Model Fairness Meets Source Tracing: Toward Trustworthy AI for Manipulated Speech Attribution

Special Session: Odyssey 2026

Contact: jagabandhu.mishra@uef.fi | nicolas.mueller@aisec.fraunhofer.de

Overview and Objectives

Source tracing is the digital forensic process of attributing synthetic or manipulated audio to its generative origin. It seeks to answer a critical question: **"Which specific source (TTS/VC) system created this audio deepfake?"**

By identifying the source, whether it be Vendor A, Vendor B, or an open-source model-platform, providers and authorities can take decisive action, such as closing malicious accounts, tracking the spread of coordinated disinformation campaigns, and improving the accountability of generative AI.

However, many current models rely on **"shortcuts"**-spurious correlations like speaker identity, language, or recording conditions. While these models may perform perfectly in laboratory environments, they often fail in real-world scenarios and do not generalize to new conditions or unseen generative models.

This special session explicitly addresses these challenges. We aim to foster the development of **fair and robust attribution methods** that capture genuine system-specific fingerprints and demonstrate real-world generalization.

Scope and Topics

We invite original contributions including, but not limited to:

- **Closed-Set Attribution:** Supervised learning approaches classifying an seen audio file into one of N seen attack classes.
- **Open-Set Attribution:** Using either verification or identification style design (Clustering or embedding-based systems) to determine if two audio files originate from the same vendor/system without prior category knowledge.
- **Bias & Shortcut Mitigation:** Methods to decouple language and speaker identity from attack fingerprints.
- **Trustworthy AI:** Explainable and transparent frameworks for forensic attribution.
- **Generalization:** Cross-model, cross-language, and cross-dataset performance analysis.
- **Digital Forensics:** Practical application of source tracing in legal and cybersecurity contexts.

Resources and Benchmarking

The organizers provide a standardized evaluation protocol based on the MLAAD dataset. Participants are invited to train their systems on the provided `train` and `dev` splits and report results on the `eval` portion to ensure fair comparison.

Protocol Download: https://deepfake-total.com/sourcetracing

Submission Information

Authors should follow the official Odyssey 2026 guidelines and templates.

- **Submission Deadline:** March 15th, 2026
- **Submission Portal:** https://cmt3.research.microsoft.com/ODYSSEY2026/
- **Subject Area:** Select **6.02 "Model fairness meets source tracing: Toward trustworthy AI for manipulated speech attribution"** during submission.
- **Preparation Guidelines:** https://odyssey2026.inesc-id.pt/preparation-guidelines-and-templates/

Organizers

Nicolas Müller (Fraunhofer AISEC / Resemble AI)
Hemlata Tak (Pindrop, USA)
Adriana Stan (Technical Univ. of Cluj-Napoca)
Jennifer Williams (Univ. of Southampton, UK)
Piotr Kawa (WUST / Resemble AI)
Xin Wang (National Inst. of Informatics, Japan)
Jagabandhu Mishra (Univ. of Eastern Finland)
Ajinkya Kulkarni (Idiap Research Institute, CH)