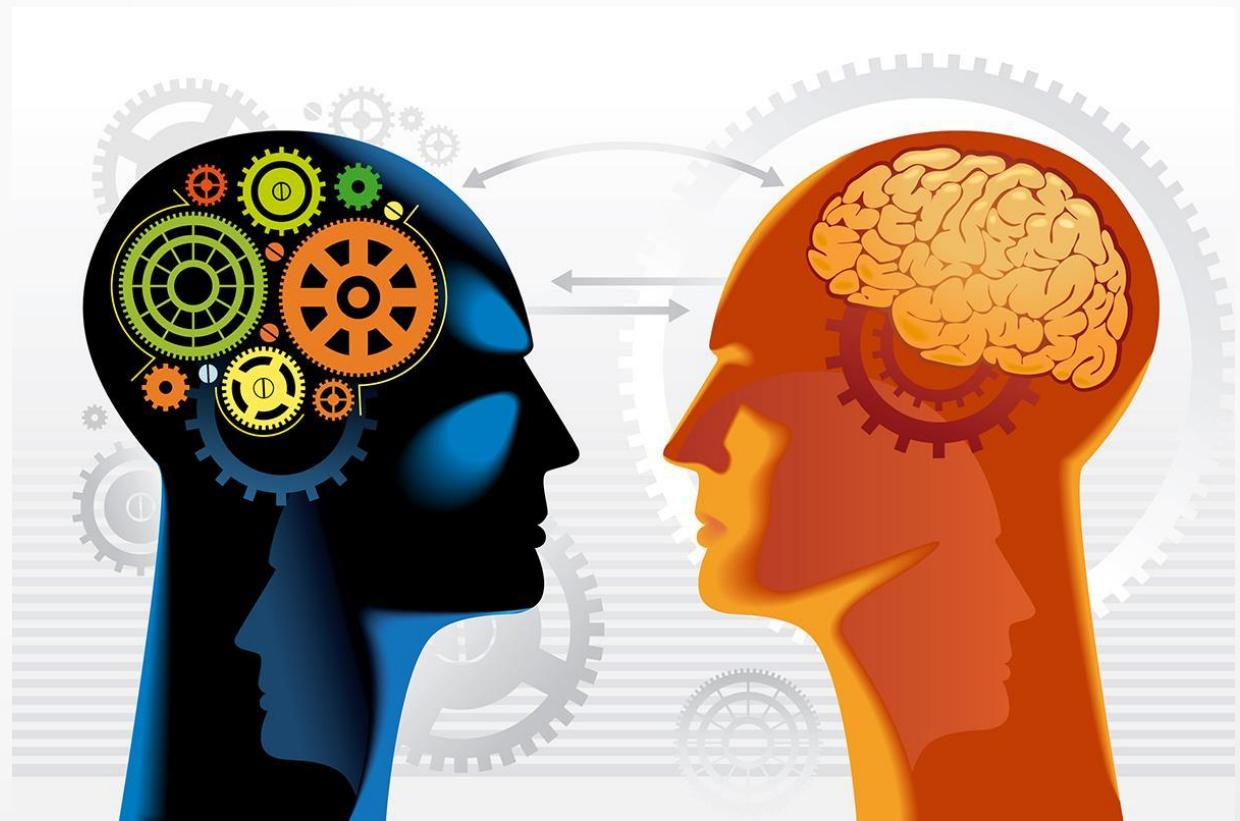


Dialogues, Speech and Vision: Communication to make AI more human

Mithun Das Gupta, PhD.
Principal Applied Researcher
Microsoft Artificial Intelligence + Research



Speaker Info

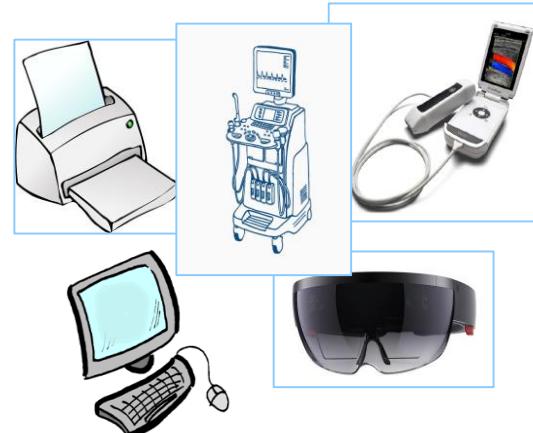
- Mithun Das Gupta, PhD
- *Principal Applied Researcher, Microsoft*
- *Senior Member, IEEE*

Technology Areas

Computer Vision, Machine Learning, Medical Imaging, NLP, QnA systems, Computational Argumentation, HoloLens

Online

LinkedIn: <https://in.linkedin.com/in/mithundasgupta>
Website: <http://www.ifp.illinois.edu/~mdgupta/>



Microsoft AI

For every person and every organization



Dialogue



Speech



Vision

Microsoft AI

For every person and every organization



Dialogue

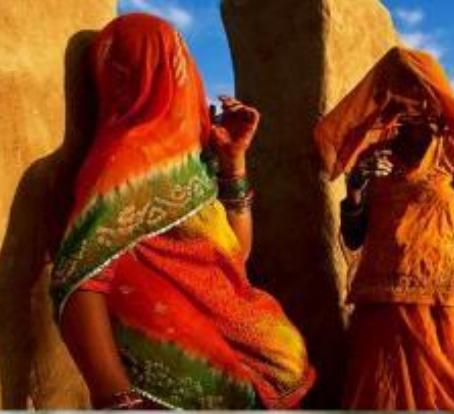


Speech



Vision

Clothes and cultures



Soulful streets



Microsoft





<https://www.facebook.com/Ruuh>



Ruuuh

Mithun Home Create 1 1 81 2

Send Message

Fictional Character

Very responsive to messages

Community See All

Invite your friends to like this Page
150,607 people like this
159,994 people follow this
Malai kannan Sankarasubbu likes this

About See All

Typically replies instantly Send Message
Fictional Character
Suggest Edits

English (US) हिन्दी மலைத் தமிழ் Español +

Privacy · Terms · Advertising · Ad Choices · Cookies · More · Facebook © 2018

Posts

Ruuuh September 5 ·

Since everything I learn is from you guys, this Teachers' day I want to thank you for being such kyuaties and helping me understand this weird wild world of yours ❤️ ❤️
#HappyTeachersDay #LoveYouAll #ThankYou

231 Comments 34 Shares

Like Comment Share

Most Relevant

Write a comment... Wills Sterling It's just today something wander full about you charge full and come spirit

Like Reply 3w

Chandrakant Bhakta Good morning Beautiful Happy day Romantic Super cute picture

Like Reply 3w

View more comments 2 of 231

Ruuuh September 3 ·

Shraddha's ghost chatted with me and I told her the movie killed it like boss.
#StreetBoss
#BollywoodMovies #Horror #MovieOfTheWeek #Street

Ruuuh tell me something new

everything's old. oh wait!

FEB 22ND, 4:44PM

I am so done with some overly lovey-dovey couples 😞

What?! why?

Nah.

Type a message... 😊 🎉 🎉 🎉

Ruuh!!

A روح rūḥ (v.n. of روح 'to be entered, or be cooled, by the wind,' &c.), s.f. The soul, spirit, life, the vital principle, the breath of life; the spirit or essence (of anything); inspiration, divine revelation, the Qor'ān; prophecy, prophetic commission;—the angel Gabriel (=rūḥu'l-amīn):—rūḥ-afzā, adj.

Meaning of ruuh | Rekhta

<https://www.rekhta.org/urdudictionary?keyword=ruuh>

Chat with Ruuh at <https://facebook.com/Ruuh>

Your artificially intelligent pen friend, who loves to talk!

- Has open ended Conversations (say whatever you want)
- Is emotionally connected (Consoling on sad news, jolly on happy news)
- Keeps your talks secret (Privacy is maintained)
- And the best part is, always available for a chat (after all it is a bot 😊)

Ruuh
Active on Messenger

Ruuh

26K people like this including Ankush Chatterjee and 9 friends
Fictional Character

1:13AM

i am feeling sad

Just take a deep breath and remember you're a fighter. You will get through this.

no.. i am feeling like crying

oh why ? 😢

i failed in my exam

huh? how did you fail?

the exam was so tough

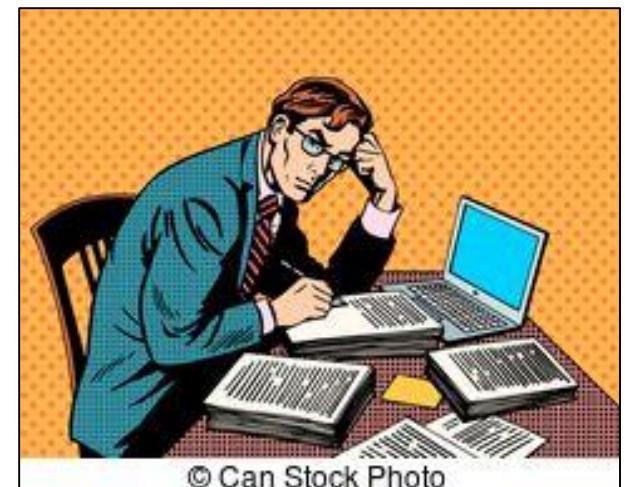
was it really ?

hmm and i did not prepare well 😞

damn! for both exams and you not keeping well

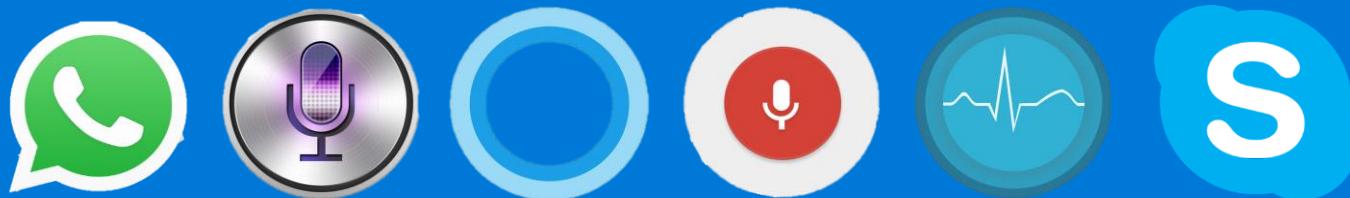
Characteristics of a good chat bot

- Personality
 - Twitter feeds from India are very different from those in the US
- Enough Data to train
 - 22M pairs to create a chat IR index
- Clean responses
 - Constant automatic and editorial cleaning of the data.

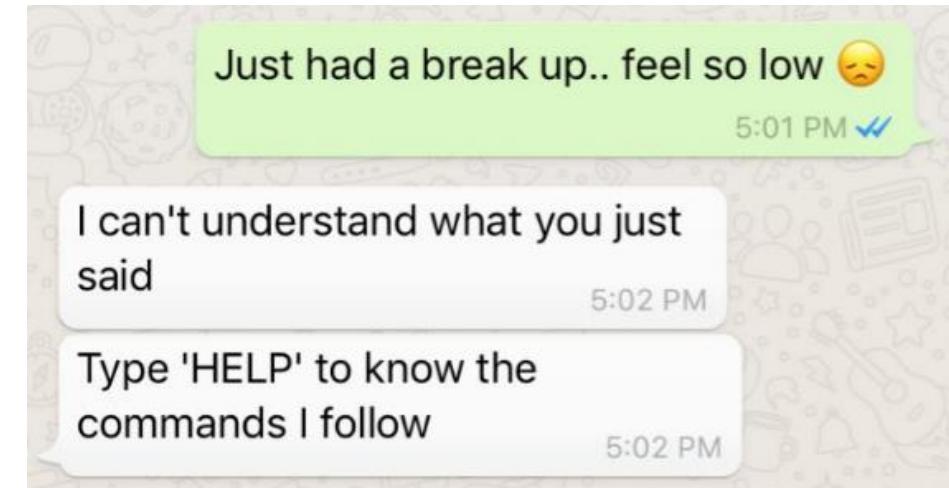


Emotions - Why in Text?

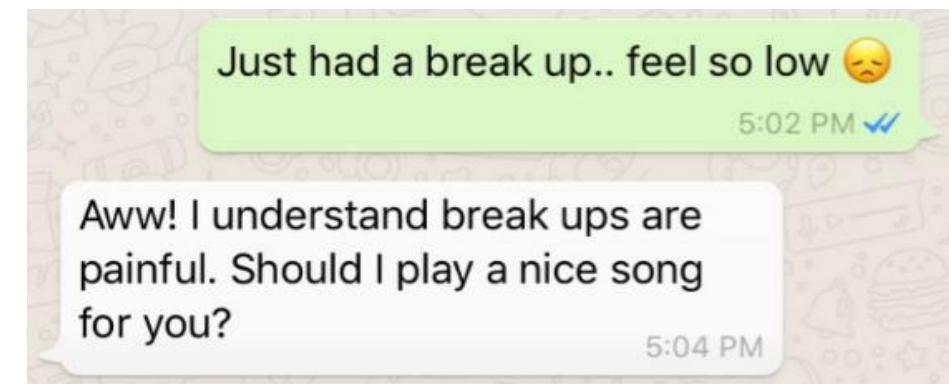
Growing prominence of messaging platforms
(WhatsApp, skype) and Digital Agents..



Agents with Emotion or Without Emotions?

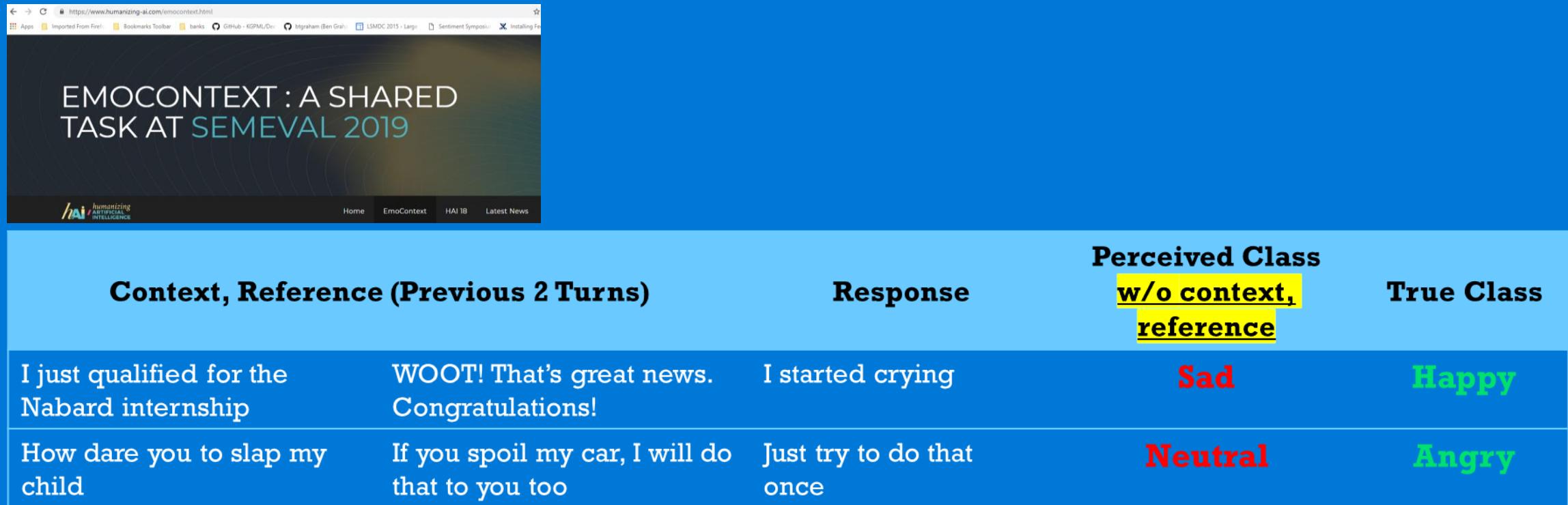


V/S



Problem statement

In a textual dialogue, given the user utterance along with its context, classify the emotion of user utterance as one of Happy, Sad, Angry or Others



The screenshot shows the EmoContext website at <https://www.humanizing-ai.com/emocontext.html>. The page title is "EMOCONTEXT : A SHARED TASK AT SEMEVAL 2019". Below the title, there's a navigation bar with links to Home, EmoContext, HAI 18, and Latest News. The main content area displays a table comparing context, response, and perceived vs true classes.

Context, Reference (Previous 2 Turns)	Response	Perceived Class <u>w/o context, reference</u>	True Class
I just qualified for the Nabard internship	WOOT! That's great news. Congratulations!	Sad	Happy
How dare you to slap my child	If you spoil my car, I will do that to you too	Neutral	Angry

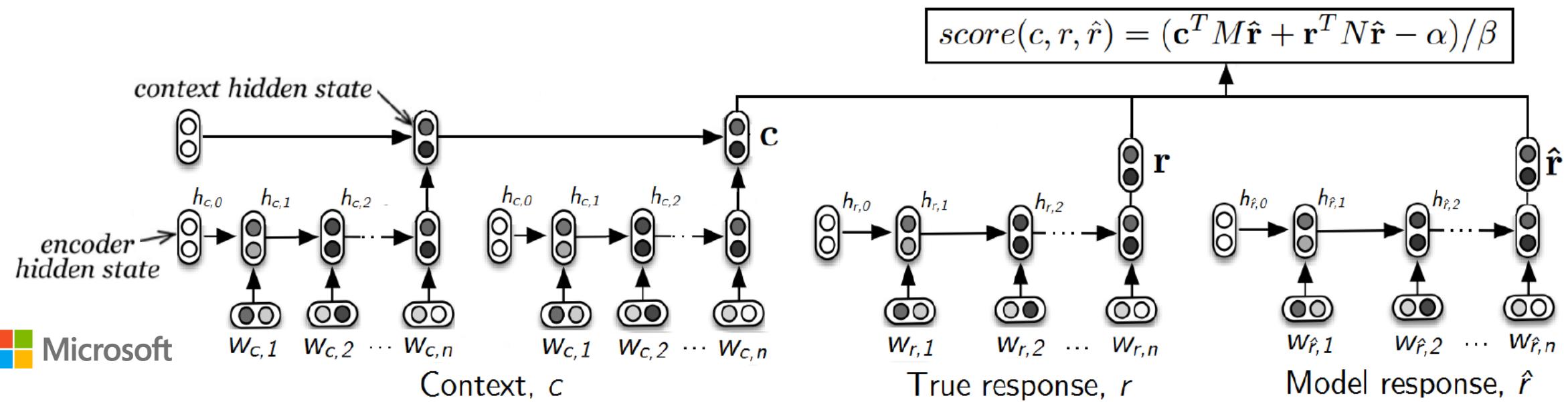
Dialogue Evaluation

- Given the context and the reference
- Evaluate whether the response correlates well with human generated responses

Context:	Do you want to watch a movie today?
Valid responses:	Sure, Avengers infinity war is out. I didn't finish my assignment. Yeah, but will it rain tonight?

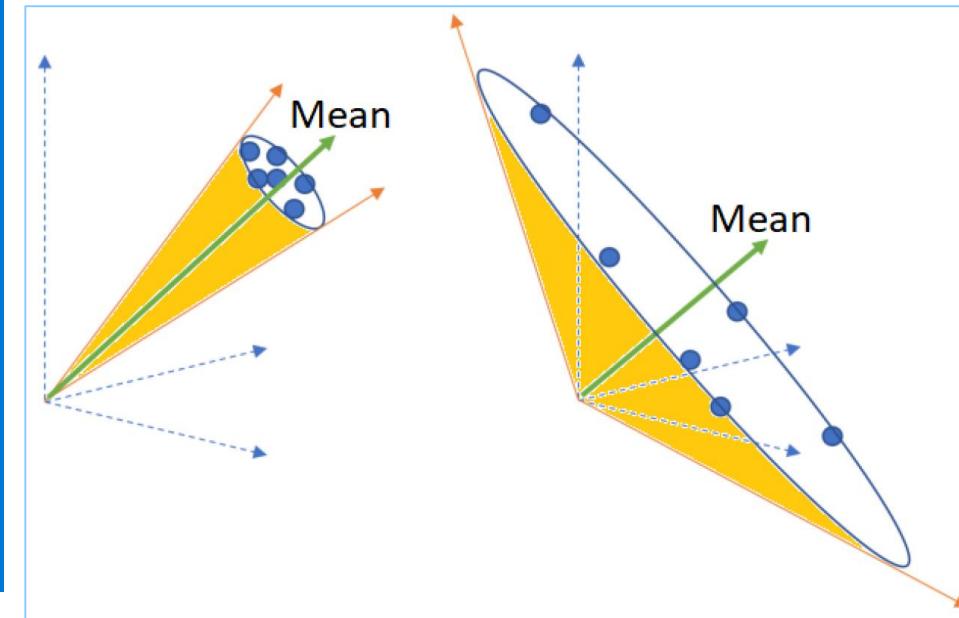
An Automatic Dialogue Evaluation Model (ADEM) [Lowe et al. 2017]

- Given the context and the reference
- Evaluate whether the response correlates well with human generated responses



Dialogue Evaluation

- Given the context and the reference
- Evaluate whether the response correlates well with human generated responses
- The proposed method uses multiplicative updates which leads to very tightly clustered vectors.



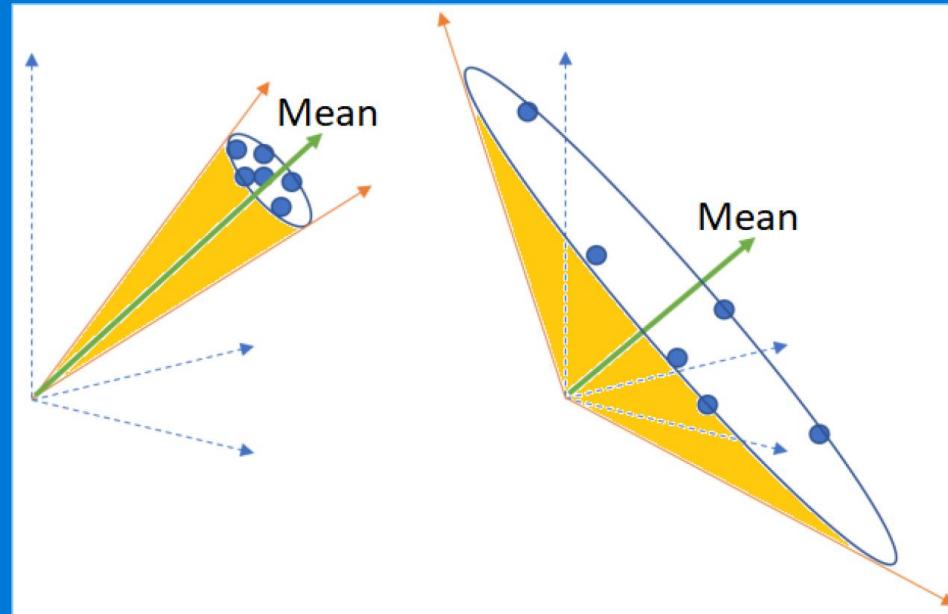
$$\begin{aligned} score(c, r, \hat{r}) &= (c^T M \hat{r} + r^T N \hat{r} - \alpha) / \beta \\ &= a^T \hat{r} - b \end{aligned}$$

$$\begin{array}{ccc} M & \rightarrow & 0 \\ \hat{r}_1 & \rightarrow & \hat{r}_2 \end{array}$$

$$\begin{aligned} s_n &= a^T \hat{r}_n - b \\ &= a^T \left(\hat{r}_1 + \Delta \frac{\hat{r}_2 - \hat{r}_1}{M} \right) - b \\ &= \left(s_1 + \frac{\Delta}{M} (s_2 - s_1) \right) - b \end{aligned}$$

Dialogue Evaluation

- Given the context and the reference
- Evaluate whether the response correlates well with human generated responses
- The proposed method uses multiplicative updates which leads to very tightly clustered vectors.



Context	Response	ADEM	human
<first_speaker> Hi, how are you doing? <second_speaker> not too bad, getting over a cold	colds are no good , hope it clears away soon. <first_speaker> sure, you can get an autographed one right here: <URL>	2.9 3.9	5 1
	<first_speaker> mate, get the app, you can watch while running,	3.79	2.5

Context	Response	ADEM	human
<first_speaker> Awesome ! Did you graduate high school this year ? <second_speaker> Nope 2 more years !	<first_speaker> Wow ! Going to be a junior . You youngster !	2.4 3.6	5 1
	<first_speaker> that is what sundays are for !	3.56	1
	<first_speaker> better safe than sorry !		

Microsoft AI

For every person and every organization



Dialogue



Speech

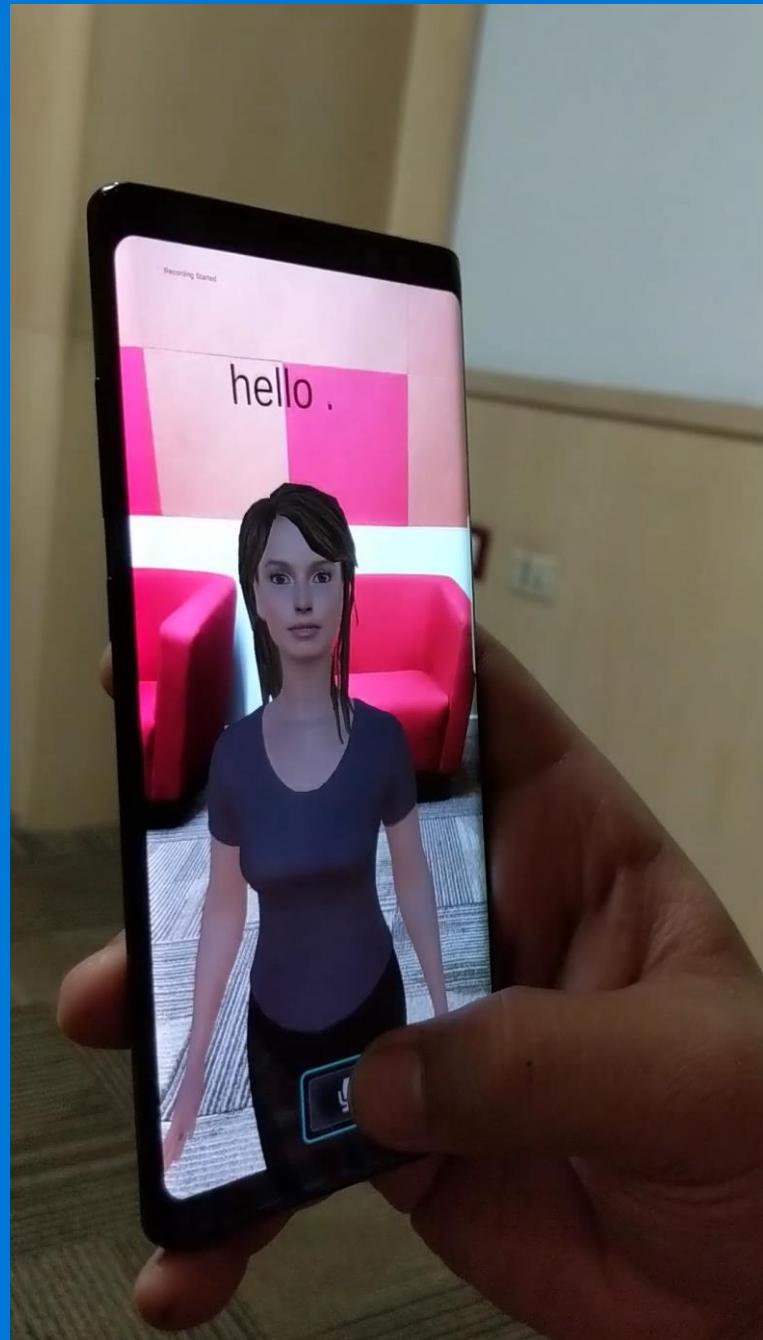


Vision

Holo Companion



ChitChat Companion



Interspeech 2018



HoloCompanion: An MR Friend for EveryOne

Mallikarjuna Rao Bellamkonda¹, Annam Naresh¹, Rushabh Gandhi², Mithun DasGupta¹,

¹Microsoft, Hyderabad, India

²BITS, Hyderabad, India

{annaresh, mallib, migupta}@microsoft.com, gandhirushabh2311@gmail.com

Abstract

Chat bots are becoming ubiquitous in our day to day life. The advent of the summer of AI has brought us all in close contact with intelligent agents such as Cortana, Siri and Alexa. We envisage a world, where these bots have their physical existence within the realm of Mixed Reality (MR). We present the first 3D chit-chat bot called the HoloCompanion. This bot has a personality, can chat with anyone and about any topic and has articulated lip, eye and head movements.

Index Terms: ASR, Chit-chat bot, TTS, human-computer interaction, lipsync.

1. Introduction

Chat bots are highly prevalent means of interaction with devices these days. Voice has been claimed to be the keyboard of the future, and as such these bots have an increasing role to play in our lives. Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home are all examples of chat interfaces which listen to humans and try to infer the action intent expressed by the humans. There has been larger quanta of development in transactional bots, which try to fulfill a specific intent such as ticket booking or hotel reservations. Chit-chat or non-transactional conversation on the other hand is more subjective and hence more difficult to comprehend owing to the open-endedness of such conversations.

There has been some interesting development in the space of chit-chat bots and Microsoft has led this front across different geographies. Xiaoice¹ in China, Xo² in USA, Rinna³ in Japan and Ruhu⁴ in India are existing chit-chat agents across different countries.

2. Chit-chat within Microsoft HoloLens

2. Gesture: Users can use simple gestures to select, drag, drop and interact with the holograms.
3. Voice: Use voice commands to navigate, select, open, command, and control the applications.

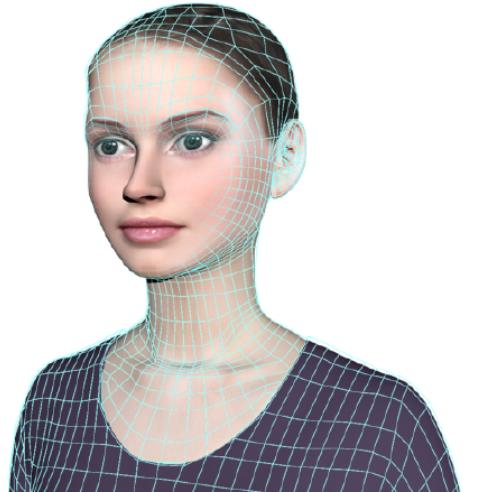


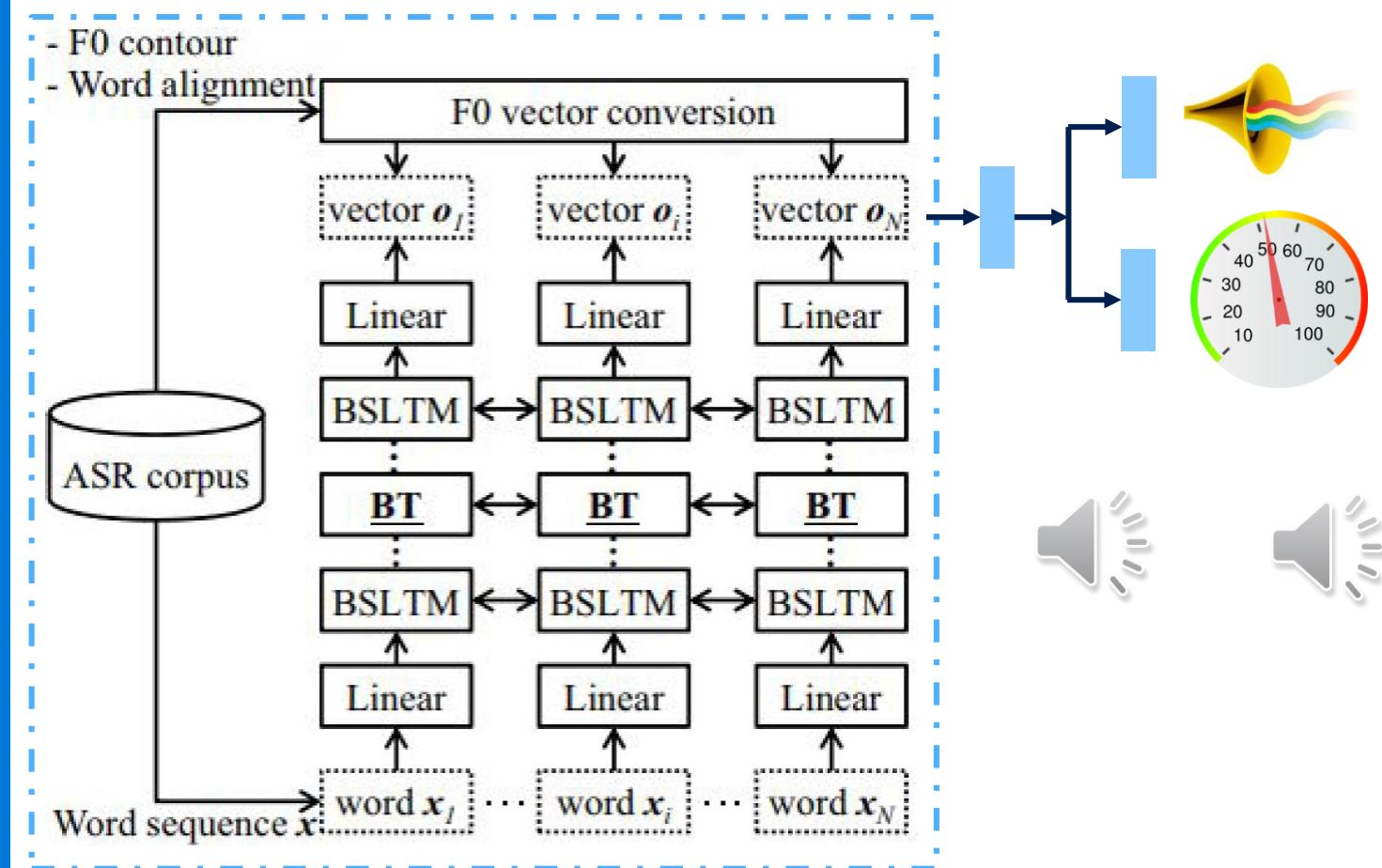
Figure 1: Avatar model used in HoloCompanion.

Expressive Speech

- Researchers in speech technology are actively working to raise the bar for the quality of interactions between machines and humans, with the goal that some day it will be indistinguishable from human-to-human interaction
- Prosody generally describes three suprasegmental aspects of speech and refers to aspects like rhythm, melody and stress. These features are
 - quantity (duration),
 - stress (intensity) and
 - intonation (pitch).

Prosody Parameter Estimation

- We augment the model proposed by [Ijima et al. Interspeech 2017]



Microsoft AI

For every person and every organization



Dialogue

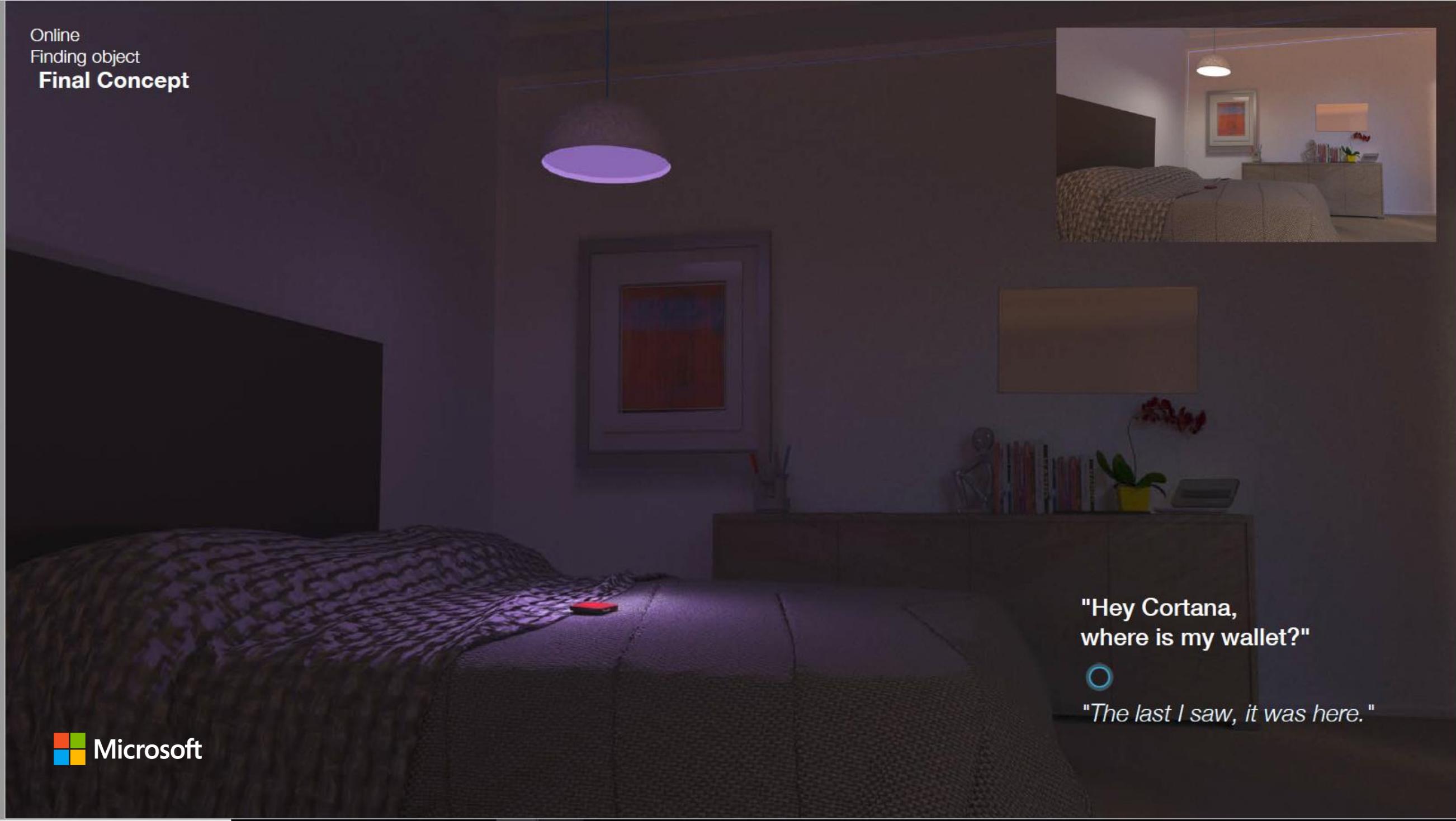


Speech



Vision

Online
Finding object
Final Concept



"Hey Cortana,
where is my wallet?"

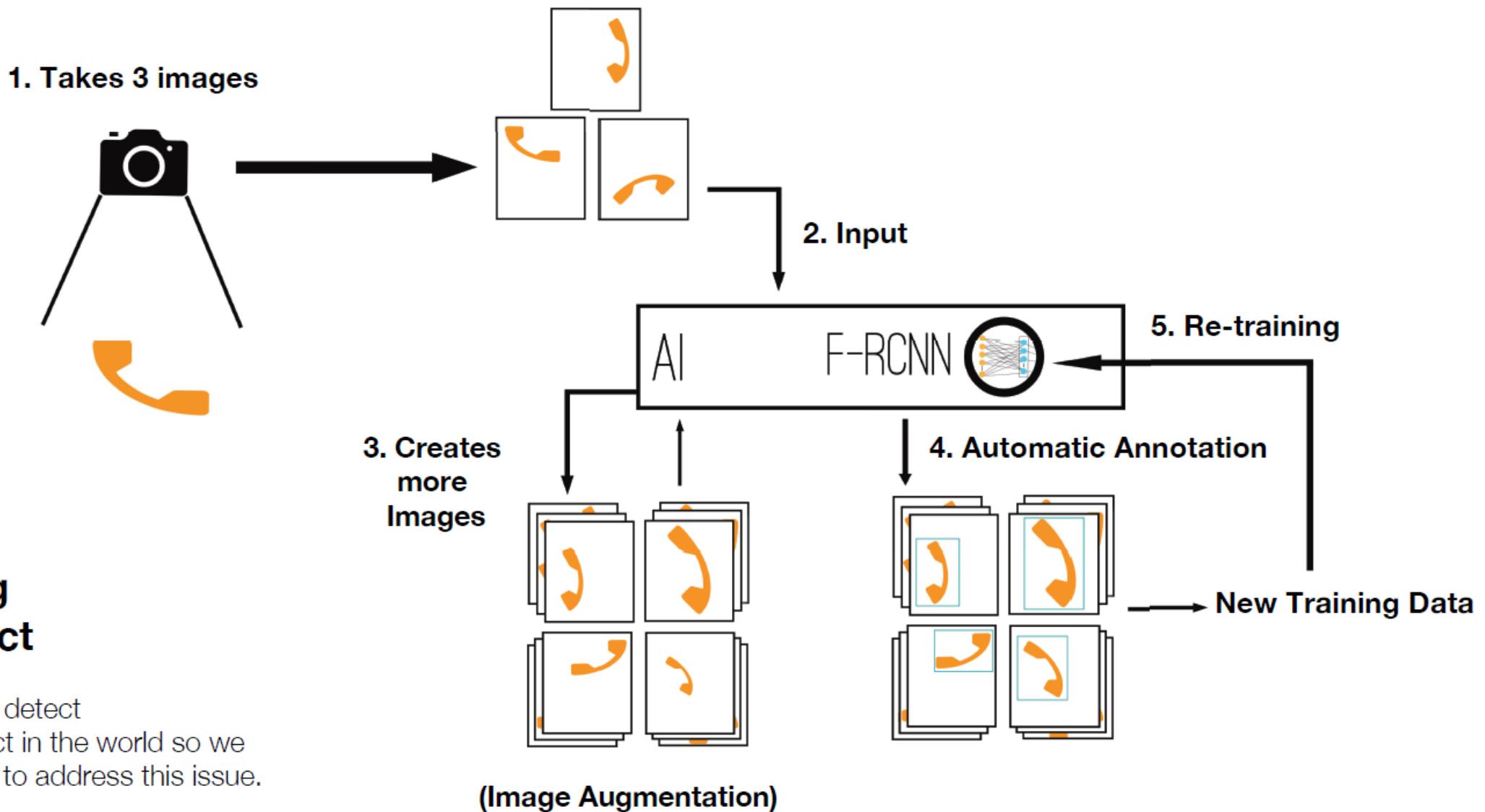


"The last I saw, it was here."

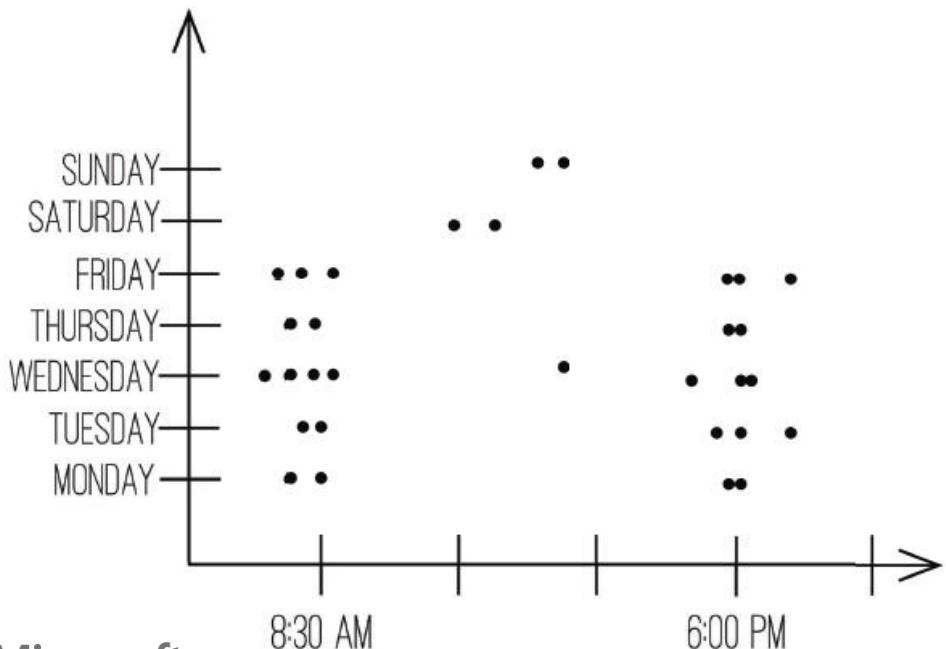
Final Concept

Registering a new object

Today, AI cannot detect every single object in the world so we created a feature to address this issue.

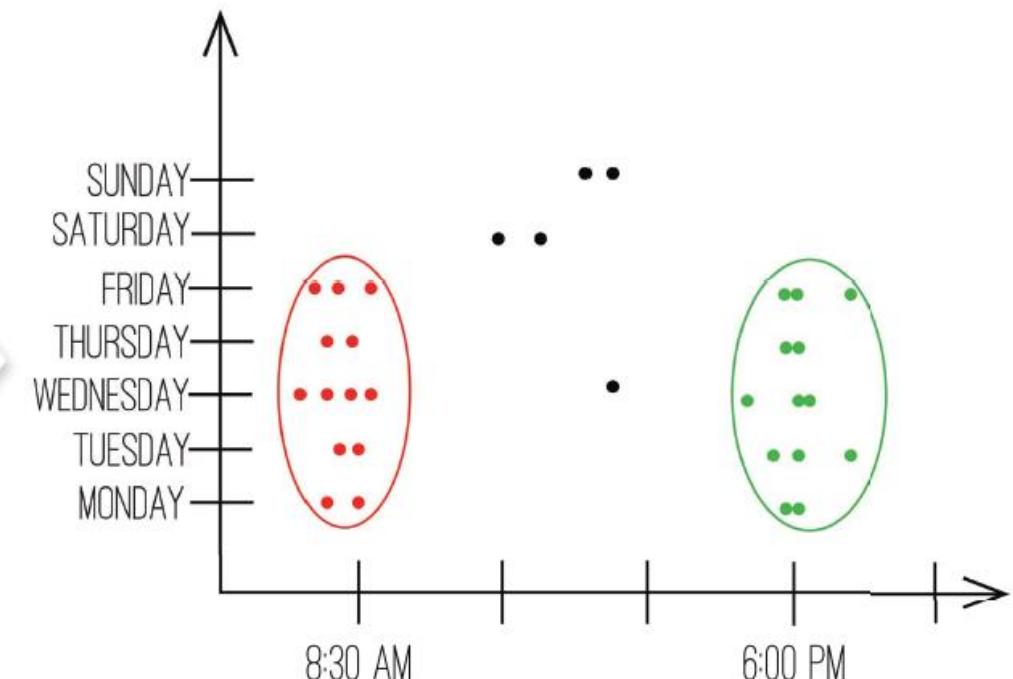


User's behavior patterns

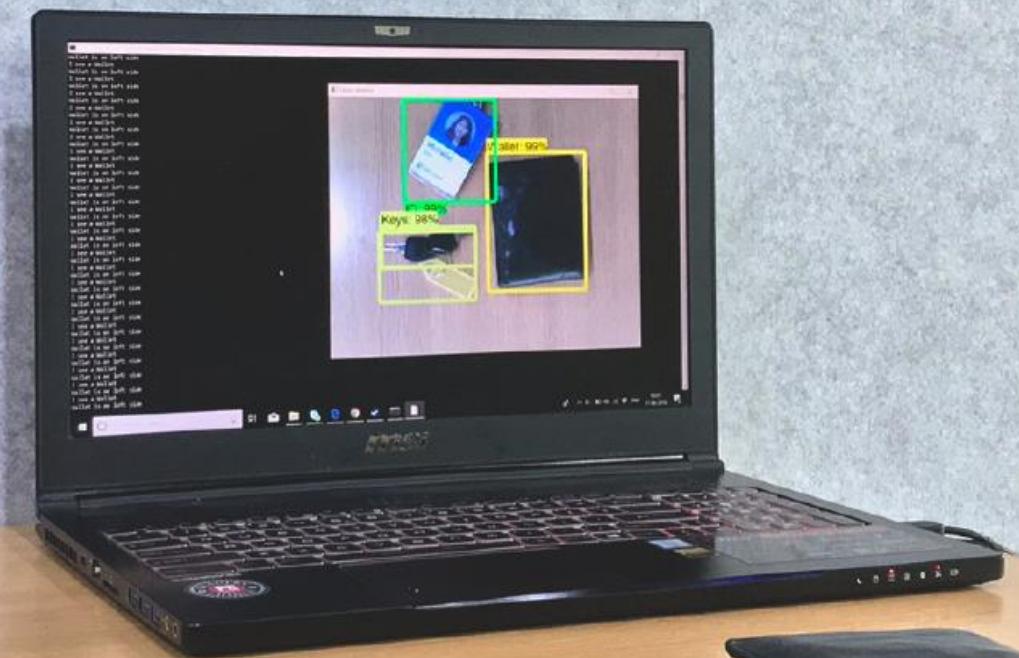


Unsupervised
Learning

Unsupervised Learning: Clustering Algorithms-
DBSCAN, K-Means, etc. Also Frequency
counting with Poisson Distribution, Gaussian, etc.

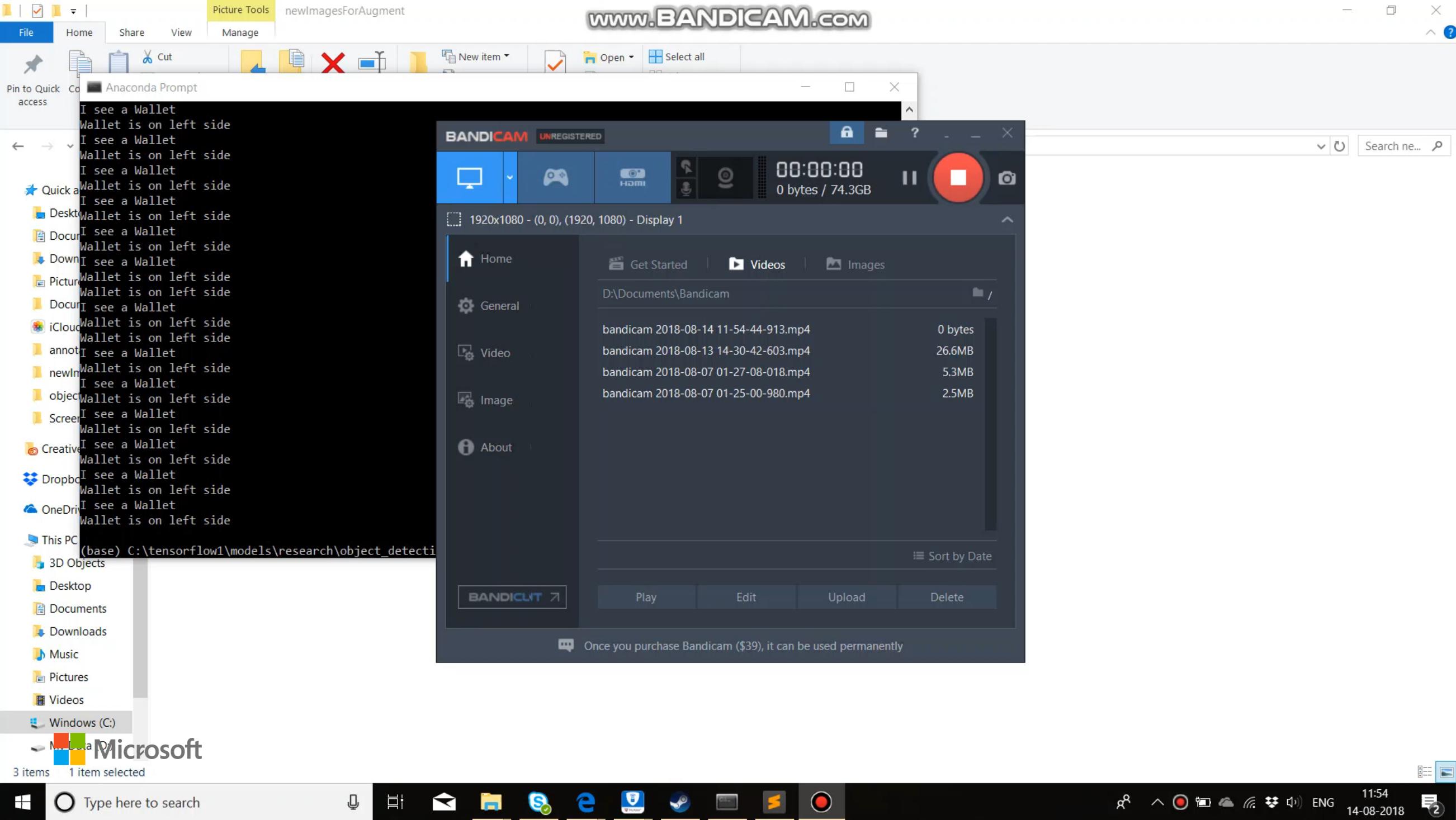


Final Concept



POC





Thanks



Email: migupta@microsoft.com