AI Chatbots – Reality V/S Hype

AI chatbots give a perception of being intelligent, but intelligence is a long way away, says Navveen Balani.

Uncover some of the real facts on chatbots and limitations associated with current AI chatbot platform and frameworks. The intent of the article is to help readers take informed decisions on how to design AI chatbots and workarounds with the existing chatbot implementation.

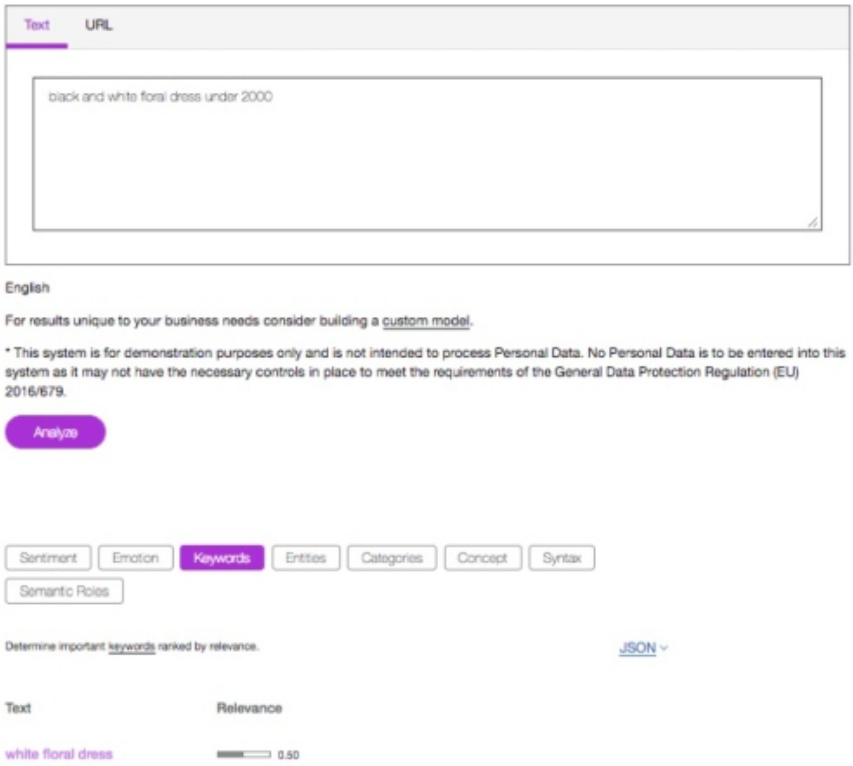
What are chatbots?

A chatbot is a software program which carries out a conversation with a human. The conversation can be through textual methods, voice or even through recognising human expressions. Chatbot interactions can range from simple questions being answered like – 'what is the outside temperature', to sophisticated use cases which requires a series of dialogue to arrive at an outcome – like a chatbot for booking holiday trips or providing financial advice.

What should I keep in mind for developing an AI chatbot?

Chatbots work well when domain is well understood by the AI system. As the AI chatbot relies on NLP to understand the semantics of the input message, unless the NLP parser is trained on the domain, the accuracy of recognising the intent and topics of interest would be very low or not as per acceptable criteria.

Take an example of a shopping chatbot which advises user what to buy based on the latest fashion trends.



Screenshot 1 – Keywords from Watson NLP.

Consider 3 queries below from a user:

Query 1 – Show me medium size trending black and white dresses for Christmas party.

Query 2 – Show me white colour, 3 inches platform heels.

Query 3 – Find AND black and white floral dress under 2000.

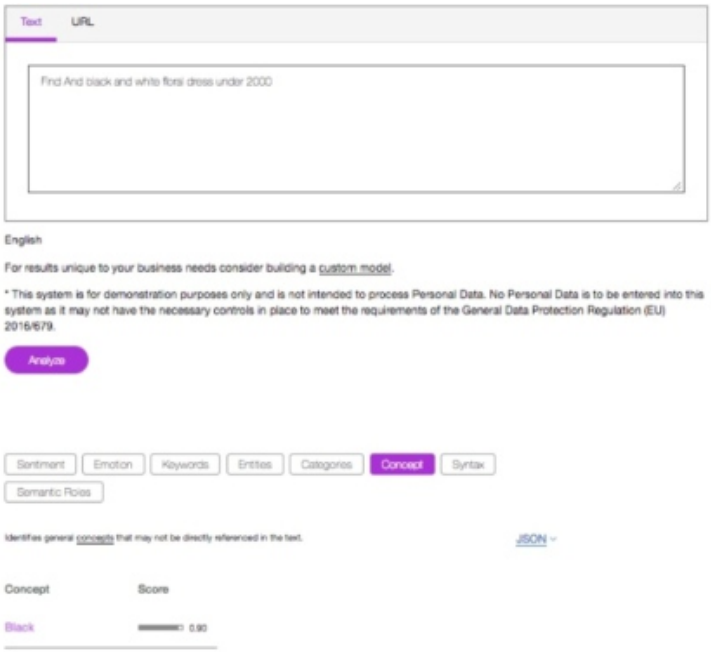
Here the chatbot needs to understand the following

1. Understand the shopping language.
2. Understand the intent – It's a shopping query.
3. Understand the domain – Its shopping query for apparel and shoes (i.e., there can be multiple domains – grocery, electronics, books, etc.)
4. Understand clothing shopping category and terminology, like:-

Category – dresses, sandals etc.-
Variants – sizes (medium/large etc.), colour (various colours and combinations like black and white), heel size (3 inches, etc.)
- Prices and ranges – 2000, etc.
- Brands like – AND, Nike, etc.

Out of the box, any chatbot implementation wouldn't understand the domain. You need to train the chatbot on the custom domain to recognise the context and the language. For instance, out of the box NLP parsers would not recognise 'AND' as a brand. Let's inspect how well some of the leading Cloud AI NLP services recognise the sentence – 'Find AND black and white floral dress under 2000'

Here is a snapshot from Watson NLP (out of the box) implementation (Screenshots 1, 2 and 3).



Screenshot 2 – Concepts from Watson NLP.

As you see, the Watson NLP recognises 'white floral dress' as keywords and 'black' as concept. Ideally it should have recognised 'black and white' as concept as we are looking for a combination of these colours. The dress could also be a concept, as it's quite generic. The floral can be a keyword, which has a dependency on dress. Identifying all the facts in the right way is important, as based on the facts you would convert this to a search query to get the required details from the data store (or from respective search indexes).

> **"A chatbot is a software program which carries out a conversation with a human.**

For instance, the above should result as:
Colour = 'black and white'
Category = 'Dress'
Gender = 'Female'
Price < 2000
Pattern = 'floral' or Keyword within category = 'floral' (where colour, category, gender, price, pattern are all the columns or indexes you are searching against.)

The Watson NLP parser doesn't recognise 'AND' as brand and recognises 'And' as a conjunction ('CCONJ') in part of speech, which is expected as it's not trained on it.

The above is true for any of the available NLP implementation (that is available today from various cloud vendors or open source), where it fails to understand all the correct context of the sentence. The use case was pretty simple. Even if we train the NLP implementation on these examples, it would fall short as you need to plugin specific NLP rules for such conditions to get the desired results. As the complexity and context that needs to be inferred increases, training would also not help as you can never come up with a generalised model for such conditions. That is the single most limitation if we only rely on today's generation of NLP implementation.

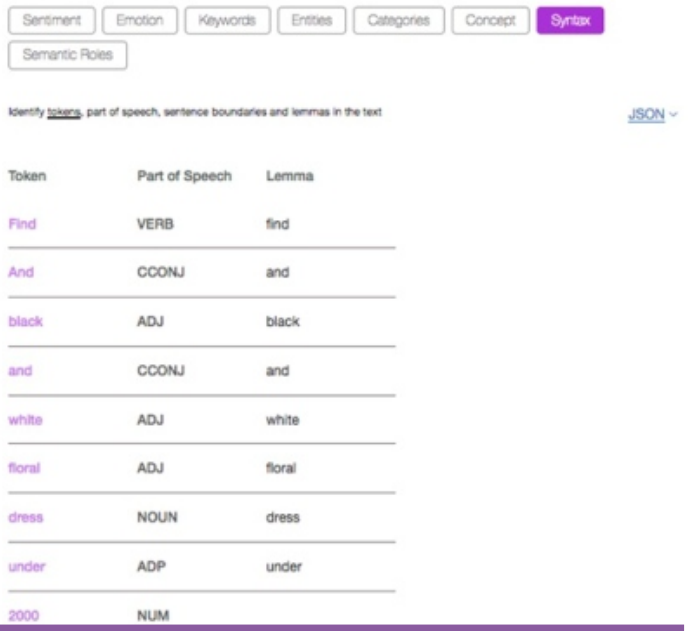
Based on my experience on building a sophisticated shopping personalised advisor, none of the out of the box AI NLP implementation fitted the requirements. A simple scenario is these 3 sets of sentences – 'black and white dress', 'and black dress' and 'blue jeans and white shirt'. In all these 3 examples, the use of word 'and' means different meaning. In the first case, its represents a combined colour 'black and white', in second instance 'and' represents a brand and in third instance two queries joined by a conjunction (i.e., and). Even with required training, a generalising model was not possible with any of the available solutions. These are just one of the many examples I am highlighting. Imagine the complexity when dealing with medical literature. In our case, we ended up building our domain specific NLP implementation which worked for all such scenarios.

In general, while designing chatbot solutions, start with a closed domain and what kind of questions the chatbot needs to answer. Don't start building a general purpose chatbot from start, as it would be difficult to get the required accuracy. Secondly, if you are using any cloud vendor or third party implementation, ensure your use cases can be simply solved by the default implementation or you need to build components to work around it.

What is not real about chatbots?

Chatbot are examples of Weak AI. Current generation of chatbot can be thought of smart dialog systems driven through techniques like NLP and fixed conversation flows. Out of the box, a chatbot doesn't understand any domain. We need to train the chatbot to understand the domain. Also, based on the complexity of the domain, you would incrementally train and add subdomains. For instance, a chatbot helping you book a cab is an example of fixed domain, while a chatbot helping assisting doctors for cancer treatment would be trained on various types of cancer incrementally. As mentioned in the shopping advisor example, understanding the meaning of the same word in different context is difficult for the current generation of NLP implementation to resolve and you need to rely on custom techniques to handle such conditions.

> **"Current generation of chatbot can be thought of smart dialog systems driven through techniques like NLP and fixed conversation flows.**



Screenshot 3 – Part of Speech from Watson NLP.

Now, let's look at some marketing gimmicks around AI chatbots:

Ingest And Know It All chatbots – These are chatbots being marketed where you can ingest millions of documents, like medical literature and can ask questions, which can provide expert assistance like diagnosis of diseases. Such kinds of systems unless trained appropriately will never provide desired accuracy.

By appropriately, I mean it can take years to train these systems. The fundamental problem with these systems is that, they still don't understand the complete language and complexity of the domain. You typically end up with custom domain adoptions and infinity language rules, which is definitely not smart enough to manage in longer run. The predictions of such systems are usually not accurate.

Self-learning chatbots – How often you have heard this terminology called self-learning chatbots. This again is a misconception, where chatbots learns on its own. You have to train chatbot on what you what the chatbot to learn. Usually you would capture the user behaviour details through their interaction with the chatbot application. This would include capturing user analytics information like capturing his likes or dislikes in some way, either through explicit or implicit means. Explicit information can be a user rating a product and implicit can be the time a user spent looking at a response.

Once you know the user well and have its data, it becomes a recommendation problem on what you want to recommend to the user. So, you end you building a recommendation algorithm to recommend something. For instance, for a fintech application, this would mean recommending similar stocks based on what stock he views regularly or his portfolio. Different domain and use cases, would need different recommendation algorithms and that needs to be developed as part of the chatbot. However, the learning is boxed, for instance, if you have a chatbot which assist you in booking restaurants, it can recommend you similar restaurants, but it can't recommend your places to stay, it only knows about your restaurants taste.

General purpose, generative chatbots – A chatbot which is capable of learning new concepts from scratch and provide responses like human. As it learns from open domain, the chatbots would start behaving similar to the famous Microsoft Tay chatbot (which was forced to shut down on its launch day), as it started learning unwanted details from tweets and started posting inflammatory and offensive tweets. This is a classic example of what I describe AI as – 'AI can learn, but can't think'. The generative chatbots are formulating the response based on probability of words and creating a grammatically correct sentence, without understanding the real meaning of it.

As I mentioned earlier, the first focus should be on getting domain specific chatbots right and with the current techniques we are far away from realising the vision.

> **"Different domain and use cases, would need different recommendation algorithms and that needs to be developed as part of the chatbot.**

Will chatbots make human agents obsolete?

To answer this question, let's understand what functionality chatbots currently provide. Current chatbot implementation do well for handling fixed set of dialogs with the user, repetitive tasks and certain initial aspects of customer service tasks. Wherever there is a fixed set of processes and flows to automate, chatbot can be used to provide 24x7 support for any queries.

If human expertise is used for answering basic set of questions where answers are readily available, it would be eventually be replaced.

> **"The chatbot would always act as assistance to an expert person to get some job done.**

But in real-life scenarios, most of the conversation usually doesn't follow a fixed flow paradigm. But if the conversation moves from basic questions to questions which need further analysis, or the topic of conversation gets changed, you need a sophisticated chatbot implementation to take care of various conversation flows, identify the context switch, identify intents which your chatbot may not be aware of and create queries to find that information from your knowledge source.

You are now moving from fixed set of flows to more dynamic flows which needs to be interpreted by your chatbot. Building such complex chatbot implementations requires sophisticated domain specific adoption using machine learning techniques and custom solutions.

Current out of the box chatbot services fall short of building such chatbot implementations.

And even if you have all the data in the world at your disposal, infinite processing and computation power, using the current generation technology and research, you can never build a system that can compete with an expert human in the field. Taking even a 5 year horizon from now, I don't think we can develop such a level of intelligent chatbots. For instance, can chatbots or an assistant, help doctor to recommend cancer treatments accurately and consistently. The answer is No.

The information provided from chatbot can aid doctors to take a clue from the answer provided, it may be right or wrong. You can never certify this. The chatbot would always act as assistance to an expert person to get some job done. Ultimately, these systems are throwing a bunch of answers based on some probabilities. The answers are limited to what you have fed into the system, you can't infer a new knowledge on the fly or can correlate information like a human expert to come to any conclusion.

While, there are research going on to use deep neural nets for conversation flows, we are still quite far away of building truly conversational interfaces which understands the nitty-gritty of language and domain. Also, the answers provided needed to be explainable and unless you have a way to backtrack on why a particular answer was provided, such deep neural systems can't be used for use cases which requires auditability and explainability.

In short, enjoy the smart chatbots that gives a perception on being intelligent, but intelligence is a long way away.

Navveen Balani leads and manages Technology at Bridgeweave. He has over 19 years of experience in building enterprise products using exponential technology, specialising in AI, Blockchain, IoT. Prior to his current role, he was the CTO and co-founder of a cognitive retail startup. He is a former IoT and Watson Lab Leader for IBM India Labs, where he was responsible for setting up the Watson Lab and worked with customers across the globe on evaluating and building cognitive and IoT solutions.