Introduction

Welcome to the 3rd Workshop on NLP for Music and Audio (NLP4MusA)! NLP4MusA aims to bring together researchers from various disciplines related to music and audio content on one hand, and NLP on the other. It focuses on topics such as:

- NLP approaches applied to music analysis and generation
- Lyrics analysis and generation
- Exploiting music related texts in music recommendation
- Use of LLMs in music and spoken audio contexts
- Taxonomy learning
- Podcasts recommendations
- Music captioning
- Multimodal representations

The workshop is held in Oakland (CA), in conjunction with ISMIR 2024, and spans half a day featuring a keynote followed by presentations of the accepted papers through short talks and a poster session. The accepted papers cover topics of high relevance to the intersection of music, audio and NLP, including the use of large language models for music tasks such as recommendation, information extraction and lyrics analysis. They also explore the creation and use of new datasets for training and evaluating these models, as well as the generalization capabilities of these models across languages, musical genres and cultural contexts.

We are honored to have Noah Smith (University of Washington, Allen Institute for Artificial Intelligence) as our keynote speaker for this edition of the workshop. His talk explores the challenge of building a language model for MIR. We include the abstract of his talks in this volume.

In response to our call for papers, we received 33 submissions. Each submission was rigorously reviewed by two Program Committee members selected for their expertise. Based on the reviewers' feedback, we accepted 18 papers (55%).

We are extremely grateful to the authors for their valuable contributions and to the Programme Committee members for their detailed and helpful reviews. We also thank our sponsor, Deezer, and our host SiriusXM, who have helped making the workshop possible in this form. We hope you find the workshop insightful and inspiring!

Anna Kruspe, Sergio Oramas, Elena V. Epure, Mohamed Sordo, Benno Weck, SeungHeon Doh, Minz Won, Ilaria Manco, Gabriel Meseguer Brocal

November 2024

Organisers:

Anna Kruspe (Munich University of Applied Sciences)

Sergio Oramas (SiriusXM)

Elena V. Epure (Deezer)

Mohamed Sordo (SiriusXM)

Benno Weck (UPF)

SeungHeon Doh (KAIST)

Minz Won (Suno)

Ilaria Manco (QMUL)

Gabriel Meseguer Brocal (Deezer)

Program Committee:

Morteza Behrooz (Worcester Polytechnic Institute)

Dimitrios Bralios (University of Illinois Urbana-Champaign)

Jonah Casebeer (Adobe Research)

Keunwoo Choi (Genentech)

Shuqi Dai (Carnegie Mellon University)

Matthew Davies (SiriusXM)

Hao-Wen Dong (University of Michigan)

Andres Ferraro (SiriusXM)

Masataka Goto (AIST)

Haven Kim (University of California San Diego)

Junghyun Koo (Seoul National University)

Jongpil Lee (Neutune)

Junwon Lee (KAIST)

Kongmeng Liew (University of Canterbury)

Pasquale Lisena (EURECOM)

Manuel Moussallam (Deezer)

Zachary Novack (University of California San Diego)

Julian Parker (Native Instruments)

Lorenzo Porcaro (Joint Research Centre, European Commission)

Bruno Sguerra (Deezer)

Rosa Stern (Sonos)

Pragati Verma (SiriusXM)

Kento Watanabe (AIST)

Shuo Zhang (Georgetown University)

Yixiao Zhang (QMUL)

Ge Zhu (University of Rochester)

Invited Speaker:

Noah Smith (University of Washington, Allen Institute for Artificial Intelligence)

Invited Talk

Noah Smith: Imagining a Music Language Model

Language models and their multimodal variants now present many exciting new opportunities for advancing music processing applications. As a researcher in natural language processing for over twenty-five years, and as a musician for even longer, I've recently started learning about music IR and related challenges with some of my students. In this talk, I'll offer some opinionated observations, technical ideas, and lessons learned from NLP that I believe could be useful for the music processing community. These include matters of evaluation methodology, the roles of data and theory, the framing of problems, and guiding questions about who we are building technology for. I'll also reflect a bit on our efforts to improve the state of the art in open (multimodal) language models and speculate about a grand challenge: building a (hopefully open) language model for MIR.

Table of Contents