

IncluSet: A Data Surfacing Repository for Accessibility Datasets

Hernisa Kacorri, Rie Kamikubo, Utkarsh Dwivedi, Sravya Amancherla, Urja Thakkar, Mayanka Jha, and Riya Chanduka

Overview

Public datasets has often served as a mean of attracting, nurturing, and challenging data scientists to work on specific problems. However, this approach has seen limited use in the field of accessibility.

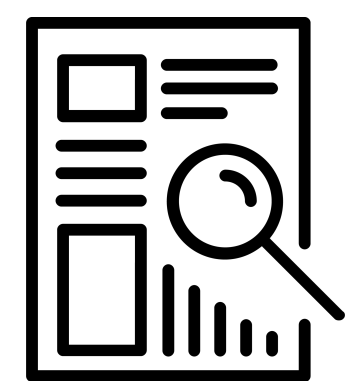
We aim to promote data science for accessibility by increasing awareness on challenges of accessing datasets sourced from people with disabilities and older adults and creating a repository to surface currently available resources.

Populating the repository

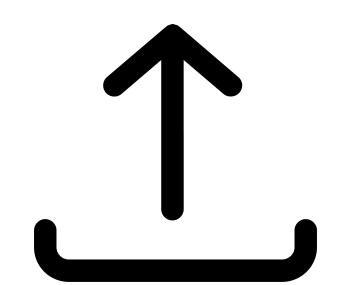


- Search existing repositories (e.g., Kaggle and UCI)
- Google Scholar search with keywords, venues and authors
- Coding and analyzing datasets
- Creating the Includset repository.

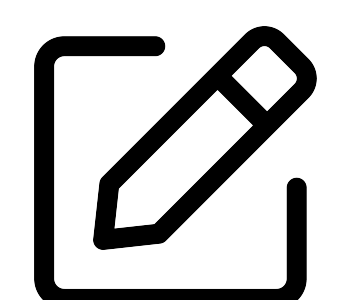
What can you do on IncluSet?



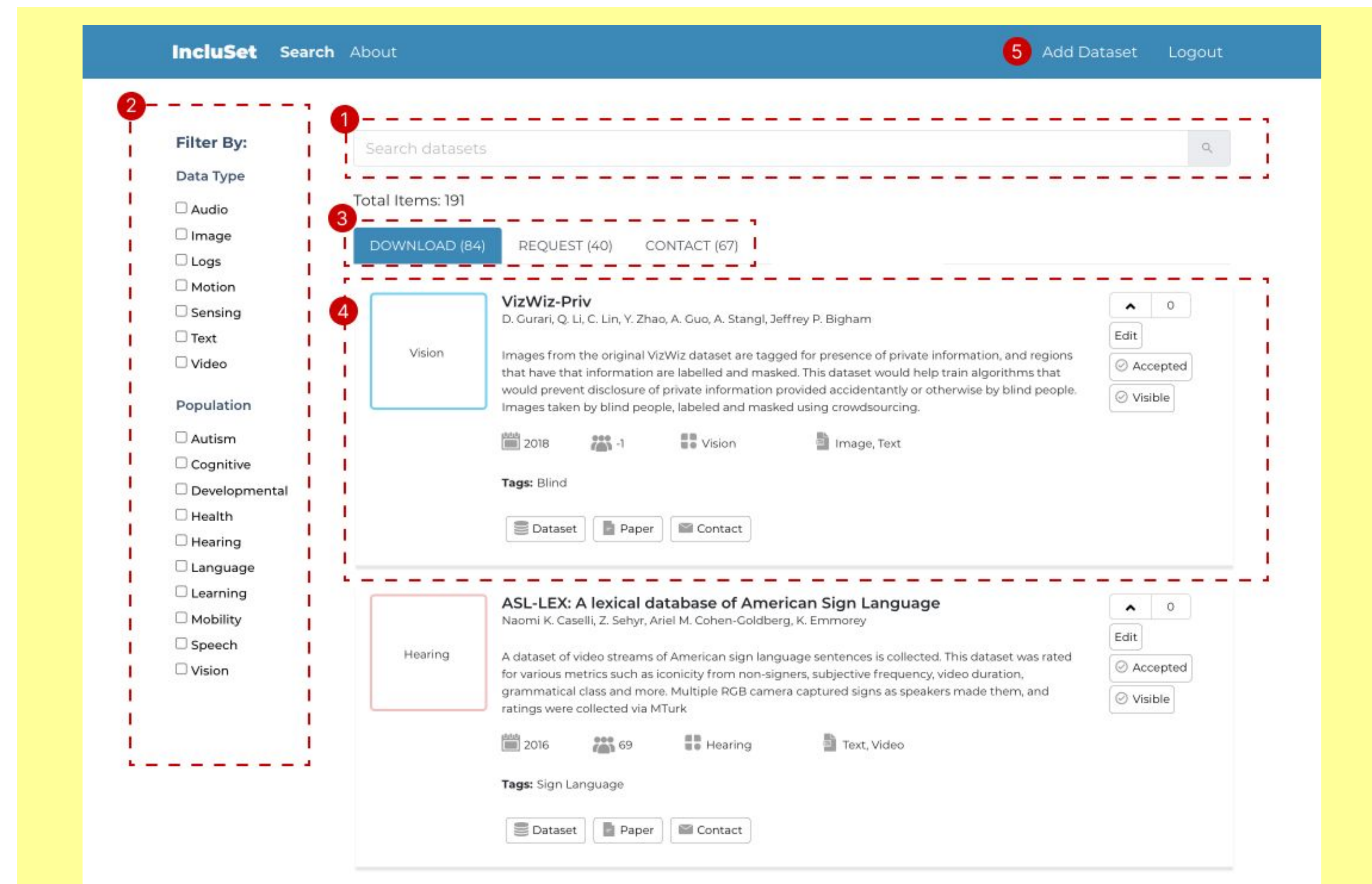
Discover accessibility datasets from Search page



Link new accessibility datasets via Add Dataset request form



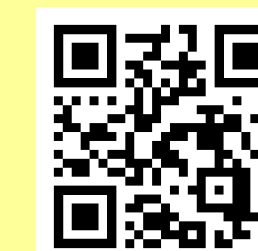
Community can moderate new or existing datasets



IncluSet: Interface

User can surface accessibility datasets by ① searching or ② filtering through data types and represented populations, which are ③ grouped by data access methods. To promote transparency and discovery, ④ each dataset is listed with metadata linking to the data source and description, and ⑤ user can also send a request to add a new dataset.

Available at <https://includset.com/>

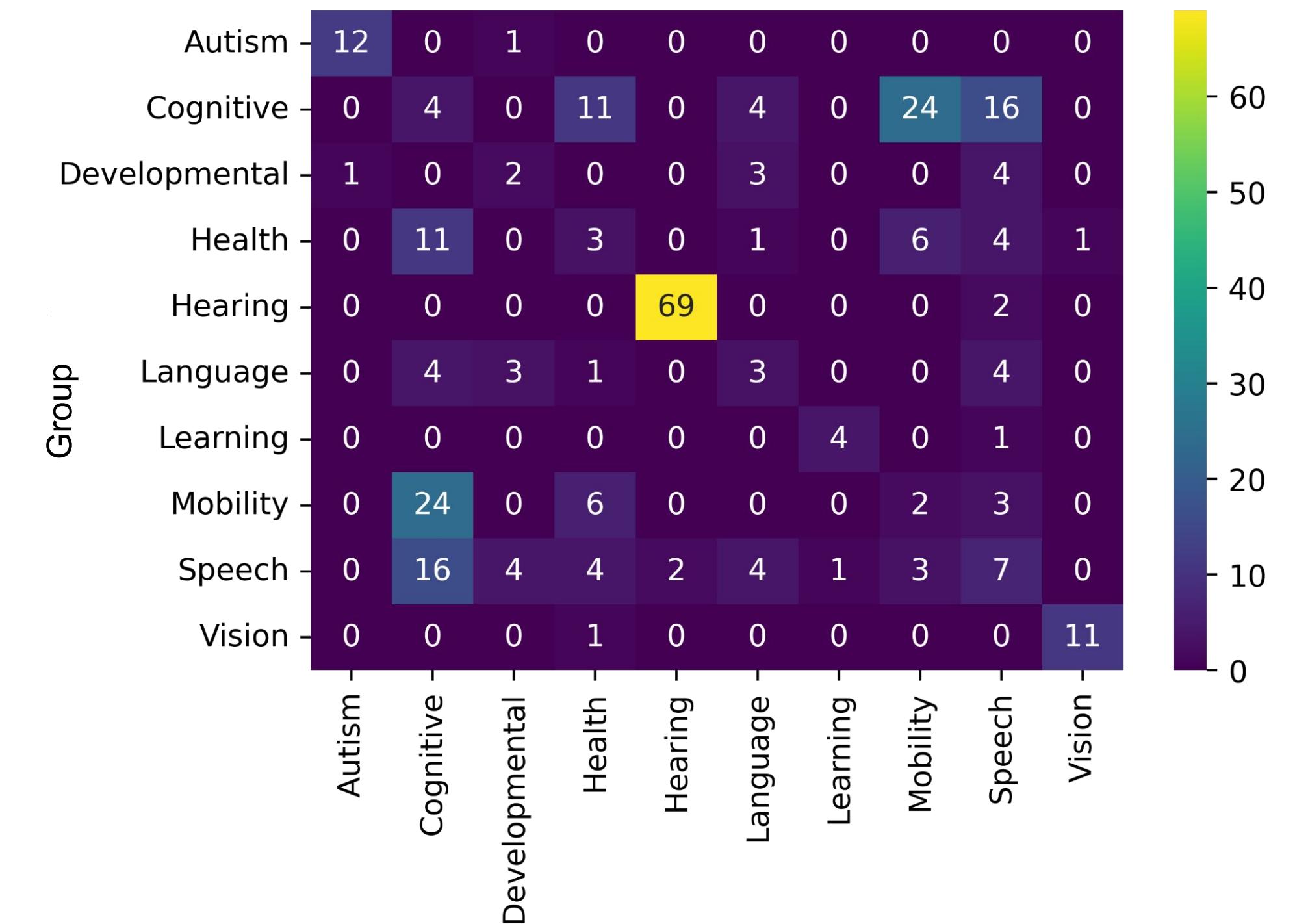


Examples of Datasets in Our Collection

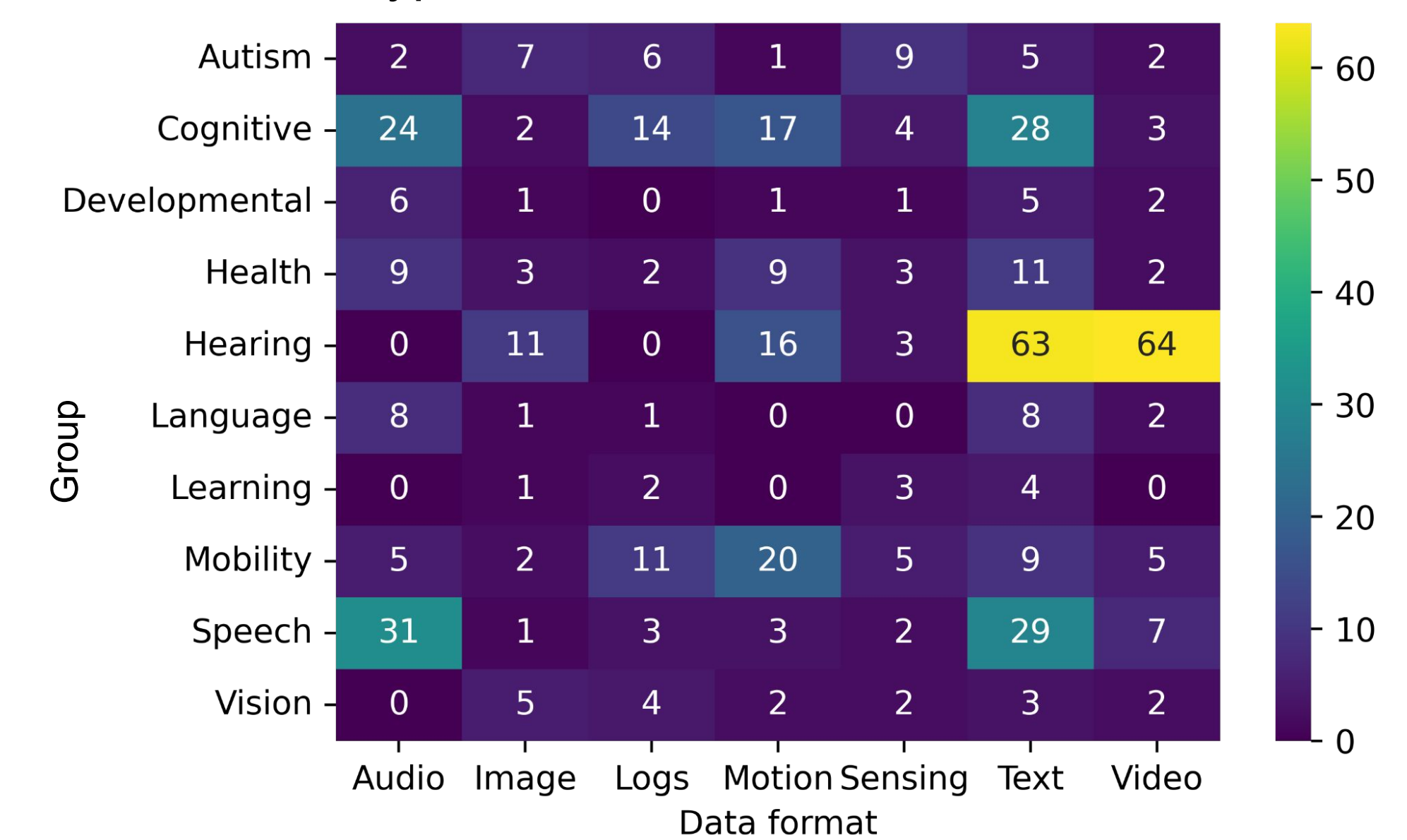
From left to right: photos of objects taken by a blind individual [4], motion captured signs [3], a human action dataset containing depth data sourced from older adults [5], eye-tracking measurements of autistic children [2], voice recordings of people with speech impairments such as dysphonia [1], and stroke gestures by people with motor impairments [6].

Dataset Distribution

Across Communities of Focus



Across Data Types



Relevant Publications

- Rie Kamikubo, Utkarsh Dwivedi, Hernisa Kacorri. 2021. Sharing Practices for Datasets Related to Accessibility and Aging. In The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21).
- Hernisa Kacorri, Utkarsh Dwivedi, Rie Kamikubo. 2020. Data Sharing in Wellness, Accessibility, and Aging. In NeurIPS Workshop on Dataset Curation and Security.
- Hernisa Kacorri, Utkarsh Dwivedi, Sravya Amancherla, Mayanka Jha, and Riya Chanduka. 2020. IncluSet: A Data Surfacing Repository for Accessibility Datasets. In The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20).

References

1. U. Cesari et al. 2018. A new database of healthy and pathological voices. Computers & Electrical Engineering.
2. H. Duan et al. 2019. A dataset of eye movements for the children with autism spectrum disorder. In Proceedings of the 10th ACM Multimedia Systems Conference.
3. S. Hassan et al. 2020. An Isolated-Signing RGBD Dataset of 100 American Sign Language Signs Produced by Fluent ASL Signers. In Proceedings of the LREC 2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives.
4. K. Lee and H. Kacorri. 2019. Hands Holding Clues for Object Recognition in Teachable Machines. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).
5. D. Leightley et al. 2015. Benchmarking human motion analysis using kinect one: An open source dataset. In 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA).
6. R. Vatavu and O. Ungurean. 2019. Stroke-Gesture Input for People with Motor Impairments: Empirical Results & Research Roadmap. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).