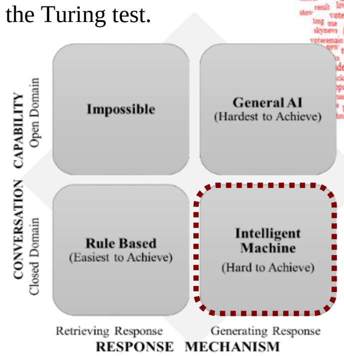
Literature Review

2015 Google

An end-to-end open domain dialogue system using sequence-to-sequence model. Dataset: 62 million sentences for training purpose and 26 million sentences for validation purpose. Predicted the next sentence on the basis of previous sentence/sentences but failed to pass the Turing test.

McGill University · Université de Montréal **2016**

Designed a dialog system by using 1 million multi-turn Ubuntu Dialogue Corpus to train the neural language model - to generate the next best possible response in a conversation. Reports that LSTM performs better than RNN

Intent: *#Leave*

facebook

ROBOSTOP Facebook shuts off AI experiment after two robots begin speaking in their OWN language only they can understand

Sample Tweets
#bbcbreakfast #IndependenceDay #VoteLeave why have we not seen the migrants attack british drivers in france when it is on american tv?
£10billion per year to support a dysfunctional beaurauracy! #VoteLeave #LeaveEUOfficial #IvoteLeave #LeaveEU
£17bn, the true cost of immigration to the UK every year @XXX #VoteLeave #Brexit #LeaveEU
@David_Cameron if we leave we'll still be part of a vibrant, connected world, but free to shape our own destiny #voteleave #TakeBackControl

2017

Microsoft **2017**

Designed a fully data driven knowledge grounded neural conversation model to produce a more content-rich response without slot filling. Considered 23 million tweet conversations and 1.1 million tips from foursquare to train their model. Simple entity matching approach, for open domain systems, failed to match the expectations.

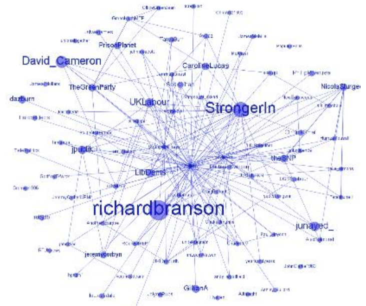
NEW YORK POST **Microsoft's chatbots keep turning racist**

THE VERGE Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

In the **Tay AI Project:** Microsoft used unfiltered tweet data for the training

Research Context

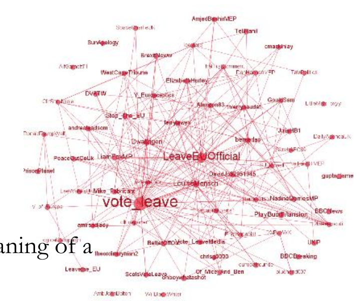


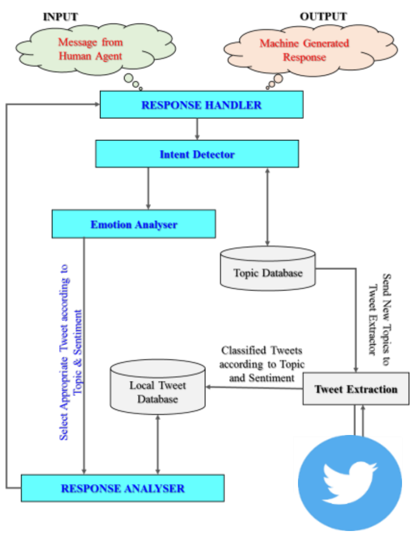
Intent: *#Remain*

Research Problem

Human: What is the topic of your poster?
Machine: Closed domain Chatterbot in a political context
Human: Why Political?
Machine: It is challenging because understand the political leaning of a human agent is crucial which is not case for IT helpdesk
Human: So, what's new?
Machine: Even if a politically opinionated human agent chats in a biased manner, then our Chatterbot agent considers the political leaning or intent of the human agent to avoid the conversation breakdown



Chatterbot Architecture



Intent Classifier in Political Contexts

"iPhone6 is expensive but nice" and *"iPhone6 is nice but expensive"* would look similar from a bag-of-words perspective. However, they bear opposite polarity because "the former is positive as the user seems to be willing to make the effort to buy the product despite its high price, the latter is negative as the user complains about the price of iPhone6 although he/she likes it". The situation is similar in political contexts. For instance,
Human: *The argument of VoteLeave lobby sounds more logical than VoteRemain.*
Human: *The argument of VoteReamin lobby sounds more logical than VoteLeave.*
Human: *I have to think twice before joining #VoteRemain campaign.*
Machine: *What is the Intent?* ☹
So, we are employing emotional intelligence to mimic human-like behavior on Chatterbot platform!

Nigel Farage has done more to destroy the #VoteLeave campaign than anyone could hope for

David Cameron has misled people about immigration. The truth is, we can only control it if we #VoteLeave

Sample Tweets
#Brexit fallacy: UK great enough to stand alone in world but not big enough to influence EU #VoteRemain : UK big influence in EU #StrongerIn
I'm a European, born in Durham in the British Isles. I believe we are #StrongerIn and better off #INtogether. Please #VoteRemain in #EURef!
#VoteRemain for an equal britain. for a strong economy. for a diverse and interesting place to live. #LabourInForBritain
#iVoted #VoteRemain To protect the NHS. To protect jobs. To protect people #MoreInCommon
"A country isolated from its neighbours is one that struggles to progress." #StrongerIN #GreenerIN #EUref

Proposed Model

If the human agents asks a question $\mathbf{Q}_i$ then our neural model will tag this question $\mathbf{Q}_i$ with an intent $\mathbf{I}_i$. Intent $\mathbf{I}_i$ is associated with a set of appropriate response $\{ \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n \}$. So, in response to $\mathbf{Q}_i$ will select an appropriate response from the above set. The efficiency of our model depends on proper tagging of $\mathbf{Q}_i$ with $\mathbf{I}_i$. If our response handler fails to identify the intent then 'it creates its own meta-dialogue with the user' to identify the intent. For instance,
Human: *What do you think about Brexit outcome?*
Machine: *It is difficult to predict the outcome. Whom you are supporting?*
Human: *UK will not support David Cameron.*
Machine: *So, you support VoteLeave.*
Human: *Yes!*

Why Richard Branson is a prominent node?
"As an investor, an entrepreneur, a father & grandfather, I urge the UK to #VoteRemain" was retweeted many times by his followers

Why our Chatterbot is not biased?

Human: *Why we will vote for united EU?*
Machine: *#VoteRemain today and secure a stronger future for our country. We're stronger, safer and better off in Europe*
Here the machine generated response is biased. So, our proposed model attempts to maintain its neutrality by associating a response $\mathbf{r}_i$ with an user $\mathbf{u}_i$ as follows:
Human: *Why we will vote for united EU?*
Machine: @David_Cameron says "*#VoteRemain today and secure a stronger future for our country. We're stronger, safer and better off in Europe*"
In brief, a question $\boldsymbol{Q}_i$ related to intent $\boldsymbol{I}_i$ generates a response $(\boldsymbol{u}_i + \boldsymbol{r}_i)$. Associating the response $\boldsymbol{r}_i$ (which is a tweet $\boldsymbol{t}_i$ from our corpus) with user $\boldsymbol{u}_i$ allows our Chatterbot to maintain its neutrality. However, this approach might not be appealing to all human agents. Thus, we propose **text summarisation** of all tweets by David Cameron (who supported the VoteRemain campaign) or summarization of **a few appropriate tweets** with intent – remain as follows:
Human: *Why we will vote for united EU?*
Machine: *VoteRemain supporters are saying that united EU will secure a stronger future for our country.*

DS3 DATA SCIENCE SUMMER SCHOOL

Let's Chat about Brexit! A Context Sensitive Dialog System using Twitter data (DS3-506)

Aparup Khatua, Temasek Laboratories @ NTU, Singapore & Dr. Erik Cambria, School of Computer Science and Engineering, NTU, Singapore.
E: aparupkhatua@gmail.com & cambria@ntu.edu.sg