

# Next generation Chatbot using AI

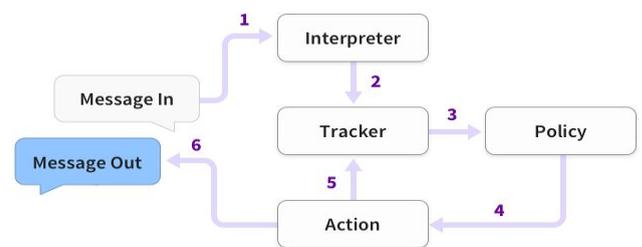
**Ajinkya Pathak**  
Software Developer, Infogen-Labs

**Kirti Bihade**  
AI Studio Partner(DL/ML), Infogen-Labs

Chatbots are software programs that perform conversation with an user in natural language, understand the user's intent and send responses based on the organization's business rules and data. Our chatbots use AI to process language, enabling them to understand human language. The goal is to provide people, an easier way to have their questions answered quickly. It will reduce the overhead of client support team as user can get answers of their queries. Chatbot can handle simple to medium level query from user which otherwise would have to be answered by human.

Our AI team at Infogen Labs have build a bot that could enhance the working of support team. We developed a chatbot that will handle FAQs. We are taking this data from our support team for training our model. We are using supervised learning approach. We used Rasa core and Rasa NLU which are open source python libraries for creating conversational software that helps in make machine-learning based dialogue management and natural language understanding systems. Main component of the model is a recurrent neural network (an LSTM i.e Long Short Term Memory), which maps from raw dialog history directly to a distribution over system actions. Therefore, it can be optimized using supervised learning (SL), where a domain expert provides example dialogs which the LSTM should imitate.

Chatbot predicts which action to take from a predefined list of actions. An action can be a simple utterance or an call of API or trigger into database, i.e. sending a message to the user, or it can be an any function to execute. After execution of an action, it is passed to a tracker instance, and so it can make use of any relevant information collected over the history of the dialogue and the results of previous actions. Figure(1) is showing high level architecture of rasa core.

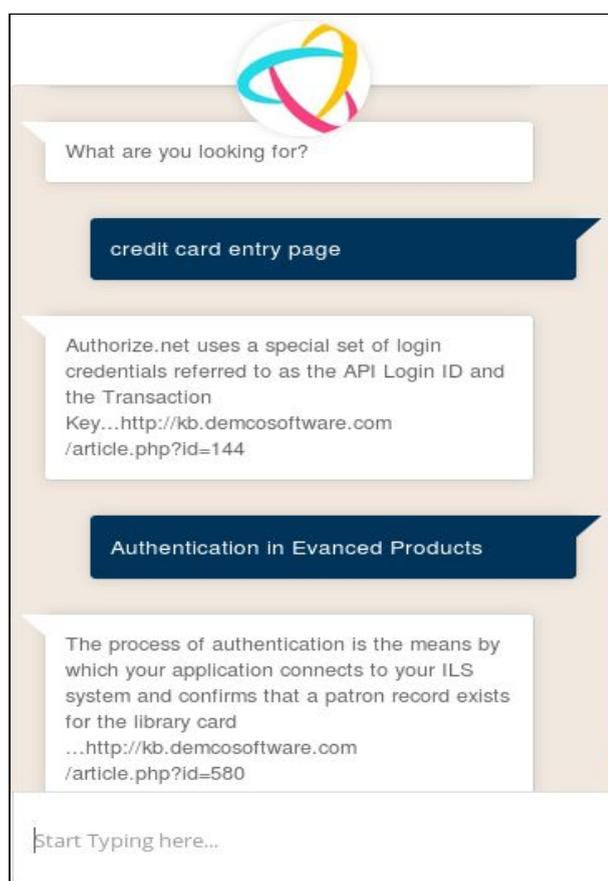


**Fig.(1)High level architecture[1]**

The text classification is loosely based on the fastText approach. Sentences are represented by pooling word vectors for each constituent token. Using pre-trained word embedding such as GloVe, the trained intent classifiers are robust to variations in phrasing when trained with just a few examples for each intent.

NLU comprises loosely coupled modules combining machine learning libraries in a consistent API, some predefined pipelines like Spacy\_sklearn, tensorflow\_embedding, MITIE\_sklearn with sensible defaults which work well for most use cases. The recommended pipeline, spacy\_sklearn (or tensorflow) processes text. Initially text is tokenized and parts of speech (POS) annotated using the spaCy NLP library.

Then the spaCy featuriser looks up a GloVe vector for each token and pools these to create a representation of the whole sentence. Then the scikit-learn classifier trains an estimator for the dataset, by default a multiclass support vector classifier trained with five-fold cross-validation. The ner\_crf component trains a conditional random field to recognize the entities in the training data, using the tokens, POS tags as base features. Since each component implements the same API and hence, it is easy to swap the GloVe vectors for custom, domain-specific word embedding, or to use a different machine learning library to train the classifier. Figure(2) displays the conversation from chatbot.



**Fig.(2) Snippet of chatbot conversation**

Natural Language Generation (NLG) is a subdivision of Artificial Intelligence (AI) which aims to reduce communicative gaps between machines and humans. The technology accepts input in non-linguistic format and turn it into human understandable formats like reports, documents, text messages, etc. A processing pipeline is very important component and a basic building block of any Rasa NLU model. The pipeline defines how user inputs should be parsed, tokenized and how the features has to be extracted. The components of the pipeline are important because they have a direct impact on how the NLU model works and performs well over other models.

True potential of the chatbot shines when you give it more and more data to train. Moreover, even after deployment of the chatbot and it will learn from user interactions. This will make it capable of handling multiple cases which are not yet fed to the training data and eventually improve the model.

#### **References:**

- [1] [Ilya Sutskever, Oriol Vinyals, Quoc V. Le @ Google. Sequence to sequence learning with neural networks. Proceedings of the 27th International Conference on Neural Information Processing Systems-Vol:2](#)
- [2] [Conversational-ai-chatbot-using-rasa](#)
- [3] <https://rasa.com/docs/>
- [4] <http://blog.rasa.com/>