Building a Question Answering Artificial Intelligence Chatbot using Open-source Technologies

Charith Gunasekara, PhD. P.Eng

Defence Scientist

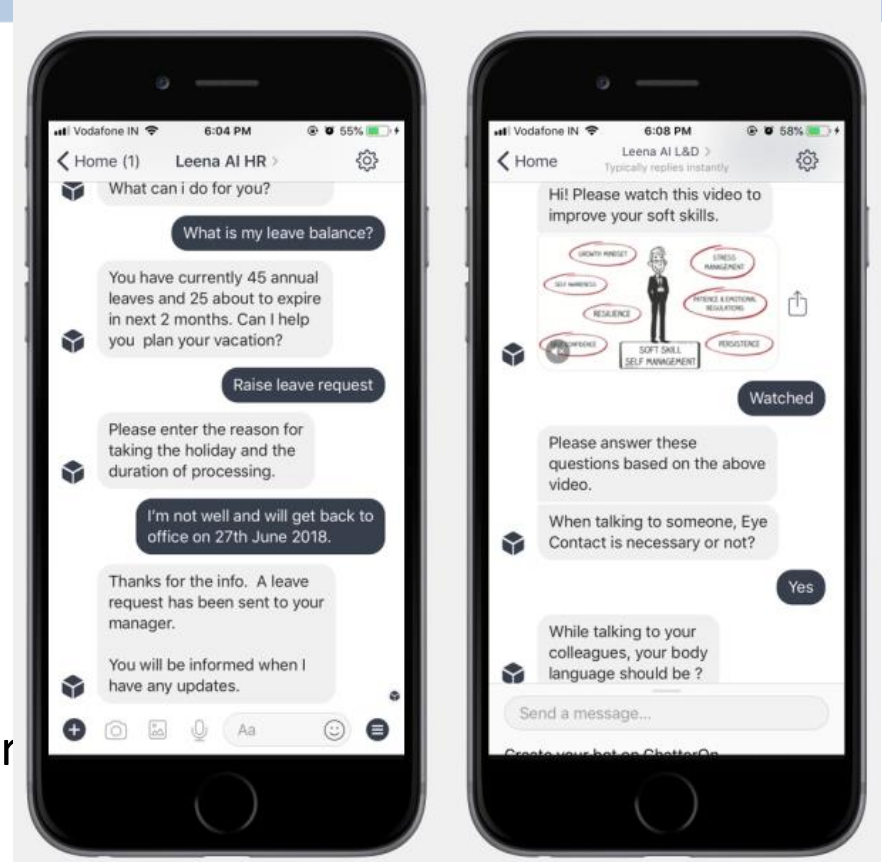
[1]ADM(Data, Innovation, Analytics)

[2]DRDC Centre for Operational Research and Analysis

National Defence

Project Overview

- Gov policy documents can be extremely long
- Need an easier way to get answers for users' questions
- Provide a natural language interface to complex policy documents
- Explore feasibility of using NLP to automatically analyze policies and answer questions without pre-written scripts

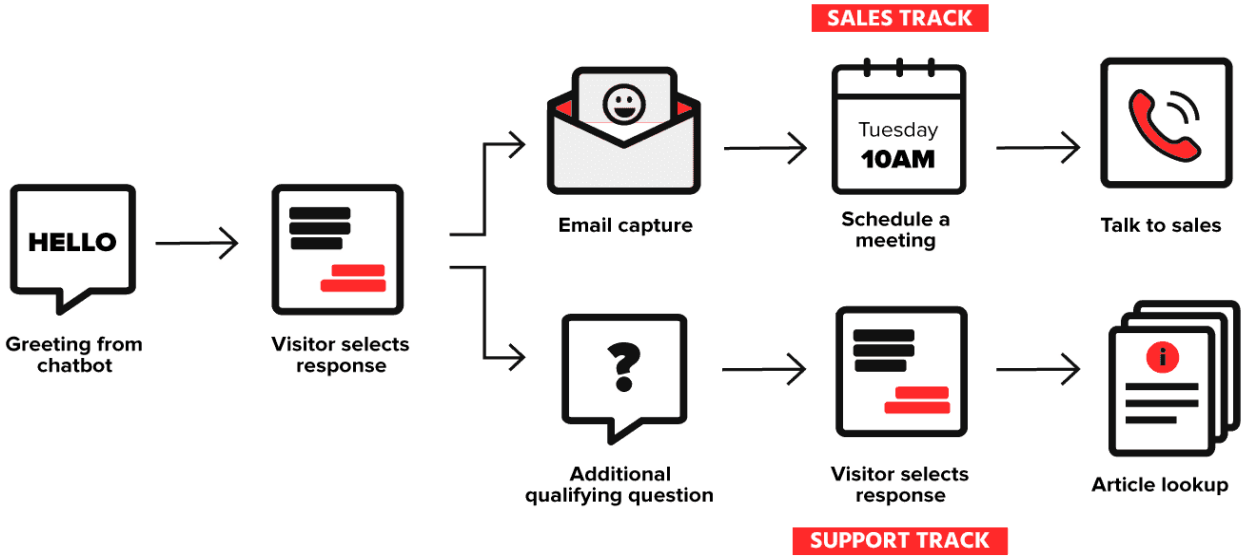


https://techcrunch.com/wp-content/uploads/2018/06/leena.png?w=680

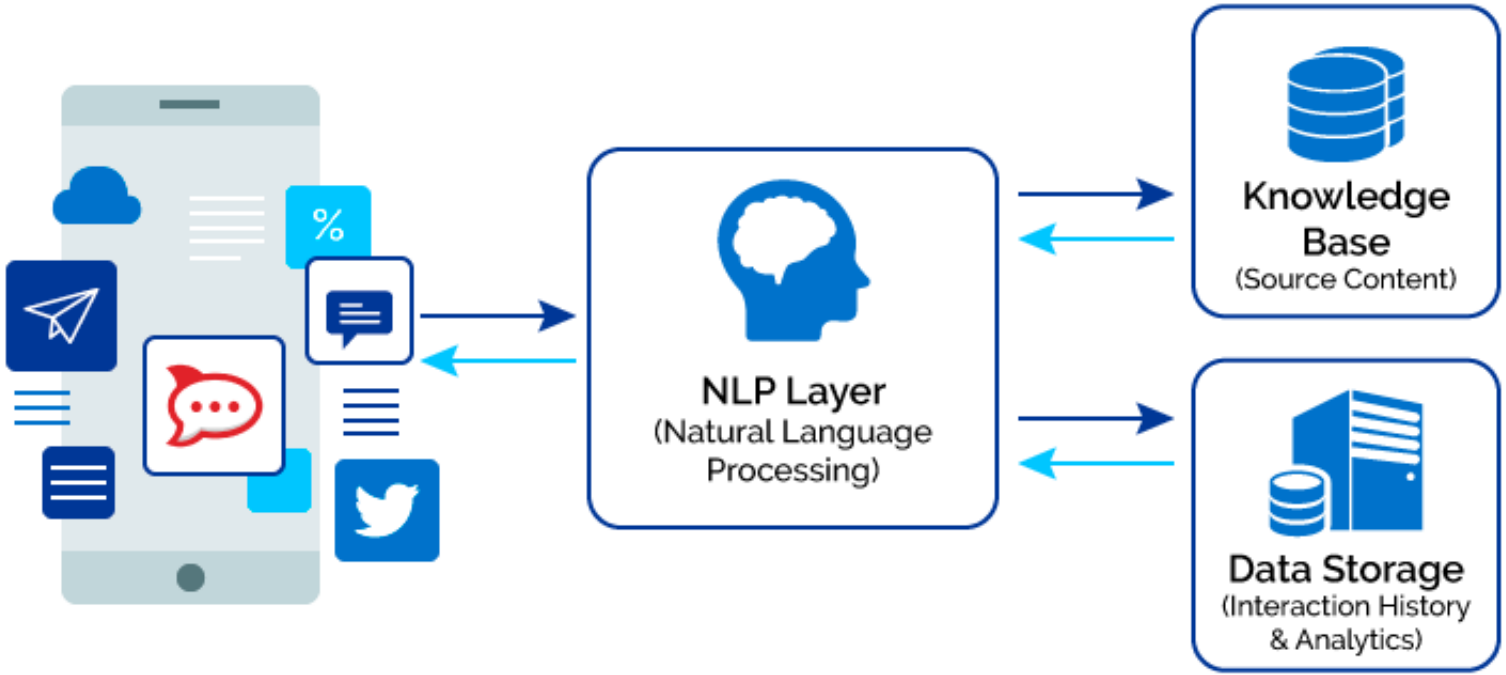
Two Types of Chatbots

- Rule Based
  - ✓Keyword search
- AI Powered
  - ✓Semantic search

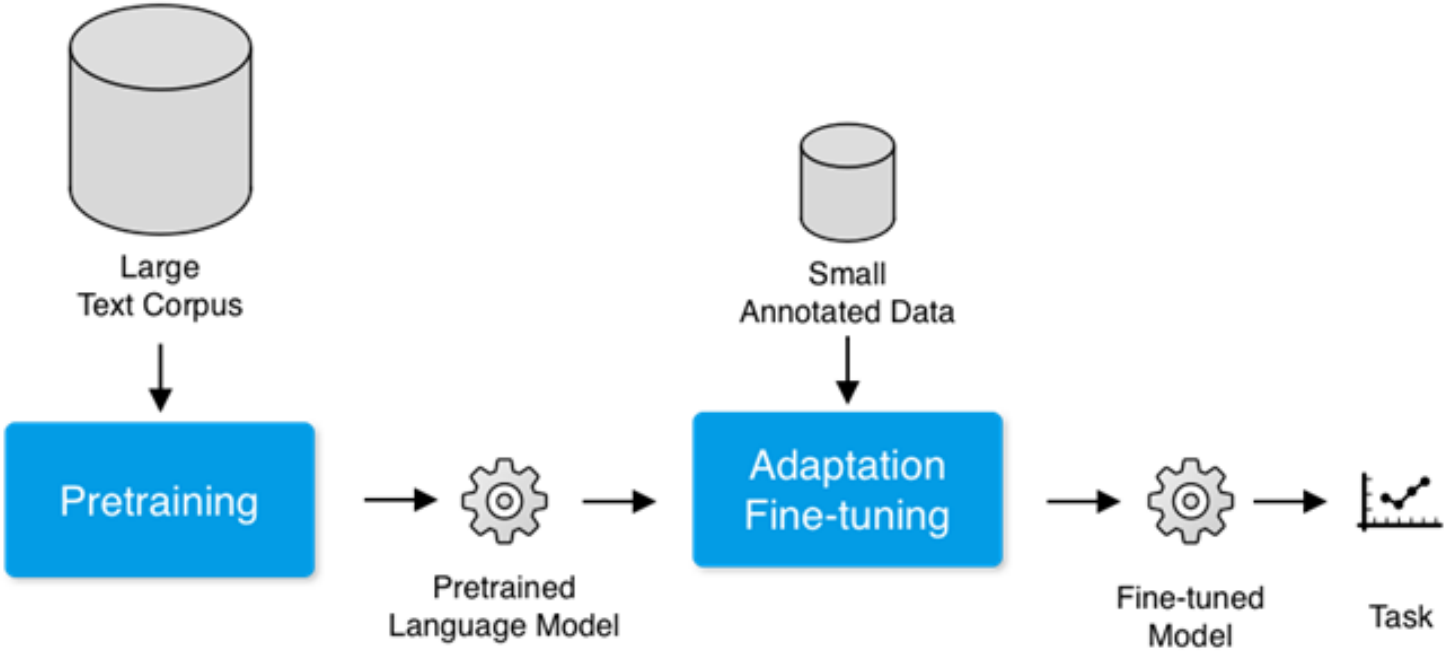
Rule Based Chatbot



AI Based Chatbot



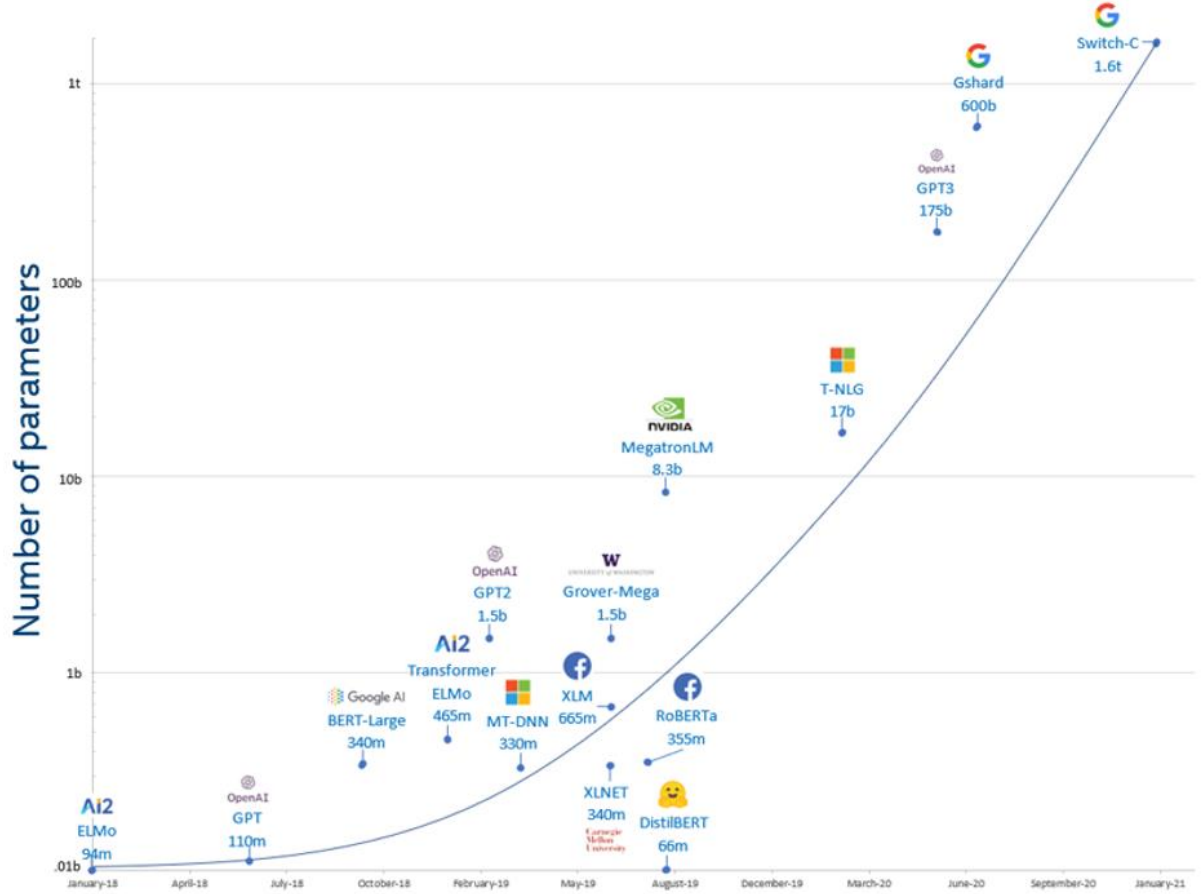
NLP Model


Large
Text Corpus
Pretraining
Pretrained
Language Model
Small
Annotated Data
Adaptation
Fine-tuning
Fine-tuned
Model
Task

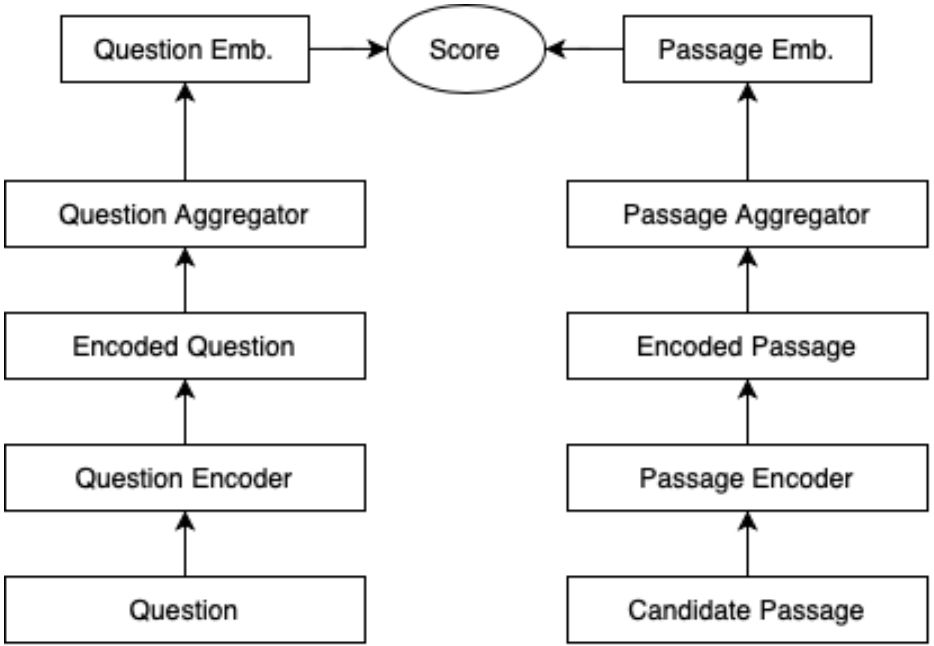
State-of-art Conversational AI Technologies

HUGGING FACE

FACEBOOK AI


Number of parameters
1t
100b
10b
1b
.01b
January-18
April-18
July-18
October-18
February-19
May-19
August-19
December-19
March-20
June-20
September-20
January-21
Ai2
ELMo
94m
OpenAI
GPT
110m
Google AI
BERT-Large
340m
Ai2
Transformer
ELMo
465m
OpenAI
GPT2
1.5b
MT-DNN
330m
XLM
665m
W
Grover-Mega
1.5b
XLNET
340m
NVIDIA
MegatronLM
8.3b
RoBERTa
355m
DistilBERT
66m
T-NLG
17b
OpenAI
GPT3
175b
Gshard
600b
Switch-C
1.6t

Semantic Search



Current State of AI Chatbot

- Trained on the Military Dress Manual
- Currently being expanded on other policies and standards
- A user asks a question and the system retrieves the top passage that it believes contains the answer
- Does pretty well at answering questions that have enough detail
- 85% accuracy before fine tuning.
- Reproducible technical details published in a peer-reviewed scientific report.

Demo

Questions?

Charith.Gunasekara@forces.gc.ca