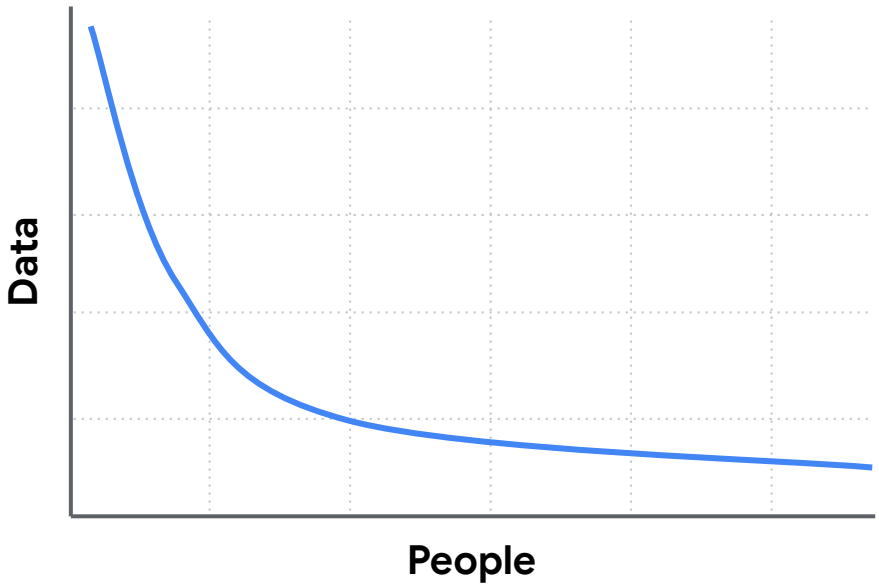
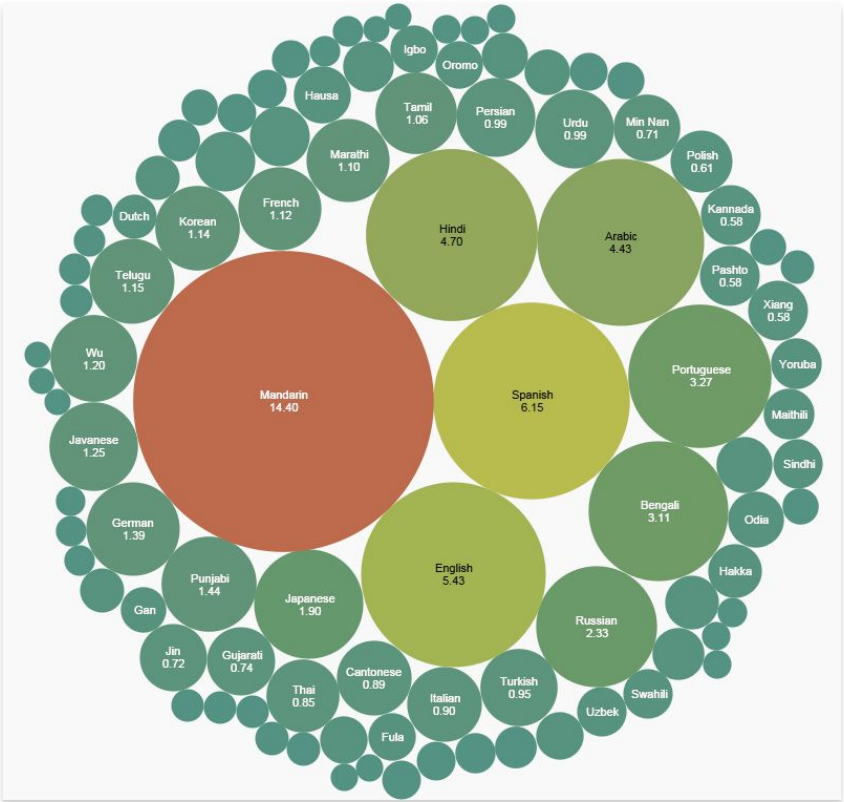
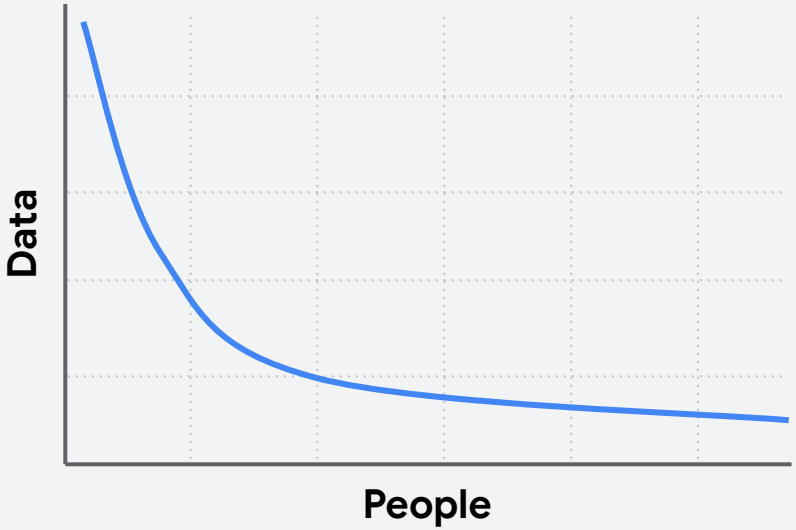
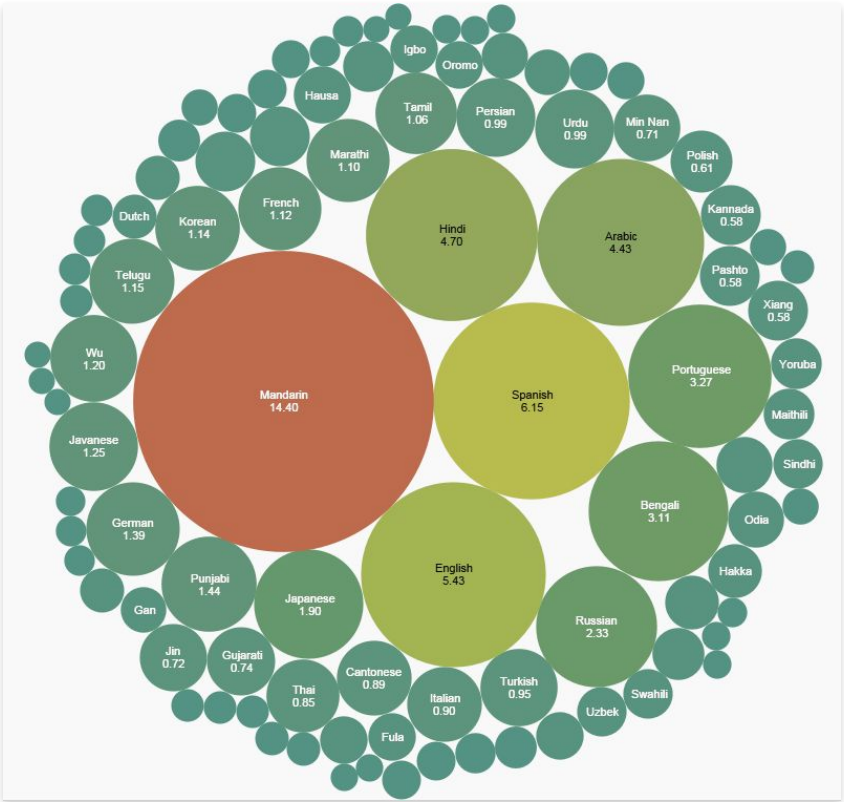
Crowdsourcing Data for the Long Tail


Data
People


Igbo
Oromo
Hausa
Tamil
1.06
Persian
0.99
Urdu
0.99
Min Nan
0.71
Polish
0.61
Marathi
1.10
French
1.12
Dutch
Korean
1.14
Hindi
4.70
Arabic
4.43
Kannada
0.58
Telugu
1.15
Pashto
0.58
Xiang
0.58
Wu
1.20
Mandarin
14.40
Spanish
6.15
Portuguese
3.27
Yoruba
Maithili
Javanese
1.25
Sindhi
Bengali
3.11
German
1.39
Odia
Punjabi
1.44
English
5.43
Hakka
Japanese
1.90
Gan
Russian
2.33
Jin
0.72
Gujarati
0.74
Cantonese
0.89
Thai
0.85
Italian
0.90
Turkish
0.95
Swahili
Uzbek
Fula

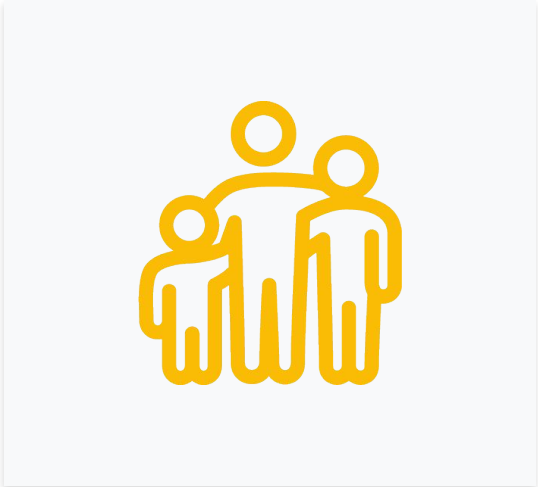

Data
People



- Speech commands for the **whole planet**?
- For **more than** just voice assistants

Cost Model v. **Community** Model?

Limited Scale



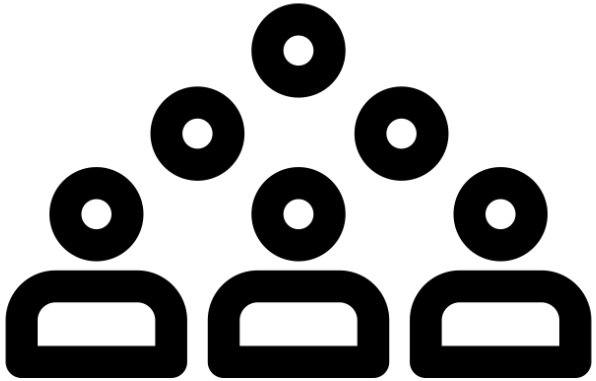
Social Good

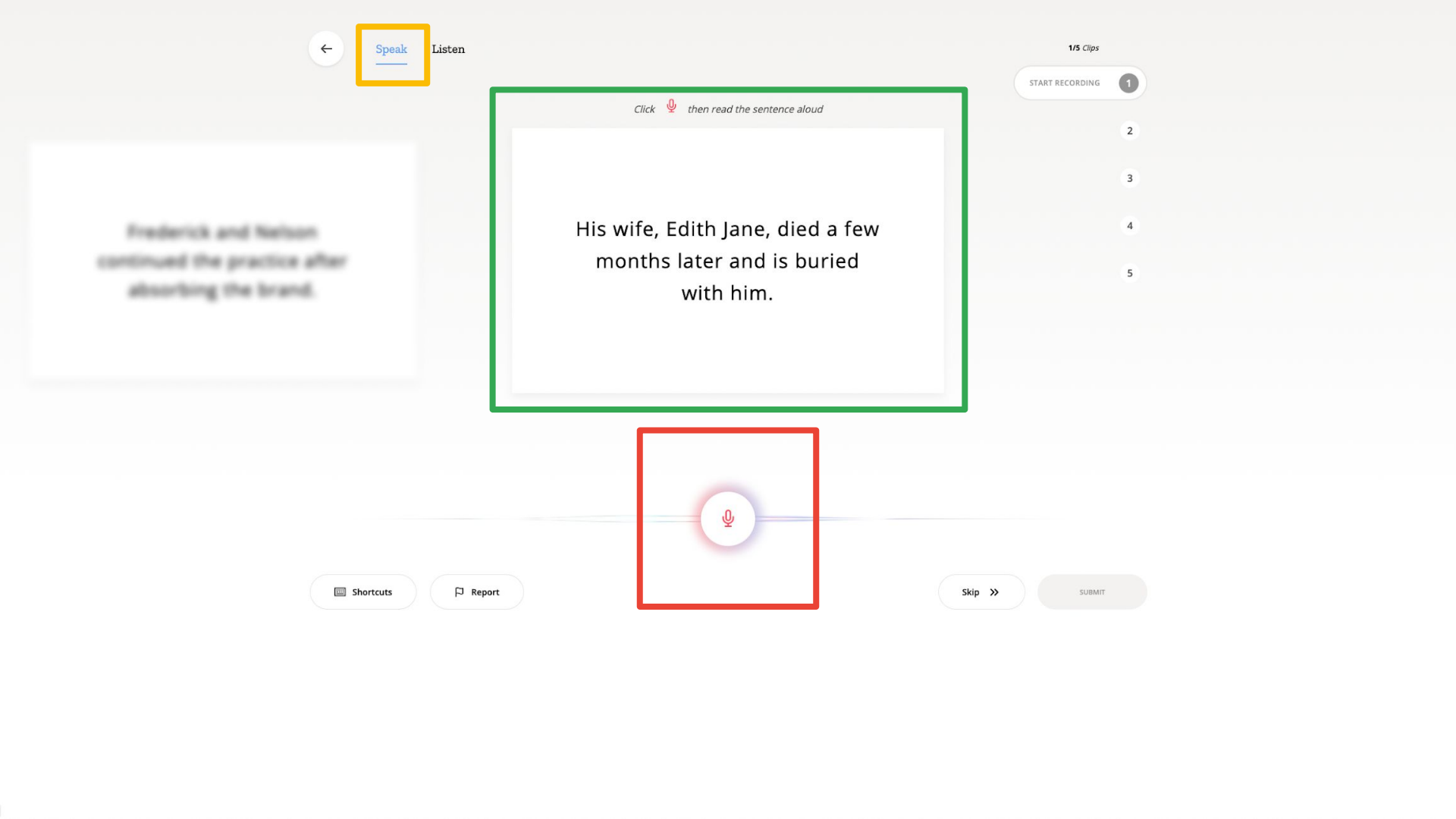
Common Voice

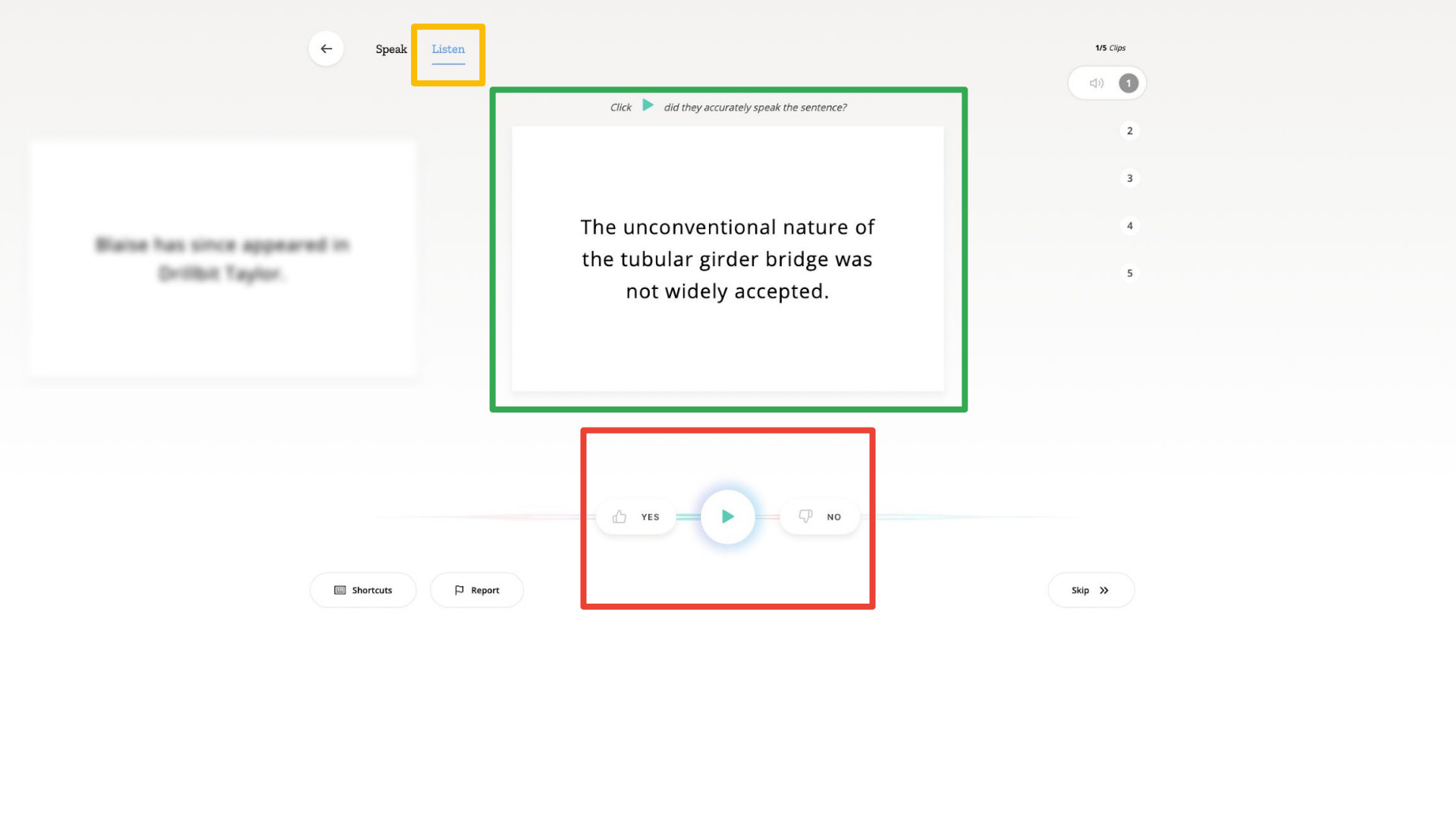
https://commonvoice.mozilla.org

Common Voice

- **Crowdsourcing** platform

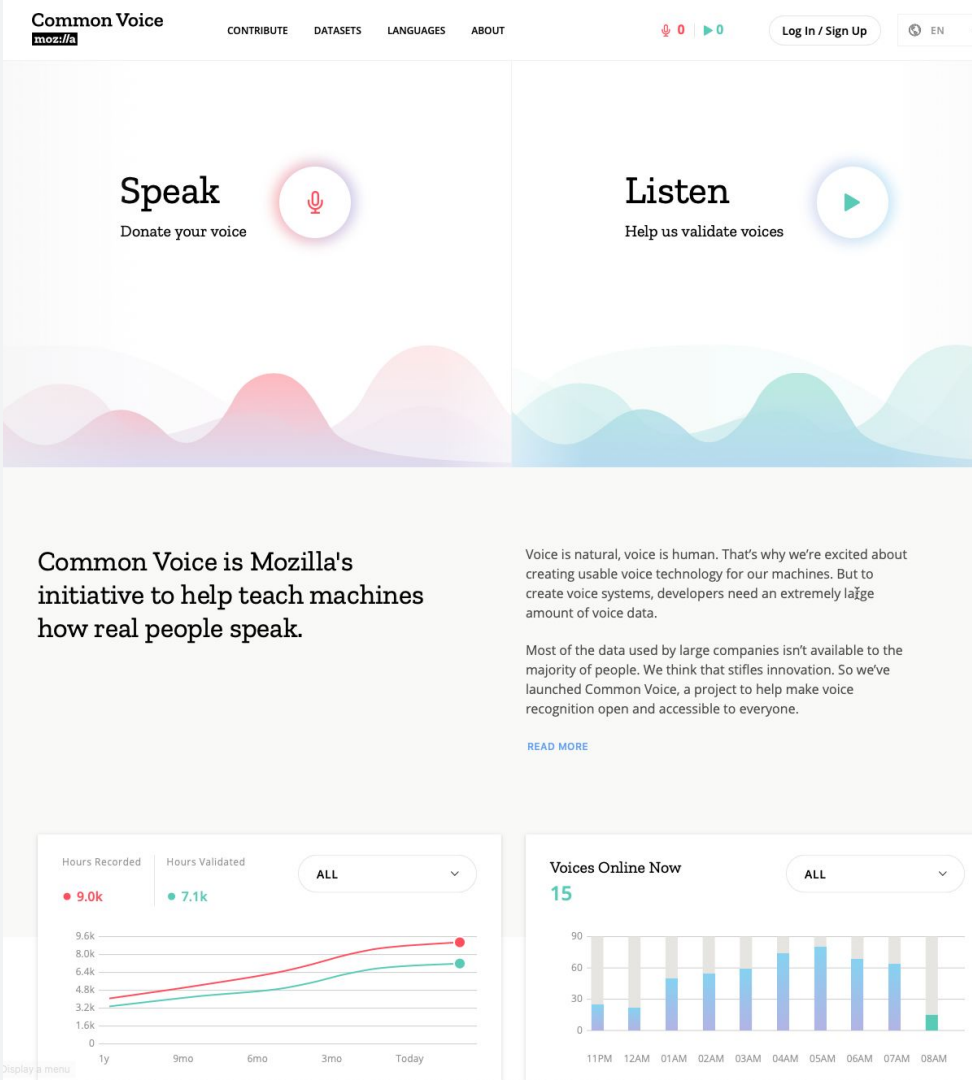



Speak
Listen
1/5 Clips
START RECORDING
1
2
3
4
5
Click
then read the sentence aloud
His wife, Edith Jane, died a few months later and is buried with him.
Shortcuts
Report
Skip
SUBMIT


Speak
Listen
1/5 Clips
1
2
3
4
5
Click
did they accurately speak the sentence?
The unconventional nature of the tubular girder bridge was not widely accepted.
YES
NO
Shortcuts
Report
Skip

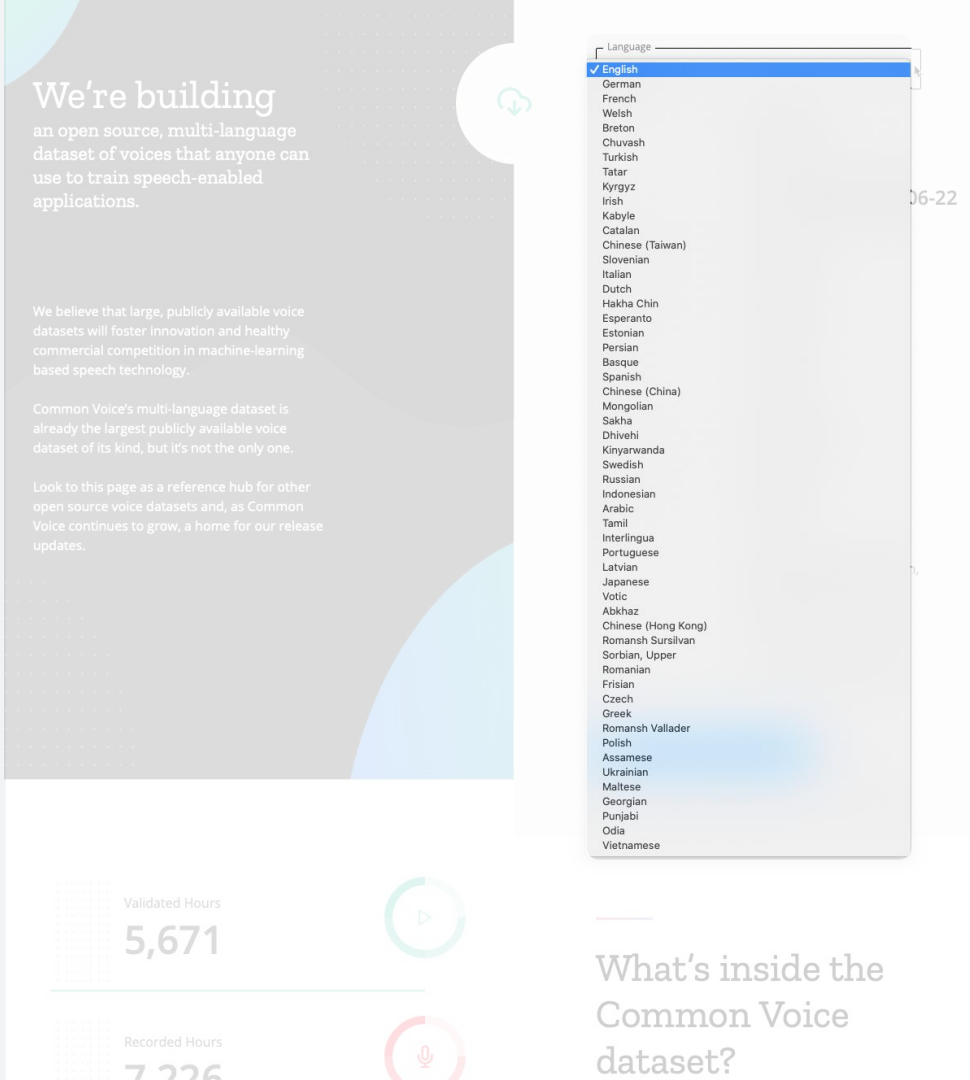
Common Voice

- **Crowdsourcing** platform
- Over **50,000 volunteers**



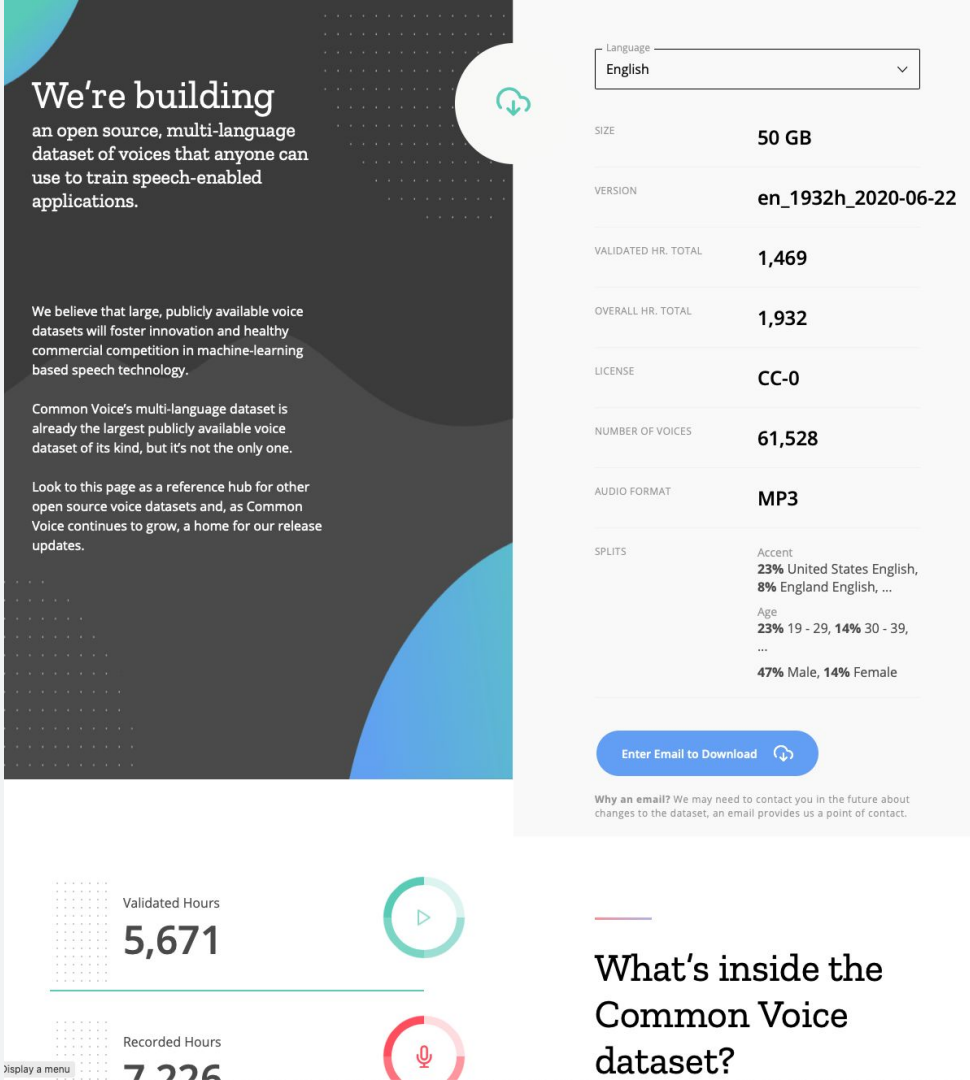
Common Voice

- **Crowdsourcing** platform
- Over **50,000 volunteers**
- **54 different languages**



Common Voice

- **Crowdsourcing** platform
- Over **50,000 volunteers**
- **54 different languages**
- Goal: **speech recognition** for ***all languages*** on the planet

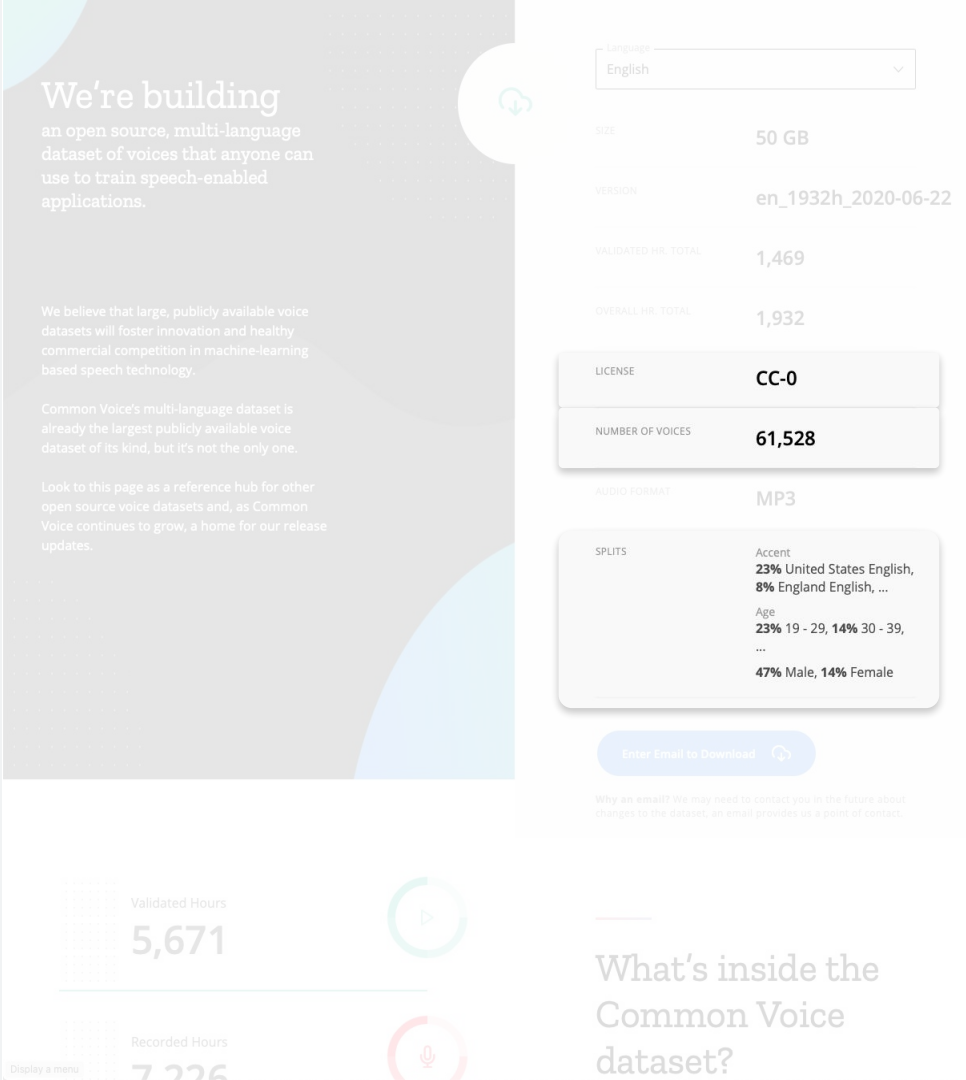


File **Structure**

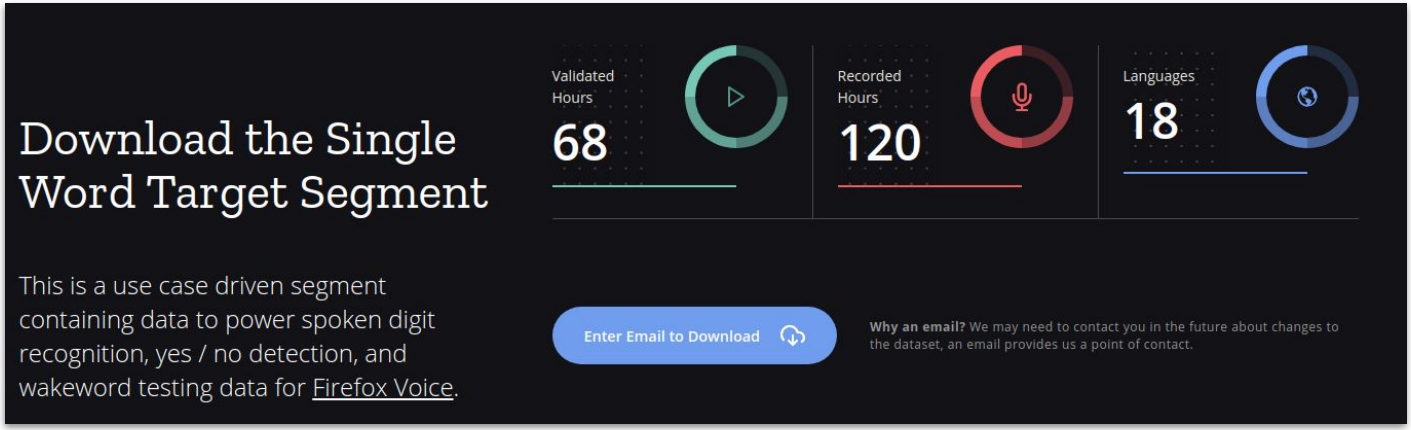
- **Valid**
  - At least 2 people listen to them, and the majority of those listeners say the audio matches the text
- **Invalid**
  - At least 2 listeners, and the majority say the audio does not match the clip
- **Other**
  - All other clips, i.e., fewer than 2 votes, or those that have equal valid and invalid votes, are labelled “other”

Interesting Attributes

- Permissive license
- Many contributors
- Comes with metadata



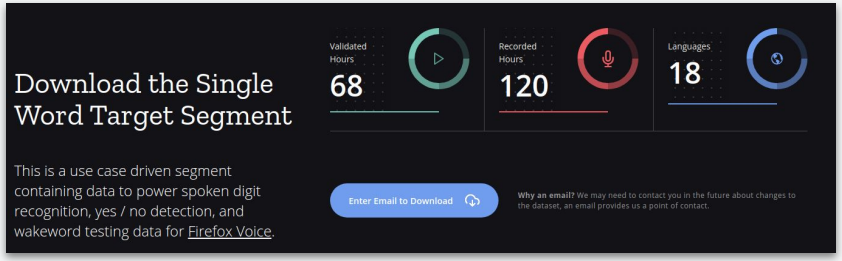
Common Voice



Single Word Target Segment

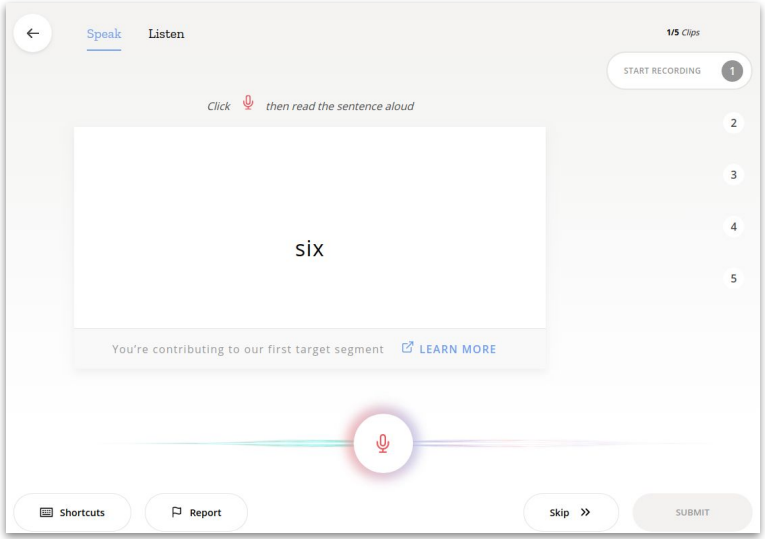
A ***speech commands-style*** dataset for **18 languages**

- “Yes” // “no”
- “hey” & “Firefox”
- **digits** 0-9



ASR **Diversity** and **Reach**

- **Common Voice**
  - Permissive license
  - Minority languages
- **Ease-of-use, wide reach**
  - Browser-based
  - Community can add new languages
- **You can contribute!**



Common Voice **Data Structure**

“yes”				
Prompt	Record	Attributes	Validation	Data Splits