MASRI — Los corpus abiertos de reconocimiento de voz de la Universidad de Malta para una lengua de unos 500 000 hablantes, publicados solo para uso no comercial

Digital Inclusion · Buena práctica · Malta

Puntuación de calidad: 72/100

Solidez de la evidencia: Sólida

Resumen

MASRI creó los primeros corpus de reconocimiento de voz en maltés para que las interfaces de voz funcionen en maltés y no solo en inglés: un corpus de 8 horas revisado por pares (25 hablantes) creció a unas 142 horas. La licencia no comercial limita su uso en productos.

Dimensiones clave

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

Evaluado por C-NAPSE

Fuentes de datos

C-NAPSE web research — consultado el 2026-11-19

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

University of Malta — MASRI project page — consultado el 2026-11-19

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

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

Citar — Evidoria. *MASRI — Los corpus abiertos de reconocimiento de voz de la Universidad de Malta para una lengua de unos 500 000 hablantes, publicados solo para uso no comercial*. ID persistente: 20dcd978-f906-46c6-9479-e19936878200.