MASRI — University of Malta's Open Maltese Speech-Recognition Corpora for a Language of About 500,000 Speakers, Released Only for Non-Commercial Use

Digital Inclusion · Good practice · Malta

Quality score: 72/100

Evidence strength: Strong

Summary

MASRI built the first Maltese speech-recognition corpora so voice interfaces can work in Maltese, not just English: an 8-hour peer-reviewed headset corpus (25 speakers) grew to about 142 hours of data. The non-commercial licence limits uptake in products.

Key dimensions

Dimension	Score
Innovation level	3/5
Sustainability (ongoing)	1/1
Evaluation evidence	1/1
Demonstrated reach	0/1
Documented learning	1/1

Evaluated by C-NAPSE

Data sources

C-NAPSE web research — accessed 2026-11-19

University of Malta — Institute of Linguistics and Language Technology — accessed 2026-11-19

University of Malta — MASRI project page — accessed 2026-11-19

Mena et al. (2020), MASRI-HEADSET: A Maltese Corpus for Speech Recognition, LREC 2020 (arXiv) — accessed 2026-11-19

MASRI-Dev evaluation set on Hugging Face (MLRS) — accessed 2026-11-19

Cite this — Evidoria. *MASRI — University of Malta's Open Maltese Speech-Recognition Corpora for a Language of About 500,000 Speakers, Released Only for Non-Commercial Use*. Persistent ID: 20dcd978-f906-46c6-9479-e19936878200.