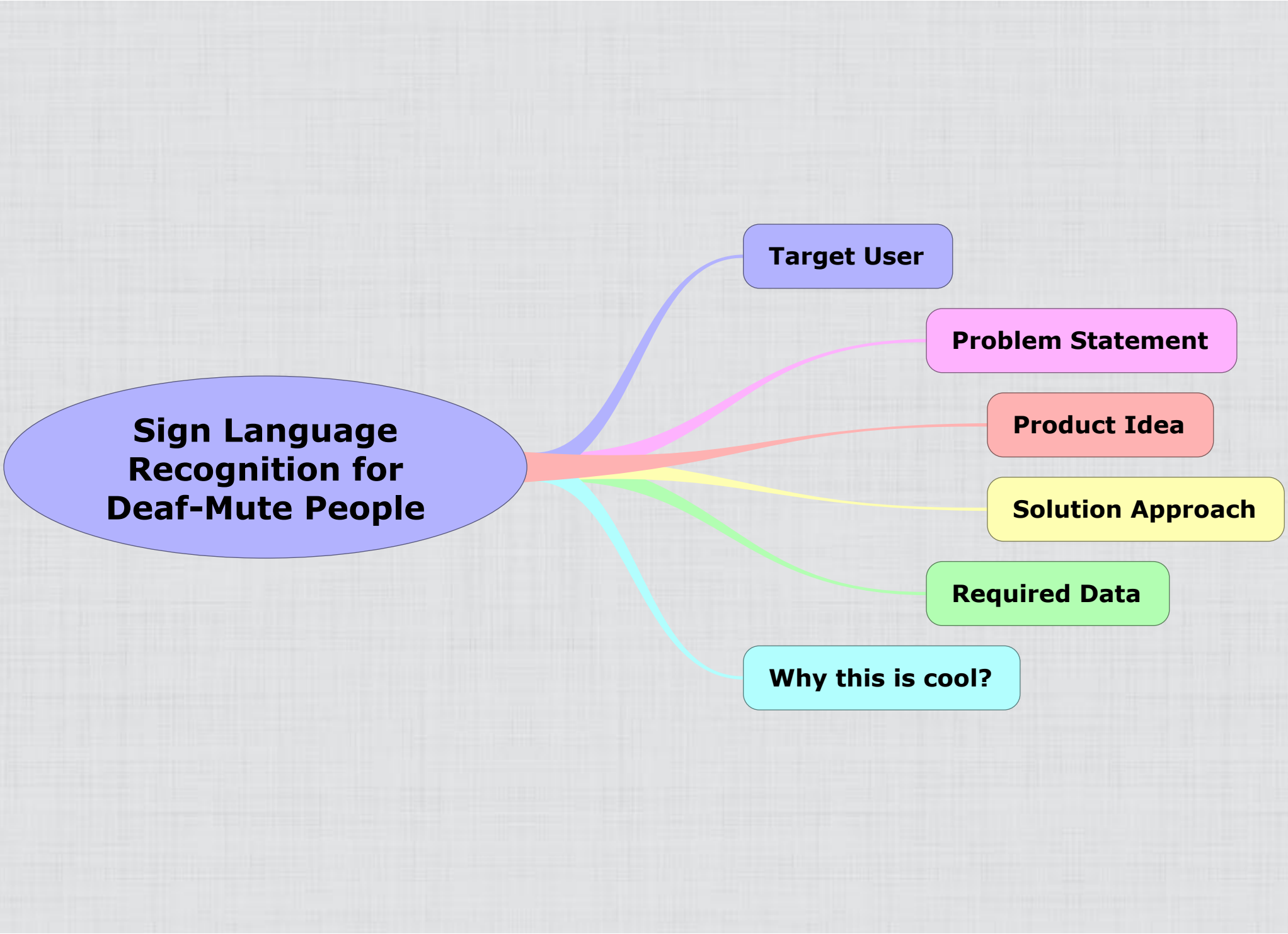
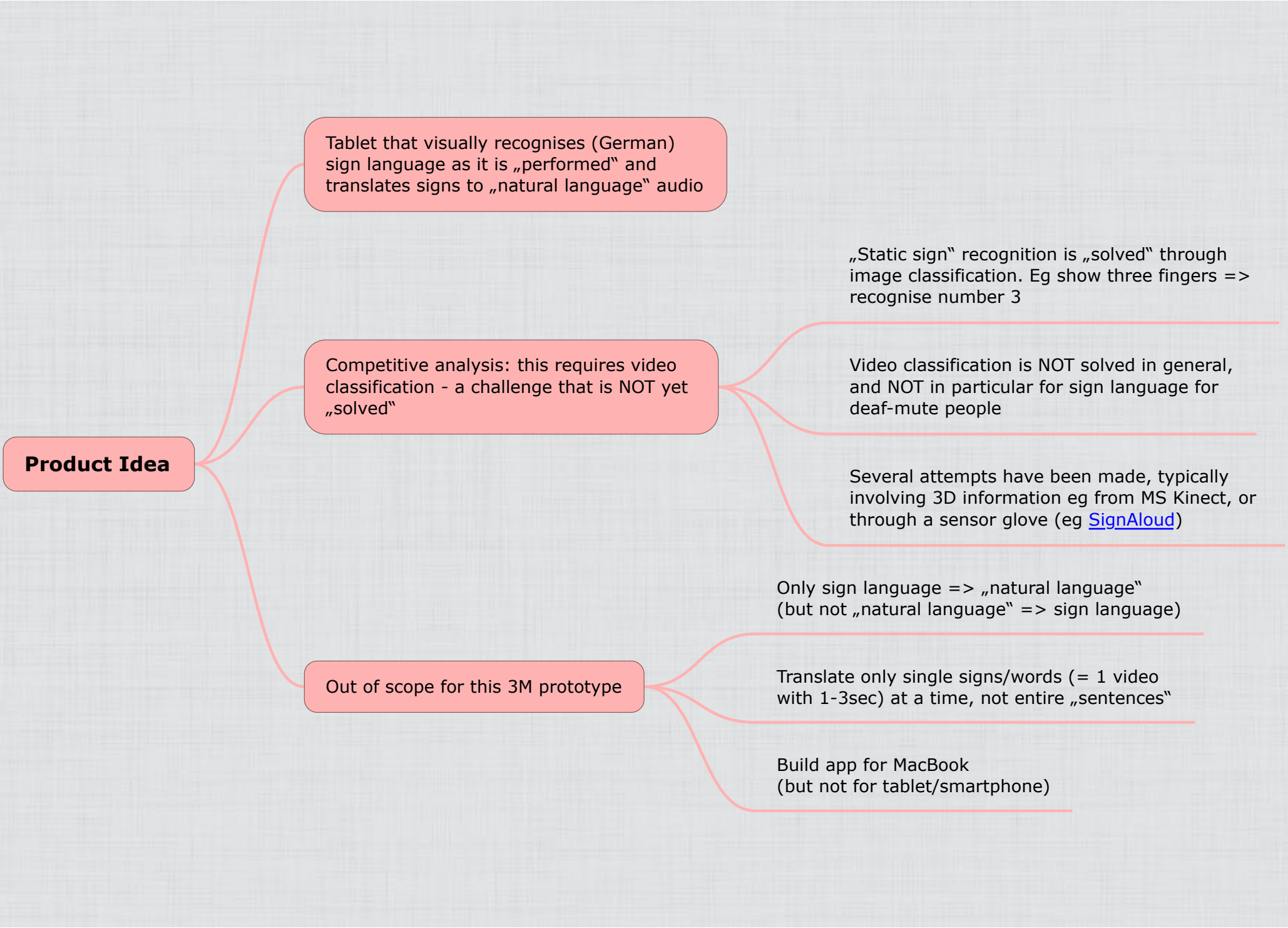

Sign Language Recognition for Deaf-Mute People
Target User
Problem Statement
Product Idea
Solution Approach
Required Data
Why this is cool?

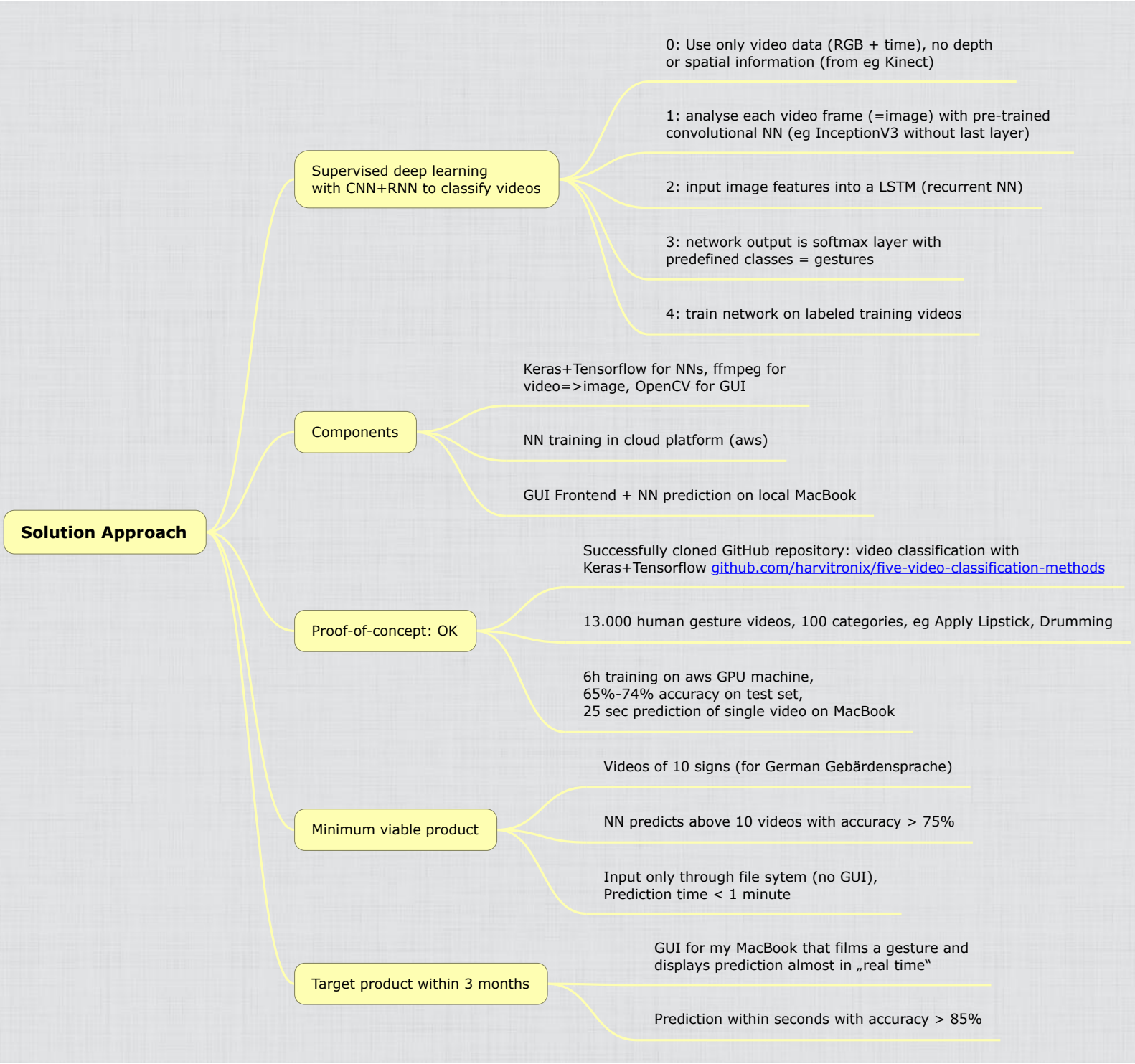
Target User

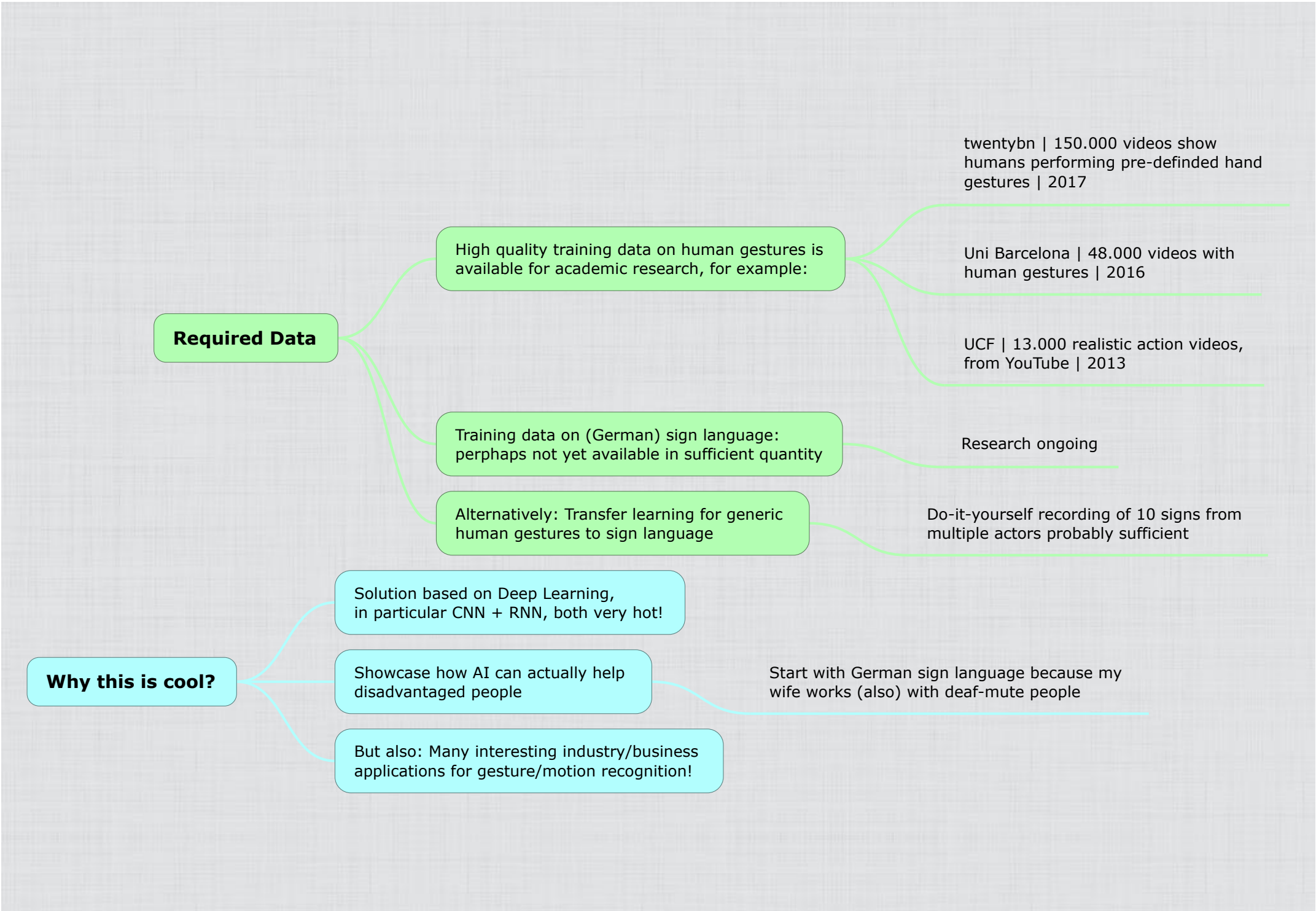
- Deaf-mute person who communicates only through sign language
- Approx 1-3 deaf-mute in 1000 persons
  In Europe approx 700.000 - 2.1M

Problem Statement

- Sign language for deaf-mute people: complex body gestures
  - Different sign languages all over the world. „Gebärdensprache" in Germany is very different from American sign language
  - Have their own grammar/syntax (not copied from their naturual language cousins)
  - Complex body movements required. The „flow" is essential, not only static signs
  - Example: Gebärdensprachkurs (DGS) 1. "Begrüßung"
- Hard to understand by „natural language speakers" - typically only relatives and teachers make the effort to learn sign language
- Deaf-mute people often need/want to „talk" to other „natural language" people


Product Idea
Tablet that visually recognises (German) sign language as it is „performed" and translates signs to „natural language" audio
Competitive analysis: this requires video classification - a challenge that is NOT yet „solved"
„Static sign" recognition is „solved" through image classification. Eg show three fingers => recognise number 3
Video classification is NOT solved in general, and NOT in particular for sign language for deaf-mute people
Several attempts have been made, typically involving 3D information eg from MS Kinect, or through a sensor glove (eg SignAloud)
Out of scope for this 3M prototype
Only sign language => „natural language" (but not „natural language" => sign language)
Translate only single signs/words (= 1 video with 1-3sec) at a time, not entire „sentences"
Build app for MacBook (but not for tablet/smartphone)


Solution Approach
Supervised deep learning with CNN+RNN to classify videos
0: Use only video data (RGB + time), no depth or spatial information (from eg Kinect)
1: analyse each video frame (=image) with pre-trained convolutional NN (eg InceptionV3 without last layer)
2: input image features into a LSTM (recurrent NN)
3: network output is softmax layer with predefined classes = gestures
4: train network on labeled training videos
Components
Keras+Tensorflow for NNs, ffmpeg for video=>image, OpenCV for GUI
NN training in cloud platform (aws)
GUI Frontend + NN prediction on local MacBook
Proof-of-concept: OK
Successfully cloned GitHub repository: video classification with Keras+Tensorflow github.com/harvitronix/five-video-classification-methods
13.000 human gesture videos, 100 categories, eg Apply Lipstick, Drumming
6h training on aws GPU machine, 65%-74% accuracy on test set, 25 sec prediction of single video on MacBook
Minimum viable product
Videos of 10 signs (for German Gebärdensprache)
NN predicts above 10 videos with accuracy > 75%
Input only through file sytem (no GUI), Prediction time < 1 minute
Target product within 3 months
GUI for my MacBook that films a gesture and displays prediction almost in „real time"
Prediction within seconds with accuracy > 85%


Required Data
High quality training data on human gestures is available for academic research, for example:
twentybn | 150.000 videos show humans performing pre-definded hand gestures | 2017
Uni Barcelona | 48.000 videos with human gestures | 2016
UCF | 13.000 realistic action videos, from YouTube | 2013
Training data on (German) sign language: perphaps not yet available in sufficient quantity
Research ongoing
Alternatively: Transfer learning for generic human gestures to sign language
Do-it-yourself recording of 10 signs from multiple actors probably sufficient
Why this is cool?
Solution based on Deep Learning, in particular CNN + RNN, both very hot!
Showcase how AI can actually help disadvantaged people
Start with German sign language because my wife works (also) with deaf-mute people
But also: Many interesting industry/business applications for gesture/motion recognition!