

Language Trainer

Team 1

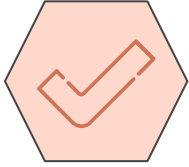


Introduction

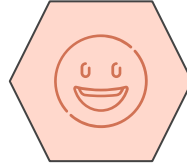
This project explores the critical role of deep learning on new language learning through mobile devices.



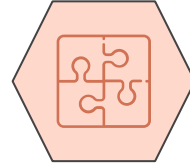
Advantages of learning apps



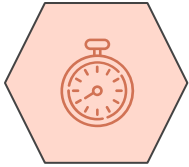
Easy to use



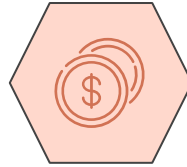
**Content of
one's choice**



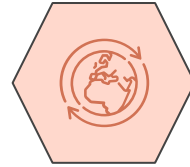
Own pace



Learn anytime

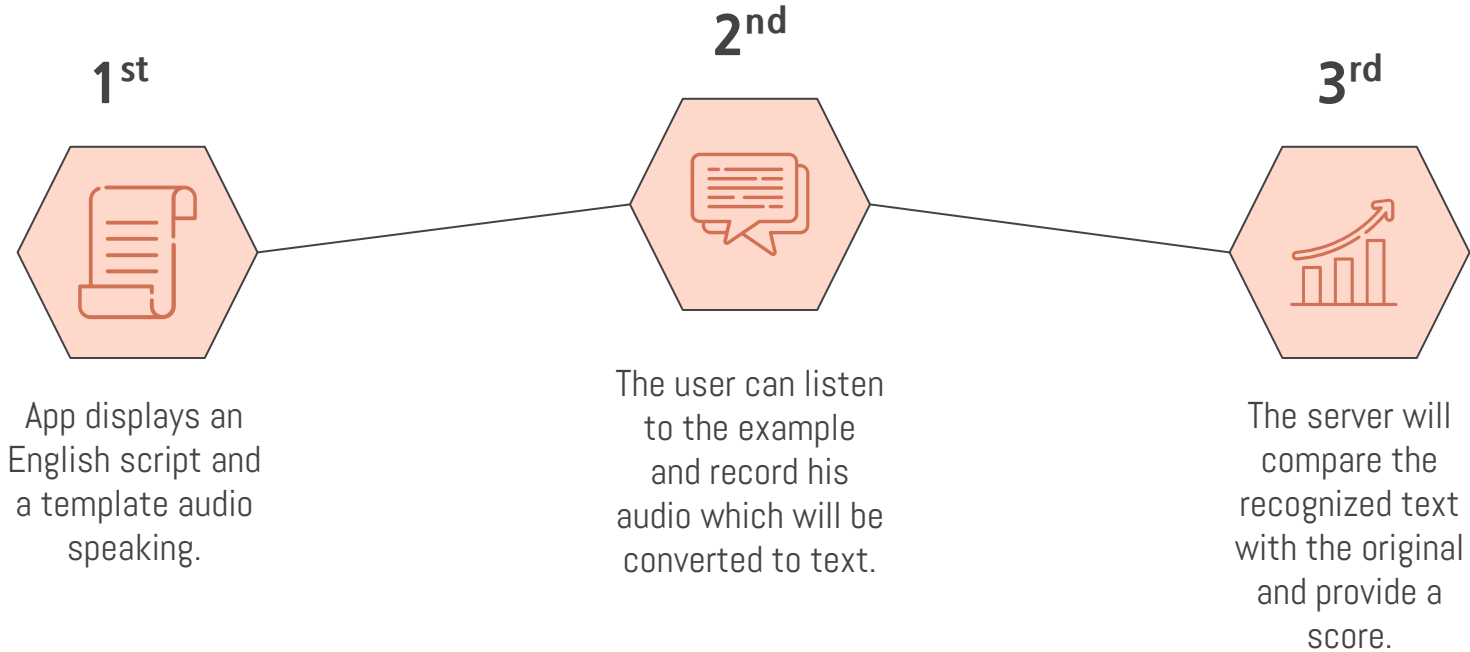
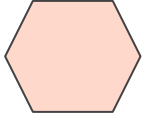


Affordable



**Learn
anywhere**

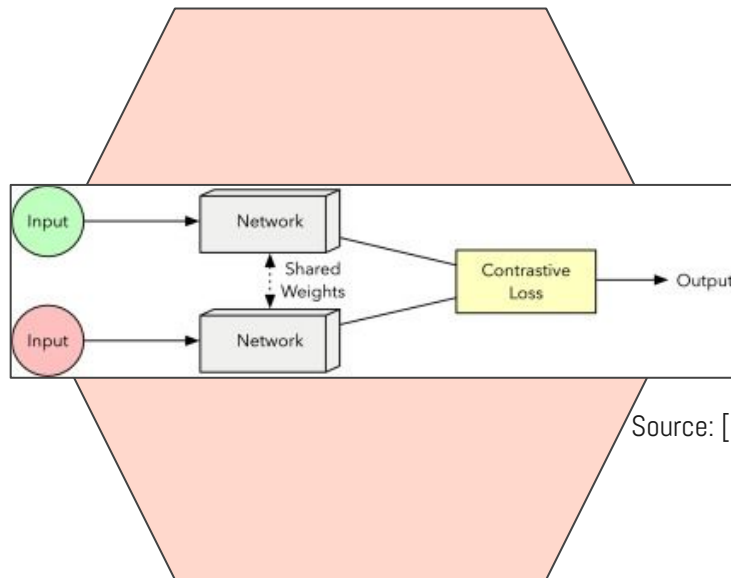
Procedure



Siamese neural networks

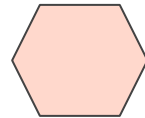
This algorithm is the key to calculating the similarity between the script and the recognized text. They have been proven successful in other tasks like:

- Signature and face verification
- Image and sentence similarity



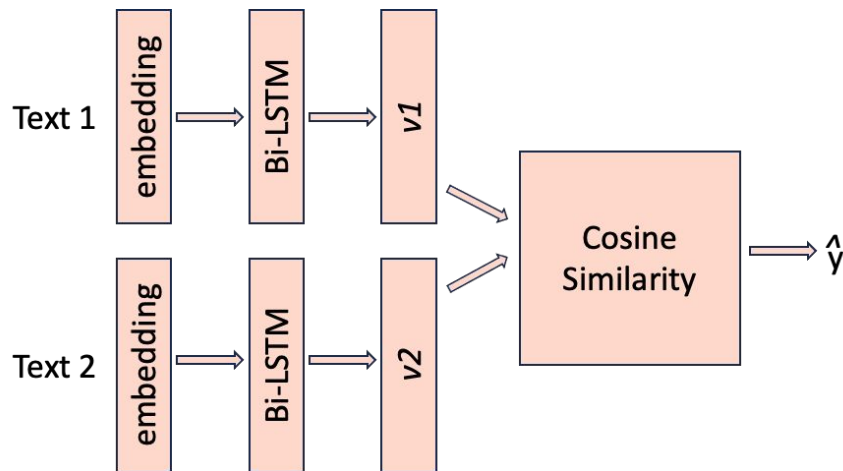
Source: [1]

Model in our Task



In our task, for each subnetwork, a embedding layer and a bidirectional long short-term memory (BiLSTM) layer is used to extract the feature of the input sentence. The feature vectors are compared by the cosine-based similarity.

- Embedding layer
- Bidirectional long short-term memory (BiLSTM)
- Mean - Averages the embedding features across all words for each sample in the batch
- Normalization - L2
- Loss Function - Triplet loss function



Model Training

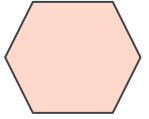
We test our model in Quora's first public dataset regarding the problem of recognizing duplicate questions. The total amount of samples in this dataset is over 400,000. Each sample has a duplicate question pair, ID, and a binary Integer of whether the question pair is duplicate.

- Dataset information
- Data preprocess - splitting the dataset, tokenizing and padding the text sequences.
- Hyperparameter is shown in the right table.

	id	qid1	qid2	question1	question2	is_duplicate
0	0	1	2	What is the step by step guide to invest in sh...	What is the step by step guide to invest in sh...	0
1	1	3	4	What is the story of Kohinoor (Koh-i-Noor) Dia...	What would happen if the Indian government sto...	0
2	2	5	6	How can I increase the speed of my internet co...	How can Internet speed be increased by hacking...	0
3	3	7	8	Why am I mentally very lonely? How can I solve...	Find the remainder when 23^{24} $[/math] i...$	0
4	4	9	10	Which one dissolve in water quikly sugar, salt...	Which fish would survive in salt water?	0

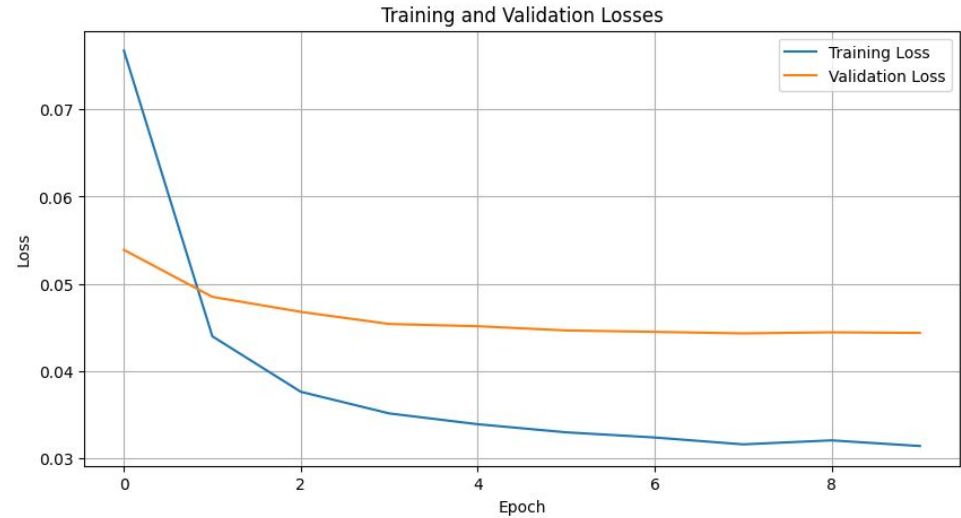
Hyperparameter	Value
Learning rate	0.01
Epochs	10
Optimizer	Adam
Loss function	Triplet Loss
Embedding dimension	128
LSTM hidden size	128
Batch size	256

Testing Result

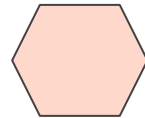


We test the trained model on the splitted test dataset

- Accuracy: 70.29%
- Training and Validation loss line chart

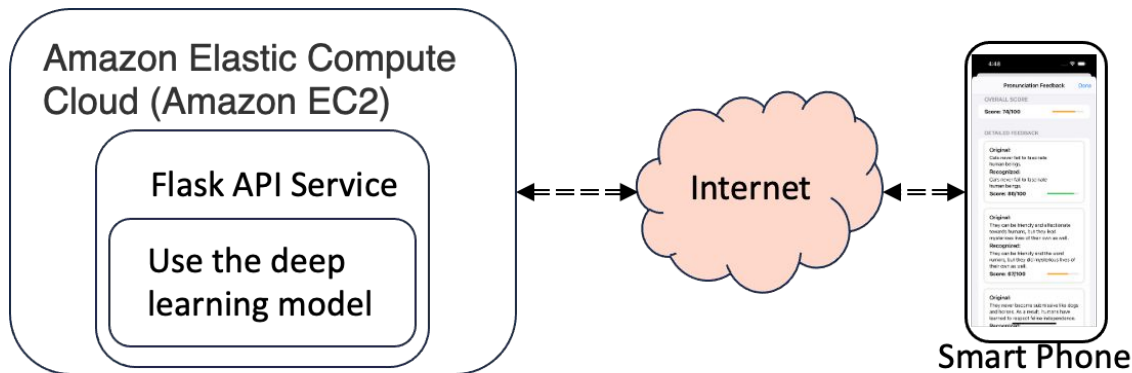


Model Deployment

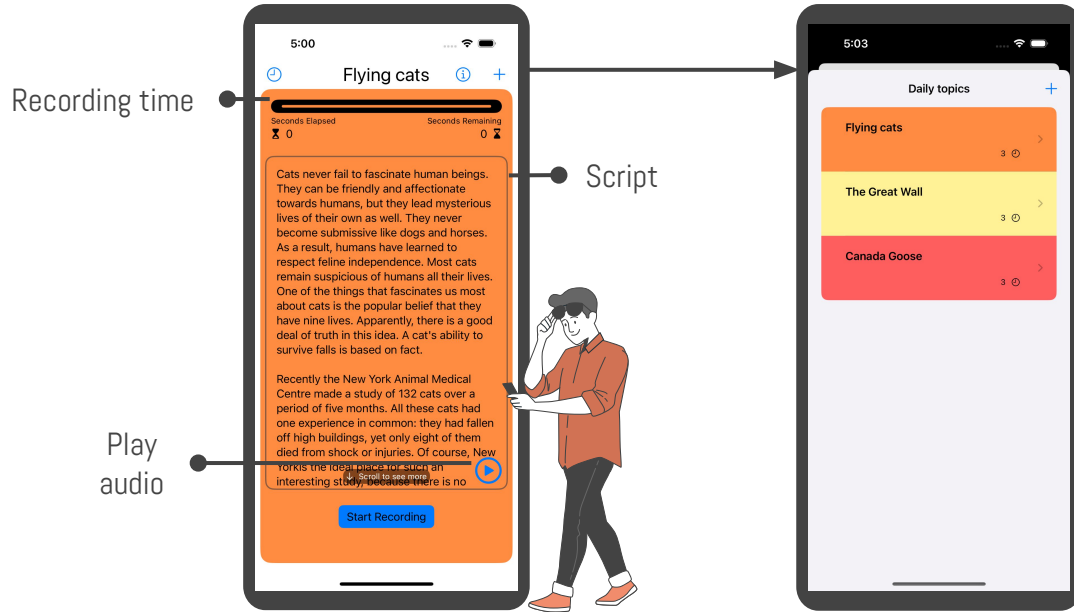


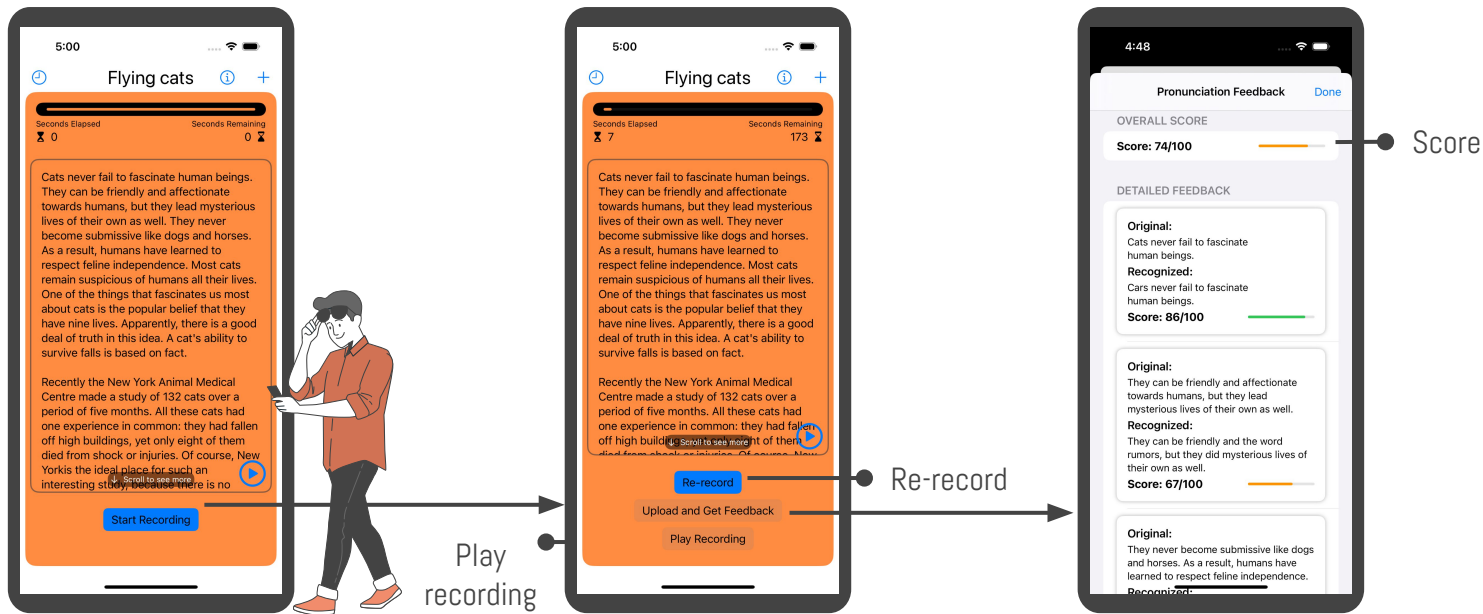
We deploy a trained model on the AWS EC2 cloud host.
And Flask API framework is used to implement a http server for mobile app to request.

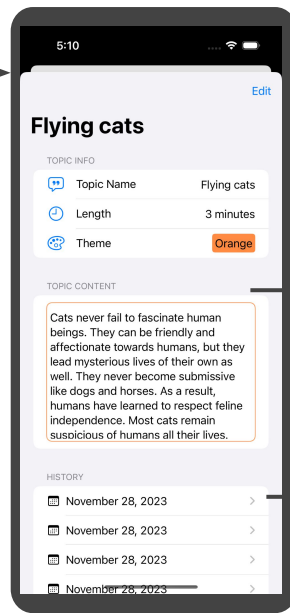
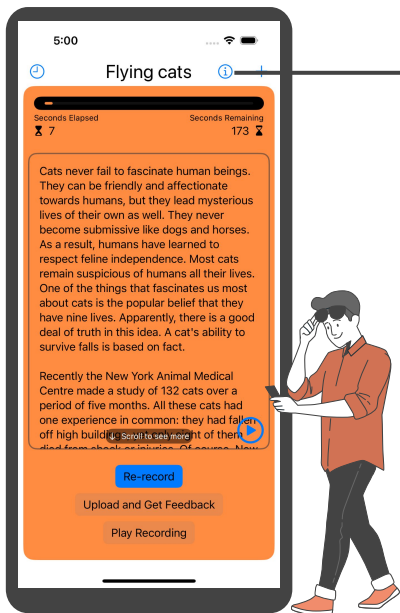
- AWS EC2
- Flask API



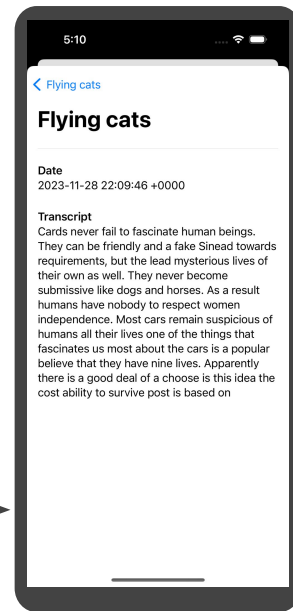
App flowchart



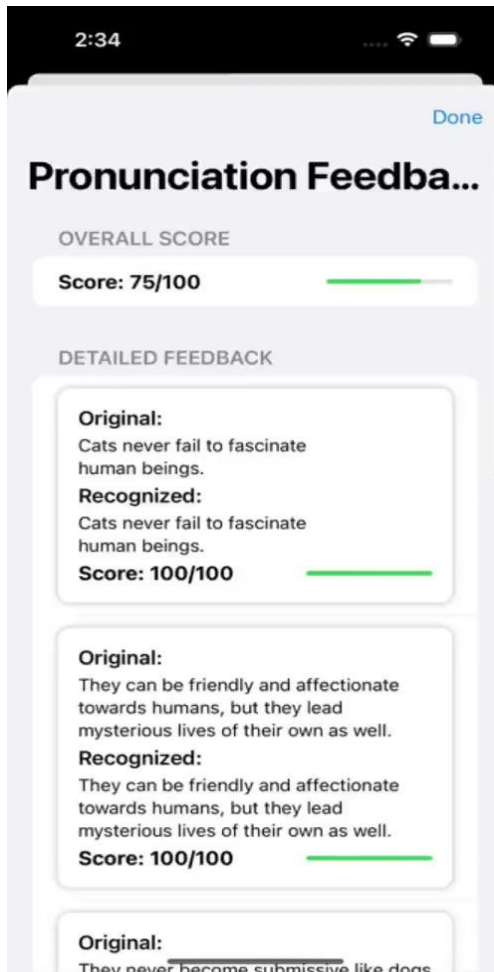




History
recording



App Demo



References

[1] Gustavo H. de Rosa, João Paulo Papa, Chapter 7 - Learning to weight similarity measures with Siamese networks: a case study on optimum-path forest, Editor(s): Alexandre Xavier Falcão, João Paulo Papa, Optimum-Path Forest, Academic Press, 2022, Pages 155-173, ISBN 9780128226889, <https://doi.org/10.1016/B978-0-12-822688-9.00015-3>.



Thanks

CREDITS: This presentation template was created by **Slidesgo**,
including icons by **Flaticon**, infographics & images by **Freepik**
and illustrations by **Storyset**

