

Mozilla Common Voice — Crowdsourced Open Speech Corpus Extending Voice Technology to 250+ Languages

Digital Inclusion · Good practice · United States of America

Quality score: 76/100

Evidence strength: Strong

Summary

Mozilla's CC0 crowdsourced speech corpus, launched in 2017, lets volunteers record and validate sentences so under-resourced languages can get speech technology; the 2020 LREC paper reports ~2,500 hours from 50,000+ people.

Key dimensions

Dimension	Score
Innovation level	4/5
Sustainability (ongoing)	1/1
Evaluation evidence	1/1
Demonstrated reach	1/1
Documented learning	0/1

Evaluated by C-NAPSE

Data sources

[C-NAPSE web research](#) — accessed 2026-10-11

[Mozilla Foundation \(Common Voice\)](#) — accessed 2026-10-11

[Common Voice: A Massively-Multilingual Speech Corpus \(arXiv/LREC 2020\)](#) — accessed 2026-10-11

[Common Voice \(Wikipedia\)](#) — accessed 2026-10-11

[Common Voice](#) — accessed 2026-10-11

Cite this — Evidoria. *Mozilla Common Voice — Crowdsourced Open Speech Corpus Extending Voice Technology to 250+ Languages*. Persistent ID: 522361ba-22d4-430f-afbd-dafb924334f3.