Skill Based LU

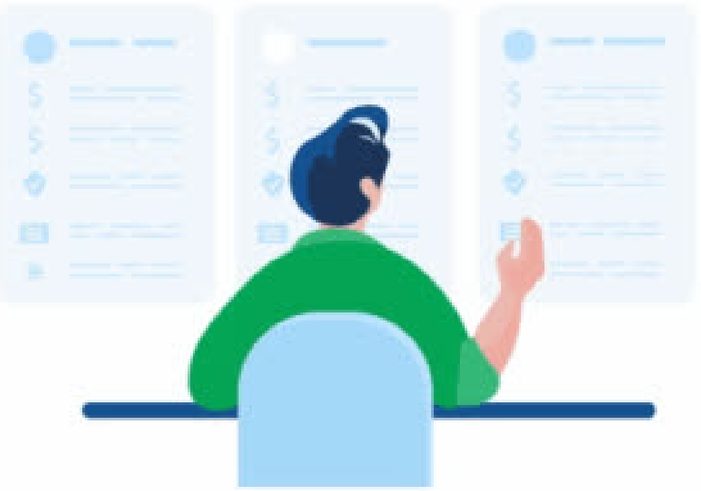
CANVAS ENG PROJECT IDC

The Team

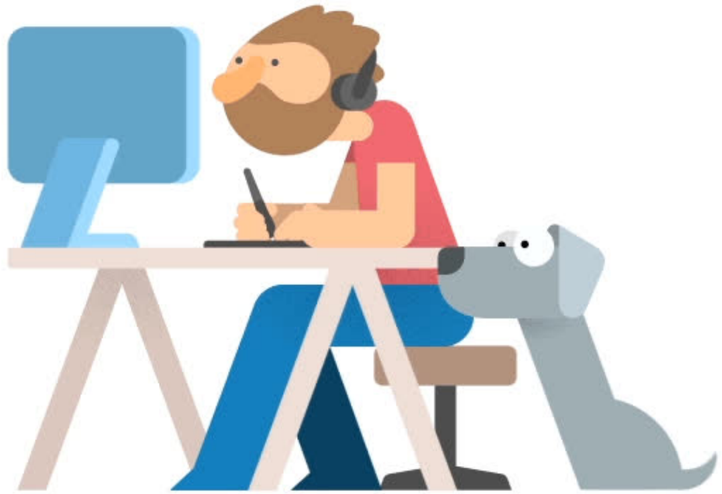
Dev Interns	:	Abhirupa Mitra, Koushiki Dasgupta Chaudhuri, Parth Shettiwar, Shivam Goel
Dev Coaches	:	Ankit Jain, Amit Kumar Yadav
PM Coach	:	Abhishek Agarwal
UX Coach	:	Shourya Mehrotra

Objectives

- Enable a voice command mode which takes commands in the offline mode
- An ML based module running on client must be able to infer the right command from set of available commands.
- Accurate distinction between positive and negative intents.

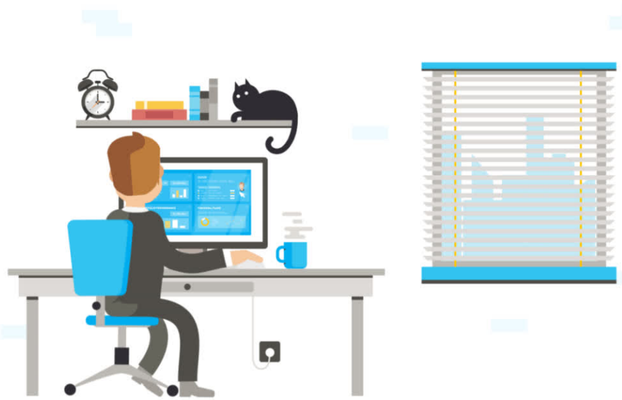


Requirements



- Switch to move from transcription to command mode.
- Size (<250 kb)
- Accuracy (>95%)
- The model should be downloadable dynamically based on scenario and language.

Final Models Chosen



- Average Word2Vec (TFlite) : Android(<300kB size)
- Linear Regression : Android(~12kB size)
- ML Text Classifier : iOS(<50kB size)

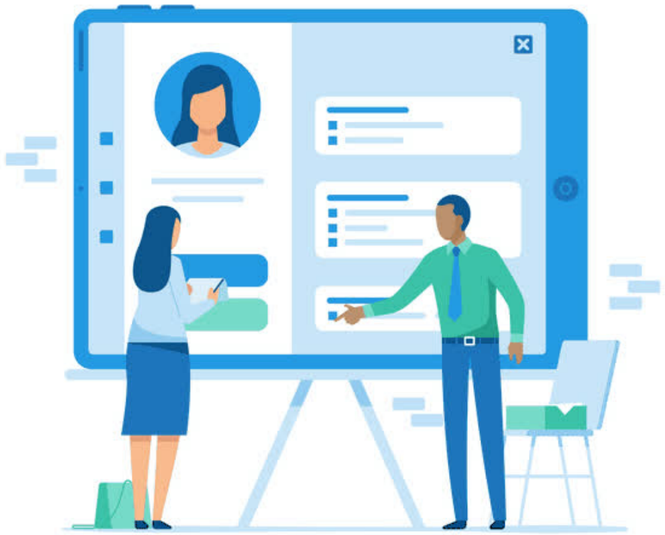
Parameter extraction from commands

- Regex matching was used to identify parameters like last word, last sentence, that etc.
- Another text classifier to classify parameters on top of the command classifier was explored.
- A named entity recognition model to identify categories of parameters like context, number and object was also explored.

Improving Speech-to-Text Accuracy

- **22 commanding intents** were included in the dataset .
- An exhaustive **voice dataset** of these commands were fed into offline speech-to-text API.
- **Similar sounding words** were included in the dataset to improve accuracy.
- Handled offline **Speech to text Conversion errors** for both Google and Siri.

Dynamic download of models on device



- Dynamic download of models on device was explored to reduce size.
- Download using FTP, Azure Blob Storage and direct download from URL was explored.
- Easy switch between models as per user requirement was implemented.

Implemented Commands

UNDO

BOLD

REMOVE_BOLD

ITALIC

REMOVE_ITALIC

UNDERLINE

REMOVE_UNDERLINE

SUPERSCRIPT

REMOVE_SUPERSCRIPT

SUBSCRIPT

REMOVE_SUBSCRIPT

STRIKETHROUGH

REMOVE_STRIKETHROUGH

DELETE

REMOVE_FORMATTING

INSERT_COMMENT

ALIGN_LEFT

ALIGN_RIGHT

ALIGN_CENTER

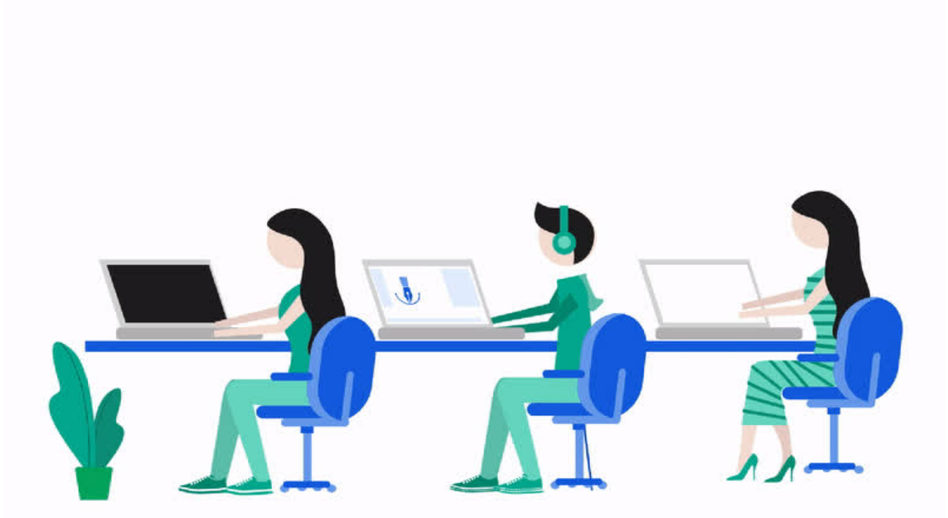
INSERT_BULLET

NEXT_BULLET

END_BULLET

Future Prospects

- Our models, being easily scalable, can be easily shifted to other clients like Outlook.
- Offline dictation can be implemented.
- Own speech to text recognizer can be built to further improve accuracy .



Thank you