MissionX

AI-ENABLED CHATBOT FOR ANSWERING COMPLEX TECHNICAL QUESTIONS

Presented by:
Chetna Khanna
Raghuram Anur Srinivas

Guided by:
Parag Kulkarni

In association with: CITAP and SEWA International

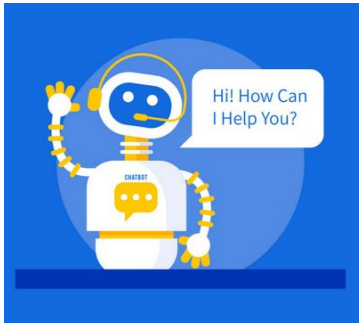
Date: September 3, 2020

CONTENTS

- Problem
- Knowledge Base
- Architecture
- Approach
- Training Data
- Modelling
- Results
- Conversations
- Conclusion
- Future Work

PROBLEM

- To develop a Question Answering chatbot (using Artificial intelligence) that can answer technical questions
- Provide assistance for Nutanix’s Licensing application
- Reside in a Slack channel of the company



slack

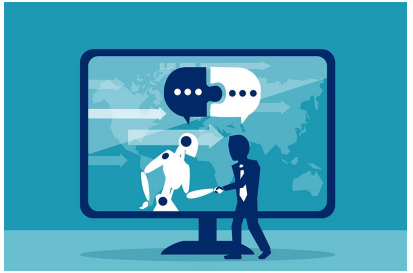
WHAT IS AN AI-CHATBOT?

Chatbot

- A **chatbot** is a software application used to conduct an on-line chat conversation via text or text-to-speech, in lieu of providing direct contact with a live human agent

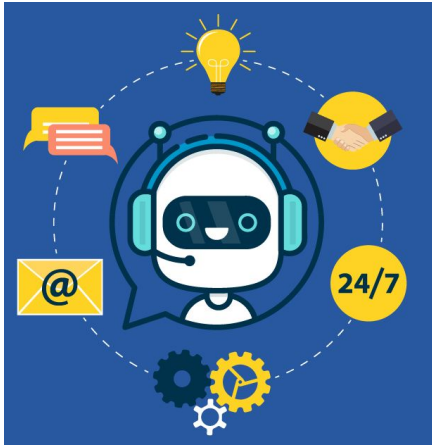
AI Chatbot

- Mimic human conversation
- Identify underlying intent behind the query a user is asking
- Provide intelligent human-like responses
- Constantly learns and improves over time



BENEFITS OF AN AI-CHATBOT

- Human and AI can work together to deliver the best customer experience
- Use of AI can make conversations engaging and lively
- Provide 24/7 uninterrupted human-like conversational service to customers
- Reduce burden on customer support executives



THE CHATBOT MARKET

- The Chatbot market is estimated to be worth around 2.6 Billion USD in 2019 and is expected to grow to 9.4 Billion USD by 2024
- eCommerce Chatbot Statistics indicate that by employing Chatbots, companies can reduce the operational costs by 30 percent
- The customer service cost reduction across the retail, banking, and healthcare sectors is estimated to amount to $11 billion annually by 2023
- According to Gartner, 25% of tech interactions of customer service operations will use virtual assistants by 2020

WHY THE PROBLEM INTERESTS US?

- One of the most researched problem
- Lot of advancements in the field of Natural Language Processing
- Challenging task to train a machine learning model to understand the language and context
- Opportunity to understand latest state-of-the-art like LSTM, GRU and BERT
- Opportunity to develop an end-to-end product

KNOWLEDGE BASE

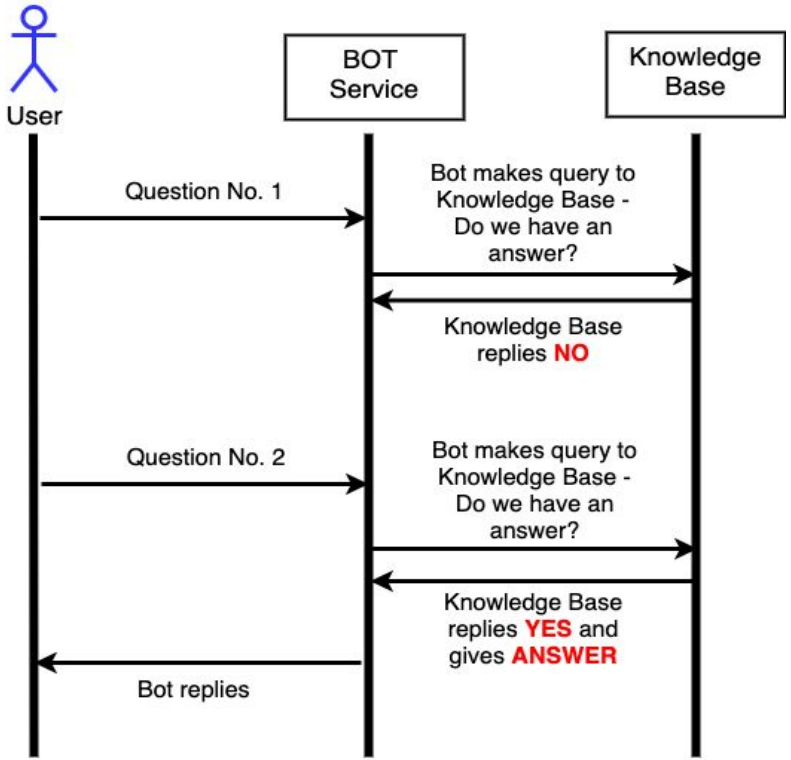
Dataset

- Data for the project is obtained from the public documentation of Nutanix for Licensing application
- Documentation has articles on various areas of Licensing

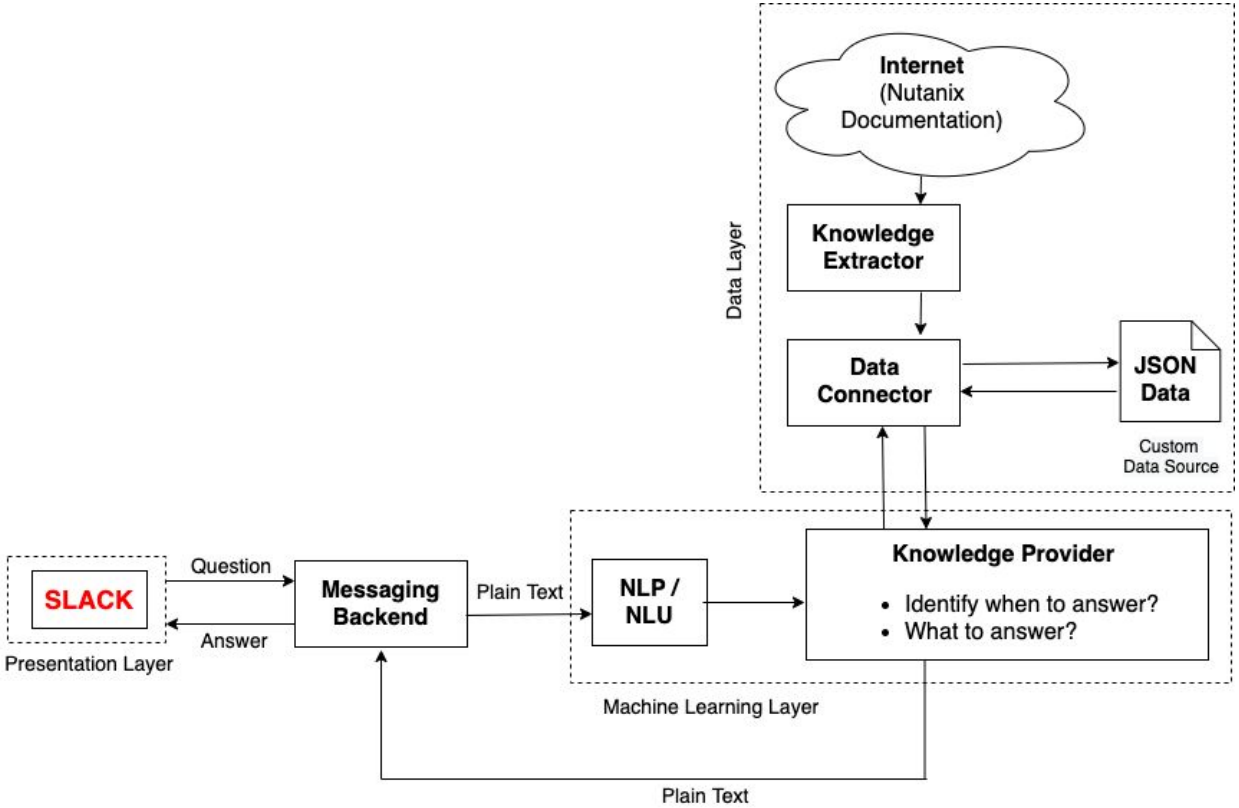
Challenges and Approach

- Small dataset
- Not in FAQ format
- HTML scrapped to create JSON using the articles and its attributes
- Use of JSON in Deep Learning models challenging

INITIAL IDEA



FINAL ARCHITECTURE



APPROACH

- Document scraping
- Dataset creation
- Pre-processing
- Creation of training data
- Modelling
- Response prediction
- Integration with Slack
- Bot as a service running on cloud

TRAINING DATA

Challenge

- Modelling text is messy and difficult and algorithms prefer well-defined inputs

Approach

- Stop words removal, stemming and tokenization
- Feature extraction done using bag-of-words model
- Texts are converted into vectors of numbers
- Pre-processing applied to article's headings
- Training data created which measure the presence of vocabulary words in the text

MODELLING

Artificial Neural Network (ANN)

- Three layered artificial neural network (256, 512 and 50 neurons)
- Dropout technique to avoid overfitting
- Hidden layer activation function - ReLU (not all neurons are activated at the same time)
- Output layer activation function - Softmax (to give responses with probability)
- Optimizer - Stochastic gradient descent with momentum
- Loss function - Categorical cross entropy
- Performance metric - Accuracy
- Epochs - 100, Batch size - 16

MODELLING

Long Short Term Memory (LSTM)

- Three layered recurrent neural network - Embedding, LSTM and Dense
- Embedding layer with a maximum vocabulary dimension of 50,000
- Dropout technique to avoid overfitting
- Output layer activation function - Softmax (to give responses with probability)
- Optimizer - Adam
- Loss function - Categorical cross entropy
- Performance metric - Accuracy
- Epochs - 100, Batch size - 16

RESULTS

Artificial Neural Network Model

- Trainable parameters - 175,666
- Accuracy - 92%

Long Short Term Memory Model

- Trainable parameters - 2,878,242
- Accuracy - 80%

PREDICTIONS

Approach

- Query asked by the user is pre-processed
- Predictions along with their probability obtained from model
- Predictions sorted to receive rank by relevance to the query
- Predicted probability matched against a threshold
- First 100 words from relevant article along with article link returned by bot
- Else directed to customer support team

SLACK INTEGRATION

- **SLACK**
  - workplace chat app
  - well defined and user-friendly
  - lets user extend, expand and automate workspaces

- **FLASK**
  - micro web framework used as Python app server
  - provides tools, libraries and mechanics with little dependencies

- **INTEGRATION**
  - to help Slack bot respond using predictions made by the deep learning models via Flask application

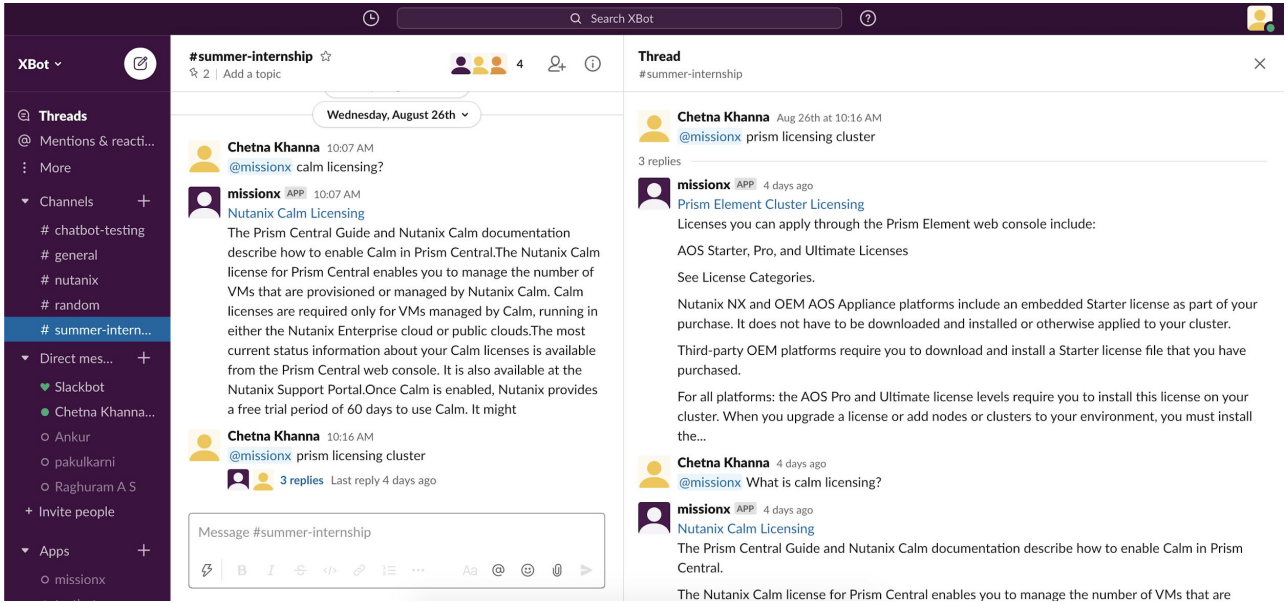
CLOUD SERVICE

- **Cloud computing**
  - computing based on internet
  - provide flexibility to grow with the demand
  - greater security than the local machine
  - disaster recovery
  - reach users worldwide without disruption

- **Google Cloud Platform (GCP)**
  - time-tested service that Google relies on
  - friendlier hand-holding for beginners
  - provide security with by default encryption
  - low cost

CHATBOT CONVERSATIONS

- Use of thread to organize discussions
- Formatted text to enhance readability



https://app.slack.com/client/T016RH3RRQX

CONCLUSIONS

After a lot of learning and working on this project, the chatbot:

- resides in a slack channel of the workspace
- wakes up with the word “@missionx”
- answers user’s queries for Nutanix’s licensing application
- responds to the query if the prediction probability surpasses a threshold value
- directs to the customer support team if threshold condition is not met
- provides link of the relevant article along with some portion of article content
- provides link of related articles too
- uses thread to organize discussions

FUTURE WORK

- Humanoid experience - with no wake word
- Better predictions using language modelling BERT
- Use of approach different from bag-of-words to help model learn the context by not discarding the word order
- Learn from past conversations and incorporate feedback
- Use articles together with the headings to create vocabulary
- Expand dataset with documentations of multiple domains
- Personalized response
- Scalable service architecture
- Extendable to be used for other applications

REFERENCES

1. **Images:**
   - https://slack.com/
   - https://snapengage.com/info-capture-chatbot-available-now/
   - https://hrdailyadvisor.blr.com/2019/05/30/conversational-ai-vs-recruiter-chatbots-whats-the-difference/
   - https://haptik.ai/blog/using-conversational-ai-customer-support/
   - https://www.newbreedmarketing.com/blog/how-to-build-a-knowledge-base-for-self-service-support
   - https://www.negotiations.com/articles/negotiation-approach/
   - https://www.shiksha.com/boards/articles/board-results-blogId-39627
   - https://www.thebluediamondgallery.com/handwriting/p/predict.html
   - http://dhhs.michigan.gov/course212/sco_1301/page15.htm
2. **Slack documentation:** https://api.slack.com/
3. **DZone:** https://dzone.com/articles/python-chatbot-project-build-your-first-python-pro
4. **Research papers:** https://arxiv.org/pdf/1809.08267.pdf , https://arxiv.org/pdf/2004.13637.pdf , https://arxiv.org/pdf/2001.09977.pdf

THANK YOU!