

NLP4MusA 2026

The 4th Workshop on NLP for Music and Audio

Proceedings of the Workshop

March 28, 2026

The NLP4MusA organizers gratefully acknowledge the support from the following sponsors.

Sponsors



©2026 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-369-2

Introduction

Welcome to the 4th Workshop on NLP for Music and Audio (NLP4MusA 2026)! The workshop is held in Rabat, Morocco, in conjunction with EACL 2026, and spans half a day, featuring four keynote talks followed by lightning presentations of accepted papers and a poster session.

Building on a tradition of cross-disciplinary impact, the intersection of NLP with music and audio-based creative media presents a frontier full of unique challenges and exciting opportunities. The Fourth Workshop on Natural Language Processing for Music and Audio (NLP4MusA) aims to explore the multimodal synergies between language, music, and sound. As NLP increasingly enables domains where language and interaction converge, the entertainment industry offers a particularly compelling case: most audio content - such as songs or podcasts - contains an inherent linguistic dimension, while user engagement often occurs through language, from search queries to social media conversations. Recent advances in language modeling and multimodal representation learning have deepened the connection between NLP and music, giving rise to a new generation of applications. These range from music transcription and captioning to the generation of lyrics, symbolic compositions, and even complex audio signals from text-based prompts. At the same time, techniques from Question Answering and Information Retrieval are transforming how people search for, recommend, and interact with music and audio. Because language also reflects culture and multilingual diversity, NLP-driven approaches open pathways for cross-cultural discovery and creativity - enriching the way we experience and understand music and sound.

The workshop focuses on topics such as:

1) NLP for Music and Audio Understanding

- Music Tagging and Auto-tagging, Knowledge Graph Construction, Semantic Ontologies
- Information Extraction, Named Entity Recognition, and Entity Linking
- Multimodal Representation Learning, Lyrics and Symbolic Representation Analysis
- Emotion and Sentiment Analysis, Culture-specific Music Understanding, Corpora Bias
- Music Captioning and Description Generation

2) NLP for Music Retrieval or Recommendation

- Conversational Interfaces, Query understanding and Intent Prediction
- Multimodal, Cross-modal Music Information Retrieval and Recommender Systems
- Natural Language User Modeling
- Music Question Answering
- Fairness and Transparency

3) NLP for Music and Audio Generation

- Lyrics Generation, Audio/Symbolic Query-driven Music Generation
- Synthetic Music Content Detection

In response to our call for papers, we received 19 submissions. After an initial screening, two submissions were desk-rejected and one was eventually withdrawn by the authors. Of the remaining submissions, each was rigorously reviewed by at least two Program Committee members selected for their expertise. Based on the reviewers' feedback, we accepted 10 papers (53% acceptance rate), of which one was not included in the proceedings due to a double submission.

We are extremely grateful to the authors for their valuable contributions and to the Program Committee members for their detailed and constructive reviews. We also thank our sponsors and the EACL Workshop Chairs for their support and prompt responses.

We hope you find the workshop insightful and inspiring!

Elena V. Epure, Sergio Oramas, SeungHeon Doh, Pedro Ramoneda, Anna Kruspe, Mohamed Sordo

Organizing Committee

Organizing Committee

Elena Epure, Deezer (France) / Idiap Research Institute (Switzerland)

Sergio Oramas, SiriusXM (Spain)

SeungHeon Doh, KAIST (South Korea)

Pedro Ramoneda, Universitat Pompeu Fabra / Songscription (Spain)

Anna Kruspe, Munich University of Applied Sciences (Germany)

Mohamed Sordo, SiriusXM (USA)

Program Committee

Reviewers

Anna Aljanaki, University of Tartu
Hayeon Bang, KAIST
Morteza Behrooz, Meta AI
Hans-Ulrich Berendes, AudioLabs Erlangen
Dimitrios Bralios, University of Illinois Urbana-Champaign
Eunjin Choi, KAIST
Matthew Davies, Sirius XM
SeungHeon Doh, KAIST
Robie Gonzales, Dalhousie University
Simon Hachmeier, Humboldt-Universität zu Berlin
Mikaela Keller, Université de Lille
Haven Kim, University of California, San Diego
Soham Korade, IIT Hyderabad
Daeyong Kwon, KAIST
Dinh-Viet-Toan Le, Université de Lille
Jongpil Lee, Neutune
Junwon Lee, KAIST
Linxi Li, University of Warwick
Pasquale Lisena, EURECOM
Benno Weck, Universitat Pompeu Fabra
Zachary Novack, UC San Diego
Julian D Parker, Stability AI
Amit Namburi, UC San Diego
Yusuke Sakai, NAIST
Simon Schwär, AudioLabs Erlangen
Suswara Pochampally, IIITH
Rosa Stern, Sonos
Giulio Prevedello, Sony CSL
Yongyi Zang, Smule, Inc.
Shuo Zhang, Tufts University
Yixiao Zhang, ByteDance Inc.
Kongmeng Liew, University of Canterbury
Kento Watanabe, AIST
Aleksandr Petrov, Spotify
Mucun Tian, Amazon

Keynote Talk

Invited 1

Willem Zuidema
University of Amsterdam

Bio: Willem Zuidema is associate professor of NLP and Explainable AI at the Institute for Logic, Language and Computation (ILLC) in Amsterdam. He has published widely on computational models of language, including comparisons with music, animal communication and explorations of their evolutionary origins. He has done pioneering work in deep learning for NLP and interpretability methods for LSTMs and Transformers. He leads the InDeep consortium, focused on interpretability for text, speech and music, involving 7 PhD students and 5 universities in the Netherlands.

Keynote Talk

Invited 2

Harin Lee
University of Cambridge

Bio: Harin Lee is a multidisciplinary researcher combining large-scale data analysis with cross-cultural experiments to study the psychological foundations of music cognition and musical diversity worldwide. His research includes field experiments with Tsimané villagers in the Bolivian Amazon and developing human-in-the-loop online paradigms to investigate cultural evolution in artificial worlds. Currently a Junior Research Fellow at University of Cambridge, he earned his PhD at the Max Planck Institute for Human Cognitive and Brain Sciences.

Keynote Talk

Invited 3

Ilaria Manco
Google DeepMind

Bio: Iaria Manco is a Research Scientist in the Magenta team at Google DeepMind. Her research spans music generation and understanding, with a current focus on new forms of musical interaction via controllable, real-time generative models. Iaria received her PhD from Queen Mary University of London, where she developed multimodal representation learning approaches to connect music and language. During her doctoral work she also collaborated with Universal Music Group on large-scale audio-caption datasets and audio-language models.

Keynote Talk

Invited 4

Enrico Palumbo
Spotify

Bio: Enrico Palumbo researches and builds Generative AI features for Search and Recommendations at Spotify, with a focus on agentic technologies and generative recommendations. Before joining Spotify, he was a Research Scientist at Amazon, developing language understanding models for Alexa in non-English locales. He holds a PhD on Knowledge Graph Embeddings for Recommender Systems, carried out jointly between the Polytechnic University of Turin, EURECOM, and Links Foundation.

Table of Contents

<i>From Novice to Expert: Generating Audience-Dependent Concert Moderations with RAG-LLMs</i> Kerstin Denecke	1
<i>LabelBuddy: An Open Source Music and Audio Language Annotation Tagging Tool Using AI Assistance</i> Ioannis Prokopiou, Ioannis Sina, Agisilaos Kounelis, Pantelis Vikatos and Themis Stafylakis ..	7
<i>Stochastic Parrots or True Virtuosos? Digging Deeper Into the Audio-Video Understanding of AVQA Models</i> Sara Pernille Jensen, Hallvard Innset Hurum and Anna-Maria Christodoulou	13
<i>Beyond Musical Descriptors: Extracting Preference-Bearing Intent in Music Queries</i> Marion Baranes, Romain Hennequin and Elena V. Epure	20
<i>How Far Can Pretrained LLMs Go in Symbolic Music? Controlled Comparisons of Supervised and Preference-based Adaptation</i> Deepak Kumar, Emmanouil Karystinaios, Gerhard Widmer and Markus Schedl	27
<i>MIDI-PHOR: Multi-View Distillation for Music Understanding and Captioning</i> Steven Au	33
<i>Read Between the Tracks: Exploring LLM-driven Intent-based Music Recommendations</i> Anna Hausberger, Petra Jósár and Markus Schedl	44
<i>Learning When to Personalize: LLM Based Playlist Generation via Query Taxonomy and Classification</i> Fedor Buzaev, Ivan Sukharev, Rinat Mullahmetov, Roman Bogachev, Ilya Sedunov, Oleg Pavlovich and Daria Pugacheva	51
<i>HumMusQA: A Human-written Music Understanding QA Benchmark Dataset</i> Benno Weck, Pablo Puentes, Andrea Poltronieri, Satyajeet Prabhu and Dmitry Bogdanov	58