Rethinking Data Practices for 'Intelligent' Musical Interfaces

ELIZABETH WILSON, Creative Computing Institute, University of the Arts London, UK

Abstract

As artificial intelligence continues to become embedded in musical instrument design, this Alt.NIME paper proposes examining how data is used in any such system. It extends the concept of small data beyond dataset size alone, proposing a framework that emphasises limited, self-curated, marginalised, and situated datasets. To explore current practices, we conduct an exploratory audit of submissions to the NIME 2026 proceedings that employ artificial intelligence or data-driven techniques. The audit investigates how datasets are described, whether their sources are documented and accessible, and the extent to which smaller or situated datasets are used. Through this investigation, we aim to highlight the visibility (or absence) of dataset transparency within NIME research.

1 Context

As data-driven processes become increasingly embedded in musical instrument design, new forms of interaction between performers and technology are emerging. These systems range from those that are explicitly using AI-based instruments to interfaces that rely on data mappings, archives, sensor histories, or rule-based transformations. While artificial intelligence represents a prominent strand of this development, it is only one instance within a broader ecology of data practices in NIME. As a result, musicians are increasingly engaging with instruments that learn from performance behaviour and evolve over time, expanding both expressive possibilities and creative workflows.

Many artificial intelligence algorithms deal with data in some form or another: as the raw material from which AI systems derive patterns, make decisions, and learn. Without data, which is most often about or made by humans, the systems could not learn, and thus cannot gain the required knowledge to function.

Most artificial intelligence systems developed by larger corporations tend to make use of vast datasets which are often hidden or proprietary. The scale and origin of such data raise important questions regarding ownership, consent, and bias. Because the data often reflects existing social and cultural patterns, AI systems may reproduce or amplify those patterns in their outputs [2].

Small data is currently being proposed as a means to counteract some of the wider issues with current AI music practices [3]. Small data, when referred to size of the dataset, looks at specific niches and genre/sub-genres to perform specialised analysis or generation techniques rather than assuming a one-size-fits-all approach with extensive datasets.

Some of the issues with working with larger data, as identified in Bryan-Kinns et al. [3], include the marginalisation of genres, cultures and traditions that exists with the big data framework; the infringement of creative rights where data is scraped from musicians and artists without their consent; and the requirement for extensive computational power to train models which has extreme ecological costs attached.

For this work, we propose extending 'small' data approaches beyond datasets that are exclusively small in size. Instead, this concerns approaches that use datasets that are:

- **Limited**: Prioritising datasets of limited scope of things like recordings or scores, perhaps solely a single performer's gestures, a specific environment's sounds
- **Self-Curated**: Intentionally chosen by the artist, with known provenance and clear authorship
- **Exploring "marginalised" genres/traditions**: This pertains to those existing outside mainstream commercial or academic canons, these could be from oral, folk, experimental or non-Western traditions that are under-represented in large-scale datasets
- **Situated**: Situated data, comes from Donna Haraway's concepts of "situated knowledge" this data is explicitly *context-bound and partial* [4]. For Haraway, she conceives of knowers as situated in particular relations to what is known and to other knowers. What is known, and how it is known, reflects the situation and perspective of the knower. If we know that artificial intelligence is dependent on data, and that data can commonly be understood as "factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation" [1], here lies a problem: understanding data as a fact, or as zeros and ones, flattens their constructed, situated, and timely

Author's Contact Information: Elizabeth Wilson, Creative Computing Institute, University of the Arts London, London, UK.

NIME '26, June 23–26, 2026, London, UK

aspects. Consequently, the concept of data remains categorically different from—and in a sense opposed to—the very idea of process. Instead, this approach does not see data as disembodied and place-less, and instead chooses to adopt a feminist techno-science situated approach.

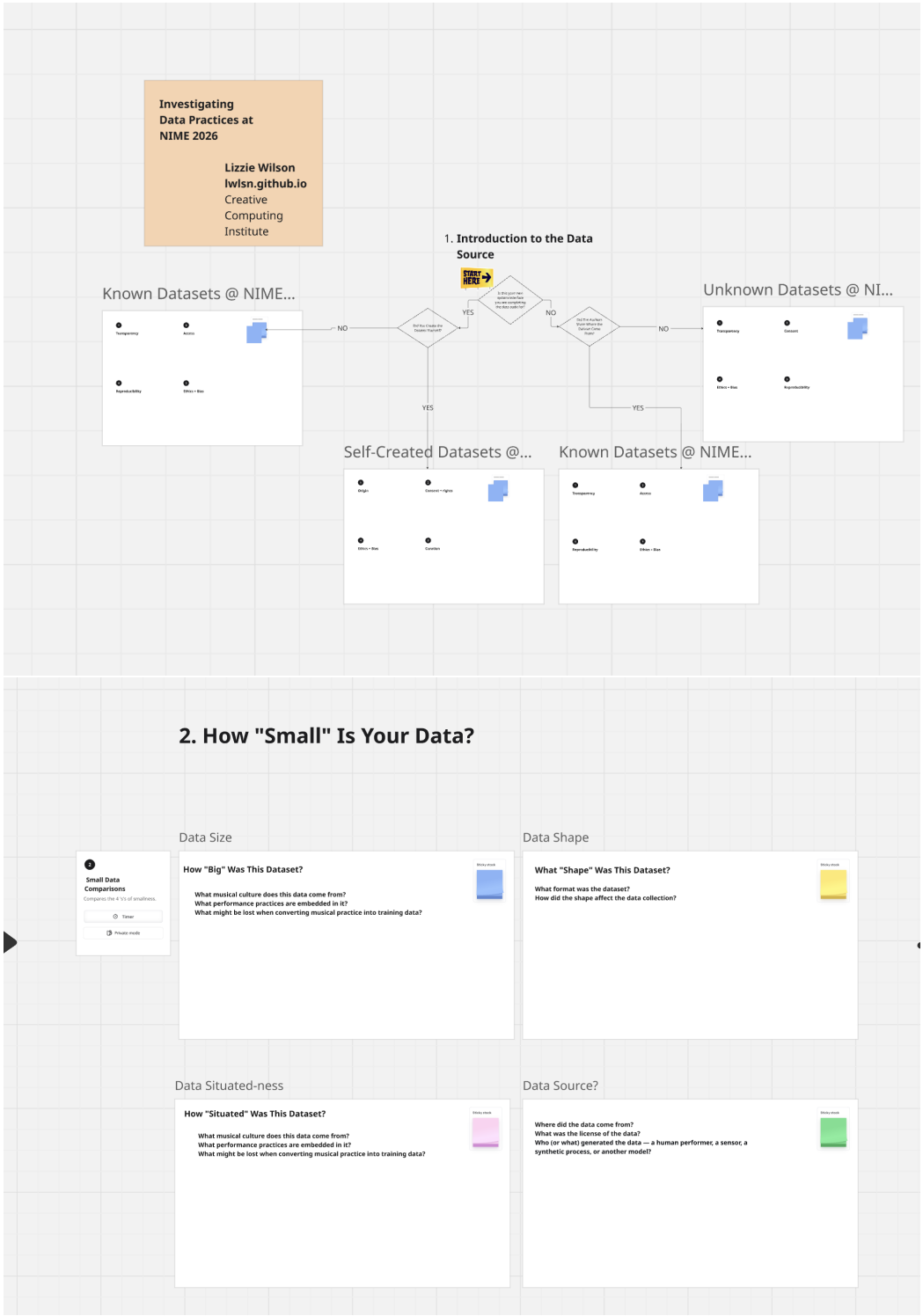

Fig. 1. Miro Board: (Top) A screenshot of the miro board with a flow-diagram to determine the source of the data set as either self-curated, known or unknown.

This expanded understanding of small data motivates an investigation not only of AI-driven systems, but of data practices across NIME more broadly. In particular, this work asks whether notions of "small" or situated data emerge differently within AI contexts compared to non-AI data-driven musical systems.

2 Data Investigations

This Alt.NIME submission proposes an exploratory audit of data practices within the NIME 2026 conference proceedings. The project investigates how data is described, contextualised, and operationalised within musical systems that rely on datasets, archives, performer recordings, sensor histories, machine learning, or other forms of accumulated or curated information.

For the purposes of this audit, "data-based" systems are defined as musical interfaces or systems that rely on stored, recorded, curated, or dynamically collected data as part of their operation, interaction, analysis, or generation processes. This includes, but is not limited to:

- Machine learning and AI systems
- Gesture recognition systems trained on examples
- Corpus-based or retrieval-based systems
- Generative systems using musical archives or datasets
- Sensor-data driven mappings and adaptive interfaces
- Performer-specific or self-curated datasets

Systems that rely solely on deterministic signal processing without the use of accumulated or curated data will not be included in the audit.

2.1 Methodology

The proposed audit of all the submissions of the NIME 2026 conference proceedings wille examine those that use any form of data-based musical practice, especially those which claim 'intelligence'. In particular, we propose the following questions for enquiry:

- Are all systems that use data reporting on the dataset that was used in any training process? Are these datasets easily accessible to the public?
- What proportion of systems that use data are using "small" datasets (with the aforementioned description)?
- What is the "smallest" dataset that has been used at NIME 2026?
- To what extent are datasets self-curated, performer-specific, or otherwise situated within a particular musical practice?
- How frequently do submissions address ethical, legal, or ecological considerations relating to data collection, ownership, and model training?
- How do conceptions and uses of "small data" differ between AI-based systems and non-AI data-driven instruments?

Through this audit, the work seeks to foreground questions of data transparency, provenance, and scale within NIME research. By examining how data is framed and used across the proceedings, we aim to highlight opportunities for more situated, transparent, and artist-led data practices in the design of musical interfaces and AI-driven musical systems. This expanded understanding of small data motivates an investigation not only of AI-driven systems, but of data practices across NIME more broadly.

3 Media Links

A Miro board will be used to document this data investigations, asking participants of the conference who engage with data in their practice to complete reflections. Meanwhile, the author will also attempt to see how transparent data practices are in the NIME submissions.

Link to the Miro board can be found here: tinyurl.com/NIME-smalldata and screenshots can be seen in Figure 1.

4 Ethical Standards

This work combines document analysis of publicly available NIME 2026 conference papers with voluntary participant reflections collected through an online interactive board. The document analysis focuses exclusively on publicly accessible academic materials and examines them solely for the purpose of understanding how datasets and data practices are described within NIME research.

Participant reflections are voluntary, may be submitted anonymously, and are intended as short responses to prompts concerning data practices in musical work. Participants will be encouraged not to disclose sensitive or commercially confidential information.

Care has been taken to approach both the audit and participant reflections in a constructive and reflective manner. The aim is not to critique individual authors or projects, but to identify broader trends regarding transparency, provenance, and situated-ness within data-driven musical interface research.

References

[1] Cambridge University Press & Assessment 2026. *Data*. Cambridge University Press & Assessment. https://dictionary.cambridge.org/dictionary/english/data Definition of "data" from the Cambridge Advanced Learner's Dictionary & Thesaurus.

[2] Jean-Pierre Briot and François Pachet. 2020. Deep learning for music generation: challenges and directions. *Neural Computing and Applications* 32, 4 (2020), 981–993.

[3] Nick Bryan-Kinns, Anna Wszeborowska, Olga Sutskova, Elizabeth Wilson, Phoenix Perry, Rebecca Fiebrink, Gabriel Vigliensoni, Rikard Lindell, Andrei Coronel, and Nuno N Correia. 2025. Leveraging small datasets for ethical and responsible AI music making. In *Proceedings of the 20th International Audio Mostly Conference*. 70–81.

[4] Donna Haraway. 2013. Situated knowledges: The science question in feminism and the privilege of partial perspective 1. In *Women, science, and technology*. Routledge, 455–472.