection requirements (data minimisation, pseudonymisation, need for a legitimate legal ground underlying the data processing, etc.) starts applying. In cases where datasets feature named vocalists such as in the named contributors of the Saraga (discussed in Section 3.5) and Rembetiko dataset (discussed in Section 3.6), this means that recordings will fall within the material scope of the GDPR. Even when vocalists are not named but voices are recognizable as belonging to identifiable vocalists, the GDPR also applies. One consequence of the application of the GDPR is that the data processing needs to be based on a legal ground from Article 6(1), such as consent, which can always be withdrawn, or a legitimate interest that outweighs the legitimate interest of the data subject against processing. There is a research exemption in Article 89, which prevents the data subject from withdrawing consent if that would hinder the research, but the data controller still needs to comply with pre-emptive and structural data protection measures such as data minimization, as outlined in Article 5(1)(c), which stipulates that data be limited to what is necessary for carrying out the stated purpose. In our MIR context, these recordings would fall under the research exemption, and there is a lawful basis for their collection and analysis. However, the interpretation of the data minimisation principle in this context is problematic, because it is hard to assess ***how much*** singing data (both per data-contributor and across the whole dataset) is sufficient for a minimised collection that is usable within research contexts.

A similarity between data protection and copyright protection is a focus on consent and control. A difference is the subject matter: unlike data protection copyright does not focus on personal data and how unlawful, unfair and opaque data processing could limit the individual in the identity related exercise of fundamental rights. Instead it is focused on the moral and economic protection of original works that are an author's own intellectual creation. In practice there is often a strong focus on the question of remuneration. If voice

training data can be classified as copyright-protected original works and their use in the training process is qualified as copyright relevant use, depends what the voice expresses: for example, an original song created by the singer is a protected work, but there is no copyright protection for recordings of non-original content such as syllables, numbers or facts. If a voice recording is a protected work it can only be used if a copyright exception applies (for example, fair use in the US or the TDM exception in the EU) or if there is a licensing agreement between the copyright holder and the third party that permits certain uses for a period of time. Often such an agreement will involve some form of remuneration (royalties). Because entering individual licensing agreements is likely to be a costly and time-consuming procedure there is an increased interest in collective licensing possibilities, where a collective rights management organization collects a fee for a whole database and distributes it to all authors whose work is included based on an opt-in or opt-out mechanism (Axhamn 2025; Creamer 2025; U.S. Copyright Office 2025). While the main focus of the licensing debate is on copyright protected works - because the term of protection granted by copyright is typically longer - very similar issues emerge in relation to performances protected under so-called neighboring rights benefitting performers, producers and record labels. If the data can be classified as a musical performance, then performance rights apply.

Copyright protection in relation to AI models is currently a highly contested topic. One example of the ongoing debate is the EU parliamentary resolution on copyright and generative AI (Voss 2026). The resolution was adopted in March 2026, and establishes the EU Parliament's position on how EU copyright should adapt in response to generative AI. It outlines the need for AI developer transparency (see also below, our discussion of the AI Act) when it comes to the use of copyrighted material in training; collective licensing frameworks; standardising opt-out mechanisms; and compensation for past use. Within the scope of this paper, the Voss resolution's focus on copyrighted and related rights framework protected material poses a limitation. Some forms of singing data–especially sacred, collective heritage, traditional or culturally significant material–are not necessarily protected under copyright or related rights. Although the resolution has been met with criticism (Geiger et al. (Geiger 2026)), it does provide important pointers for further legislative reform.

Whereas data protection focuses on personal data, and intellectual property protection on protected original works or performances, the new AI Act 2024/1689 has the product safety of AI systems and models as its focus. Within our context two types of transparency obligations are relevant. First, the labelling obligation for deepfakes in Article 50 (the end user has to be informed when content has been generated by AI). This transparency obligation is oriented towards the enduser, and has no impact on the performer. A potentially more useful transparency obligation can be found in Article 53 which relates to so-called general-purpose AI models, that is: large generative AI models that allow for "flexible generation of content, such as in the form of text, audio, images or video, that can readily accommodate a wide range of distinctive tasks" (Recital 99), which requires a certain level of transparency about the training data. The transparency about the training data can be an important tool in giving bite to the aforementioned data protection and intellectual property rights and prevent disputes from getting stuck in evidentiary debates. However, the Article 53 transparency requirements do not apply if the model does not qualify as a GPAI model because it is small, that is, trained on less than 10^{23} floating-point operations, and only applies to specific tasks, or qualifies as an open-source GPAI. Moreover, the AI Act does also not apply "to any research, testing or development activity regarding AI systems or AI models" (Articles 2(6) and 2(8)). Flexibility is thus afforded in a context where we are already lacking precise approaches for how this data might be subsequently licensed, potentially adding ambiguity as to best-practices. As soon as research is commercialised and put on the market or into service, the AI Act will apply. This is relevant for research outputs that move into industry or commercial production (Cunningham and Lee 2013).

2.4 Licensing agreements

While data protection can be an empowering way to stop or allow certain uses of one's voice, potentially even resulting in administrative fines, it is only through intellectual property and contract law that a performer or author has the possibility to get some remuneration. Some licensing models draw inspiration from the open source movement (see Section 2.4.1), while others are focused on fair remuneration in a commercial context (see Section 2.4.2). The applicability of both approaches to singing data-contributors varies considerably and is dependent on whether the recorded singing constitutes a protectable original work (see Section 2.3). This is especially complex to determine for singing data from collective, sacred or culturally embedded traditions (see Section 2.2.2).

Both approaches focus mainly on training data that can be qualified as original works or performances, because voices as such are not protected. As mentioned above, the preferred legal vehicle is licencing, acting as an umbrella term for different contractual agreements. Even in cases where there is uncertainty about the exact legal protection of voice recordings, a licencing agreement gives de facto access to a database behind a paywall. Apart from collecting fees for training a model, the use of a service based on that model can also generate a source of income for voice performers.

Some organisations (STIM 2025) even explore if it would be possible to collect royalties over AI-generated output. However, attribution technologies mostly function on a superficial level of similarity measures instead of establishing a solid pedigree relation between training data and output.

2.4.1 Licensing agreements drawing inspiration from the open source movement and beyond

Regarding data licensing, some notable approaches are observed in previous research (Benjamin et al. 2019; Benhamou and de Rosnay 2023; Chen et al. 2023). Benjamin et al (2019) propose a taxonomy for data licensing equivalent to open source software licensing. The resultant *Montreal Data Licence* establishes definitions; descriptions of data usage with respect to the data in isolation and its utilisation with models. Benhamou and de Rosnay (2023) challenge the current terminology and philosophy underpinning how contractual agreements and rights subject data to different protections. They propose a re-definition of common terms, seeking to address "power imbalance[s] and information asymmetry". They model their term re-definitions in line with "the open data philosophy initially developed for Free and Open Source software and Creative Commons licences for artistic and other copyrighted works". They further suggest how these standardised terms can be used within collective governance initiatives, and present a template for their Open Data Commons Licence (ODCL). Chen et al (2023) investigate licensing of publicly available visual (image or video) datasets on GitHub. Although focused primarily on image datasets, their discovery of licensing violations and ambiguities of non-commercial datasets in commercial systems are applicable to audio datasets. A more recent approach is the CC signal project, which utilises a series of iconography to "help content stewards express how they want their works used in AI training" (Creative Commons 2025). This is of particular relevance for singing datasets published under standard CC licences–such as Saraga–where the licence terms do not explicitly address model training of AI voice synthesis. Additional progress with regards to improving transparency through labelling AI-generated audio–especially AI-generated voice–is evident in Transparent Audio's development of technical solutions for ensuring compliance with global AI content regulations (Transparent Audio 2025).

2.4.2 Voice licencing agreements in commercial sectors

Insights into how voice licensing agreements in audio industries are formulated remain difficult to obtain, as these are often confidential documents between parties, with no systematic studies mapping such agreements (Ding et al. 2024). Companies that sell AI voice technologies are generally non-transparent about their training data, yet openly market their products as replacements for human voice talent. Speechify, for instance, formulates this in a typically blunt way: "Skip the hassle of hiring voice talent and long production times with Speechify AI Voice Generator. Streamline your voice over production with instant access to over 1,000 humanlike AI voices across 60+ languages, featuring a range of dialects and accents." (Speechify 2025). From an industry and commercial perspective, the commodity is understood not as voices or voice datasets, but as the trained speech synthesis technology itself (Dale 2022). The voice data seems to be considered available, as just 'there'. This implies that the training has already been done, meaning that there are sufficient voice data resources available for these companies to train their models, epitomising a potential instance of the downstream usage we discussed in 2.1.

In recent years, voice actors and their guilds have become increasingly aware of unregulated use of human voices for training AI models, leading to heated debates and efforts to establish just licensing frameworks. In creative industries, various efforts have been made to provide voice actors with just licensing agreements in relation to AI model training, such as the proposed standard agreement between the Screen Actors Guild-American Federation of Television and Radio Artists (SAG-AFTRA) and Replica Media (SAG-AFTRA and Replica Studios 2024). The SAG-AFTRA and Replica Agreement establishes a contract template for voice artists to explore employment opportunities to create and license a digital replica of their voice. The Agreement establishes clear terms of remuneration; clear terms and conditions for the usage of vocal replicas in new applications; and clear terms of external usage of resultant trained voice models. Notably, there are also provisions in the Agreement document concerning the use of pre-existing recordings of voice artists (living or deceased) to train a voice model, which are similarly helpful for data-collectors involved with the collection and curation of historical singing data.

For individual voice actors and vocalists, judging whether legal documents and complex licensing agreements are fair in an AI model training context remains extremely difficult. Debates in voice actor forums (such as "Svenska Rösters Öppna forum") suggest this must be negotiated case-by-case, which is problematic for voice actors often working under precarious conditions (Ganguzza 2021). When individuals depend on voice acting for their livelihood, declining work due to problematic AI training clauses becomes difficult, particularly when potential harms from these models are typically long-term and diffused rather than immediate and concrete. While various services and guidance have emerged to support individual voice actors in navigating these agreements (Ganguzza 2021), placing responsibility for fair licensing on individuals working in precarious conditions is a significant burden. We argue that these burdens and

concerns as equally relevant to singing data-contributors. The temporal asymmetry is particularly acute—the nature of data-contributors' involvement in contributing singing data for a dataset may be event-centric, but their vulnerability to harm from downstream usage is temporally unbounded, extending far beyond the initial moment of data contribution.

3 EASE applied to case studies

To conduct a critical survey of our case studies, we applied the Ethically Aligned Stakeholder Elicitation (EASE), developed by Kaila et al (2023) to assist in pragmatically identifying "power relations and value tensions" within creative AI-tools.[1] EASE utilises a 2-part process to first identify stakeholders (using a series of targeted questions[2]), and then to establish their respective power-to-interest stake. Kaila et al (2023) provide guidance on how decision-making authority (power) and investment (interest) may be determined, acknowledging the potential subjectivity of this process. EASE was initially developed to be utilised in a workshop format as an activity for project teams that actively develop musical-AI applications. We followed the two-part protocol by Kaila et al (2023)–stakeholder identification and power-to-interest positioning–with our adaptations for retrospective dataset analysis (outlined in Section 3.3).

3.1 Rationale for selecting EASE

We have utilised EASE rather than alternative ethical analytical frameworks for three reasons, which correlate with our research aims.

Firstly, EASE's emphasis on visualising the power-interest stake of involved stakeholders directly addresses our central research question: identifying which stakeholders involved in singing datasets are most vulnerable to harm from downstream usage. This stakeholder-centred approach differs substantially from other ethics analytical approaches discussed in Section 2.2 and summarised in Table 1: FAIR prioritises the interests of data *users*, and does not account for other stakeholders that may be adversely affected by the accessibility and reusability of data; CARE principles "shift[s] the focus from regulated consultation to value-based relationships" with a particular consideration for Indigenous worldviews, peoples, and cultures. Our implementation of EASE expands upon CARE, by foregrounding such relationships as requiring ongoing care and maintenance in light of the evolving capabilities of downstream singing models.

Secondly, EASE was originally developed as a workshop activity for project teams involved in musical AI development. Its provenance in musical AI contexts makes it directly relevant to our context, as the stakeholder elicitation questions formulated by Kaila et al (2023) are formulated with a view towards the applications which depend on high quality musical data. Further, the power-interest mapping proposed by Kaila et al is relevant to identifying the asymmetries between data-collectors and data-contributors when it comes to the governance and access decisions around musical datasets.

Thirdly, EASE is adaptable for retrospective analysis and can be applied to evaluate the power-interest stake of stakeholders of existing and already published datasets. This adaptation is discussed in greater depth in Section 3.3.

3.2 Selection of datasets for EASE

The selection of datasets for our EASE was motivated by several considerations. First, we sought to examine datasets featuring vocal practices that were not reflective of the mainstream MIR research communities' focus on Western popular music (Chen et al. 2019). Second, our selection of the datasets for EASE prioritised examples that demonstrate different approaches to data stewardship—allowing us to assess EASE across datasets with varying ethical commitments and access models. We sought datasets where there was some sort of evidence (by way of licences, access forms) of consideration for the data-contributor's interests. These two criteria led us to select three datasets that exemplify different historical moments and approaches to data stewardship:

Together, these three datasets illustrate different historical moments and approaches to singing data stewardship: from contemporary access-controlled sacred recordings (VocalNotes), to permission-based art music recordings with explicit research licences (Saraga), to a dataset reflecting earlier MIR conventions where licensing and permissions were not systematically addressed (Rembetiko). Each dataset features vocal practices underrepresented in mainstream MIR research and demonstrates different relationships to the ethical concerns we outlined in Sect. 2.2.2.

3.3 Adaptation of EASE for retrospective dataset analysis

As previously discussed, our application of EASE differs from its original context, as we apply it to the evaluation of datasets containing singing data in order to establish the power-to-interest stake of implicated stakeholders (Phase 2 of EASE). Our adaptation of EASE is thus centred on the

[1] We use the same definitions for power and interest that are established in Kaila, Jääskeläinen, and Holzapfel 2023.

[2] Please see Kaila, Jääskeläinen, and Holzapfel 2023 for the full question protocol.

identification and relationship-mapping of implicated stakeholders on the basis of their power and/or interest in downstream usage of the 'completed', or published singing datasets. As opposed to the original use-context of EASE (which is ideally conducted prior to commencing the development of musical-AI projects), we operate within a particular temporal context where it is no longer possible to intervene in *how* these datasets came to be. We therefore acknowledge the limitations of our utilisation of EASE with respect to already-published or already-accessible singing data. However, we emphasise that such identification of power-interest stakes in already-published singing datasets may elicit valuable insights as to how data-collectors may implement our suggestions (see Sect. 4) in the present and future care-taking of the data corpus. Our argument here is for developing more sustainable and data-contributor-centred approaches to dataset collection, aligned with extant positions on data maintainability (Wilkinson et al. 2016; Chen et al. 2019).

Accordingly, we approached the positioning of stakeholders along the power-interest axes by considering:

1. The stakeholders' formal and/or legal authority to decide how the dataset is licensed, accessed or used.
2. The nature of the stakeholders' contribution/connection to the dataset (are they involved in an event-centric activity like contributing data, or is their contribution/connection of a longer duration like maintaining access to the dataset)
3. The stakeholders' level of vulnerability to potential harms emergent from downstream usage.

The first factor—legal authority over licencing, access and usage decisions—is drawn from data protection and legislative frameworks discussed in Sect. 2.3, where legal authority to determine how data is governed determines a stakeholder's capacity for influence (their power and interest).

The second factor—the temporal scope of stakeholder involvement—emerges from the gap we identified in existing data ethics frameworks whereby current approaches (specifically the ones discussed in the paper and presented in Table 1 in Sect. 2.3) are concerned with the moment of data collection rather than engaging with more longitudinal concerns.

The third factor—vulnerability to potential downstream harm—is derived from the particular biometric properties of singing data (discussed in Sect. 2.2.1) and the downstream risks identified in the literature (Rusti et al. 2023; Hutiri et al. 2024; and Sharma et al. 2025).

Following the stakeholder identification protocol outlined by Kaila et al (2023), we generated the stakeholder list for each dataset using the EASE guided question set. These questions prompt consideration for the stakeholders involved in and with the creation, use, governance of the dataset. It further aids in identifying stakeholders involved in or affected by the dataset, encompassing those directly involved with data collection and those who become affected by the existence or availability of the dataset.

In our EASE Figures, power is positioned on the x-axis and interest on the y-axis. Stakeholders with high-power and high-interest typically have low vulnerability to downstream harms, while stakeholders with low-power and low-interest face the highest vulnerability to downstream harms. We rank stakeholders based on their *relative vulnerability*, comparing who is more or less vulnerable to harms such as unauthorised voice cloning, commercial exploitation, or decontextualisation of sacred and culturally specific vocal practices.

Each stakeholder is placed on the power-interest axes one-at-a-time, so that we can evaluate the placement of each stakeholder in relation to each other (i.e., does stakeholder A have more or less power in comparison with stakeholder B when we consider our 3 consideration factors). This comparative and relational placement is continued as more stakeholders are added to the power-interest axes. This is so that all identified stakeholders are considered in relation to each other's power and interest positioning. This whole process is qualitative, comparative and does not implement any ordinal scaling or other form of mathematical meaning. In instances of inter-author disagreement on stakeholder placement, we conducted a second round of comparing between stakeholders' power and interest by again considering our three consideration factors. This process was continued until consensus had been achieved with the revised placements. All five co-authors participated in both the stakeholder identification and placement process for each dataset. The number of rounds required for the placements varied by dataset, reflecting the complexity of each case: VocalNotes required 4 rounds, while Saraga and Rembetiko each required 3 rounds.

Our three consideration factors form an additional methodological contribution. Existing frameworks and philosophical positions concerning data ethics are largely concentrated on the data-collectors, and focus predominantly on what we may describe as the initial stages of dataset development: establishing a protocol for data collection; seeking consent; collecting the data; establishing the licencing mechanism for the dataset; and disclosing details of their development process in a research publication. There is little to no discussion on the maintenance of the dataset, nor on the continuation of the researchers' responsibilities and obligations towards the data-contributors who may be adversely impacted by downstream usage of the data. Such a singular focus on one party—the researcher(s) and data-collectors—and the initial dataset development as the sole act that occurs and is affected by data ethics actively invisibilises (Hampton 2021) the data-contributors. Although the temporal context

of a data-contributors' involvement with a dataset development process is more event-centric, they are undoubtedly affected in a more temporally prolonged way.

Here, our implementation of EASE seeks to address the tensions affecting data-contributors, by introducing additional dimensions that data-collectors should consider with respect to the potential consequences of downstream usage of data. This builds upon arguments expressed in Holzapfel et al.'s paper on the importance of "decreas[ing] the separation between the research and the music practitioners" and more collaborative approaches to minimise "the remoteness of MIR from the … related people, practices and culture" (Holzapfel et al. 2018). We argue that the decrease of such separation and the minimisation of such remoteness should not solely occur at the initial dataset development stage, but should be considered as an ongoing task in dataset management.

We further provide additional information about the dataset licencing, access mechanisms, stated purpose of the dataset (as defined by the data-collectors in their publications), additional information about consent processes conducted during dataset collection and curation (where applicable), as well as author-identified potential downstream risks. This information is provided in Tables 1, 2 and 3 for our three case studies. The column structure of these tables is informed by the benchmark questions proposed by Gebru et al (2021) for dataset documentation. Their inclusion in this paper is to inform the contextual understanding around each dataset. The placement of stakeholders in each case-study is driven by our deployment of EASE as informed by our three consideration factors, not solely from these additional evidence tables. We note that conducting EASE solely from available dataset documentation as an area of future research, but that this is contingent on widespread adoption of richer documentation practices as proposed by Gebru et al (2021).

3.4 VocalNotes

The VocalNotes dataset is a corpus of voice performances across five musical traditions, annotated by experts in the musical traditions (Proutskova et al. 2023). Proutskova et al have made the dataset available under the Creative Commons Attribution Non-Commercial ShareAlike 4.0 (CC-BY-NC-SA 4.0) licence. Access to the dataset is possible by completing an online request form, with a clear stipulation that the "audio will only be released for research purposes".[3] As discussed in (Proutskova et al. 2023), the formal access request form is motivated by the dataset featuring recordings of religious, sacred and ritual music.

[3] Accessible via this hyperlink.

Table 2 Evidence table for the VocalNotes dataset, detailing licensing type, access mechanism, stated purpose of the dataset, any notes pertaining to how data-contributors were asked for consent, and the identification of potential downstream risks as proposed by the authors

License type	Access mechanism	Dataset content and stated purpose	Data-contributor consent notes	Potential downstream risks (authors' assessment)
CC_BY-NC-SA 4.0 CC-BY-NC-SA license restricts commercial use, but does not address AI model training or voice synthesis permissions Non-commercial synthesis applications are permissible under the current license terms- this represents a license gap and underscore the insufficiency of existing CC licensing to address voice synthesis risks	Gated request form Audio released only following author approval of data-requesters declared intended use Annotations available via Zenodo	The dataset contains sacred and ritual material from living communities Non-commercial MIR research Pitch and note annotation of vocal performances from five musical traditions	Not documented in dataset metadata Access form discusses researcher obligations only Dataset paper requests users treat the recordings of performers "with respect" No documented withdrawal mechanism	Performers' voices could be used to train synthesis models within the terms of the existing licence Remixing: taking samples of the voices out of context and using them outside of the music traditions' original sacred and ritual context Metadata "misuse": taking pitch annotations to generate the sound of an instrument that has very different connotations in that culture than the voice

Table columns are informed by the datasheet benchmark questions proposed by Gebru et al (2021), adapted to the context of singing data

Table 3 Evidence table for the Saraga dataset, detailing licensing type, access mechanism, stated purpose of the dataset, any notes pertaining to how data-contributors were asked for consent, and the identification of potential downstream risks as proposed by the authors

License type	Access mechanism	Dataset content and stated purpose	Data-contributor consent notes	Potential downstream risks (authors' assessment)
CC-BY-NC 4.0 (for audio and annotations); AGPL-3.0 (for code and scripts)	Open access via GitHub repository and PyCompMusic API	MIR Research tasks in Indian art music, such as melody extraction, rhythm analysis, and structural segmentation	Srinivasamurthy et al. 2021 state: "due permissions were taken from artists and the licensees for all the audio recordings to store, process, and distribute them for research purposes"	Dataset's open access and named contributor list increases the likelihood of targeted singing voice cloning applications
CC-BY_NC 4.0 restricts commercial use, but does not address AI model training or voice synthesis permissions			No documented withdrawal mechanism	

Table columns are informed by the datasheet benchmark questions proposed by Gebru et al (2021), adapted to the context of voice and singing data

The table (see Table 2 below) details the license type, mechanism(s) for access, the stated dataset content purpose and research intent motivating the need for the dataset, any documentation concerning the consent mechanisms of the data-contributors, and potential downstream risks we identify.

Our use of the three consideration factors within the EASE analysis of Vocal Notes (Fig. 1) helped to map the power-to-interest of stakeholders. In the upper right quadrant (high power, high interest), we positioned the paper authors, MIR researchers, and research institutions. These stakeholders maintain sustained engagement with the dataset, particularly by granting access and utilising it in their research, and bear primary responsibility for its governance and sharing. These stakeholders also derive immediate benefits from the dataset's availability through citations and the ability to implement the dataset within their research. In contrast, we positioned the data-contributors—the performers—in the lower left quadrant. Their contribution and involvement with the dataset is event-centric despite the more longitudinal capacity for harm emergent from downstream usage of their data. This disparity stems from the nature of their contribution: they contribute their data once, while the curatorial stakeholders maintain sustained responsibility to manage access to the dataset due to the sacred and ritual singing traditions documented.

The positioning of stakeholders in Fig. 1 is instructive: those who can access, use and benefit from the dataset are clustered in the higher power, higher interest regions whilst the data-contributors and musical listeners occupy lower power, lower interest positions. This spatial relationship in the visualisation reflects a substantive ethical concern: those most likely to be affected by applications such as singing voice cloning (the performers whose voices would be synthesised) are positioned with least power to prevent or control such applications. The curatorial team's gatekeeping role—their high-power position—thus becomes critical for protecting contributors' interests, as contributors themselves lack formal mechanisms to influence downstream usage once their data is contributed.

3.5 The Saraga dataset

The Saraga Dataset is an open data corpus of audio recordings of Carnatic and Hindustani music, with time-aligned melody, rhythm and structural annotations (Srinivasamurthy et al. 2021). The Carnatic sub-collection includes recordings of live performances with some recordings in multitrack. The Hindustani sub-collection contains recordings from prominent Indian art music artists. Permissions for usage of the audio data for research purposes are licensed under Creative Commons (CC), with the annotations made publicly available. The Saraga dataset was formulated with the

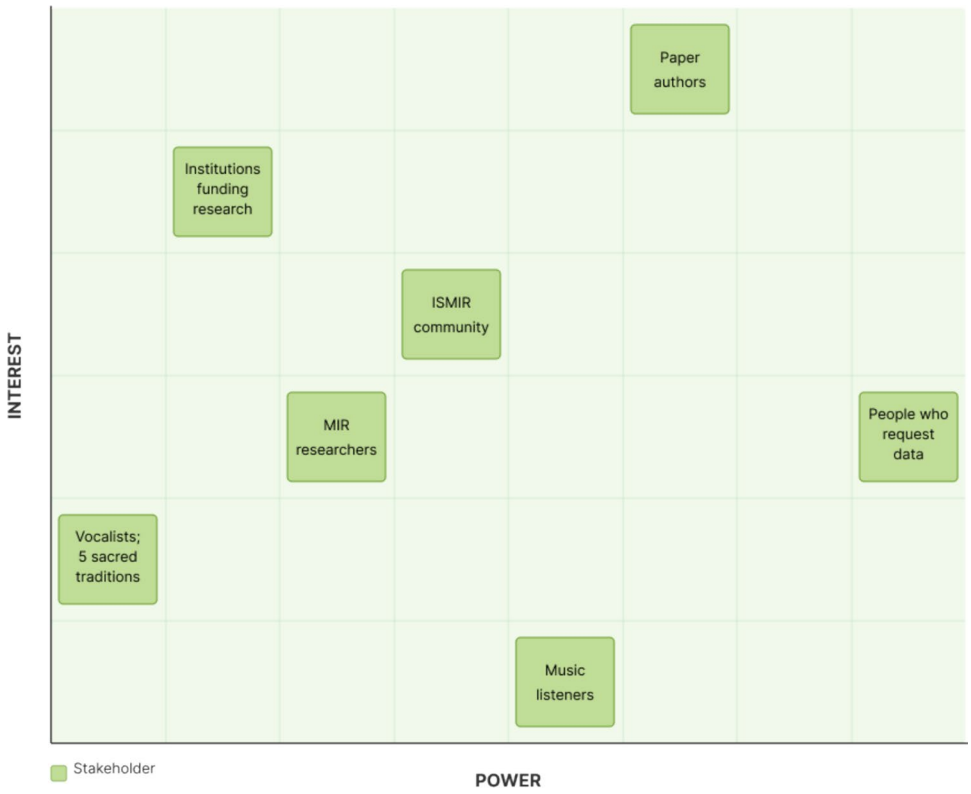


Fig. 1 Diagram depicting Part 2 of the EASE conducted on the VocalNotes Dataset, power-to interest positioning of implicated stakeholders

intention to be utilised for MIR tasks, and has been curated according to the proposed data corpus guidelines in (Serra et al. 2013).

The table (see Table 2 below) details the license type, mechanism(s) for access, the stated dataset content purpose and research intent motivating the need for the dataset, any documentation concerning the consent mechanisms of the data-contributors, and potential downstream risks we identify.

Our power-to-interest mapping of the stakeholders implicated by and involved in the creation of the Saraga dataset can be seen in Fig. 2. As in the VocalNotes dataset, we observe that stakeholders involved in the curatorial processes or who immediately benefit from the availability of the dataset have a higher power and interest stake compared with data-contributors. Data-contributors and performance licensees have a comparatively lesser power and interest stake. In this particular dataset we note that the nature of the dataset licensing and the publicly available GitHub repository affords wider access than the gated VocalNotes dataset, creating pathways for downstream non-commercial AI synthesis applications and model training. As with VocalNotes, Saraga's CC-BY-NC 4.0 licence restricts commercial use but does not address AI model training or voice synthesis permissions- meaning non-commercial synthesis applications are technically permissible. We therefore positioned AI developers (of non-commercial synthesis applications and models) in closer proximity to high-power stakeholders (researchers, institutions). This visualises a structural concern: those who can authorise and implement applications are distant on the power-interest axes from those whose voices would be exploited.

Our positioning of data-contributors in the lower-left region of Fig. 2 is motivated by the documentation noting that many are "prominent Indian art music artists." Their public recognition and established reputations increases their vulnerability to voice cloning- as their voices are more recognisable and potentially more commercially valuable.

3.6 The Rembetiko Dataset

The Rembetiko dataset is a compilation of 290 recordings of songs by 21 singers, in the style of Greek Rembetiko music (Holzapfel and Stylianou 2007). The age range of the recordings span from the 1910s- 1990s. All of the recordings were available as commercial recordings, many of them first on analog media and later as digital re-issues. As observed by the dataset curator: "At the time, I published the paper with the last sentence stating, "*All data along with the labeling can be obtained on request from the authors.*" No effort was made to think about potential licensing options, and the complexity that at least some of the recordings would be protected by copyright or neighboring rights was not something I reflected upon at the time. Instead, I followed the

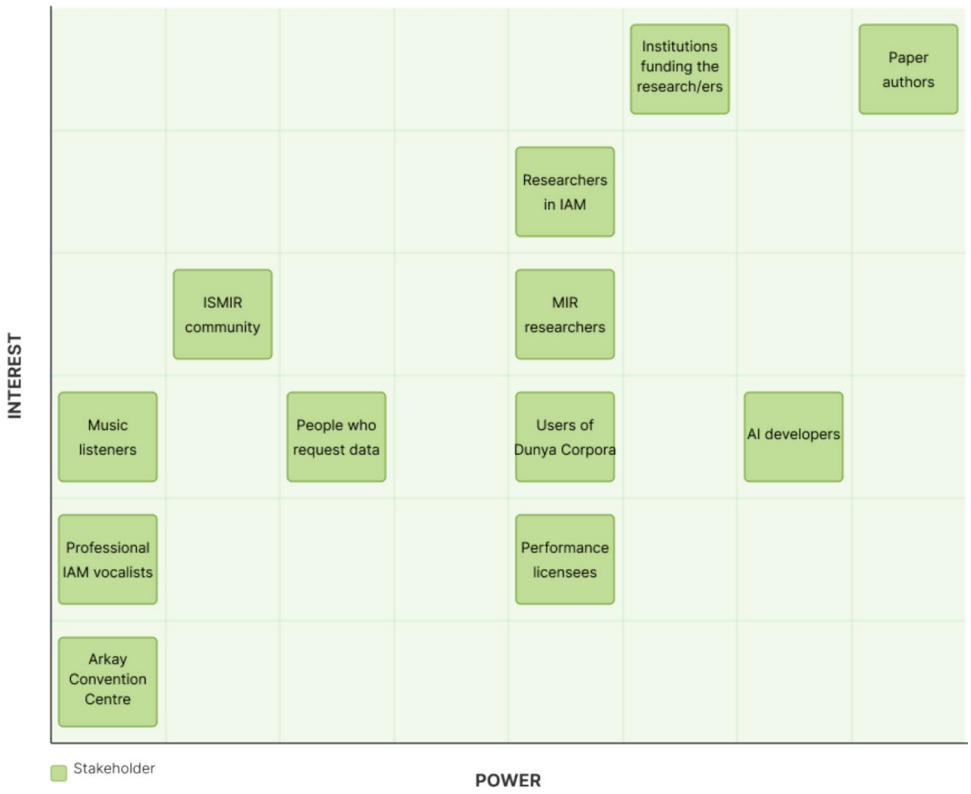


Fig. 2 Diagram depicting Part 2 of the EASE conducted on the Saraga Dataset, power-to-interest positioning of implicated stakeholders

unwritten rule in MIR - at least at that time - that informal peer-to-peer sharing between authors should be tolerated as a fair use for research purposes."

The table (see Table 3 below) details the license type, mechanism(s) for access, the stated dataset content purpose and research intent motivating the need for the dataset, any documentation concerning the consent mechanisms of the data-contributors, and potential downstream risks we identify (Table 4).

The authors' acknowledgement is valuable for framing how past MIR data practices may be reflexively analyzed, especially in light of how these data may now be protected under certain copyright protections or may be in the public domain.

The temporal span of the Rembetiko recordings creates specific ethical challenges. For recordings from the 1910s through the mid-20th century, most of the featured singers are now deceased. Greek copyright law and neighboring rights frameworks would therefore govern whether these recordings remain under copyright protection or whether they have entered the public domain. However, even if select recordings are no longer under copyright, the question of personality rights remains ethically relevant. The voices of deceased performers can be analyzed, modeled, and potentially synthesised using contemporary AI technologies that did not exist during their lifetimes. While these artists consented to commercial recording and distribution within the technological and cultural context of early-to-mid 20th century Greek music industry, they could not have anticipated computational voice modeling. The temporal distance between the original recording contexts (spanning over eight decades) and potential contemporary applications (such as generating synthetic Rembetiko performances using cloned voices of deceased artists) represents an extreme case, yet it raises important questions about the relationship between cultural heritage preservation and contemporary AI training data. These recordings are a valuable resource for research, and there is legitimate scholarly value in making such recordings accessible. However, when such recordings become available for computational analysis and potential AI model training, a tension emerges: does treating historical cultural documentation as "fair game" for voice synthesis research constitute a form of data extractivism? The research value of the dataset must be weighed against the potential harms of cloning the voices of deceased artists, or living artists who may not have anticipated how their commercial recordings might be used in downstream tasks. Accordingly, we positioned the data-contributors in the lower-left region with minimal power and minimal ongoing interest (see Figure 3). This positioning reflects not merely the event-centric nature of their contribution (as in VocalNotes and Saraga) but also their temporal distance from the dataset- many are deceased, and none could have anticipated that their

Table 4 Evidence table for the Rembetiko dataset, detailing licensing type, access mechanism, stated purpose of the dataset, any notes pertaining to how data-contributors were asked for consent, and the identification of potential downstream risks as proposed by the authors

License type	Access mechanism	Dataset content and stated purpose	Data-contributor consent notes	Potential downstream risks (authors' assessment)
No formal licence applied at time of publication	Informal peer-to-peer sharing via request to authors, in line with prevailing MIR field norms at the time of dataset collection	Singer identification research using historical Rembetiko recordings	Not documented in dataset metadata Recordings were sourced from existing commercial releases spanning the 1910s-1990s: reflects field-wide practices at the time of dataset collection	Recordings span deceased and living artists, who may not have anticipated computational voice modelling on their commercial music releases The estates of deceased artists may retain rights under Greek copyright law and neighbouring rights frameworks, yet may lack awareness of and access to formal mechanisms to restrict or condone downstream uses

Table columns are informed by the datasheet benchmark questions proposed by Gebru et al (2021), adapted to the context of voice and singing data

commercial recordings would later be used for computational voice analysis.

We positioned paper authors, MIR researchers, and research institutions in the upper-right quadrant with high power and high interest—consistent with the other two datasets. Figure 3 also positions AI developers, Singer identification researchers, and Amateur audio editors as stakeholders with varying degrees of power and interest in the dataset's downstream usage. AI developers are positioned in closer proximity to high-power stakeholders to reflect the absence of formal licensing restrictions that would limit their access to the dataset. Singer identification researchers and Amateur users are positioned at lower power and interest regions, reflecting their more constrained capacity to govern or direct dataset usage and access. Unlike VocalNotes and Saraga, there is no formal licence for the Rembetiko dataset which means that stakeholders face no formal barriers to accessing and using the data from downstream purposes such as voice synthesis and cloning. This reflects the prevailing norms in MIR data sharing practices at the time of the dataset's publication, wherein licensing considerations were not yet a standard practice.

4 Results

Comparing our three EASE mappings (Figs. 1, 2, and 3) reveals consistent patterns of stakeholder distribution that transcend the specific characteristics of individual datasets. Across all three case studies, we observe a persistent clustering of paper authors, MIR researchers, and research institutions in the upper-right quadrant (high power, high interest), while data-contributors consistently occupy the lower-left region (low power, low interest). This consistent pattern suggests that current practices in singing data collection systematically advantage those engaged in computational research while systematically disadvantaging those whose voices constitute the raw material for that research. The consistency of this pattern across datasets with different licensing frameworks, different musical contexts, and different temporal spans indicates a structural issue rather than isolated oversights.

To make the placement reasoning transparent, we positioned our stakeholders across all three datasets according to the following logic.

Paper authors and research institutions occupy the high-power, high-interest quadrant because they hold formal and legal authority over dataset licensing and access decisions; they maintain longitudinal involvement with the dataset through publication of data and making the dataset accessible to prospective users; and they bear low vulnerability to downstream harms as they have not contributed their own voice data. Data-contributors occupy the low-power,

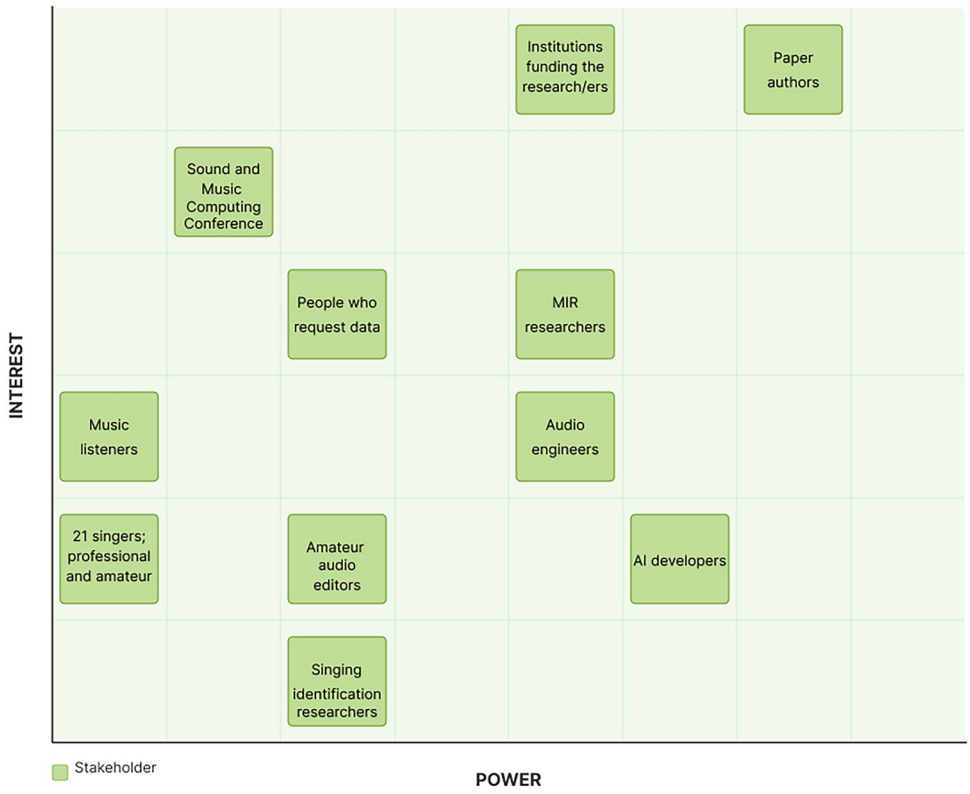


Fig. 3 Diagram depicting Part 2 of the EASE conducted on the Rembetiko Dataset, power-to interest positioning of implicated stakeholders

low-interest quadrant because they hold no formal authority of licensing or access decisions; their involvement in the dataset is even-centric (they contribute their data); and they face greater vulnerability to downstream harms.

Variation in placements across the datasets (for example, the positioning of funding institutions or performance licensees) reflect differences in the degree of influence of these stakeholders in each dataset context. Beyond the immediate data-contributors, the remaining stakeholders depicted across Figs. 1,2 and 3 are positioned according to their formal governance authority, the temporal scope of their involvement with the dataset, and the implications they face through downstream data usage. At the higher-power end, funding institutions occupy elevated power positions due to their authority over research funding conditions, though their interest in any individual dataset is indirect and reputational rather than sustained — a funding body whose grant enabled the collection of singing data may, for instance, face reputational consequences if the funded research is found to have engaged with inadequate consent or licensing practices. The ISMIR community occupies a mid-range position on both axes, as they assert a degree of influence over research norms and publications standards but do not exercise direct authority over data-collectors licensing or access decision-making. Performance licensees (identified in the Saraga dataset) hold moderate legal authority over specific recordings, but are positioned in a lower interest position as their involvement in the dataset is transactional rather than ongoing. At the lower end of both axes, music listeners may be affected by the downstream application of voice datasets (such as encountering synthetic likenesses of deceased artists. They hold no formal stake in dataset governance decisions.

Finally, AI developers occupy a high-power position in relation to downstream usage as they are the stakeholders who may utilise voice and singing data in downstream applications, but they are external to the dataset governance and access decision-making. This thus limits their formal interest stake within the EASE mapping.

With regard to legal authority: curators and researchers control licensing terms, access mechanisms, and usage permissions while contributors typically lack formal authority over how their voices are used. This vulnerability is not unique to singing data collected within a research context, indeed it is a reality of any digitised audio recording in the current AI landscape. However, this vulnerability is particularly acute for our three case studies. VocalNotes features sacred and ritual vocal practices from living communities, whose cultural and religious practices may be directly implicated in non-consensual downstream applications. Rembetiko features the performances of deceased singers, whose estates may lack awareness of, or access to, mechanisms to restrict downstream use. Saraga features musically and culturally specific vocal practices where decontextualised

use through downstream synthesis may misrepresent or disrespect living artistic traditions. In each case, the cultural and musical specificity; limited legal recourse; and the identifiability of data-contributors creates a heightened exposure to harm. A performer who contributed recordings to VocalNotes will remain vulnerable to cloning- which may not have existed at the time of contribution. A deceased Rembetiko singer's estate, though temporally distant from the original recording, remains vulnerable to synthetic performances using the singer's cloned voice. This temporal asymmetry—wherein contributors' agency is bounded but their vulnerability is persistent—represents the core ethical challenge visualised across Figs. 1, 2, and 3 and justifies their placement within a lower power, lower interest stake.

Second, with respect to temporal involvement: researchers maintain longitudinal relationships with datasets through ongoing publication and analysis while contributors' involvement is event-centric—they provide recordings once and may not be involved in decisions around how, where, or to whom their data is made available. This temporal dimension therefore informs how a stakeholder with more longitudinal involvements (paper authors, MIR researchers) occupy a higher power, higher interest stake whilst stakeholder with more event-centric involvement (data-contributors) have a lesser power, lesser interest stake.

Third, and most critically, in vulnerability to harm: despite contributors' lower positioning on power and interest axes, they face the highest propensity for harm from downstream applications such as unauthorised voice cloning, decontextualisation of sacred practices and cultural context (especially in cases where the data-collectors are not themselves from the communities where data is being collected), or commercial exploitation without consent or compensation.

Our proposals in the forthcoming section seek to address the differing concerns of potential dataset stakeholders. These are further informed by our observations of how higher-power-to-interest stakeholders and lower-power-to-interest are implicated by longitudinal factors concerning the dataset curation; the tensions between the interest versus harm propensity of various stakeholders; and what the respective needs of mapped stakeholders may be.

5 Proposed research guidelines for singing data collection

In this section, we propose research guidelines for singing data. These guidelines are conceptualised on two layers: a more specific layer which requires more procedural guidelines to guide individual interactions between data-contributors and data-collectors; and a layer which requires more substantive discussion on approaches to archiving cultural heritage or sensitive material. The procedural guidelines offer more concrete and implementable points for researchers to integrate into their existing research processes for singing data collection. The substantive discussion points provide potential aspects to deliberate upon, and are intended to elicit critical evaluation of how they relate to the researcher's existing procedures for collecting singing data.

Foremost, we agree with existing approaches to data collection in their emphasis on fidelity of the data quality (Serra et al. 2013), reusability and accessibility (Wilkinson et al. 2016), and musical diversity (Serra 2011). The continuation of open-access datasets is integral to the development of high-fidelity and flexible synthesis and generation models. Outside of MIR research, such synthesis and generation models are integral to areas such as voice and speech rehabilitation and supportive therapies/products (Yang et al. 2023); customer service (Amoozad Mahdiraji et al. 2022; Buhalis and Moldavska 2022; Grewal et al. 2022); improving accessibility and disability support (Esquivel et al. 2024; Huq et al. 2024; Semary et al. 2024; Volochtchuk et al. 2024; Vonica 2021; Wald 2021).

We note that these proposed guidelines are emergent from our particular scope. We observe that future work in including more diverse legislative scope is vital.

5.1 Procedural guidelines

Critical stakeholder mapping: apply EASE or other critical dataset exploration tools to formulate a list of stakeholders prior to commencing dataset curation. When implementing such exploration tools, do not take the use of the tool as a given, but extend it and question it in all its aspects. Where applicable, researchers should consider the three consideration factors introduced in Section 3.3: legal authority over licensing and access decisions; the event-centric versus longitudinal nature of stakeholder involvement; and the vulnerability to potential downstream harms, as part of this mapping process. This process should be conducted in a comparative and relational manner, wherein each stakeholder is considered in relational to all others. Instances of inter-author disagreement should be resolved through an additional round of comparison, informed by the three consideration factors until consensus is reached.

When critical stakeholder mapping indicates instances of data-contributors in low-power, low-interest positions this should be taken as an action point for researchers and data-collectors to address. In these instances, researchers and data-collectors should evaluate:

– What licenses for data usage would better protect data-contributors' interests and increase their power-interest position?

– What mechanisms for access (such as a data access request form) would better protect data-contributors' interests and increase their power-interest position?

What longitudinal governance changes could ensure data-contributors have more power-interest in the dataset they have contributed to?

Dataset Licence**:** Clarify under which licence the data you collect will be published, and consider explicitly how the biometric nature of the collected data, the musical communities and culture the data comes from, and identifiability of data-contributors may inform licence selection. Consider also what kinds of downstream data use should be prohibited, and utilise the knowledge of who your affected stakeholders are to inform the choice of licensing. Data-contributors should be informed as to what the dataset licensing permits and prohibits. Well-intentioned licensing may leave data-contributors exposed to risks from model training and downstream tasks, and researchers and data-collectors should not assume that standard CC licences are sufficient for singing data. Where possible, data-collectors should consult with legal specialists on how the publication of the dataset may be implicated in downstream model training contexts and which licence best protects data-contributors' interests.

The use of emerging tools such as the CC signal project (Creative Commons 2025) to make AI training permissions explicit can be beneficial here. For datasets predating current AI voice synthesis capabilities (such as those presented by our Saraga and Rembetiko case studies), researchers should consider whether the existing licence terms adequately address AI training and voice synthesis risks. Where they do not, updating licence documentation to explicitly address AI training permissions should be considered (this is discussed further in Section 5.2). The choice of licence should also consider AI Usage Permissions, informed by the explicit communication of the potential associated harms with data-subjects. The stakeholder mapping should directly inform dataset licensing decisions, especially where data-contributors occupy lower power positions and have greater vulnerability to downstream harms.

Weighing legitimate interest against potential harms: consider not publishing data. There may be numerous reasons (such as the vulnerability of stakeholders) that prevent the publication of a singing dataset. In these cases, highly restricted sharing of data may be a compromise between a standard license and complete disclosure of data. This can work in the way of e.g. an agreement between the dataset compiler and a specific researcher wanting to use the data for a specific project.

In instances where such agreements are in place, data-contributors should be informed as to the terms of such an agreement and be able to give explicit and informed (insofar as is possible) consent. Just as criticised (Morreale et al. 2023) in the case of music generation datasets, the stakeholders involved in creating the contained data should be asked for detailed consent on what use of their data they give consent to. This should inform the selection of the dataset licence, as outlined above.

Researchers should consider the following when weighing legitimate interest against potential harms: how identifiable data-contributors are within the dataset with respects to their identifiability both within their own community and outside of it; the cultural, musical or sacred significance of the singing data; whether existing licensing mechanisms adequately address the risks posed by downstream AI training; and the temporal vulnerability of data-contributors to harms that may not exist at data collection time.

Here, researchers could consider and communicate the potential harms to data-contributors concerning downstream production of non-consensual voice models of themselves to promote products and services from which they do not financially benefit (Cucci 2023; Parkinson 2015; Savov 2011; Warman 2011; Ravitz 2013); are unwilling to be affiliated with (Geerling 2024a,b); or which may pose reputational damage (Norms 2024, 52 Card Laboratory 2024). We further note that consent obtained at the moment of data collection carries limitations as demonstrated by our EASE: vulnerability of data-contributors extends as AI technology continues to develop. Researchers should therefore not treat consent as a one-time decision, but one requiring periodic review.

De-identification of data: Our EASE analysis demonstrates that the identifiability–and in some cases the explicit naming of data-contributors–may contribute to their vulnerability to downstream harms, through targeted use of their data. Consider de-identifying the collected data through the anonymisation of data-contributors. We note that there are challenges to the de-identification of singing data, where the same characteristics such as vibrato or tonal quality which make a data-contributor identifiable also makes their data immensely valuable for research. As has been demonstrated by (Rusti et al. 2023), even anonymised data can be re-identified using advanced speaker verification models, which may be similarly applicable within the context of singing data. As many singing datasets involve a risk of misuse by using the data to create a model for a specific person without the knowledge or consent of the data-contributor, consider limiting the amount of data related to each person so that such misuse is impeded.

Dataset documentation: Besides documenting the content and format of a dataset, its provenance should be documented in line with proposals from Gebru et al. (2021). This should involve all steps, from decisions of what was included and excluded to ways the data was obtained and formatted. The clarification of what specifically data-subjects consented

to should also be disclosed. More detailed documentation of how the dataset came to be would then become vital for future attribution or remuneration, which is discussed further in Sect. 5.2. Further, the mechanisms through which data-contributors gave and can withdraw consent should be specified.

Dataset maintenance and review: Consider treating dataset governance as an ongoing responsibility (Swedish Research Council 2024), rather than a decision made at the initial creation and establishment of the dataset. As AI voice synthesis capabilities and applications continue to evolve, licensing terms that were considered sufficient at the time of releasing and publishing the dataset may become insufficient to protect data-contributors from new forms of downstream harm. Researchers should periodically review their datasets' licensing and access mechanisms as new technology develops. Where possible, data-collectors should maintain and document the means through which data-contributors can inquire about the license their data is issued under, and request a withdrawal of their data from the dataset after the initial collection.

5.2 Substantive discussion

5.2.1 Attribution and fair remuneration

There are several challenges to meaningful attribution and fair remuneration for data-contributors affected by downstream data usage. As our EASE demonstrates, the temporal asymmetry of data-contributor's involvement makes the tracking of downstream usage incredibly difficult. Further, there may be a further temporal disconnect between the moment of contribution and the moment of value creation: downstream applications may emerge years after the initial contribution of data. Further, singing data from communities or data which features cultural or sacred material poses challenges as to *who* receives remuneration. Of additional concern is how paying data-contributors from under-represented or economically vulnerable communities enforces a colonial view of someone's singing data as ownable, transferable, and commodifiable (Duarte 2020). This is of particular concern where the data collected features sacred or culturally significant material, and we require greater coordination between research bodies, research funding institutions and legislative bodies to formulate more robust practices. A ground-up approach that can be implemented by researchers is to adopt and maintain robust and detailed documentation practices (Gebru et al 2021), as fairer attribution and remuneration cannot be achieved without traceability of data provenance.

5.2.2 Community-involved, community-led and community-guided governance approaches

Where singing data is collected from specific cultural and musical communities, or from religious or sacred practices, governance decisions of the dataset should not be determined solely by the data-collectors. Engaging data-contributors in the governance decisions of dataset contributions beyond an event-centric contribution of their data may help to increase their power-interest stake, especially when it comes to the considerations we identified in Section 3.3. Specifically, community involvement in governance decisions may increase the legal authority stake of data-contributors, and may increase their longitudinal involvement with a dataset beyond the moment of contributing their data. This has the potential to empower data-contributors to make more informed decisions about how and under which terms they wish their data to be accessed, licensed and maintained (Ding et al. 2024). This is in line with perspectives from decolonial data practices and ethics (Christen et al. 2018; Carroll et al. 2020; Ricaurte 2019; Arora et al. 2024). Data-collectors working with under-represented or potentially vulnerable musical traditions and communities have a heightened responsibility to prioritise and actively engage with community concerns in data collection practices.

Unforeseen use and retrospective licensing: as noted in Section 2.4.2, the inclusion of provisions for pre-existing recorded material outlined in the SAG-AFTRA and Replica Agreement is relevant for researchers to consider. This is relevant considering the amount of readily available historical datasets that do not have clear licensing terms for model training. The provisions in the SAG-AFTRA agreement concerning conditions of use for recordings of deceased artists offer a helpful framework to data-collectors of historical singing data. Researchers who have previously published datasets might consider adding or updating the licensing terms to consider AI model training as a potential use case for these datasets. In practice this may encompass: adding explicit AI training restrictions to the existing dataset license; and updating the access mechanisms of the dataset. In circumstances where well-known, named or highly identifiable data-contributors are featured in a dataset, consider whether additional restrictions may be helpful to protect data-contributors' interests.

Consult archival scientists: considering consulting archival specialists for their advice on initiating the recording of culturally sensitive or cultural heritage singing material. This is especially pertinent for researchers conducting work in collecting data from sacred, religious or ceremonial singing traditions. Archivists bring specialist knowledge to the preservation of cultural heritage which is often distinct from, and complements, data ethics and legal knowledge. Culturally-aware and informed decisions on what to record;

how to record vocalists and communities from vulnerable or under-represented cultures; the management of sensitive cultural material (Christen et al. 2018); as well as how to best store for greatest digital longevity may not be within the knowledge or training of MIR data-collectors.

Archival restrictions: much historical singing data has been released without specifications for outlining how this data may or may not be utilised for downstream model development.

Whilst some of these historical recordings may be within public domain under copyright law, this does not mean that ethical obligations towards the communities or estates of involved data-collectors are not applicable. This is especially relevant in the present reality of AI voice synthesis where synthetic likenesses of deceased artists can be produced.

Here, potential approaches could include the addition of specific clarifications for future usage of the dataset for singing model development; or the redaction of the dataset altogether pending authorisation to access from the researcher or institution.

Data-collectors involved with historical singing data should consider the addition of explicit restrictions on AI training informed by engagement with archival scientists and the communities represented in the recordings; reducing the access to datasets featuring highly identifiable data-contributors; and maintaining longitudinal involvement of community representatives and estates to determine the terms within which historical material may be used and accessed. Similar approaches have already been deployed by cultural institutions which archive historical media material (The North Africa Institute 2026). We emphasise that decolonial perspectives as discussed in Sect. 2.2.2 are applicable to both historical and newly collected datasets.

6 Discussion

Across our three case studies, we observe a progression in data ethics awareness over time, even as fundamental power asymmetries persist across all three datasets. The Rembetiko dataset (reflecting earlier MIR practices) operated under informal peer-to-peer sharing with no explicit licencing—a practice the curator (and co-author of this paper) acknowledges as ethically insufficient. Saraga (2021) implemented Creative Commons licencing and followed established corpus guidelines, providing more structure than informal sharing. VocalNotes (2023) introduced access request forms and CC-BY-NC-SA 4.0 licencing restrictions in recognition of the sacred nature of the material. Yet despite this progression toward more ethically aware practices, the fundamental power-to-interest distribution remains remarkably similar across all three datasets as depicted in Figs. 1, 2, and 3. This suggests that while the field has developed greater reflexivity about data ethics (Chen et al. 2019), the structural positioning of data-contributors in our three case studies relative to researchers has not substantially shifted. Contributors remain in the lower-left region regardless of whether datasets implement informal sharing (Rembetiko), standard Creative Commons licencing (Saraga), or access restrictions (VocalNotes). Whether this pattern holds more broadly across the MIR community would require further analysis beyond the scope of this study. As the data-contributors across VocalNotes, Rembetiko and Saraga occupy lower power-to-interest stakes than the datasets' data-collectors, our findings point to a need to reflect upon how current practices for singing data collection might place data-contributors more "in the centre" of decisions concerning dataset licencing. What licencing-related steps could researchers take to protect the personality or identity rights of human participants within an AI climate in which public data can (and often is) treated as "free to take and is transformed into a capital commodity" (Cotton et al. 2024; Tatar et al. 2023)? Further, could we correlate these emerging needs with the extant values of the MIR community, which (rightfully) prioritise diversity, inclusion, and open-access data?

Decolonial perspectives offer a helpful approach for navigating these questions through foregrounding community authority and sovereignty in data governance decisions: asking both what data is collected and how communities can retain meaningful agency over how this data is used (Ricaurte 2019; Carroll et al., n.d.; Christen et al. 2018). MIR community values around diversity, inclusion and open access should be considered with respect to how these values are enacted in data collection practices (Ricaurte 2019; Duarte 2020). Meaningful engagement with data-contributors to position them more explicitly in the centre of dataset licensing, access and governance decisions, as identified in our EASE analyses, is a crucial first step in addressing power-interest asymmetries between stakeholders and developing more robust approaches to data collection in light of downstream AI voice synthesis. The singing datasets analysed in this paper are representative of valuable contributions to musical and cultural diversity in MIR research, with each dataset representing music from outside Western popular music (Chen et al. 2019) and reflecting the broader scope that research on MIR ethics has long called for (Holzapfel et al. 2018; Bauer et al. 2020; Morreale et al. 2023). As the MIR field continues to conduct, publish and share more musical and culturally diverse singing datasets, our findings indicate that more structural engagement with how data-contributor authority and involvement is formalised and sustained will be critical. The stakes and potential harms faced by singers and vocalists from under-represented communities and musical cultures are higher than ever in the present landscape of downstream AI voice synthesis.

As we noted in Sect. 3.3, stakeholder placement is informed by our three consideration factors: legal authority; the event-centric versus longitudinal nature of stakeholder involvement; and the vulnerability to potential harms. We note two inherent limitations in how this process is represented. First, the power-interest axes in Figs. 1, 2 and 3 present power and interest as orthogonal dimensions, yet in practice these dimensions are not fully independent: a stakeholder's legal authority (a dimension of power) is often directly related to the sustained nature of their involvement (a dimension of interest). Second, placement involves weighing incommensurable factors, which means that researchers from differing disciplinary backgrounds may arrive at slightly different positions for peripheral stakeholders. The per-stakeholder placement reasoning is detailed in Sect. 3.3. We assert that our analyses are sufficiently robust across alternative interpretations of our three consideration factors for the following three reasons.

Firstly, increasing a data-contributor's power does not reduce their vulnerability for potential harm from downstream usage. Although the data-contributors in VocalNotes, Saraga and Rembetiko (and indeed any singing dataset) possess a genuine cultural authority as practitioners of their respective musical traditions, weighing their cultural authority heavier than their legal authority over the dataset does not change the realities of what their singing data is capable of being used for. As an example, if we were to weight a data-contributor's (low-interest, low-power) cultural authority more heavily and move them to a higher-power position, this does little to change the realities of how their contribution of singing data to a dataset exposes them to potential harms from downstream usage. Cultural authority carries no formal mechanism to restrict downstream usage. We acknowledge that mechanisms for increasing legal authority (our first consideration factor) through licencing arrangements to empower contributors to define specific permitted uses of their data could reduce their vulnerability in a meaningful way. However, such mechanisms are not present, documented or discussed in our examined case studies.

More power does not mean that a data-contributor becomes less vulnerable to harm—the harms may still exist. Singing data can still be extracted, modelled and synthesised in downstream and unauthorised usage, regardless of how much cultural authority the data-contributor is determined to have. The biometric realities of data–its capacity to identify–are not affected by a data-contributor's power.

Secondly, variations in the placement of peripheral stakeholders (funding institutions, AI developers or performance licencees) are possible, but still do not affect the central findings of our analysis. Our adaptation of EASE is towards the identification of stakeholders most implicated by non-consensual downstream usage of singing data. Peripheral stakeholders have not contributed data, and thus are not implicated in the same way as data-contributors are. These stakeholders are difficult to position due to their involvement being less visible in the dataset documentation, and researchers implementing this adapted EASE may place them differently depending on how they weigh institutional influence versus direct decision-making authority over dataset licensing and access. Repositioning these peripheral stakeholders towards the upper and middle regions of the diagram could be plausible in light of this, but we do not consider institutional influence in our consideration factors. These peripheral stakeholders have not contributed data themselves, and therefore are simply not exposed to the same potential harms as data-contributors. The potential harms they may be exposed to are more concerned with reputational risk, such as a funding institution financing research that has been found to have engaged with inadequate or questionable consent practices.

Thirdly, our three consideration factors inform a power-interest positioning for implicated stakeholders through an aggregative process. That is, the three consideration factors reinforce each other even if one of them is considered lesser in impact than another factor. A researcher who lessens vulnerability as a consideration factor would still find that legal authority and temporal involvement still place data-contributors in a comparatively lower-power, lower-interest position. In all three datasets we surveyed, data-contributors hold no formal authority to make licensing or dataset access decisions. In VocalNotes, the access form outlines researcher obligations only and there is no documented means through which the data-contributors can influence how their data is shared and used. For Saraga, permissions were obtained for a specific and defined purpose, with no documented withdrawal mechanism available to contributors. In Rembetiko, no formal licence or consent withdrawal mechanism is documented. The event-centric involvement of data-contributors does little to improve their power and interest positions, whereas the data-collectors maintain longitudinal responsibility and obligations towards the datasets which thus affords them greater power and interest. Our three consideration factors therefore reinforce each other, to the degree that the placement of data-contributors is stable even if we vary individual factors.

7 Conclusion

There is much work to do in regards to developing and implementing new research practices for singing data collection, considering the dearth of studies mapping licensing approaches and the sufficiently available datasets that are seemingly just 'there' - and available to train hi-fidelity, high-performing models. The unforeseen consequences of this can be vast. The lack of licensing agreements specifying

AI model training usage or consideration towards how datasets may be utilised for non-consensual downstream usage of the data have significant implications for data-contributors. Our application of EASE to three already published singing datasets–VocalNotes, Saraga and Rembetiko–demonstrates power asymmetries between data-collectors and data-contributors, especially when it comes to the potential harms posed by downstream applications that depend on such data. A more proactive approach to new singing data collection research practices is desperately needed. Our proposed guidelines in Section 5 are a first step, and may prove applicable to non-singing voice data and audio data more generally.

The persistence of power asymmetries across datasets with different licencing frameworks, musical and cultural contexts and different temporal span, indicates that incremental improvements to existing singing data collection practices are insufficient. What is required is a fundamental shift towards re-positioning data-contributors more centrally within dataset governance decisions.

Our adaptation of EASE to conduct analyses on already-published datasets forms a methodological contribution of this paper: through introducing factors which consider legal authority; the event-centric versus longitudinal nature of stakeholder involvement; and the vulnerability to potential harms. By applying these consideration factors through a comparative, relational placement process, we offer a replicable method for researchers who wish to assess the power-interest dynamics of their own already-published datasets. The evidence tables, along with the diagrams illustrating the power-interest placements, provided in each case study demonstrate how this may be achieved and the form such documentation can take.

Our EASE analysis revealed critical findings that point towards necessary changes in how the field approaches dataset governance. First, we identified a persistent temporal asymmetry: data-contributors' involvement in dataset development is event-centric, yet their vulnerability to harm is temporally unbounded (as new AI technologies emerge and create new applications). This temporal mismatch means that consent obtained at the moment of data contribution cannot adequately protect against harms that emerge years or decades later.

Dataset governance must therefore be treated as an ongoing responsibility, not a decision made once at publication time.

Second, within our three case studies, we observed that while the MIR community has developed increasing ethical awareness over time—progressing from informal peer-to-peer sharing (Rembetiko) to standard licencing frameworks (Saraga) to access restriction mechanisms (VocalNotes)—the fundamental power distribution between researchers and contributors has not substantially shifted. Contributors remain positioned with low power regardless of licencing framework, suggesting that incremental improvements in practices are insufficient without structural interventions. What licencing-related steps could researchers take to better protect the personality and identity rights of data-contributors, whose data risks being treated as a freely available commodity for downstream and non-consensual voice AI training (Cotton et al. 2024; Tatar et al. 2023)? We argue that this question needs to become a standing obligation for researchers collecting singing data—embedded throughout the lifecycle of a dataset and not addressed singularly at publication.

7.1 Ethics statement

This paper intends to contribute to constructive and critical dialogue around the collection of singing data that considers the implications of downstream AI model development and training. We therefore consider our paper as contributing constructively to societal implications within this area. Similarly, as this work did not implement any user studies or actively engage with human participants, there is no potential risks or harms to individuals. The datasets discussed in this paper are accessible, with varying levels of public access. VocalNotes is accessible upon a research request application following the procedures outlined in the dataset paper (Proutskova et al. 2023). Saraga is publicly accessible. The Rembetiko dataset is available upon request to the original dataset authors. The prior societal context of this work is situated within an ongoing discourse on legislation and protections for vocalists affected by AI voice model development. This paper seeks to contribute to this discourse by presenting new guidelines for researchers working with and collecting data from vocalists. No machine learning or deep learning models were trained or utilized for content (voice, audio, image, graphs etc.) generation for this study. No Large Language Models were involved in manuscript preparation of this paper. The computational impact of writing this paper and conducting the evaluation of the selected case studies is comparable to daily personal computer usage- the predicted environmental impact of this work is thus considered minimal.

Acknowledgements This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program—Humanities and Society (WASP-HS) funded by the Marcus and Amalia Wallenberg Foundation. This work was partially supported by the Swedish Research Council (2024-01832). This work was supported by the Wallenberg AI, Autonomous Systems and Software Program—Humanity and Society (WASP-HS). This work was partially supported by Formas—a Swedish Research Council for Sustainable Development (2024-01579)

Author contributions The conceptualisation of this manuscript was done by K.C, A.H, K.d.V, K.B and K.T, with the content of the paper

informed by discussions between all authors. The data curation for this paper involved the sourcing of open access material, and datasets-which was conducted by K.C. The formal analysis of this data, and the Investigation was led by K.C with input, suggestions and feedback from co-authors. A.H contributed domain knowledge and feedback on the adaptations to the Ethically Aligned Stakeholder Elicitation (EASE) proposed by Kaila et al. (2023). The visualisation component of this paper involved the EASE power-to-interest graphs, done by K.C with feedback from co-authors on stakeholder positioning. The writing—original draft was led by K.C, with feedback and contributions from all co-authors. K.d.V wrote Sect. 2.3 and contributed domain knowledge and insight pertinent to AI legislation throughout the manuscript. K.B wrote Sect. 2.4 and contributed domain knowledge and insight pertinent to voice licensing throughout the manuscript. A.H contributed domain knowledge and insight pertinent to MIR data collection practices. K.T contributed domain knowledge pertinent to data de-identification and dataset maintenance and review (Section 5.1). Subsequent writing—review and editing was led by K.C with feedback and contributions from co-authors. The final preparation of the manuscript for publication was done by K.C.

Funding Open access funding provided by Chalmers University of Technology. Wallenberg AI, Autonomous Systems and Software Program—Humanity and Society, Formas—a Swedish Research Council for Sustainable Development, 2024-01579), Swedish Research Council, 2024-01832, 2024-01832, 2024-01832, 2024-01832.

Data availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare that there are no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

52 Card laboratory, (2024). My voice was stolen - a warning to creators and voice actors. AI Sucks. https://www.youtube.com/watch?v=JxO9gP541No

Adam A (1995) A feminist critique of artificial intelligence. Eur J Women's Stud 2(3):355–377. https://doi.org/10.1177/135050689500200305

Altulahi A, Gyani J, Altamimi B. (2022). A framework for Deepfake voice synthesis

Arik SO, Chrzanowski M, Adam C, et al. (2017). Deep voice: real-time neural text-to-speech. arXiv:1702.07825. Preprint, arXiv, Mar 7. https://doi.org/10.48550/arXiv.1702.07825

Arik SO, Chen J, Peng K, Ping W, Zhou Y. (2018). Neural voice cloning with a few samples. arXiv:1802.06006. Preprint, arXiv, Oct 12. https://doi.org/10.48550/arXiv.1802.06006

Arora P, Revelo MC, Koenitz H (2024) Creative data justice: a decolonial and indigenous framework to assess creativity and artificial intelligence. Info Commun Soc. https://doi.org/10.1080/1369118X.2024.2420041

Axhamn J. (2025). Extended collective licensing for use of copyrighted works for machine learning. The Columbia Journal of Law & the Arts 48 (4): 523–45. https://doi.org/10.52214/jla.v48i4.13933

Bäckström T (2025) Privacy in speech technology. Proc IEEE 113(7):668–692. https://doi.org/10.1109/JPROC.2025.3632102

Baird A, Parada-Cabaleiro E (2019) EmoSynth: the emotional synthetic audio dataset. Zenodo. https://doi.org/10.5281/zenodo.3727593

Barthes R, (1978). The grain of the voice. In Image, Music, Text, Nachdr., translated by Stephen Heath. Hill and Wang

Bauer C, Cruz L, Ferraro A, et al. (2020). Ethics Guidelines in Music Information Retrieval: WiMIR Workshop. WorkingPaper. https://drive.google.com/file/d/1DDEMyLIvX_gm7p9XCPP2UiqnfMZn6Pe8/view?usp=sharing

Benhamou Y, de Rosnay MD. (2023). Open Data Commons Licenses (ODCL): licensing personal and non personal data supporting the commons and privacy. SSRN Scholarly Paper No. 4662511. Rochester, NY, December 12. https://doi.org/10.2139/ssrn.4662511

Benjamin R. (2019). Race after technology: abolitionist tools for the New Jim Code. 1st edition. Polity.

Benjamin M, Gagnon P, Rostamzadeh N, Pal C, Bengio Y, Shee A. (2019). Towards standardization of data licenses: the Montreal Data License. arXiv:1903.12262. Preprint, arXiv, March 21. https://doi.org/10.48550/arXiv.1903.12262

Beranekt B (2013) Voice biometrics: success stories, success factors and What's next. Biom Technol Today 2013(7):9–11. https://doi.org/10.1016/S0969-4765(13)70128-0

Birhane A, Prabhu VU, Kahembwe E. (2021). Multimodal datasets: misogyny, pornography, and malignant stereotypes. arXiv, https://doi.org/10.48550/arXiv.2110.01963

J. Blacking 1976. How musical is man? Faber and Faber.

Buhalis D, Moldavska I (2022) Voice assistants in hospitality: using artificial intelligence for customer service. J Hosp Tour Technol 13(3):386–403. https://doi.org/10.1108/JHTT-03-2021-0104

Canavan A, Graff D, Zipperlen G (1997) CALLHOME American english speech. Linguist Data Consort. https://doi.org/10.35111/EXQ3-X930

Carroll SR, Garba I, Figueroa-Rodríguez OL et al (2020) The CARE principles for indigenous data governance. Data Sci J. https://doi.org/10.5334/dsj-2020-043

Cavarero A (2005) For more than one voice: toward a philosophy of vocal expression. Translated by Paul A. Kottman. Stanford University Press

Chen W, Keast J, Moody J, et al. (2019). "Data usage in MIR: history and future recommendations." In Proceedings of the 20th International Society for Music Information Retrieval Conference. ISMIR, Nov 4. https://doi.org/10.5281/zenodo.3527733

Chen J, Yoshida N, Takada H. (2023). an investigation of licensing of datasets for machine learning based on the GQM model. arXiv:2303.13735. Preprint, arXiv, March 24. https://doi.org/10.48550/arXiv.2303.13735

Cho Y-P, Yang FR, Chang YC, Cheng CT, Wang XH, Liu YW. (2021). A survey on recent deep learning-driven singing voice synthesis systems. arXiv:2110.02511. Preprint, arXiv, October 6. https://doi.org/10.48550/arXiv.2110.02511

Christen K, Davis L, Griffith Z, Neely J. (2018). Traditional Knowledge and digital archives: an interview with Kim Christen. https://doi.org/10.13023/DISCLOSURE.27.02

Christodoulou AM, Jensenius AR. (2024). Navigating challenges in multimodal music data management for AI systems. Proceedings of the AI Music Creativity Conference (AIMC 2024). https://aimc2024.pubpub.org/pub/4y99xcm3

Coscarelli J. (2023). An A.I. hit of fake 'Drake' and 'The Weeknd' rattles the music world. arts. The New York Times, April 19. https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html

Cotton K, De Vries K, Tatar K. (2024). Singing for the missing: bringing the body back to AI voice and speech technologies. Proceedings of the 9th International Conference on Movement and Computing (New York, NY, USA), MOCO '24, June 27, 1–12. https://doi.org/10.1145/3658852.3659065

Cowen AS, Laukka P, Elfenbein HA, Liu R, Keltner D (2019) The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. Nat Hum Behav 3(4):369–382. https://doi.org/10.1038/s41562-019-0533-6

Creamer E. (2025) Collective licence to ensure UK authors get paid for works used to train AI. Books. The Guardian, April 23. https://www.theguardian.com/books/2025/apr/23/collective-licence-to-ensure-uk-authors-get-paid-for-works-used-to-train-ai

Creative Commons (2025). "CC Signals implementation." Creative Commons. https://creativecommons.org/ai-and-the-commons/cc-signals/implementation/

Cross C (2022) Using artificial intelligence (AI) and deepfakes to deceive victims: the need to rethink current romance fraud prevention messaging. Crime Prev Commun Safety 24(1):30–41. https://doi.org/10.1057/s41300-021-00134-w

Cucci D. (2023). I was the original voice of Siri. Even though apple used my voice without my knowledge, it's been a fun ride. Business Insider, February 20. https://www.businessinsider.com/original-voice-of-siri-voice-actor-apple-used-her-voice-2023-2

Cuesta H, Gómez E, Chandna P. (2019). A framework for Multi-F0 modeling in SATB choir recordings. arXiv Preprint. http://arxiv.org/abs/1904.05086v1

Cuesta H, McFee B, Gómez E. (2020). Multiple F0 estimation in vocal ensembles using convolutional neural networks. arXiv Preprint. http://arxiv.org/abs/2009.04172v1

Cui J, Gu Y, Weng C, Zhang J, Chen L, Dai L. (2024). Sifisinger: a high-fidelity end-to-end singing voice synthesizer based on source-filter model. ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April, 11126–30. https://doi.org/10.1109/ICASSP48485.2024.10446786

Cunningham SJ, Lee JH. 2013. Influences of ISMIR and MIREX research on technology patents. In Proceedings of the 14th International Society for Music Information Retrieval Conference. ISMIR, November 4. https://doi.org/10.5281/zenodo.1417303

Cuthrell P. (2024). AI music revolution sets stage for 'Dataset Ethics.' Fortune, April. https://fortune.com/2024/04/29/ai-music-revolution-sets-stage-for-dataset-ethics/

Dale R (2022) The voice synthesis business: 2022 update. Nat Lang Eng 28(3):401–408. https://doi.org/10.1017/S135132492200014 6

Ding J, Jäger E, Ivanova V, Bunz M. (2024). My voice, your voice, our voice: attitudes towards collective governance of a choral AI dataset. arXiv:2412.01433. Preprint, arXiv, December 2. https://doi.org/10.48550/arXiv.2412.01433

Duarte M (2020) Of course, data can never fully represent reality: assessing the relationship between indigenous data and indigenous knowledge, traditional ecological knowledge, and traditional knowledge. Hum Biol 91(3):163–178

Edwards PN, Mayernik MS, Batcheller AL, Bowker GC, Borgman CL (2011) Science friction: data, metadata, and collaboration. Soc Stud Sci 41(5):667–690. https://doi.org/10.1177/0306312711413314

Eidsheim NS (2015) Sensing sound: singing and listening as vibrational practice. Sign, Storage, Transmission. Duke University Press

Eidsheim N, Meizel K, Eidsheim N, Meizel K, (2019). The Oxford Handbook of Voice Studies. Oxford Handbooks. Oxford University Press

Esquivel P, Gill K, Goldberg M, Sundaram SA, Morris L, Ding D (2024) Voice assistant utilization among the disability community for independent living: a rapid review of recent evidence. Human Behav Emerg Technol. https://doi.org/10.1155/2024/6494944

Feld S. (2012) Sound and sentiment: birds, weeping, poetics, and song in Kaluli expression. 3rd ed. Duke University press

Ganguzza A. (2021). Understanding digital voice licensing and usage rights for voice actors. Anne Ganguzza, May 24. https://www.anneganguzza.com/post/understanding-digital-voice-licensing-and-usage-rights-for-voiceover-actors

Gebru T, Morgenstern J, Vecchione B, et al. (2021). Datasheets for datasets. arXiv:1803.09010. Preprint, arXiv, December 1. https://doi.org/10.48550/arXiv.1803.09010

Geerling J. (2024a). They stole my voice with AI. https://www.jeffgeerling.com/blog/2024/they-stole-my-voice-ai

Geerling J, (2024b). They stole my voice with AI (UPDATE: Elecrow Responded). https://www.youtube.com/watch?v=UMofZIT9FcQ

González-Rodríguez J, Toledano DT, Ortega-García J. (2008) Voice biometrics. In Handbook of Biometrics, edited by Anil K. Jain, Patrick Flynn, and Arun A. Ross. Springer US. https://doi.org/10.1007/978-0-387-71041-9_8

Gotham M, Bemman B, Vatolkin I. (2025). Towards an 'Everything Corpus': a framework and guidelines for the curation of more comprehensive multimodal music data | Transactions of the International Society for Music Information Retrieval. Feb 7. https://doi.org/10.5334/tismir.228

Gray JE (2022). "What Can Feminism Do for AI Ethics?" MLearning.AI, Feb 3. https://medium.com/mlearning-ai/what-can-feminism-do-for-ai-ethics-b7e401889441

Gray J, Witt A. (2021). A feminist data ethics of care for machine learning: the what, why, who and how. First Monday, ahead of print, Dec 8. https://doi.org/10.5210/fm.v26i12.11833

Grewal D, Guha A, Schweiger E, Ludwig S, Wetzels M (2022) How communications by AI-enabled voice assistants impact the customer journey. J Serv Manag 33(4/5):705–720. https://doi.org/10.1108/JOSM-11-2021-0452

Gu X, Zeng W, Wang Y. (2023). Elucidate gender fairness in singing voice transcription. arXiv:2308.02898. Preprint, arXiv, August 5. https://doi.org/10.48550/arXiv.2308.02898

Hampton LM. (2021). "Black Feminist Musings on Algorithmic Oppression." Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (New York, NY, USA), FAccT '21, Mar 3, https://doi.org/10.1145/3442188.3445929

Hao K. (2021). Stop talking about AI ethics. It's time to talk about power. MIT Technology Review. https://www.technologyreview.com/2021/04/23/1023549/kate-crawford-atlas-of-ai-review/

He J, Liu J, Ye Z, et al. (2023). RMSSinger: realistic-music-score based singing voice synthesis

Hillmann DI, Marker R, Brady C (2008) Metadata standards and applications. Ser Libr 54(1–2):7–21. https://doi.org/10.1080/03615260801973364

Holzapfel A, Stylianou Y (2007) Singer Identification in Rembetiko Music. Proceedings of the 4th Sound and Music Computing Conference, 326–29. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-193766

Holzapfel A, Sturm BL, Coeckelbergh M (2018) Ethical dimensions of music information retrieval technology. Trans Int Soc Music Inf Retr. https://doi.org/10.5334/tismir.13

Huq SM, Maskeliūnas R, Damaševičius R (2024) Dialogue agents for artificial intelligence-based conversational systems for cognitively disabled: a systematic review. Disabil Rehabil Assist Technol 19(3):1059–1078. https://doi.org/10.1080/17483107.2022.2146768

Hutiri W, Papakyriakopoulos O, Xiang A. (2024). Not my voice! A taxonomy of ethical and safety harms of speech generators. The 2024 ACM Conference on Fairness, Accountability, and Transparency, June 3, 359–76. https://doi.org/10.1145/3630106.3658911

Jeong E, Lim J (2019) Study on intelligence (AI) detection model about telecommunication finance fraud accident. J Kor Inst Info Sec Cryptol 29(1):149–164. https://doi.org/10.13089/JKIISC.2019.29.1.149

Johnson K. (2019). AI ethics is all about power. VentureBeat, November 11. https://venturebeat.com/ai/ai-ethics-is-all-about-power/

Johnson DG, Verdicchio M (2017) Reframing AI discourse. Mind Mach 27(4):575–590. https://doi.org/10.1007/s11023-017-9417-6

Johnson DG, Verdicchio M (2025) The sociotechnical entanglement of AI and values. AI Soc 40:67–76. https://doi.org/10.1007/s00146-023-01852-5

Kaila AK, Jääskeläinen P, Holzapfel A. (2023). Ethically aligned stakeholder elicitation (EASE): case study in Music-AI. Paper presented at New Interfaces for Musical Expression (NIME), Mexico City, Mexico, 31 May–2 June. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-328254

Kelsey F (1949) What is singing? Music Lett 30(3):216–230

Khalid N, Qayyum A, Bilal M, Al-Fuqaha A, Qadir J (2023) Privacy-preserving artificial intelligence in healthcare: techniques and applications. Comput Biol Med 158:106848. https://doi.org/10.1016/j.compbiomed.2023.106848

King J. (2023) We need to stop praising the advancement of AI art and voice acting. TheGamer, January 11. https://www.thegamer.com/future-of-ai-art-voice-acting-ethics-industry-games/

Leavy S, Siapera E, O'Sullivan B. 2021. Ethical data curation for AI: an approach based on feminist epistemology and critical theories of Race. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery. https://doi.org/10.1145/3461702.3462598

Lee K, Hitt G, Terada E, Lee JH. 2022. Ethics of singing voice synthesis: perceptions of users and developers. Proceedings of the 23rd International Society for Music Information Retrieval, 733–40. https://doi.org/10.5281/zenodo.7316768

Lee, K, Cooper AF, Grimmelmann J. 2023. Talkin' about AI generation: copyright and the generative-AI supply chain. SSRN Scholarly Paper No. 4523551. Social Science Research Network, July 27. https://doi.org/10.2139/ssrn.4523551.

Leix Palumbo D, Prey R (2024) Sounding out voice biometrics: comparing and contrasting how the state and the private sector determine identity through voice. Big Data Soc 11(4):20539517241297890. https://doi.org/10.1177/20539517241297889

Liu CLJ, Ren Y, Chen F, Zhao Z. 2022. DiffSinger: singing voice synthesis via shallow diffusion mechanism. Proceedings of the AAAI Conference on Artificial Intelligence 36 (10): 10. https://doi.org/10.1609/aaai.v36i10.21350

Longpre S, Mahari R, Lee A, et al. 2024. Consent in crisis: the rapid decline of the AI data commons. arXiv:2407.14933. Preprint, arXiv, July 24. https://doi.org/10.48550/arXiv.2407.14933

Ma M, Huang C-ZA, Toyama K, Liang B (2023) Beyond diverse datasets: responsible MIR, interdisciplinarity, and the fractured worlds of music. Trans Int Soc Music Info Retr. https://doi.org/10.5334/tismir.141

Mahdiraji A, Hannan KH, Kord H, Kamardi AA (2022) Analysing the voice of customers by a hybrid fuzzy decision-making approach in a developing country's automotive market. Manag Decis 60(2):399–425. https://doi.org/10.1108/MD-12-2019-1732

Mannie K. (2023). AI kidnapping scam copied teen girl's voice in $1M extortion attempt - National | Globalnews.Ca. Global News, April 18. https://globalnews.ca/news/9629883/ai-kidnapping-scam-teen-girl-voice-cloned-extortion-arizona-jennifer-destefano/

McEvoy C. (2023). Vocal AI deepfakes of major artists are cropping up everywhere – should artists be worried? MusicTech. https://musictech.com/features/music-deepfakes-ai-drake-grimes-weeknd/

McPherson EF. (2003). Voice misappropriation in California - Bette Midler, Tom Waits, and Grandma Burger. McPherson LLP. https://mcpherson-llp.com/articles/voice-misappropriation-in-california-bette-midler-tom-waits-and-grandma-burger/

Midler v. Ford Motor Co. (1988), 849 F.2d 460 (9th Cir. 1988) (US Court of Appeals for the Ninth Circuit - 849 F.2d 460 June 22, 1988). https://law.justia.com/cases/federal/appellate-courts/F2/849/460/37485/

Morreale F, Sharma M, Wei IC. 2023. Data collection in music generation training sets: a critical analysis. Proceedings of the 24th International Society for Music Information Retrieval Conference (Milan, Italy). https://doi.org/10.5281/zenodo.10265217

Muldoon J, Wu BA (2023) Artificial intelligence in the colonial matrix of power. Philos Technol 36(4):80. https://doi.org/10.1007/s13347-023-00687-8

Murdoch B (2021) Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med Ethics 22(1):122. https://doi.org/10.1186/s12910-021-00687-3

Naik V, Kulkarni AMS, Naik S, Verlekar SP. (2022). Voice cloning in real time. International Journal for Research in Applied Science and Engineering Technology 10 (8): 1443–46. https://doi.org/10.22214/ijraset.2022.44524

Napolitano D (2022) AI voice between anthropocentrism and post-humanism: Alexa and voice cloning. J Interdiscip Voice Stud 7:35–49. https://doi.org/10.1386/jivs_00053_1

Newcombe R (2003) From client to project stakeholders: a stakeholder mapping approach. Constr Manag Econ 21(8):841–848. https://doi.org/10.1080/0144619032000072137

Newlands G (2021) Lifting the curtain: strategic visibility of human labour in AI-as-a-service. Big Data Soc 8(1):20539517211016024. https://doi.org/10.1177/20539517211016026

Nnabueze Ajimah E, Iloanusi ON (2024) Biometric voice recognition system in the context of multiple languages: using traditional means of identification of individuals in Nigeria languages and English language. Res Stat 2(1):2362298. https://doi.org/10.1080/27684520.2024.2362298

Noble SU. (2018). Algorithms of oppression: how search engines reinforce racism. NYU Press. https://nyupress.org/9781479837243/algorithms-of-oppression

Norms L, dir. (2024). AI stole my voice, so i stole it back. https://www.youtube.com/watch?v=Srr1tYyBkn0

Olsen A. (2023). Voice cloning technology and its legal implications: an IP law perspective - Schmeiser, Olsen & Watts, LLP. Schmeiser Olsen and Watts LLP, August 26. https://iplawusa.com/voice-cloning-technology-and-its-legal-implications-an-ip-law-perspective/

Örpek Z, Tural B, Özmen S. (2024). Identity verification processes with voice biometrics and artificial intelligence

Parkinson, HJ. (2015). Hey, Siri! Meet the real people behind Apple's voice-activated assistant. Technology. The Guardian, Aug 12. https://www.theguardian.com/technology/2015/aug/12/siri-real-voices-apple-ios-assistant-jon-briggs-susan-bennett-karen-jacobsen

Pasquale F, Sun H. (2024). Consent and compensation: resolving generative AI's copyright crisis. SSRN Scholarly Paper No.

4826695. Social Science Research Network, May 1. https://doi.org/10.2139/ssrn.4826695
Patel P (2023) AI voice enters the copyright regime: proposal of a three-part framework. Fordham Intellect Prop Media Entertain Law J 34:451
Pathak S, Shreshtha S, Singh R, Vatsa M. (2025). Quantum-inspired audio unlearning: towards privacy-preserving voice biometrics. arXiv:2507.22208. Preprint, arXiv, July 29. https://doi.org/10.48550/arXiv.2507.22208
Peña P, Varon J. 2019. Consent to our data bodies: lessons from feminist theories to enforce data protection. Association for Progressive Communications. https://codingrights.org/docs/ConsentToOurDataBodies.pdf
Proutskova P, McBride JM, Ozaki Y, et al. 2023. The VocalNotes Dataset. Proceedings of the First MiniCon Conference, November 5, 3. https://ismir2023program.ismir.net/lbd_354.html
Przybocki M, Alvin M (2001) 2000 NIST speaker recognition evaluation. Linguist Data Consort. https://doi.org/10.35111/EX24-J205
Ravitz J. (2013). I'm the Original Voice of Siri | CNN Business." CNN, October 4. https://www.cnn.com/2013/10/04/tech/mobile/bennett-siri-iphone-voice/index.html
Regulation (EU) 2016/679 (2016). https://op.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1/language-en
Regulation (EU) 2024/1689 (2024). http://data.europa.eu/eli/reg/2024/1689/oj/eng
Ricaurte P (2019) Data epistemologies, the coloniality of power, and resistance. Television New Media 20(4):350–365. https://doi.org/10.1177/1527476419831640
Rosenzweig S, Cuesta H, Weiß C, Scherbaum F, Gómez E, Müller M. (2020). Dagstuhl ChoirSet: a multitrack dataset for MIR research on choral singing | Transactions of the International Society for Music Information Retrieval. Mar 5. https://doi.org/10.5334/tismir.48
Rothman JE (2018). The right of publicity: Privacy reimagined for a public world. Harvard University Press
Rusti C, Leschanowsky A, Quinlan C, Pnacek M, Gorce L, Hutiri W. (2023). Benchmark dataset dynamics, bias and privacy challenges in voice biometrics research. arXiv:2304.03858. Preprint, arXiv, August 18. https://doi.org/10.48550/arXiv.2304.03858
SAG-AFTRA and Replica Studios. (2024). Replica studios agreement for digital voice replicas. Screen Actors Guild-American Federation of Television and Radio Artists, Jan. https://www.sagaftra.org/sites/default/files/sa_documents/Replica%20Studios%20Agreement%20for%20Digital%20Voice%20Replicas_0.pdf
Savov V. (2011). British voice of Siri only found out about it when he heard himself on TV. The Verge, November 10. https://www.theverge.com/2011/11/10/2551519/british-voice-of-siri-only-found-out-about-it-when-he-heard-himself
Seeger A.(2003). I found it, How can i use it? Keynote Speech. The International Conference on Music Information Retrieval, USA. https://ismir2003.ismir.net/papers/Seeger.pdf
Semary HE, Al-Karawi KA, Abdelwahab MM. (2024) Using voice technologies to support disabled people. J Disabil Res. https://doi.org/10.57197/jdr-2023-0063
Serra, X (2011). A multicultural approach in music information research. In Proceedings of the 12th International Society for Music Information Retrieval Conference. ISMIR, https://doi.org/10.5281/zenodo.1416592
Serra X, Magas M, Benetos E, et al. (2013). Roadmap for music information ReSearch. The MIReS Consortium
Shanfeld E. (2023). Ghostwriter's 'Heart on My Sleeve,' the AI-Generated Song Mimicking Drake and the Weeknd, Submitted for Grammys. Variety, September 6. https://variety.com/2023/music/news/ai-generated-drake-the-weeknd-song-submitted-for-grammys-1235714805/
Sharma T et al. (2025) PRAC3 (Privacy, reputation, accountability, consent, credit, compensation): long tailed risks of voice actors in AI data-economy. https://arxiv.org/abs/2507.16247
Speechify. (2025) Free AI voice generator! no sign-up required. Accessed 5 Mar 2025. https://speechify.com/ai-voice-generator/
Srinivasamurthy ASG, Repetto RC, Serra X (2021) Saraga: open datasets for research on Indian art music. Empir Musicol Rev 16(1):1. https://doi.org/10.18061/emr.v16i1.7641
Sterne J. (2003). The audible past: cultural origins of sound reproduction. Duke University Press
Sterne J (2022) Is machine listening listening? Communication 9(1):1. https://doi.org/10.7275/zeqh-eg38
STIM. (2025). Stim unveils world's first AI license for music. Stim. Se. https://www.stim.se/en/news/stim-launches-the-worlds-first-ai-license-for-music
Stoever JL (2016) The Sonic color line: race and the cultural politics of listening. NYU Press. https://doi.org/10.2307/j.ctt1bj4s55
Sundberg J (1989) Science of the Singing Voice. Northern Illinois University Press
Sundberg J. (1996). The human voice. In: comprehensive human physiology: from cellular mechanisms to integration; (Eds: R. Greger, U. Windhorst). Springer
Tatar K, Ericson P, Cotton K et al (2023) a shift in artistic practices through artificial intelligence. Leonardo 57(3):4. https://doi.org/10.1162/leon_a_025
The North Africa Institute. (2026). "Swedish Media Database (SMDB)." Text. Accessed 14 Mar 2025. https://nai.uu.se/library/resources/liberation-africa/list-of-archives/sv/swedish-media-database-smdb.html
Tobin G. (2024). AI Sound-Alikes, Voice Generation and Deepfakes: The ELVIS Act. Nov 12. https://www.gtlaw.com.au/insights/ai-sound-alikes,-voice-generation-and-deepfakes-the-elvis-act
Transparent Audio. (2025). Future-proof your Gen AI audio business against transparency regulation. https://www.transparentaudio.ai/
Turino T. (2008). Music as social life: the politics of participation - Chicago Studies in Ethnomusicology CSE (Paperback Book) [Annotated Edition] (2008). https://imusic.au/books/9780226816982/thomas-turino-2008-music-as-social-life-the-politics-of-participation-chicago-studies-in-ethnomusicology-cse-paperback-book
United Kingdom Health Research Authority. (2023). Proof-of-concept: using anonymous or artificial health data. AI and Digital Regulations Service for Health and Social Care, January 20. https://digitalregulations.innovation.nhs.uk/regulations-and-guidance-for-developers/all-developers-guidance/proof-of-concept-using-anonymous-or-artificial-health-data/
U.S. Copyright Office. (2025). Copyright and Artificial Intelligence, Part 3: Generative AI Training. United States Copyright Office. https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf
Varon J, Peña P (2021) Artificial intelligence and consent: a feminist anti-colonial critique. Int Policy Rev 10(4):1–25. https://doi.org/10.14763/2021.4.1602
Volochtchuk AVL, Leite H, Vieira AD (2024) Voice assistant technology applied to populations with developmental and physical disabilities. Behav Info Technol 43(11):2300–2322. https://doi.org/10.1080/0144929X.2023.2243343
Vonica A. (2021). "AI & Tone of Voice for Disability Inclusion." SSRN Scholarly Paper No. 3850276. Social Science Research Network, May 20. https://papers.ssrn.com/abstract=3850276
Wald M (2021) AI data-driven personalisation and disability inclusion. Front Artif Intell. https://doi.org/10.3389/frai.2020.571955
Warman M. (2011). "The Voice behind Siri Breaks His Silence." The Telegraph, Nov 10. https://www.telegraph.co.uk/technology/apple/8879705/The-voice-behind-Siri-breaks-his-silence.html

Weidman A (2014) Anthropology and Voice. Ann Rev Anthropol 43:37–51. https://doi.org/10.1146/annurev-anthro-102313-030050

Weiß C, Zalkow F, Arifi-Müller V et al (2020) Schubert Winterreise Dataset. Zenodo. https://doi.org/10.5281/zenodo.4122060

Wilkins J, Seetharaman P, Wahl A, Pardo B. (2018). Vocalset: a singing voice dataset. Proceedings of the 19th ISMIR Conference, Mar 8, https://doi.org/10.5281/ZENODO.1203819

Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M et al (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3(1):160018. https://doi.org/10.1038/sdata.2016.18

Willman C. (2023). AI-generated fake 'Drake'/'Weeknd' Collaboration, 'Heart on My Sleeve,' delights fans and sets off industry alarm bells. Variety. https://variety.com/2023/music/news/fake-ai-generated-drake-weeknd-collaboration-heart-on-my-sleeve-1235585451/

Yamamoto R, Yoneyama R, Toda T. (2023). NNSVS: a neural network-based singing voice synthesis toolkit. arXiv:2210.15987. Preprint, arXiv. http://arxiv.org/abs/2210.15987

Yang Q, Jin W, Zhang Q et al (2023) Mixed-Modality speech recognition and interaction using a wearable artificial throat. Nature Mach Intell 5(2):169–180. https://doi.org/10.1038/s42256-023-00616-6

Yu CY, Fazekas G. (2023). Singing voice synthesis using differentiable LPC and glottal-flow-inspired wavetables. Nov 4. https://doi.org/10.5281/zenodo.13916489

Zhang Z (2016) Mechanics of human voice production and control. J Acoust Soc Am 140(4):2614–2635. https://doi.org/10.1121/1.4964509

Geiger, Christophe and Bossi, Ludovico and Di Lazzaro, Francesca, The Voss and Jensen Parliamentary Reports on Copyright and Generative AI: A Wrong Step in the Right Direction? (2026). Innovation Law and Ethics Observatory (ILEO) Research Paper Series No. 26-01, Forthcoming in: European Intellectual Property Review (EIPR), 2026, Issue 8, Available at SSRN: https://ssrn.com/abstract=6537278,or https://doi.org/10.2139/ssrn.6537278

European Parliament and Council of the European Union. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC. Official Journal of the European Union L 130, May 17, 2019. https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng

European Parliament. Resolution of 10 March 2026 on Copyright and Generative Artificial Intelligence – Opportunities and Challenges (2025/2058(INI)). P10_TA(2026)0066. Strasbourg, March 10, 2026. https://www.europarl.europa.eu/doceo/document/TA-10-2026-0066_EN.html.

Swedish Research Council (Vetenskapsrådet). Good Research Practice 2024. Stockholm: Swedish Research Council, 2024. https://www.vr.se/download/18.4857b2ae197a1e923f913e07/1751465049917/Good%20research%20practice%202024%20VR.pdf.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.