



ODYSSEY 2026

LISBOA

June 23-26, 2026

# TidyLang Challenge:

## Speaker-Controlled Language Recognition

The TidyLang Challenge is a Speaker-Controlled and Zero-Shot Language Recognition challenge. The only permitted data from Mozilla Common Voice is the official Tidy-X training and validation partition; all other Common Voice data is strictly forbidden. The core task is spoken language recognition at the utterance level under controlled speaker-overlap conditions.

### Primary evaluation conditions:

- Closed-set identification: Predict the correct language label among 40 languages seen during training (macro-averaged F1 / Accuracy).
- Open-set / zero-shot recognition: Handle additional languages not seen during training; evaluated with Equal Error Rate (EER) following standard protocols.

The challenge uses the Tidy-X dataset, a curated partition from Mozilla Common Voice featuring:

- Over 4,474 speakers across 40 languages
- Each speaker with utterances in between 2 and 10 languages
- Approximately 321,711 utterances totaling 457 hours
- Clearly defined training and validation splits
- Pseudonymized speaker identities for privacy

**Evaluation set:** Details about the evaluation data (including size, languages, and trial structure) are not disclosed before the evaluation phase to ensure a fair and unbiased benchmark. Information will be released when the evaluation phase opens.

**More Information:** <https://tidylang2026.github.io>

## Timeline

Feb 1, 2026  
Development  
Phase

Mar 1, 2026  
Evaluation Phase

Mar 5, 2026  
Registration  
Deadline

Mar 6, 2026  
Result Submission  
Deadline

Mar 15, 2026  
Challenge Paper  
Deadline

### Collaborating Organizations:



Universität  
Zürich<sup>UZH</sup>



OTTO VON GUERICKE  
UNIVERSITÄT  
MAGDEBURG



Australian  
National  
University



INDIANA UNIVERSITY

Mozilla  
Data  
Collective

Create. Curate. Control.

## Register here:

