

USC — ISSAI's Open Uzbek Speech Corpus (105 Hours, 958 Speakers) for Voice Technology in Uzbek

Digital Inclusion · Good practice · Uzbekistan

Quality score: 52/100

Evidence strength: Moderate

Summary

An open corpus of 105 hours of read Uzbek speech from 958 speakers, checked by native speakers, with baseline recognisers at 17.4–18.1% word error rate. Evidence is technical only; no user outcomes yet.

Key dimensions

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

Evaluated by C-NAPSE

Data sources

[C-NAPSE web research](#) — accessed 2026-10-19

[ISSAI, Nazarbayev University \(with Uzbek collaborators\)](#) — accessed 2026-10-19

[Paper \(arXiv\)](#) — accessed 2026-10-19

[ISSAI project page](#) — accessed 2026-10-19

[Code and models \(GitHub\)](#) — accessed 2026-10-19

Cite this — Evidoria. *USC — ISSAI's Open Uzbek Speech Corpus (105 Hours, 958 Speakers) for Voice Technology in Uzbek*. Persistent ID: d4089134-4e4c-488d-b5d8-f084bb3e2656.