

DEVELOPING WELSH SPEECH RECOGNITION TECHNOLOGY

CYMEN

THE CHALLENGE

The lack of data for Welsh speech recognition technologies - voice recognition models that can turn speech into text or action.

THE GOAL

To improve Welsh speech recognition technology by developing data collection techniques.



"Our project has significantly developed Welsh voice recognition technology.

By fostering partnerships, training freelance transcribers, and developing practical tools, he has laid the foundations for a prosperous digital future for the Welsh language."

Myfyr Prys, Cymen



"An **unexpected challenge** was the **lack of freelance Welsh transcribers**, so we **trained over 15** of them including students, who will now be available to work in the field in the future."

Over 15
freelance
transcriptionists
trained

The project has
contributed to a 67%
reduction in WER
(Word Error Rate) for a
Welsh speech
recognition model

IMPACT

- All the data collected during the project is now available online as open source for **companies** to **use and develop** in the future.
- The project led to contributing to an **improvement in the Welsh speech recognition models**.

THE FUTURE

- The project will lead to an improvement of the **Welsh speech models** that are being built.
- It will also be possible to **create guidelines** on how best to process data that is collected.