Talk-a-Palooza

An AI-powered app to boost children language acquisition

Observations

5-10% of preschool-aged children experiencing speech delays

+

Overburdened Speech Therapy Waitlists

=

Significant impact on a child's language progress during crucial developmental stages

Talk-a-Palooza

Combine the power of **machine learning, natural language processing** and **pediatric speech acquisition research** to quickly **identify areas** where a child needs help, and create **custom language acquisition exercises and predictions**

The Vision

01

Personalized exercises

Tailored to the child's unique needs and stage of language development

02

Speech Recognition

Assess and track the child's pronunciation and intelligibility over time

03

Predictive learning

Predict the words the child is likely to learn next, facilitating targeted vocabulary expansion.

04

Gamification

Engagement through rewards: images, stories, songs...

05

Progress Tracking

Understand the real scope and evolution of language development

Project Goals Suggestions

	Project Phase	Nice to have	Vision
Medium	**Web App**	PWA, Hosted	Tablet & Mobile App
Speech Recognition	**Word Intelligibility (Y/N)**	Intelligibility score	Pronunciation exercises recommendations
Personalization & predictions	**Predictions based on available datasets**	AI-driven initial assessment test	Trainable Spaced Repetition Model
Progress Tracking	**Line graphs of word production count**	Semantic and phonological networks stats	Full report shareable with speech professionals.
Gamification	*TBD*	Collectible pictures	Nursery rhymes, games, short stories, Video face masks filters...

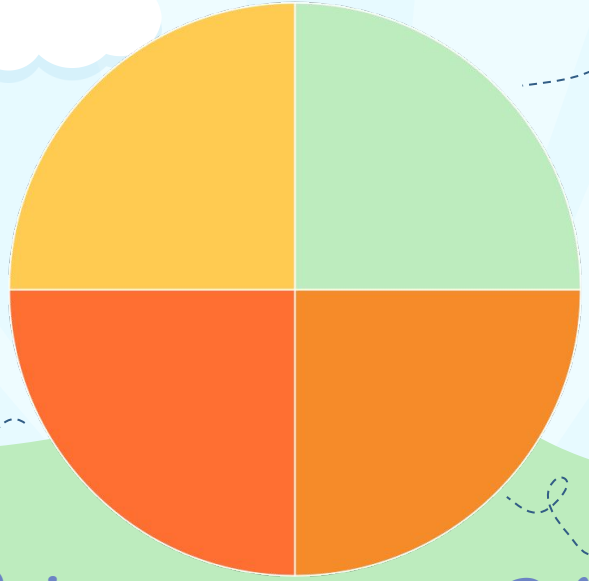
Fully "MultiTechie" Project

Deep Learning

Speech recognition, Algorithm for personalisation, Predictions!

Web Devt

Data visualisation: Interesting scope for both FE & BE



UX Design

Designing for kids requires especially fine-tuned usability and creativity

Data Science

Available datasets are significant, rich and solid.

Datasets & Resources

Datasets:

- **Wordbank** is *the* open database of children's vocabulary development. It contains data from 84,138 children and 94,451 CDI administrations, across 38 languages and 78 instruments It archives data from the **MacArthur-Bates Communicative Development Inventory (MB-CDI)**, a family of parent-report questionnaires and enables researchers to browse these data in **interactive analyses**

- **A Trainable Spaced Repetition Model for Language Learning**: an interesting research paper from Duolinguo, who also publicly released their data set and code: https://github.com/duolingo/halflife-regression

- **https://childes.talkbank.org/**: the largest database for child-language recordings and transcripts.

- **Children's Song Dataset** is an open source dataset for singing voice research which contains 50 English songs sung by one

Articles:

- **The Growth of Children's Semantic and Phonological Networks**

- **How Duolinguo AI learns what you need to learn**

- **The Amazing Ways Duolingo Is Using Artificial Intelligence To Deliver Free Language Learning**

Thank you!

Please feel free to ask any questions. 😄

By **Rose Jeantet**