

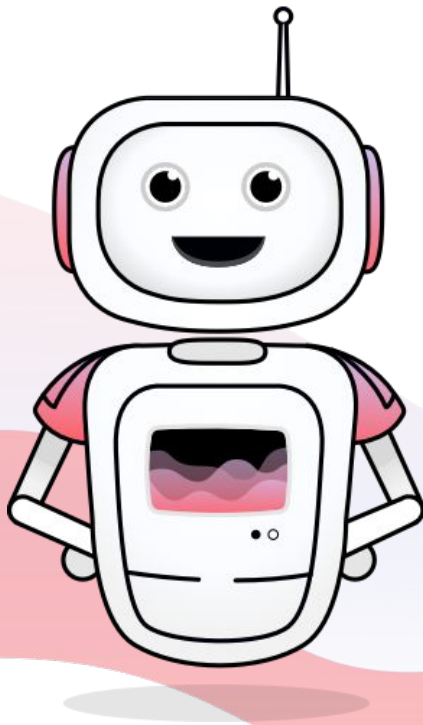
Common Voice

moz://a

Retoso

4a Aprilo 2021

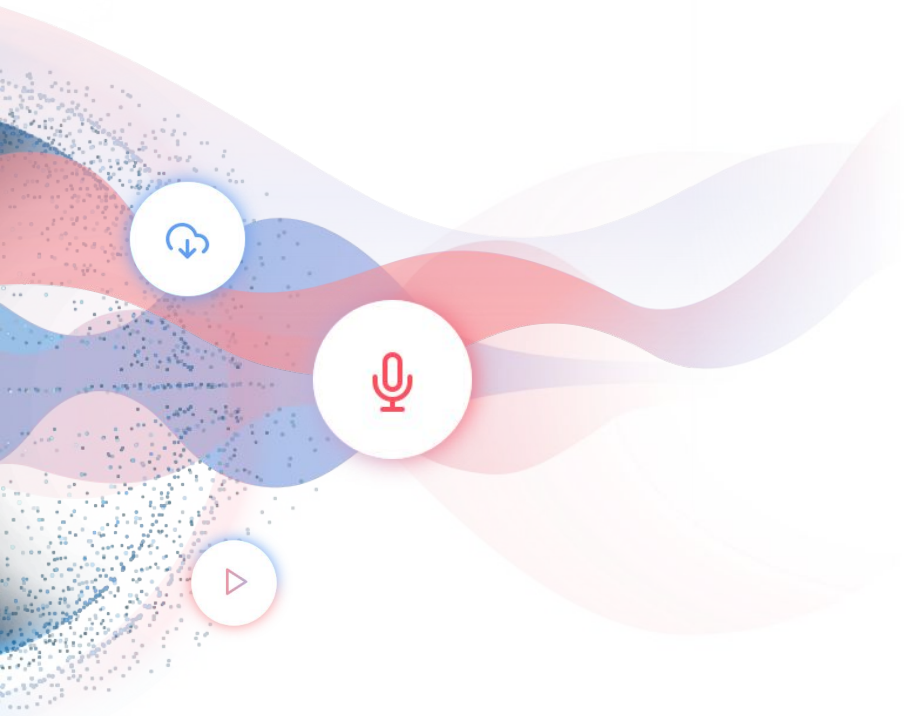
KuboF kaj Stefan



La projekto Common Voice

Kiel krei liberajn voĉajn sistemojn en
Esperanto kaj en aliaj lingvoj.

Enhavo



Kial Common Voice ekestis	01
---------------------------	----

La esperanta versio de CV	02
---------------------------	----

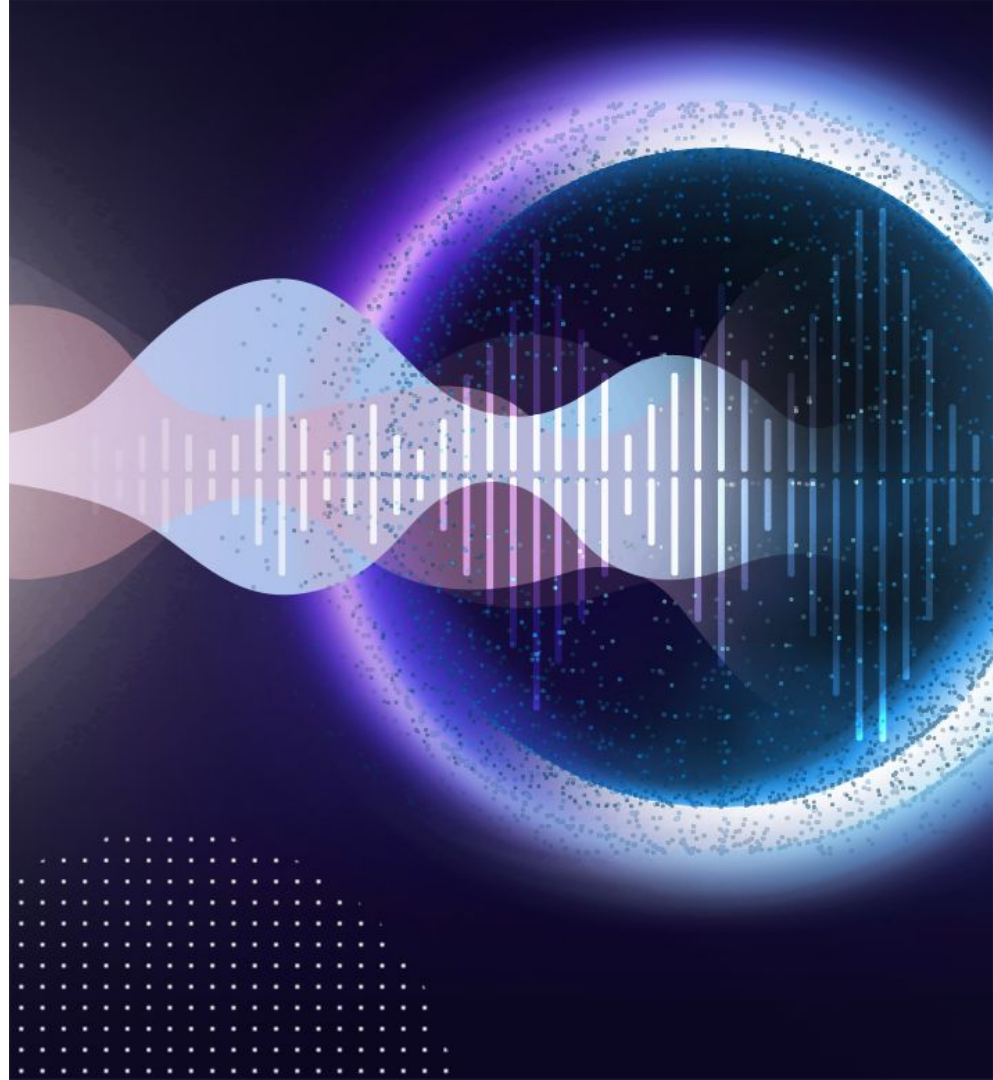
Kiel uzi la retejon	03
---------------------	----

Konkretaj projektoj kaj planoj	04
--------------------------------	----

Kunlaboru!	05
------------	----

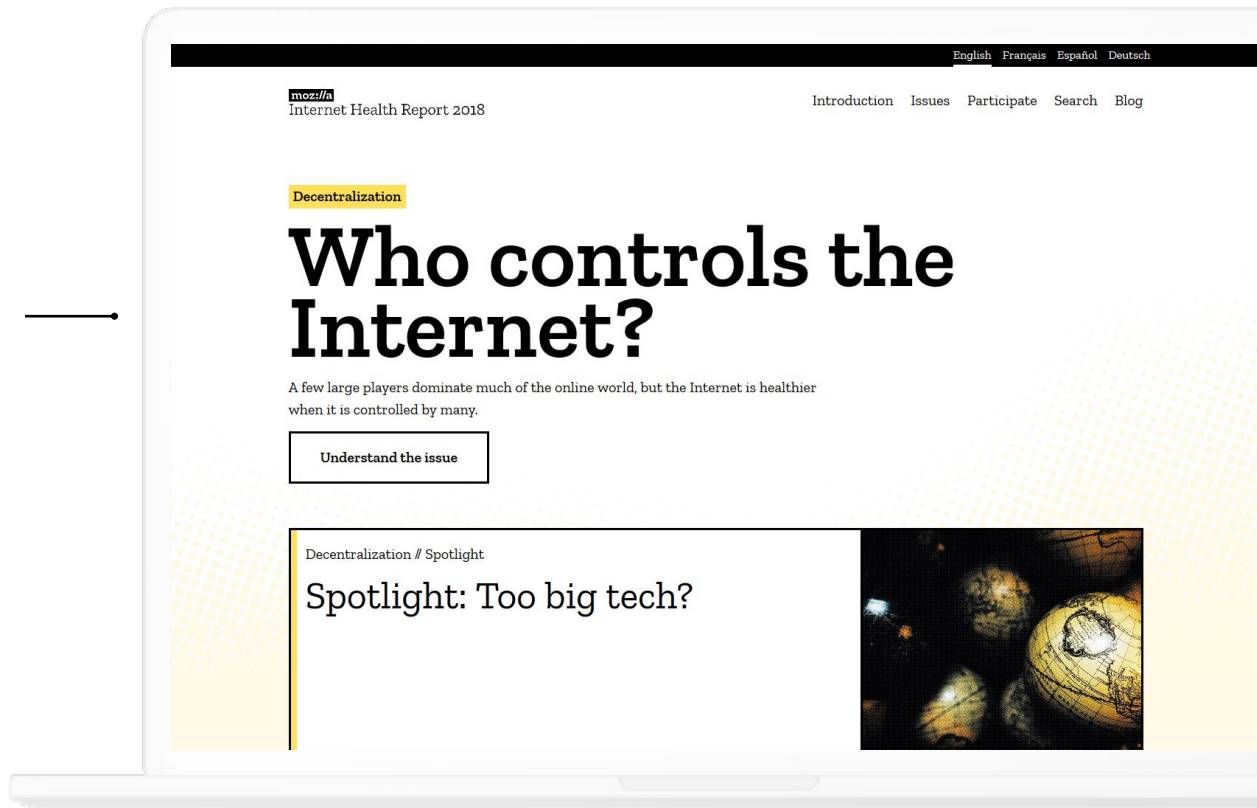
01

Kial Common Voice ekestis



moz://a

Sanraporto pri la
intereto



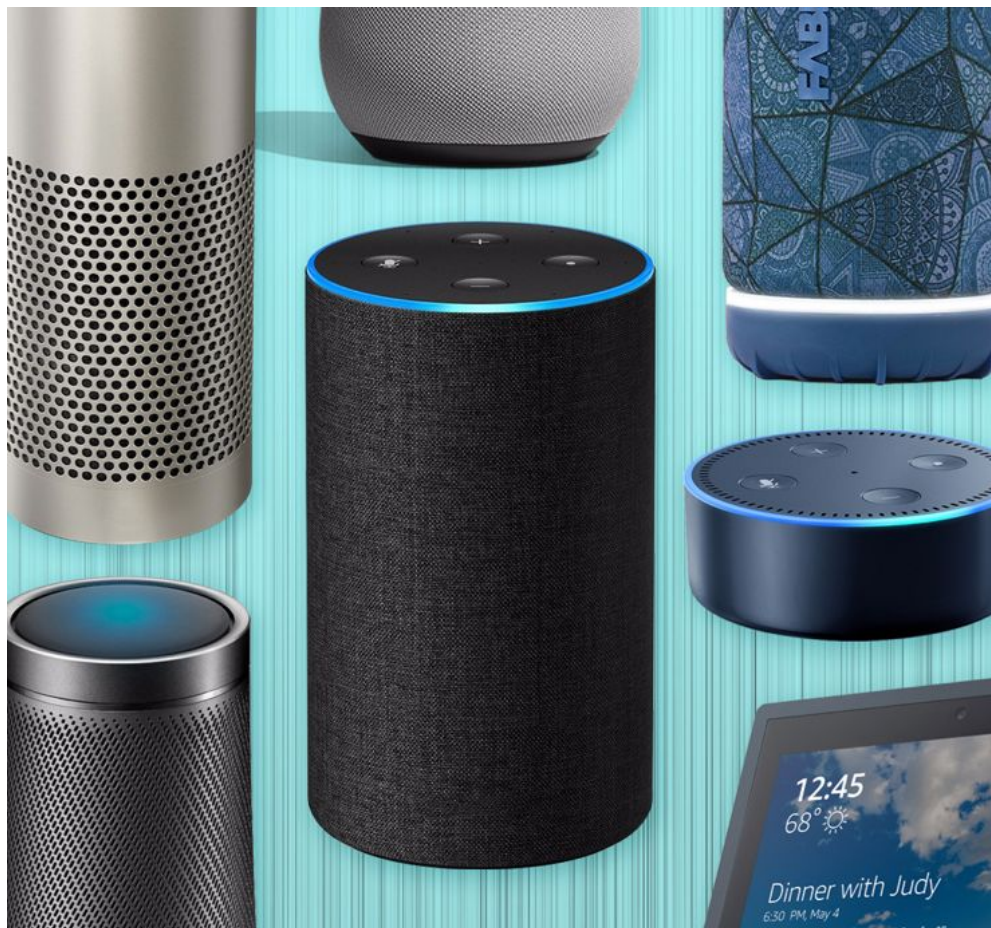
Common Voice

moz://a

Êu parolrekonado
estas oligopolo?



Common Voice
[moz://a](https://www.mozilla.org/en-US/voice-data/)



Kiel krei parolrekonadon



Voĉaj
datumoj



Voĉa
algoritmo



Komputa
potenco

PROJEKTO **COMMON VOICE**

PROJEKTO **DEEP SPEECH**



"Por krei voĉajn sistemojn, programistoj bezonas tre grandan kvanton da voĉaj datumoj. La plimulto de la datumoj uzataj de grandaj kompanioj ne estas uzeblaj de iu ajn. Ni pensas ke tio sufokas novaĵkreemon. Tial ni lanĉis la projekton Common Voice, kiu helpas igi parolrekonon malfermita kaj alirebla por ĉiuj."

- *Retejo de Common Voice*

Parolrekonado kaj Esperanto

- inteligenta virtuala asistanto
- voĉa tajpado / diktado
- transskibado de voĉregistraĵoj
(prelegoj, podkasto ktp)
- kreado de subtitoloj
- lernado de prononco

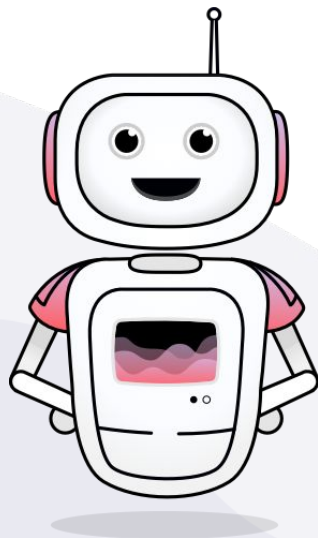


Kio estas
Common Voice?



Common Voice

Celoj de la projekto



Krei liberajn kaj
altkvalitajn
datumarojn por
multaj lingvoj



Subteni esploradon



Ebligi novajn
kreemajn projektojn



Subteni diversecon en
voĉajn sistemojn

Plej nova datumaro (12/2020)



Validigitaj horoj

9,300

(90 en Esperanto)



Lingvoj

60+

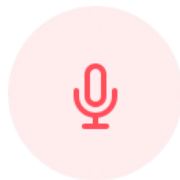
kaj **65+** planitaj lingvoj



Pli ol

200

Aprogramistoj helpis
krei programaron



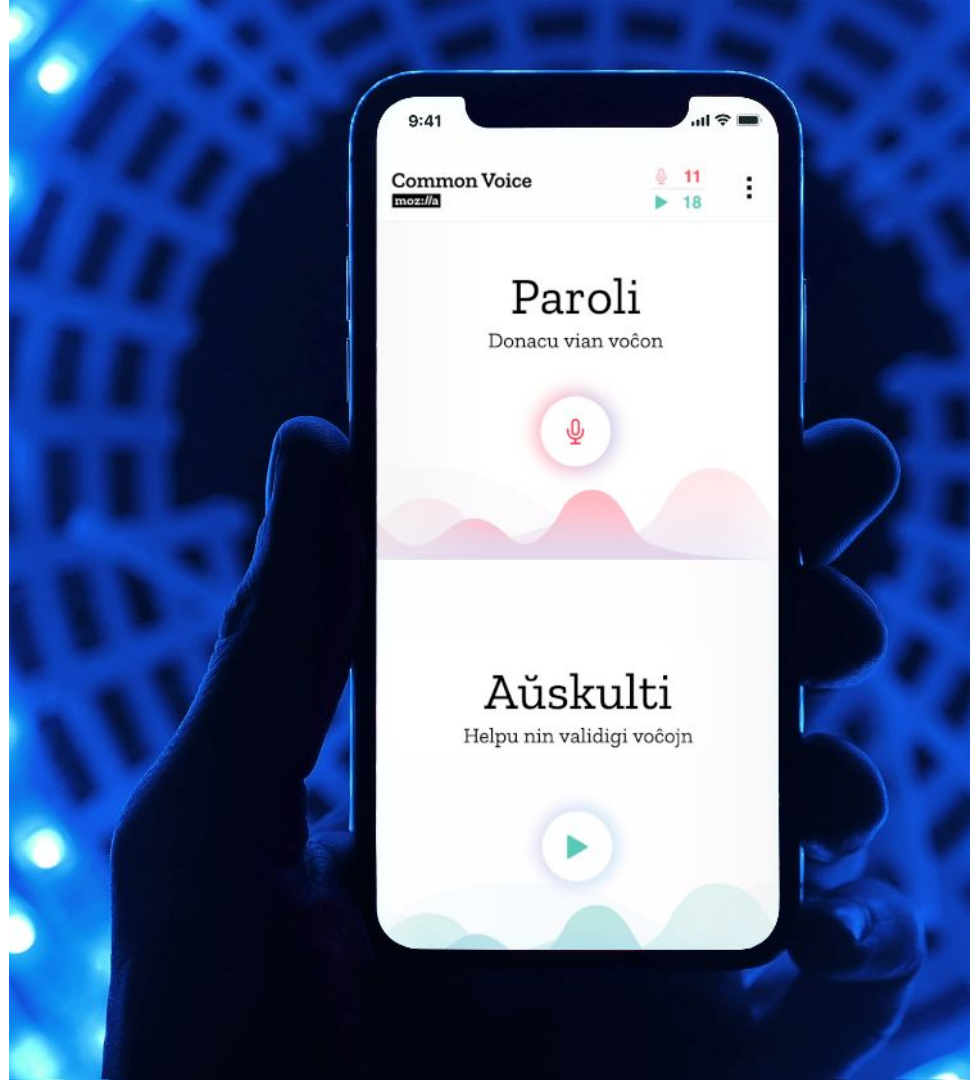
Helpantoj

393,000

homoj donacis sian
voĉon en siajn lingvojn

02

La esperanta versio de Common Voice



2018 unua versio de la esperanta retejo



1

La retejo estis
tradukita



2

Liberaĵoj (CCO aŭ publika
havaĵo) estis
kolektitaj



3

La frazoj estis
kontrolitaj



4

En 2018 CV en
Esperanto
ekestis

2019 kaj 2020 - pli da frazoj kaj varbado

- La unua eldono de la datumaro aperis en februaro 2019: 6 horojn de 53 donancantoj
- Cirkaŭ 100 000 frazoj estis ekstraktitaj el Vikipedio (maksimume tri frazoj per artikolo)
- 30 000 frazoj estis importitaj mane (dankon al Libera Folio kaj Monato)
- Unuaj eksperimentoj pri kreo de parolrekonilo
- Varbado en sociaj retoj kaj kreado de varbmaterioj

**Donacu vian voĉon kaj helpu krei
parolrekonan sistemon
en Esperanto!**

*Instruu al mi kiel
veraj homoj parolas!*

voice.mozilla.org/eo



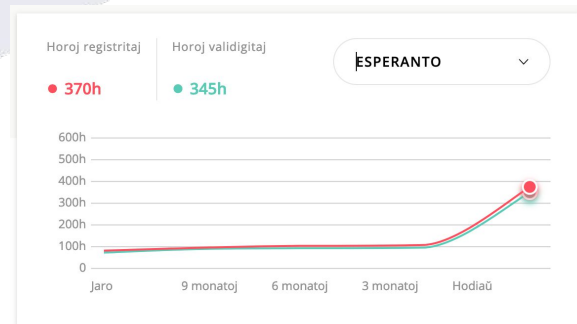
Common Voice
moz://a

2021 - trioblaj datumkvantoj en nur tri monatoj

- Ekde 2021 homoj povas gajni la ĉifromonon Miriado kontraŭ donacado por CV
- Unua uzebla "DeepSpeech"- sistemo estis kreita (kun 90 horojn) de Tim.

Ekzemploj:

- src: "tiu tamen ne konfidencis ion ajn"
- res: "tiu tamen ne kovigas misiojn"
- src: "restaĵoj de tiu agado estas videblaj en preskaŭ ĉiuj valoj"
- res: "restaĵoj de tiu agado estis videblaj impresa kiel valoj"
- src: "parizon"
- res: "la rizo"



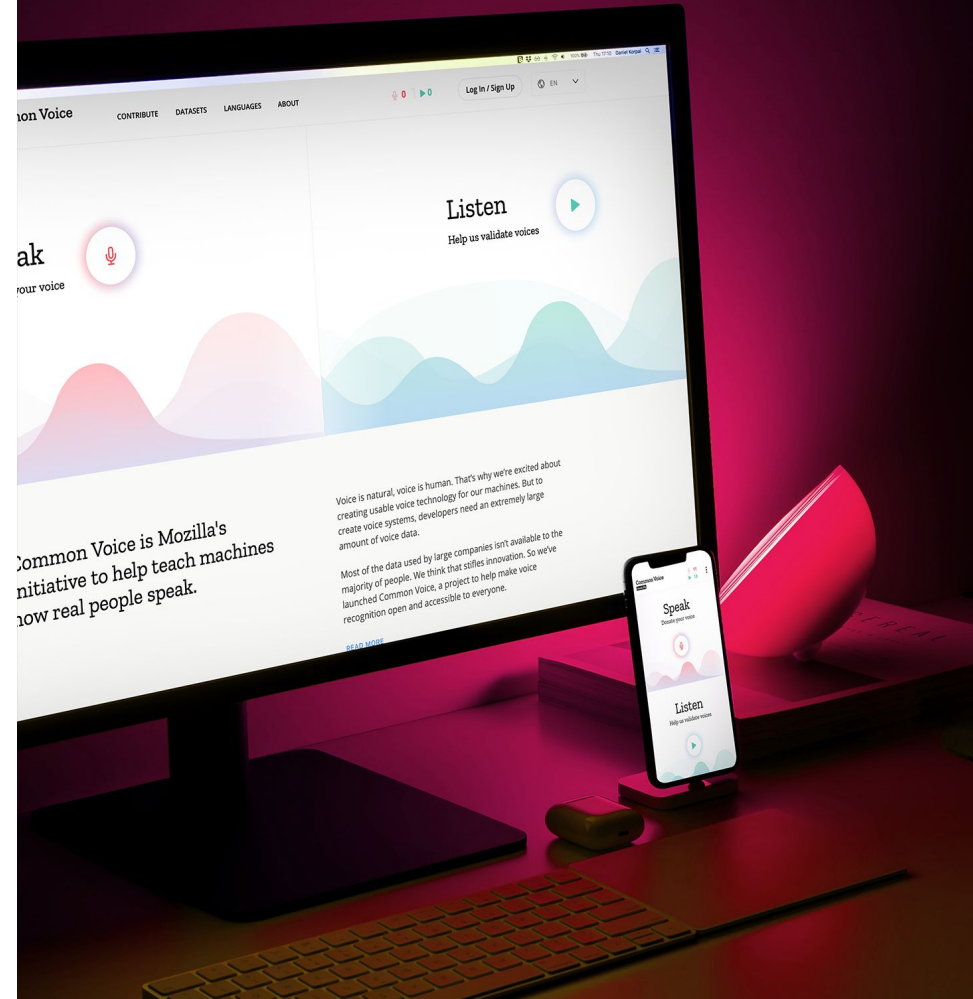
**Miriado – komercio kaj diskutejo**
106 Membroj

 <https://aperu.net/miriado/>

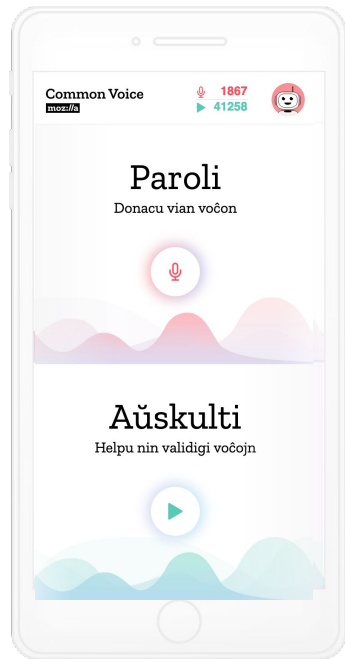
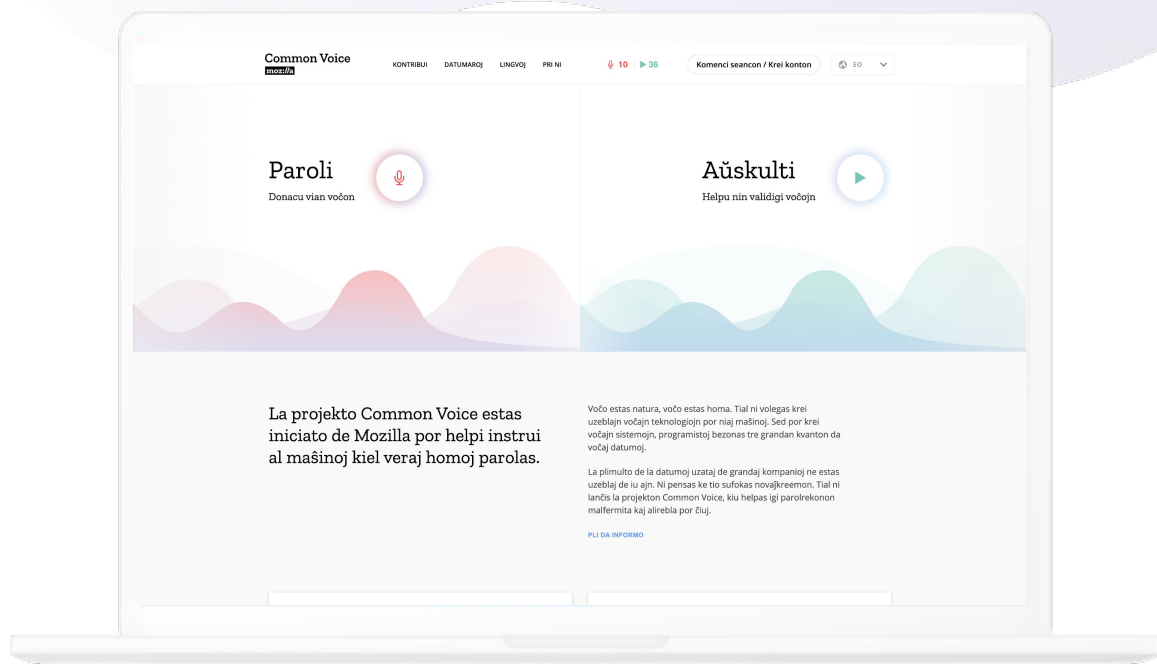
<https://t.me/miriadejo>

03

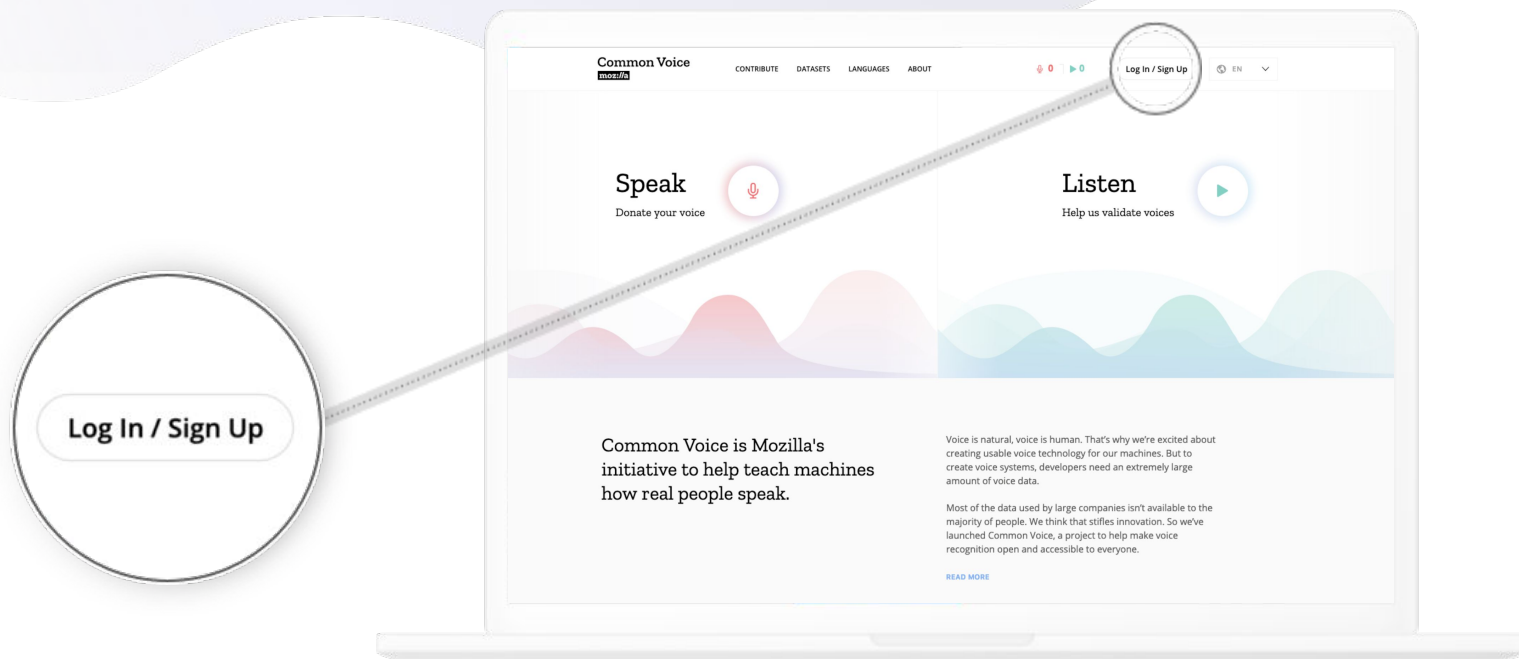
Kiel uzi Common Voice



Vizitu commonvoice.mozilla.org/eo



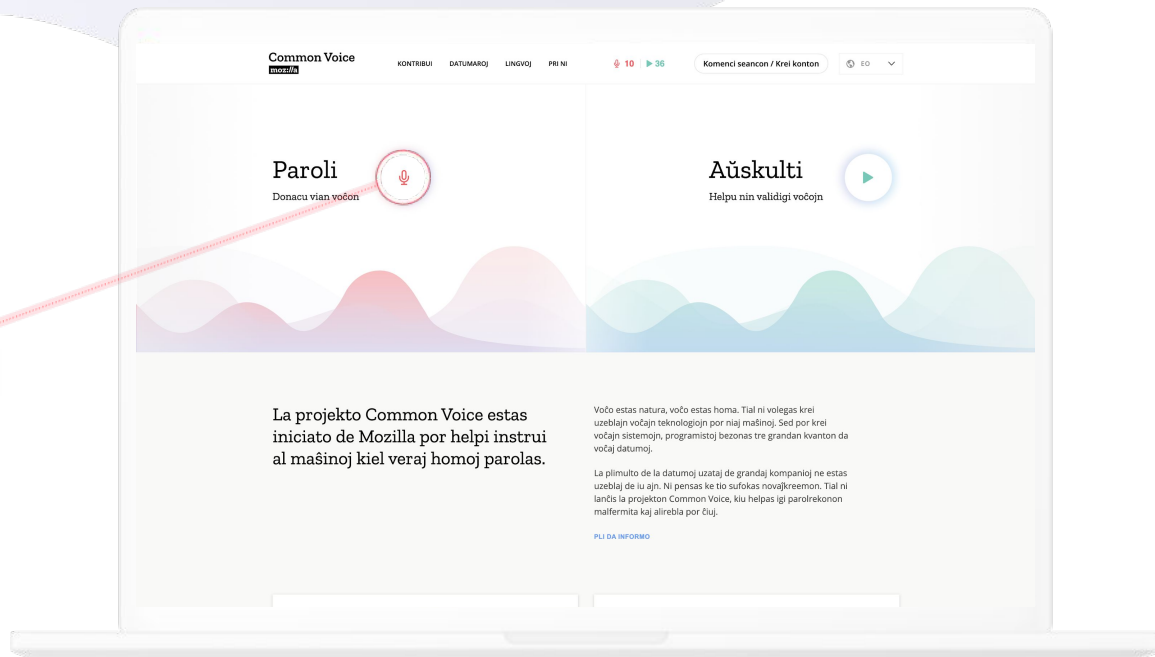
Kreu profilon (nedeviga)



Donacu vian voĉon

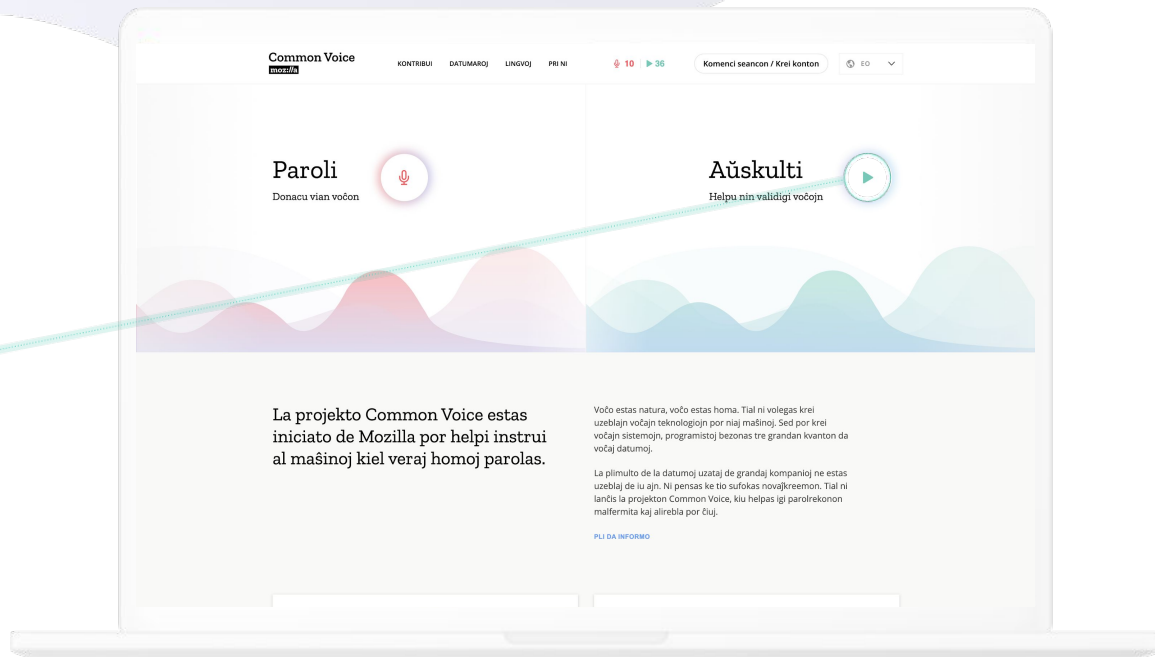
Ideala kvanto:

15 minutoj
aŭ 225 frazo



Aŭskulti voĉojn

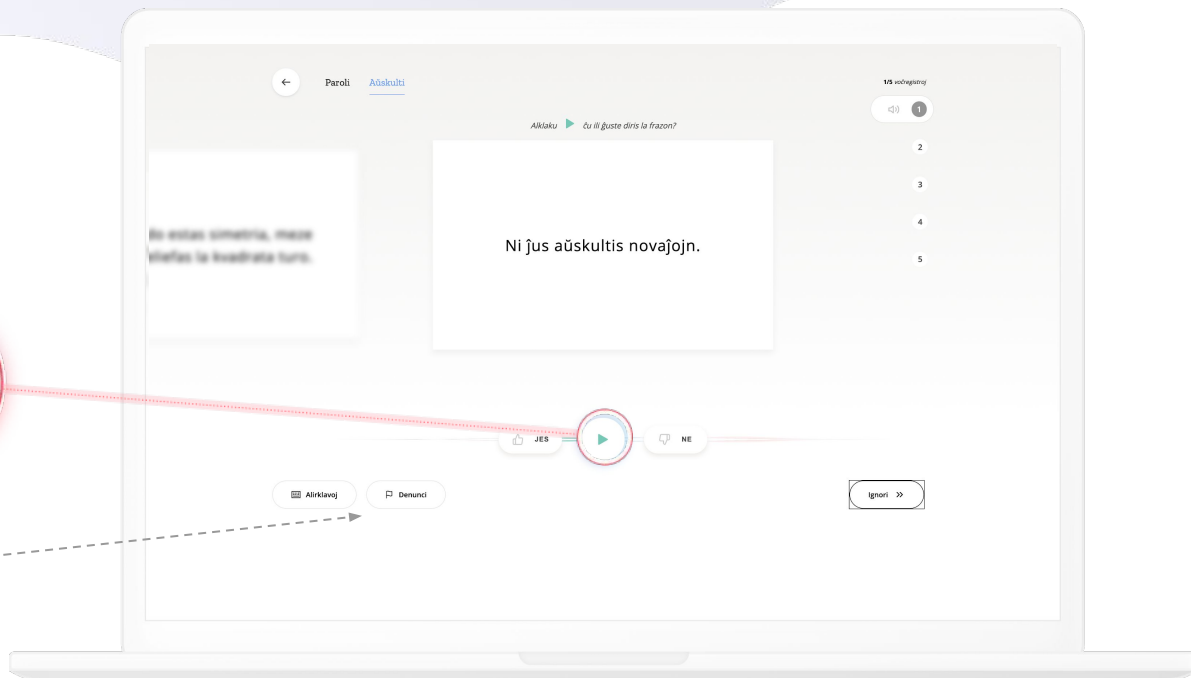
Helpu krei
altkvalitan
datumaron



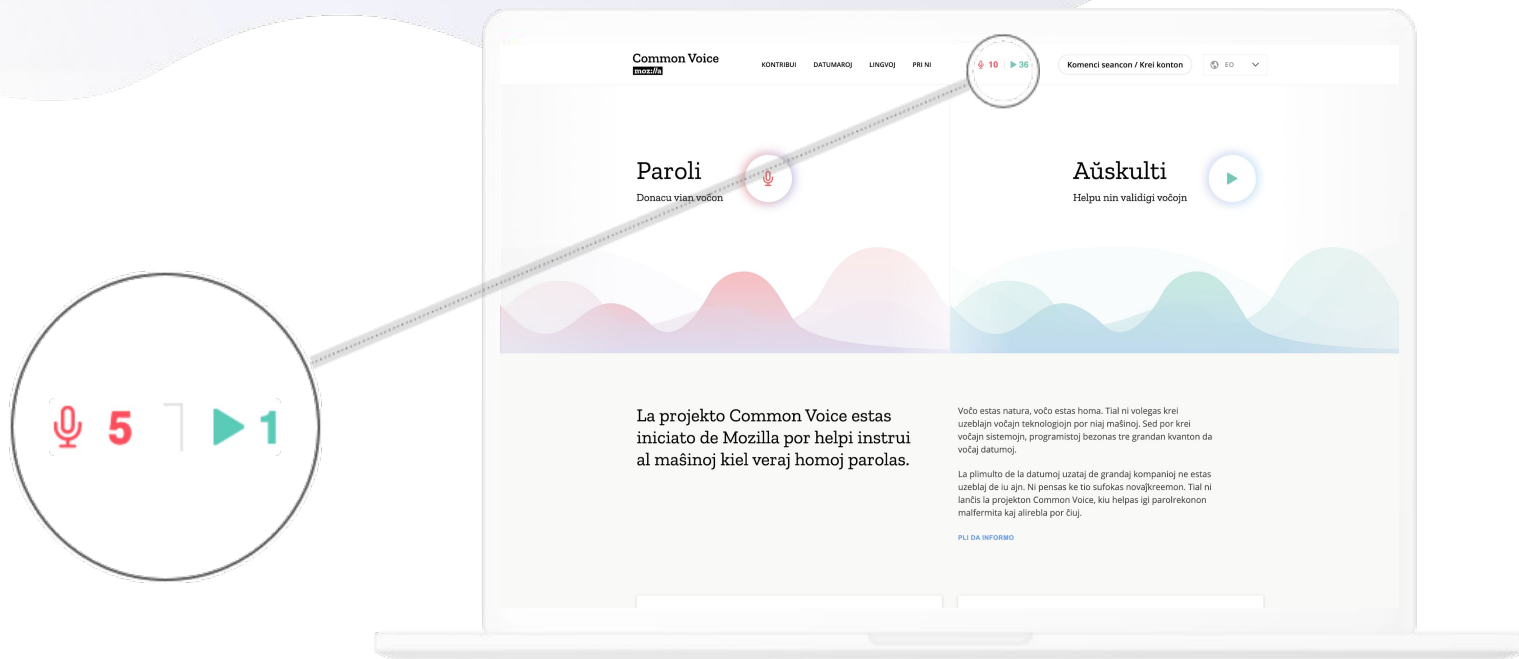
Validigi registraĵojn

Validigi

Denunci
malĝustajn
frazojn

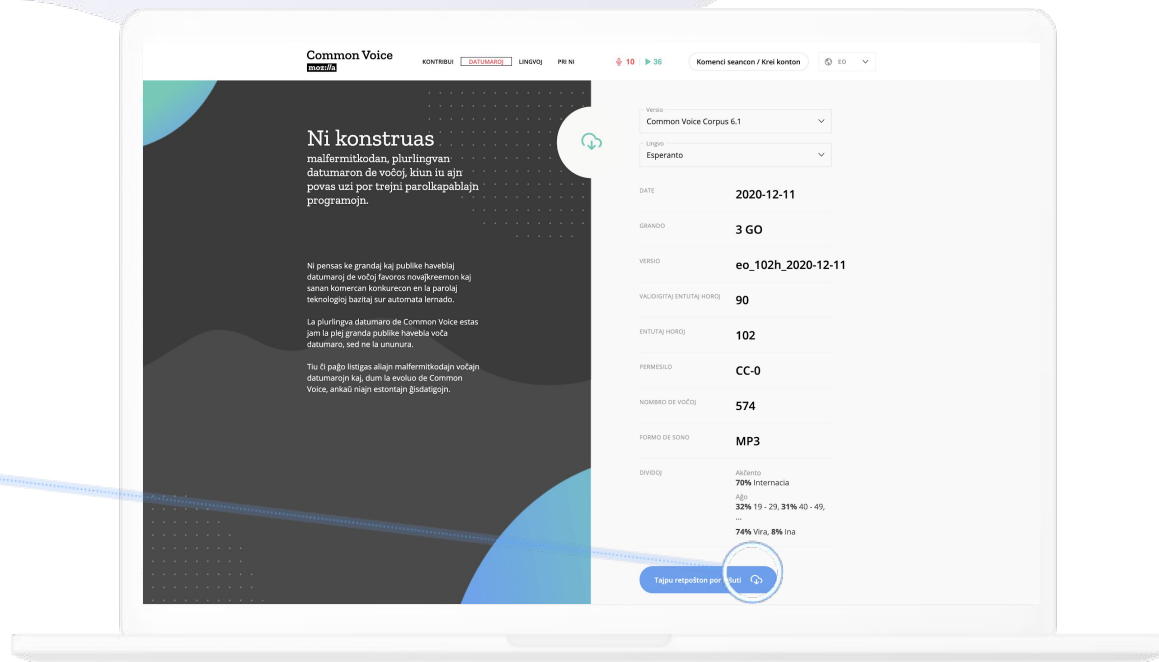


Spuri vian progreson



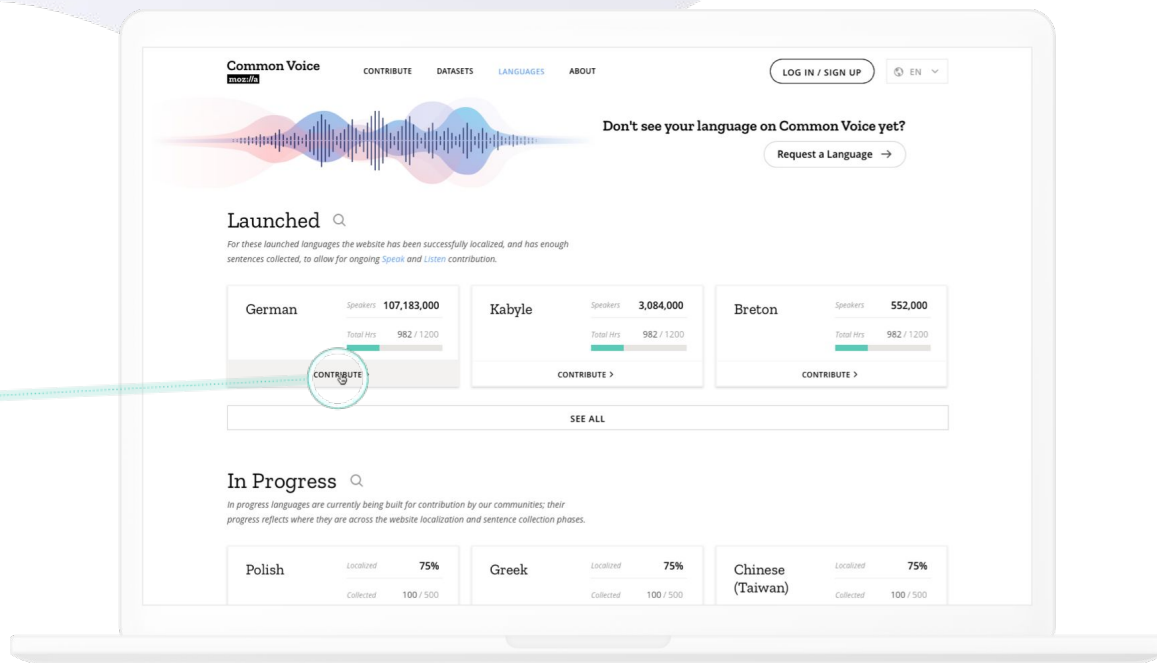
Elŝutu kaj uzu la datumaron

Elŝuti



Ankaŭ estas multaj aliaj lingvoj!

Kontribuu!



Konkretaj projektoj kaj planoj

Konkretaj aplikaĵoj:

- <https://glosbe.com> - La vortaro uzas la esperantajn frazojn kiel ekzemplojn kiel oni pronocu vortojn
- Trejnado de neŭraj retoj:
 - 90 h -> 50% de la vortoj estis ĝuste detektitaj en la sistemo de Tim
- Artisto uzis la frazojn por video

Planoj kaj ebloj:

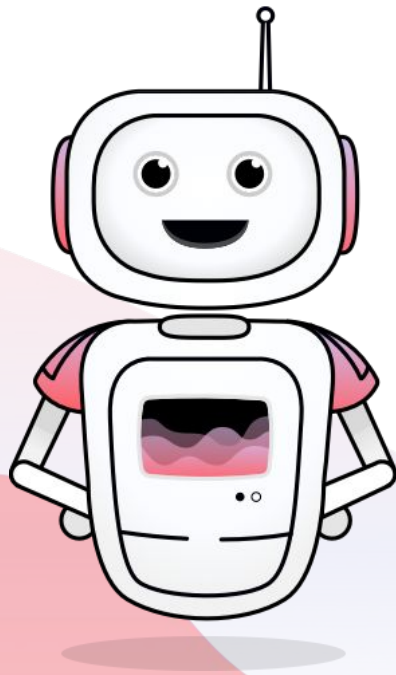
- Aŭtomata kreado de subtekstoj en Esperanto
- Transskribado de podkasto
- Voĉa tajpado / diktado
- Trejni la prononcadon en retaj lernejoj kiel Duolingo
- Faciligi plenumon de tute novaj ideoj

HEJMA AUTOMATIGO

Bildo: Hansuan_Fabregas/ Pixabay <https://bit.ly/3sRor4L>
Jeremy Geltman/ CC BY-SA 4.0: <https://bit.ly/3q268H4>

Kiel vi povas subteni la projekton?

- Donaĉu vian voĉon aŭ kontrolu la voĉojn de aliaj homj: <http://commonvoice.mozilla.org/eo>
- Helpu kolekti pli da frazoj: <https://commonvoice.mozilla.org/sentence-collector/>
- Denuncu aŭ forigu erarajn frazojn
- Traduku la retejon per pontoon: <https://pontoon.mozilla.org/eo/common-voice/>
- Kreu voĉajn sistemojn per la datumaro: <https://github.com/parolrekonado>
- Parolu pri la projekto kaj invitu viajn amikojn!
- Helpu al diverseco en la datumaro!
 - Bezonataj precipe virinoj!
- Laŭtlegado kun tre bona mikrofono.
 - Tim volas krei ankaŭ voĉ-sintezilon en Esperanto



Dankon!

[matrix]

#esperanto:mozilla.org



<https://commonvoice.mozilla.org/eo>

<https://github.com/parolrekonado>

<https://parolrekonado.github.io/>