**Chatbot Project**

Report 2

Survey about cloud computing, architecture of machine learning and neural networks, chatbot

**Submitted by:**

1. Haidy Samir
2. Esraa Mohsen
3. Menna Hussien
4. Mona Mahmoud

**Contents**

[Haidy’s task 6](#_Toc18935713)

[What is AI? 6](#_Toc18935714)

[Type 1: 6](#_Toc18935715)

[Weak AI vs Strong AI: 6](#_Toc18935716)

[Type 2: 7](#_Toc18935717)

[WHAT IS ML? 11](#_Toc18935718)

[Supervised learning 12](#_Toc18935719)

[Unsupervised learning 12](#_Toc18935720)

[Reinforcement learning 12](#_Toc18935721)

[Linear regression 13](#_Toc18935722)

[13](#_Toc18935723)

[Logistic regression 13](#_Toc18935724)

[K Nearest neighbor 14](#_Toc18935725)

[Neural Network 14](#_Toc18935726)

[Naïve Bayes 14](#_Toc18935727)

[SVM 14](#_Toc18935728)

[Why RNN? 14](#_Toc18935729)

[What is RNN? 14](#_Toc18935730)

[Long short-term memory (LSTM) 15](#_Toc18935731)

[Gated Recurrent Neural Network 15](#_Toc18935732)

[convolutional LSTM 16](#_Toc18935733)

[According to Application point of view we can classify the RNN applications into many types which are: 16](#_Toc18935734)

[RNN applications example: 17](#_Toc18935735)

[NLP 17](#_Toc18935736)

[NLP and Text mining 17](#_Toc18935737)

[speech recognition 17](#_Toc18935738)

[Natural language understanding 17](#_Toc18935739)

[Natural language generation 18](#_Toc18935740)

[Modelling 18](#_Toc18935741)

[NLP Terminology: 18](#_Toc18935742)

[**Tokenization** 18](#_Toc18935743)

[**Normalization** 19](#_Toc18935744)

[**Lemmatization** 19](#_Toc18935745)

[**Parts of speech (POS)** 19](#_Toc18935746)

[Applications of NLP: 19](#_Toc18935747)

[Esraa’s task 20](#_Toc18935748)

[Problem faces the companies before Chatbots: 20](#_Toc18935749)

[Solution: 20](#_Toc18935750)

[What is the chatbot? 20](#_Toc18935751)

[An overview about the history of chatbot: 20](#_Toc18935752)

[The example of traditional & old Chatbots: 20](#_Toc18935753)

[The Drawbacks of this traditional and old chatbots: 22](#_Toc18935754)

[*As result* 22](#_Toc18935755)

[*There are two main tasks in deep learning (DL):* 22](#_Toc18935756)

[Challenges in developing a good model: 22](#_Toc18935757)

[Chatbot applications are classified into: 22](#_Toc18935758)

[*1.* *Service chatbots* 22](#_Toc18935759)

[*2.* *Commercial chatbots* 22](#_Toc18935760)

[*3.* *Entertainment chatbots:* 23](#_Toc18935761)

[*4.* *Advisory chatbots:* 23](#_Toc18935762)

[*1.* *Task-oriented chatbots:* 23](#_Toc18935763)

[*2.* *Non-task-oriented chatbots:* 23](#_Toc18935764)

[*1.* *Goal-based chatbots:* 23](#_Toc18935765)

[*2.* *Service-based chatbots:* 23](#_Toc18935766)

[*3.* *Knowledge-based chatbots:* 23](#_Toc18935767)

[*4.* *Response Generated-based chatbots:* 24](#_Toc18935768)

[NOW, Focus on Response Generated-Based Chatbot as it is related to our Chatbot project. 24](#_Toc18935769)

[a. Template- based Model: 24](#_Toc18935770)

[b. Generative-Model: 25](#_Toc18935771)

[c. retrieval-based Model: 25](#_Toc18935772)

[The way to find responses based on Un structured data: 26](#_Toc18935773)

[An overview of recent chatbots: 27](#_Toc18935774)

[Mona’s task: 28](#_Toc18935775)

[What is Neural Network? 28](#_Toc18935776)

[What is ANN? 28](#_Toc18935777)

[Different types of NN 29](#_Toc18935778)

[What is ALEXA? 30](#_Toc18935779)

[Menna’s task 32](#_Toc18935780)

[Cloud Computing 32](#_Toc18935781)

[Cloud computing Definition 32](#_Toc18935782)

[The cloud computing model 32](#_Toc18935783)

[Advantages of the cloud 33](#_Toc18935784)

[From Cloud Computing to Fog and Edge Computing 33](#_Toc18935785)

[FOG Computing 34](#_Toc18935786)

[Edge Computing 34](#_Toc18935787)

[Examples for free cloud platforms 35](#_Toc18935788)

[AZURE vs AWS 35](#_Toc18935789)

[Microsoft AZURE 36](#_Toc18935790)

[As per our project 36](#_Toc18935791)

[IBM Watson 37](#_Toc18935792)

[Data Mining 37](#_Toc18935793)

[The stages of development of searching 38](#_Toc18935794)

[References 39](#_Toc18935795)

# Haidy’s task

**AI components**

## What is AI?

AI is the simulation of human intelligence processes by machines especially computers. It enables us understand information hidden in large volumes of unstructured data. Machine learning can be used to process text or images in seconds and annotate them with positive, negative or neutral sentiment.

During searching about AI main structure, I found two types of classifications for AI systems which are:

## Type 1:

Weak AI vs Strong AI:

-**Weak AI** (also known as narrow AI) is an AI system that is designed and trained for a particular task .It is trained and developed to perform this task and cannot do unfamiliar tasks to it which has no previous training experience with it.

Here is an important formula established for weak AI system (The dominant AI system till now) which describe the main components of weak AI systems.

A screenshot of a cell phone

Description automatically generated

-**Strong AI** (also known as Artificial general intelligence) is AI system which has it ‘s own conscious it has the ability to find solution without human intervention when it is exposed to unfamiliar task. It can make a decision, self-awareness and learn from itself. (NOT PRACTICAL YET)

## Type 2:

It is classified based on functionalities into:

1. **1.Reactive machines:**

It doesn’t have past memory and cannot use past information for future actions

1. **2.Limited memory:**

Can use experience to inform future decisions

1. **3.Theory of mind:**

This type can understand people emotions, belief, thoughts expectations and able to interact socially “NOT PRACTICAL YET”

1. **Self-Awareness:**

It has it ‘s own conscious, super intelligent, self-awareness & sentient “NOT PRACTICAL YET”

These types have common components which may all exist in one system or some of them which are as follows:

A close up of text on a white background

Description automatically generated

Talking about each component briefly:

1. **Machine Learning (ML):**

(will be discussed later in details)

1. **Natural Language Processing (NLP):**

(will be discussed later in details)

1. **Vision:**

It can be said as the machine eyes, machine vision captures and analyses visual information using a camera, analog to digital conversion and digital signal processing

1. **Robotics:**

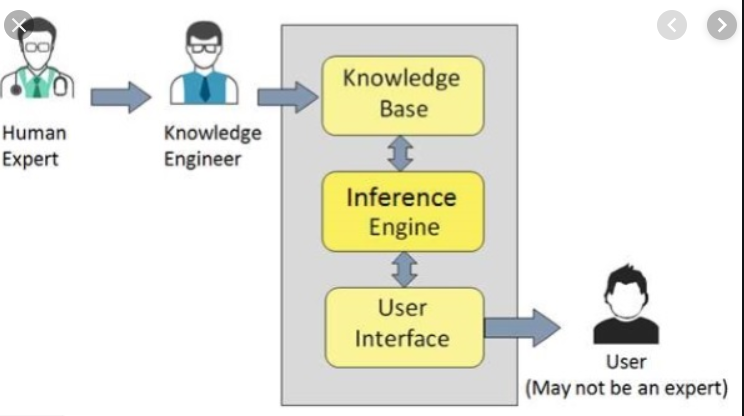
Field of engineering focused on the design and manufacturing of robots

1. **Expert Systems:**

An expert system is a software which use knowledge stored in a knowledge base to solve problems that usually require a human expert knowledge.

-The knowledge base contains files in a particular domain and rules to solve a problem procedure and all the intrinsic data relevant to that particular domain

-Inference Engine role is to inferences data from the knowledge base, rules and apply it to the problem or request of the user to return the required answer



1. **Speech:**

(will be discussed later in details)

-By Simple Architecture we can say that AI main components are:

A picture containing melon

Description automatically generated

*The previous explanation was close enough to describing the elements composing most of AI system around and it feeds up the concept of simulating human intelligence, but we have a different approach for describing our AI system which is illustrated in the fig. below*

A close up of a logo

Description automatically generated

-This was the general conclusion concerning what are the AI elements or in other words what is the involved technologies and methods in AI. Here is the answer *Machine learning is a branch of artificial intelligence where a class of data-driven algorithms enables software applications to become highly accurate in predicting outcomes without any need for explicit programming.* Data science, on the other hand, employs computer science disciplines like mathematics and statistics and incorporates techniques like data mining, cluster analysis, visualization. So, both of ML and DATA SCIENCE do all the required analysis and involve all the required functionalities required by the system. Important to note that Machine learning is a subset of artificial intelligence. While data science is an interdisciplinary field to extract knowledge or insights from data.

**Machine Learning Algorithms**

## WHAT IS ML?

**“A computer program is said to learn from experience E with respect to some task T and some performance measure P, if it ‘s performance on T as measured by P improves with experience E”**

**-TOM Michell**

Since 1950s, machine learning has revolutioned several fields in the last few decades. neural networks is a subset of ML and deep learning is a sub field of neural network. Deep learning has been showing outstanding success in almost most of the applications domain

Machine learning enables human to program computers so that the machines can recognize patterns or learn from what it is being input to it

Machine learning involves three major types of learning and each type has its methods of solving and used algorithms.

A close up of a map

Description automatically generated

## Supervised learning

SL decide the problem of classification, when finite groups of somewhat determined objects are distinguished in a potentially infinite aggregate of objects. As a rule, groups are formed by an expert. Furthermore, the expert can explain or not the reasons of the initial classification.

## Unsupervised learning

UL techniques solve the problem of clustering, when the range of initially undetermined objects is split into groups by means of automatic procedure based on the properties of these objects. Whereas the number of groups (clusters) may be given in advance or generated automatically.

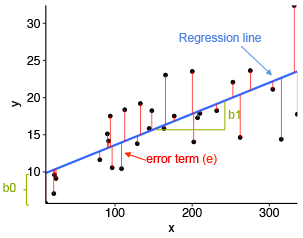
## Reinforcement learning

Reinforcement Learning is a learning technique for use in unknown environments. we do not have a straightforward loss function, thus making learning harder compared to traditional supervised approaches. (Simply we can say that our machine learns through a trial and error approach)

-To solve a problem using machine learning you have to go through a certain flow to achieve the output you want, first you have to analyze data and find patterns and then make predictions through equations and formulas and learn from feedback this is the cycle of machine learning process that you have to go through it.

-There are many algorithms used in machine learning as illustrated in the above fig the algorithms based on classification and regression is used in supervised learning while those based on clustering and association is used in Unsupervised learning.

- Here are some algorithms concepts:



## Linear regression

It is used for estimating relationships among variables. An independent variable and dependent variable where the dependent variable must be a continuous value .the relation between the independent and dependent variables is linear that is why it is called linear regression So we draw a regression line to give minimum error in predicting the values of these independent variables.

Linear regression formula is:

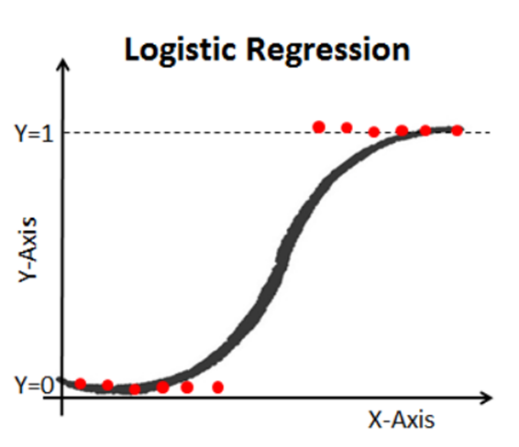
Y =bo+b1x+e

Where bo: is Y intercept,

b1: slope,

x: independent variable,

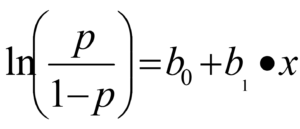
e: Error.



## Logistic regression

It is a method used to predict a dependent variable given a set of independent variables such that the dependent variable must be a categorical (ex: have a binary value 0 or 1, Yes or No etc..). We always use a sigmoid function to map the relation between dependent and independent variables.

Logistic regression formula:

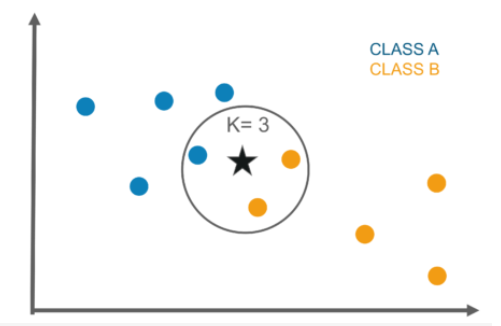
[](https://www.data-mania.com/blog/wp-content/uploads/2017/05/logistic-regression-example-in-python.png)

Where P: probability of an event to happen

bo: numeric constant,

b1: coefficient of the independent variable

x: independent variable



## K Nearest neighbor

The algorithm is based on calculation of the number of objects in each class of the sphere (hypersphere) with the center in the recognized object. The object belongs to the class, which objects dominate in this sphere. This technique supposes that weights have been chosen individually for every object. If weights are not same, instead of calculation of the number of objects their weights can be added together

## Neural Network

(will be discussed in Mona ‘s part in details)

## Naïve Bayes

It is based on conditional probability and it is particularly suited when the complexity of the inputs is high (ex: Spam mails)

## SVM

given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two-dimension space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

**Recurrent Neural Networks**

## Why RNN?

In the case where you had to memorize previous information traditional neural network approaches including CNN and DNN cannot deal with this due to the following reasons. first, because these approaches only handle fixed size vector as an input and produce a fixed size vector as an output. second, because those models operate with a fixed number of computational steps (number of layers).so we had to find a new approach to deal with such problem which is RNN.

## What is RNN?

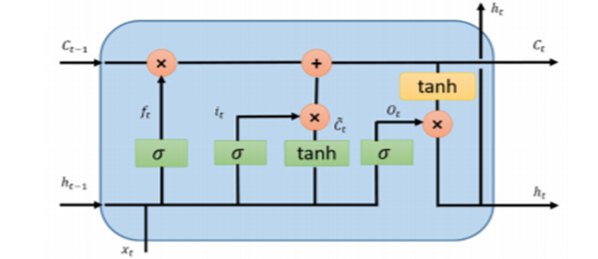
RNN is the network with the loops. Which allow information to persist in network. RNN has feedback connection the network itself which allow activations to flow back in a loop, learn sequences and information to persist. RNN can be thought of as multiple copies of the same network each network passing a message to a successor.

-RNN different architectures especially LSTM and GRU are used in most of the types of deep leaning systems, deep supervised learning, deep semi supervised learning, deep unsupervised learning. While standard NN only used in the deep supervised learning only.

-The main problem with the RNN approach is the vanishing gradient problem. Until the arouse of LSTM approach which is then developed into many algorithms whose idea is based on having a memory cell to carry certain information and parse them through the layers.

## Long short-term memory (LSTM)

The key idea of LSTMs is the cell state. LSTMs remove or add information to the cell state called gates: An input gate (i), forget gate (f) and output gate (o) the equations can be defined as:



f = σ (W. [h, x] + b),

i = σ (W. [h, x] + b),

C~ = tanh (W. [h, x] + b),

C = f ∗ C + i ∗ C~,

O = σ (W. [h, x] + b),

h = O ∗ tanh(C).

## Gated Recurrent Neural Network

GRU also came from LSTMs with slightly more variations. The main reason for the popularity is the computation cost and simplicity of the model. This technique combines the forget and input gates into a single update gate and merges the cell state and hidden state along with some other changes.

A picture containing sky, clock

Description automatically generated

z = σ (W. [h, x]),

r = σ (W. [h, x]),

h~ = tanh (W. [r ∗ h, x]),

h = (1 − z) ∗ h + z ∗ h~

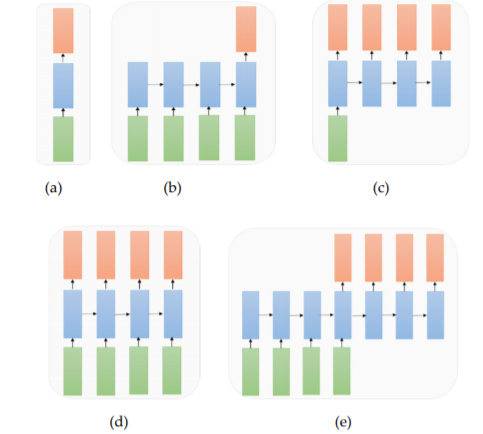
## convolutional LSTM

convolutional LSTM+FCN combines the property of LSTM to preserve or memorize certain information and the ability of CNN to recognize frames and videos, to be able to predict next frames in the video. so instead of having convolutional network that may do sampling in a single step you also have LSTM units that allow information to propagate through the next step.

A close up of a clock

Description automatically generated

According to Application point of view we can classify the RNN applications into many types which are:

1. **One to One**: Standard mode for classification without RNN (e.g., image classification problem)
2. **Many to One**: Sequence of inputs and a single output (e.g., the sentiment analysis where inputs are a set of sentences or words and output is a positive or negative expression)
3. **One to Many**: Where a system takes an input and produces a sequence of outputs (Image Captioning problem: Input is a single image and output is a set of words with context)
4. **Many to Many**: Sequences of inputs and outputs (e.g., machine translation: machine takes a sequence of words from English and translates to a sequence of words in French)
5. **Many to Many**: Sequence to sequence learning (e.g., video classification problem in which we take video frames as input and wish to label each frame of the video)

RNN applications example:

1.Tensor processing

2. Natural language processing

3.Time series data analysis

4.Speech and audio processing

**Natural language processing (NLP) & Speech recognition**

## NLP

Deep learning and natural language processing (NLP) is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages. NLP is an automated way to understand and analyze natural human language and extract information from such data by applying machine algorithms.

## NLP and Text mining

Text mining is a vast field that makes use of NLP to derive high quality information from the text, so text mining is the process and NLP is the method to carry out text mining

-- Challenges in natural language processing frequently involve speech recognition, natural language understanding and natural language generation

**Talking briefly about each of them:**

## speech recognition

Speech recognition is the problem of understanding what was said. The task of speech recognition is to map an acoustic signal containing a spoken natural language utterance into the corresponding sequence of words intended by the speaker. Given an utterance of text as audio data, the model must produce human readable text

-Particularly a good performance of speech recognition is achieved through the use of speech recognition by the statistical language modelling methods Currently, however the field of natural language processing is shifting from statistical methods to neural network methods. There are still many difficult problems to solve in natural language processing.

## Natural language understanding

NLU is the understanding the meaning of what the user or the input which is given means. That is nothing but the understanding of the text given and classifying it into proper intents.

-it is a method for Mapping the given input into a useful representation. Different level of analysis required:

1. morphological analysis.
2. syntactic analysis.
3. semantic analysis.
4. discourse analysis.

## Natural language generation

Producing output in the natural language from some internal representation. Different level of synthesis required:

deep planning (what to say),

**syntactic generation**

-NLU is harder than NLG due to some ambiguity that NLU must deal with to understand the natural languages it is dealing with these ambiguities are:

***Lexical ambiguity***

This ambiguity is due to the presence of two or more single meanings within a **single word**

***Syntactic ambiguity***

It is also called structural or grammatical ambiguity, it is due to having **a sequence of words** that may give more than one meaning, which need to be understood by the machine

***Referential ambiguity***

This type arises when we are referring to a certain word using pronouns where the machine sometimes cannot understand which word is referred to using this pronoun

## Modelling

Natural language processing is shifting from statistical methods to neural network methods. There are still many difficult problems to solve in natural language processing. Nevertheless, deep learning methods achieve the most modern results for some specific language problems.

Language modelling is really a subtask of more interesting natural language problems, specifically those that condition the language model on some other input. The main problem of language modelling is to predict the next word given the previous words

Language modelling can be useful for various speech and language processing applications, including automatic speech recognition before it used different equations based on conditional probability then now it is using NN approaches specifically RNN

A deep learning language model can predict the probability of the next word in the sequence, based on the words already observed in the sequence. This kind of language modelling is named as word-based language modelling

## NLP Terminology:

**Tokenization**

is an essential strategy for most NLP tasks. It parts a sentence or archive into tokens which are words or expressions. For English, it is minor to part words by the spaces

**Normalization**

Before further processing, text needs to be normalized. Normalization generally refers to a series of related tasks meant to put all text on a level playing field: converting all text to the same case (upper or lower), removing punctuation, expanding contractions, converting numbers to their word equivalents, and so on. Normalization puts all words on equal footing and allows processing to proceed uniformly.

**Stemming**

Stemming is the process of eliminating affixes (suffixed, prefixes, infixes, circumfixes) from a word in order to obtain a word stem.

**Lemmatization**

Lemmatization is related to stemming, differing in that lemmatization is able to capture canonical forms based on a word' lemma.

**Parts of speech (POS)**

labeling and parsing are strategies that investigate the lexical and syntactic data. POS labeling is utilized to decide the comparing POS tag for each word. Like word division, it is additionally a successive labeling issue. The POS labels, for example, descriptive word, thing, are very useful in light of the fact that opinion words are normally modifiers and opinion targets (i.e., entities and aspects) are things or mix of things. While POS labeling gives lexical data, parsing acquires syntactic data. Parsing produces a tree which speaks to the linguistic structure of a given sentence with the comparing relationship of various constituents. Contrasting with POS labeling, parsing gives wealthier structure data.

Applications of NLP:

1.Sentimental analysis

2.Chatbot

3.Speech recognition

4.Machine Translation

5.Text categorization

6.Spam filtering

7.Information Extraction

# Esraa’s task

**Concept of Chatbot**

## Problem faces the companies before Chatbots:

*Customer satisfaction with a company’s services is often seen as the key to success and long-term competitiveness for a company*

*For example, as customers sometime cannot understand how to use a specific product in a such company so they direct to call customer service for asking about what they want to know, so the employees in customer service (specifically in call centers) had to sift through long documents to find the answer.*

*As a result, the only way to get help quickly was to pick up the phone and talk to underwriting or sales support – even for answers to FAQs or to basic “how-to” questions. This overloaded the call centers, resulting in long wait times as it takes a long time to process a single request.*

*As a result, customer experience their interactions disappointed and dissatisfied which reduces the throughput and business performance drastically. Research showed that nearly 75% of customers have experienced poor customer service.*

## Solution:

*In order to truly be effective and make business processes automated an alternate system is required. An advance dialogue system known as AI chatbot application system could automate the entire business processes.*

## What is the chatbot?

*A chatbot is a computer program which conducts conversation in natural language via speech or text, understands what the user is trying to say and reply with relevant message based on business rules and data of organization.*

*And also, chatbot application system must have natural language processing (NLP), deep neural networks (DNN) so that it can understand what customers are looking for and can analyze data better than humans.*

## An overview about the history of chatbot:

*This technology started in the 1960’s; the aim was to see if chatbot systems could fool users that they were real humans.*

## The example of traditional & old Chatbots:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Chatbot | Year | Description | Open source/ proprietary | Technology/  Approach | Self-learning | Type |
| Eliza | 1966 | Simulation of psychotherapist,  rephrasing her response with few grammar rules | Open source | Basic Pattern matching | NO | Text |
| Parry | 1972 | PARRY attempt to simulate the behavior of person with paranoid schizophrenia | proprietary | Like Eliza chatbot | NO | Text |
| Jabberwacky | 1982 | This Chatbot aim to simulate natural human chat in an interesting, entertaining and humorous manners | proprietary | Contextual pattern matching ,learns by association ,storing replies to inputs in a database | Yes | Text |
| Alice | 1995 | A natural language processing chatbot-a program that engages in a conversation with human by applying some heuristic pattern matching | Open source | AIML | NO | Text |

## The Drawbacks of this traditional and old chatbots:

1. No self-learning.
2. The traditional chatbot’s dialogue capability is too inflexible. It can answer to the user only if there is a pattern (lexical) matching between the user query and set of question-answer stored in its knowledge base. The answers are given using a set of predefined responses.
3. Traditional chatbots are lacking in the intuitive capability of human beings to see the meaning, relationships and possibilities beyond the reach of sense.

As result , this traditional chatbots are not effective and customers are not satisfied , as there is no sense between human and computer . So nowadays, developers make an innovation by using DL (deep learning) in the chatbot applications.

There are two main tasks in deep learning (DL):

1. The first is to extract meaning from the input.
2. The second is to generate an output from that, either a translation or a response in the case of a chatbot application.

## **Challenges in developing a good model:**

The major challenge is that creates an adequate sense of context and effectively related inputs to outputs. The sequence-to-sequence (seq2seq) model in deep recurrent neural networks (DNN) with attention mechanism provides an appropriate architecture to meet these challenges.

## Chatbot applications are classified into:

**Chatbot applications can be grouped into four different categories, namely service, commercial, entertainment and advisory chatbot.**

1. Service chatbots:

are designed to provide facilities to customers. For example, logistics firm to respond to questions about deliveries and provide copies of dispatch documents through instant messaging channel rather than emails or phone calls.

1. Commercial chatbots:

are designed to streamline purchases for customers. For example, a pizza company can take delivery orders or notify promotions via messaging interface.

1. Entertainment chatbots:

are designed to keep customers engaged with sports, favorite band, movies or other events. It offers the option of placing bets, detail on upcoming events and ticket deals.

1. Advisory chatbots:

are designed to provide suggestions, give recommendations on service, offer maintenance or repair goods. This type of chatbot can contact people, offer support and advice tips when it is needed.

**chatbot application can be classified into two groups such as task-oriented and non-task-oriented.**

1. Task-oriented chatbots:

aim to assist the customers to complete certain tasks and have short conversations. For example, Siri, Google Now, Alexa dialogue agents can give travel directions, find restaurants and help to make phone calls or texts.

1. Non-task-oriented chatbots:

focus on conversing with customers to answer questions and entertainment.

**chatbot applications can be also divided into four groups such as goal-based, knowledge-based, service-based and response generated-based.**

1. Goal-based chatbots:

are classified based on the primary goal aim to achieve. They are designed for particular task and setup to have short conversations to get information from the user to complete the task. For example, a company deploys chatbot on their websites to help the customer to answer their question or address problems.

1. Service-based chatbots:

are classified based on facilities provides to the customer. It could be personal or commercial purpose. For example, logistics company could provide copies of dispatch documents through chatbot rather than phone calls or customer can make a meal order from MacDonald.

1. Knowledge-based chatbots:

are classified based on the knowledge they access from the underlying data sources or the amount of data they are trained on. The two main data sources are open-domain and closed-domain. Open-domain data sources answer depends on general topics and respond appropriately. Closed-domain data sources focus on a particular knowledge domain. All information required for answering the question is provided in the dataset itself such as Daily Mail.

1. Response Generated-based chatbots:

are classified based on what action they perform in response generation. The response models take input and output in natural language text. The dialogue manager is responsible for combining response models together. To generate a response, dialogue manager follows three steps. First, it uses all response models to generate a set of responses. Second, returns a response based on priority. Third, if no priority response, the response is selected by the model selection policy.

# NOW, Focus on Response Generated-Based Chatbot as it is related to our Chatbot project.

there are various response models that are based on four categories namely—**Template-based Model**, **Generative Model**, **Retrieval-based Model** and **Search Engine ModeI**.

## Template- based Model:

This Model uses AIML and Pattern Matching: Artificial Intelligence Markup Language (AIML) is most popularly used for writing patterns and response in the process of chatbot development.

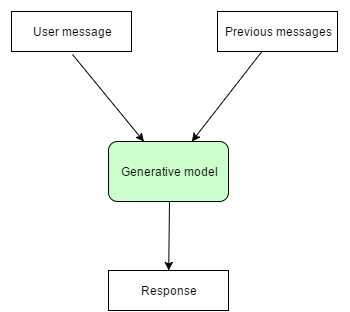
Example of AIML:



With a natural language processing pipeline and predefined rich pattern, AIML can be used to build a smart chatbot. These bots parse user message, find synonyms and concepts, tag parts of speech and find out which rule matches the user query. However, these bots do not run machine learning algorithms or any other APIs unless specially programmed.

## Generative-Model:

This model is used for the development of smart bots that are quite advanced in nature. This type of chatbot is very rarely used, as it requires the implementation of complex algorithms.



Generative models are comparatively difficult to build and develop. Training of this type of bot requires investing a lot of time and effort by giving millions of examples. This is how the deep learning model can engage in conversation. However, still, we cannot be sure what responses the model will generate. where they are able to generate new messages based on the context and/or current states of the conversation

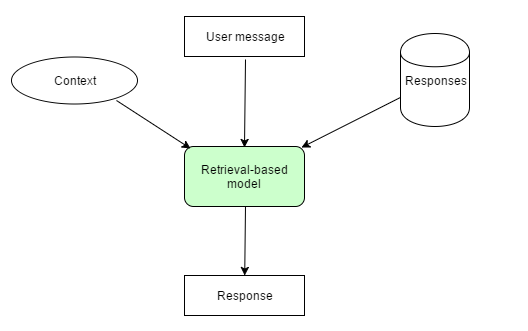
## retrieval-based Model:

Retrieval-based models are more in use at the moment. Several algorithms and APIs are readily available for developers to build chatbots on this architectural model.

This bot considers the message and context of the conversation to deliver the best response from a predefined list of messages.

The aim of the system is to learn how to select the best argument from a pre-defined topic that matches the current user’s response and the history of the conversation.

retrieval-based method is straightforward and guarantee to produce high quality messages, it is limited to only arguments that are available in the dataset and cannot adapt or tailor to every new responses from the users.



All these models are used to find responses based on structured documents. such approaches also depend on existing Q-R pairs as training data. Like other language generation tasks, such as machine translation and paraphrasing, the fluency and naturality of machine generated text is another drawback.

# The way to find responses based on Un structured data:

a novel response retrieval approach**, DocChat**, to find responses based on unstructured documents. For each user utterance, instead of looking for the best Q-R pair or generating a word sequence based on language generation techniques, our method selects a sentence from given documents directly, by ranking all possible sentences based on features designed at different levels of granularity.

## An overview of recent chatbots:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **chatbot** | **IBM Watson** | **Google Assistant** | **Cortana** | **Alexa** | **Siri** |
|  | 2006  Watson uses IBM’s Deep QA software and Approach UIMA (Unstructured information Management Architecture) | 2016  It is the most accessible chatbot, as it is available on all Android and IOS devices (Mobile App) and also in chromebooks  (**Chromebooks** are laptops, detachables and tablets powered by Chrome OS: the operating system that is speedy, smart and secure.) (Desktop App)  Google has its own line of google home speakers | 2014  Second accessible chatbot.  Microsoft's AI, Cortana, comes standard on Windows machines (desktop App), but it's also available for download on Android and iOS (Mobile App). | 2015  Third accessible chatbot.  Amazon's smart assistant is accessible through the Amazon Echo line of speakers as well as through Amazon's line of Fire tablets and Fire TV. Of all the AI assistants, Alexa is available on the widest variety of devices in terms of design. | 2011  fourth accessible chatbot.  You can access Siri on nearly any Apple device, including its line of laptops (Desktop App), desktops, phones (Mobile App) and tablets, and smartwatches. |

**Conclusion:**

*According to our customer, who are developers, we decided to create our chatbot, which is web-based, as Web Apps, are accessed via the internet browser and will adapt to whichever device you are viewing them on. They are not native to a particular system and are not required to be downloaded or installed.*

**Extra:**

*If you’re developing a chatbot, you need to choose a bot-building framework. A bot framework is like a workshop full of tools you can use to define your bot’s behavior, give it a personality, teach it what to say, and help it generate valuable responses to user inquiries. It’s where your bot comes into its own. You need to choose the one that makes the most sense for your objectives, and the best way to make an informed choice is to have a solid understanding of different frameworks. In this guide, we cover the nuts and bolts of the most popular bot-building frameworks:*

* *Amazon Lex*
* *Botkit*
* *Botpress*
* *Dialogflow*
* *Gupshup Bot Platform*
* *IBM Watson Assistant*
* *Microsoft Bot Framework*
* *Pandorabots*
* *Rasa*
* *Wit.ai*

# Mona’s task:

**Neural Network Architecture**

## What is Neural Network?

Neural networks take their inspiration from the notion that a neuron’s computation involves a weighted sum of the input values. These weighted sums correspond to the value scaling performed by the synapses and the combining of those values in the neuron.

Neural network models have provided a powerful learning method for use in many natural language problems recently. There are two major types of neural networks architectures that can be combined in two ways: feed-forward networks and recurrent networks. While convolutional feed-forward networks are able to extract local patterns, recurrent neural networks are able to capture long-range dependency in the data by abandoning the Markov assumption.

## What is ANN?

It Strands for Artifical Neural Network ANNs or general NNs consist of Multilayer Perceptron’s (MLP) which contain one or more hidden layers with multiple hidden units (neurons) in them

**DNN is a type of artifical neural network which has many approachs which are:**

1. **Gradient Descent**

The gradient descent approach is a first-order optimization algorithm which is used for finding the local minima of an objective function. This has been used for training ANNs in the last couple of decades successfully.

1. **Stochastic Gradient Descent (SGD)**

Since a long training time is the main drawback for the traditional gradient descent approach, the SGD approach is used for training Deep Neural Networks (DNN).

1. **Back-Propagation (BP)**

DNN is trained with the popular Back-Propagation (BP) algorithm with SGD. In the case of MLPs, we can easily represent NN models using computation graphs which are directive acyclic graphs. For that representation of DL, we can use the chain-rule to efficiently calculate the gradient from the top to the bottom layers with BP, as shown in References.

## Different types of NN

1. **Feedforward Neural Network (FFN)**

***Definition***

This is one of the simplest types of artificial neural networks. In a feedforward neural network, the data passes through the different input nodes till it reaches the output node.

In other words, data moves in only one direction from the first tier onwards until it reaches the output node. This is also known as a front propagated wave which is usually achieved by using a classifying activation function.

***Model***

Feed-forward Neural Network (FNN) consists of multiple layers of nodes. Each layer is fully connected to the next layer in the network. Nodes in the input layer represent the input data. All other nodes map inputs to outputs by a linear combination of the inputs with the node’s weights w and bias b and applying an activation function. This can be written in matrix form for FNN with ℓ + 1 layers as follows: y(x) = fℓ(· · · f2(w⊤ 2 f1(w⊤ 1 x + b1) + b2)· · · + bℓ).

1. **Convolutional Neural Network (CNN)**

***Definition***

CNNs are a subtype of the discriminative deep architecture and have shown satisfactory performance in processing two-dimensional data with grid-like topology, such as images and videos. The concept of CNNs is inspired by time-delay neural networks (TDNN). In a TDNN, the weights are shared in a temporal dimension, which leads to reduction in computation. In CNNs, the convolution has replaced the general matrix multiplication in standard NNs. In this way, the number of weights is decreased, thereby reducing the complexity of the network. Furthermore, the images, as raw inputs, can be directly imported to the network, thus avoiding the feature extraction procedure in the standard learning algorithms

***Model***

Convolutional Neural Network (CNN) is a class of FNN which is designed to require minimal preprocessing. The network learn filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature engineering is a major advantage of CNN. We build our CNN upon that of which is originally proposed for sentence classification. Our CNN consists of six main layers:

(1) a look-up tables to encode words in sentences by their embeddings,

(2) a convolutional layer to recognize w-grams,

(3) a nonlinear layer with the rectifier activation function,

(4) a max pooling layer to determine the most relevant features,

(5) a fully connected layer with drop-out and

(6) a logistic regression layer (a linear layer with a softmax at the end) to perform classification.

1. **Recurrent Neural Network (RNN)**

***Definition***

A Recurrent Neural Network is a type of artificial neural network in which the output of a particular layer is saved and fed back to the input. This helps predict the outcome of the layer.

***Model***

Given an input sequence [x1, x2, . . . , xn], a standard Recurrent Neural Network (RNN) computes the hidden vector sequence [h1, h2, . . . , hn] and outputs vector sequence [y1, y2, . . . , yn] by iterating the following equations from t = 1 to n: ht = σ Wxt + Uht−1 + b h yt = Vht + b y where W, U, V denote weight matrices (e.g., W is the input-hidden weight matrix, U is the hidden-hidden weight matrix, and V is the hidden-output weight matrix); the b terms denote bias vectors; and σ is the hidden layer function, which is usually an elementwise application of a sigmoid function.

**ALEXA**

## What is ALEXA?

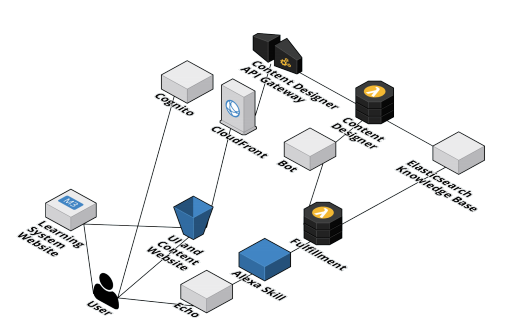
One of voice technology using Amazon Alexa along with Lex Chatbot technology to provide advice to new and current students using a mock educational platform. Alexa Skills Kit (ASK) is used to build skills for use in the Alexa ecosystem and devices and lets developers take advantage of all Alexa capabilities. Amazon Lex bots support both voice and text and can be deployed across mobile and messaging platforms

Chatbot Design

**Amazon Lex uses AWS Lambda functions to implement the business logic for a chatbot. At the conceptual level, it includes six logical functions:**

1. Bot – A bot contains all of the components of a conversation.
2. Intent – An intent represents a goal that the bot’s user wants to achieve (buying a plane ticket, scheduling an appointment, or getting a weather forecast, and so forth).
3. Utterance – An utterance is a spoken or typed phrase that invokes an intent. “I want to book a hotel” or “I want to order flowers” are two simple utterances.
4. Slots – Each slot is a piece of data that the user must supply in order to fulfill the intent. Slots are typed; a travel bot could have slots for cities, states or airports.
5. Prompt – A prompt is a question that asks the user to supply some data (for a slot) that is needed to fulfill an intent.
6. Fulfillment – Fulfillment is the business logic that carries out the user’s intent. Lex supports the use of Lambda functions for fulfillment.

**Implementation of the Learning System Customer Chatbot is as shown in Figure. This figure displays 12 components of the implementation as described:**

1. User – The end-user or an administrator of the system.
2. Echo – A smart speaker created by Amazon.
3. Learning System Website- A mock site to simulate the front page of Learning Circles website.
4. Alexa Skill- An intelligent voice service using Amazon Alexa Skills Kit. The Alexa Skills Kit (ASK) is a collection of self-service APIs, tools, documentation, and code samples that makes it fast and easy for you to add skills to Alexa. ASK enables designers, developers, and brands to build engaging skills and reach customers through tens of millions of Alexa-enabled devices.
5. User Interface (UI) and Content Website- SDK provides the ability to interface with the Bot through a website.
6. Fulfillment- The Lamda interface for the application that provides a response back to the user based on their input.
7. Elastic Search Knowledge Base- A distributed search and analytics engine built on Apache Lucene.
8. Bot- An Amazon Lex Bot that provides natural language processing.
9. Cloud Front- Amazon CloudFront is a global content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to your viewers with low latency and high transfer speeds.
10. Cognito- Amazon Cognito provides solutions to control access to AWS resources from your app. With this tool, an administrator can define roles and map users to different roles so your app can access only the resources that are authorized for each user.
11. Content Designer- SDK based web administrator interface to input questions and potential responses.
12. 12. Content Designer API Gateway- Provides Rest API for the web user interfaces.

# Menna’s task

**Cloud Computing**

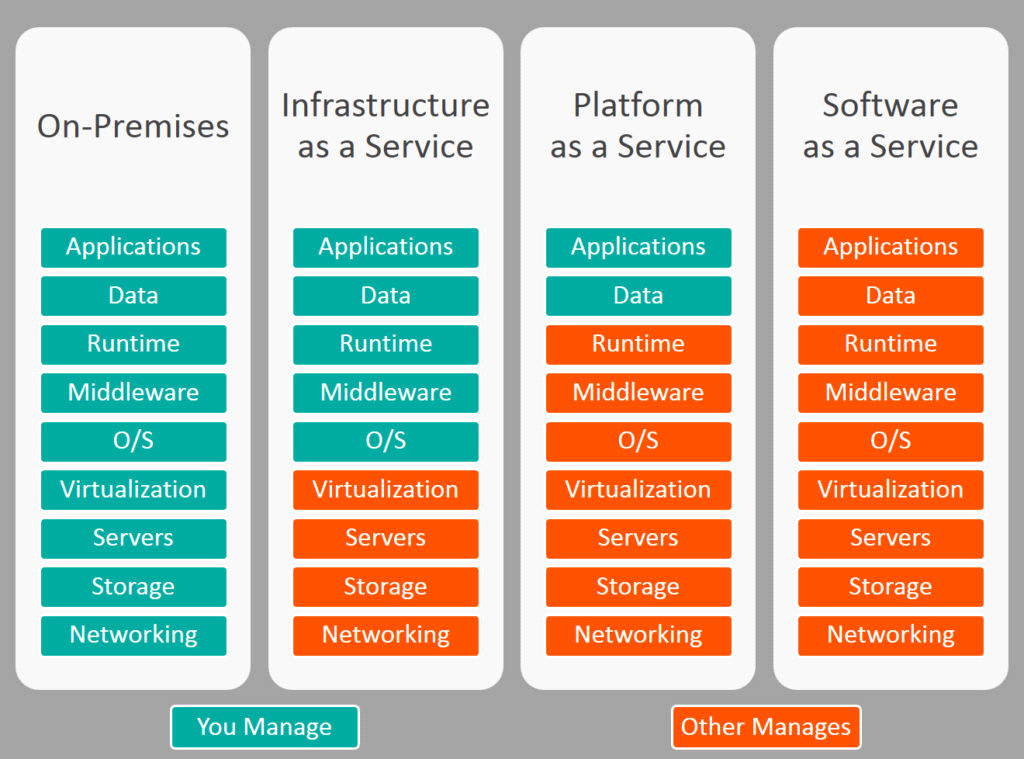
## Cloud computing Definition

NIST (The National Institute of Standards and Technology)describes cloud computing as “a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”

## The cloud computing model

It has mainly consisted of service models and deployment models.

* Deployment models: The way you can deploy your application on cloud and according to NIST they are:
* Public: the cloud infrastructure is provisioned for open use by the general public .it may be owned, managed and operated by a business, academic, or government organization or some combination of them.
* Private: the cloud infrastructure is provisioned for exclusive use by single organization comprising multiple consumers.it may be owned, managed, and operated by the organization, a third party or some combination of them
* Hybrid: the cloud infrastructure is a composition of two or more distinct cloud infrastructure that remain unique entities but are bounded by standardized or propriety technology that enables data and application portability
* Community: the cloud infrastructure is provisioned for exclusive used by specific community of consumers from organizations that have shared concerns, it may be owned, managed or operated by one or more of organizations in the community.
* Service models: How do you want the cloud provider to give you a particular service, & they are three:
  + IAAS (Infrastructure as a service)
  + PAAS (platform as a service)
  + SAAS (software as a service)

This picture simply demonstrates the difference between them & just note on-premises means you create your own private cloud and manages it totally (ex: availability, maintenance…etc.)

## Advantages of the cloud

It mainly offers:

* lower cost as Cloud computing eliminates the expense of setting up and running on-site data centers
* Decrease the workforce
* Focus mainly on app development instead of managing servers and focusing on maintenance
* Higher security as Cloud service providers typically also offer a broad set of policies, compliance, technologies and controls that strengthen your security posture by protecting your data, apps and infrastructure from threats

## From Cloud Computing to Fog and Edge Computing

Cloud computing is the computing paradigm that enables ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., computing and storage facilities, applications, services, etc.) . Through virtualization technology, cloud computing shields the diversity of underlying devices and provides users with a variety of services in a transparent way, including IaaS (Infrastructure-as-a-Service), PaaS (Platform-as-a-Service) and SaaS (Software-as-a-Service) . Due to the increasing number of access devices, cloud computing may face some problems in the bandwidth, latency, network unavailability, security and privacy, etc. Fog computing is considered as an extension of cloud computing to the edge network, providing services (e.g., compute, storage, networking, etc.) closer to near-user devices (e.g., network routers, various information systems, etc.), instead of sending data to cloud.

## FOG Computing

In fog computing paradigm, data storage and processing rely more on local devices, rather than on cloud data center as it mainly works on extending the cloud closer to the things that generate and act on data which helps in:

* Analyzing sensitive data locally instead of sending it to the cloud for analysis. Your IT team can monitor and control the devices that collect, analyze, and store data.
* Conserving network bandwidth by processing selected data locally instead of sending it to the cloud for analysis.

## Edge Computing

Similar to fog computing, edge computing also allows computation to be performed at the edge of the network, but at closer proximity to the data sources. The difference between fog computing and edge computing is that fog computing relies on interconnection capabilities among nodes, whereas edge computing runs in isolated edge nodes. Edge computing provides edge services near the source of data to meet the critical requirements in agile connectivity, real-time optimization, smart applications, security and privacy. As the supplementary, fog computing and edge computing, which provide compute, storage, and networking services between end devices and traditional cloud computing [37], offer bright prospects for smart manufacturing applications

The following figure illustrate the differences

A close up of a map

Description automatically generated

## Examples for free cloud platforms

|  |  |  |  |
| --- | --- | --- | --- |
| Point of comparison | AWS | Microsoft Azure | Google cloud |
| Caching | Elastic Cache | Redis Cache | Cloud CDN |
| Processor | In AWS, 128 can be the maximum processor in VM | In Azure, it can be 128 | In Google cloud, it is only 96. |
| Marketplace | In this, AWS marketplace | Azure Marketplace | G suite Marketplace |
| App Testing | In AWS, device farm is being used. | In Azure, DevTest labs are being used | Cloud Test lab is being used in this. |
| GIT Repositories | AWS source repositories | Azure source repositories. | Cloud source repositories. |
| Platform as service | Elastic Beanstalk | Cloud Services | Google App Engine |
| Storage of Object | S3 | Block Blob | Cloud Storage |
| Managed data warehouse | Redshift | SQL warehouse | Big Query |
| Kubernetes Management | EKS | [Kubernetes service](https://www.educba.com/kubernetes-operators/) | Kubernetes engine |
| File Storage | EFS | Azure Files | ZFS and Avere |
| Serverless computing | Lambda is being used for serverless computing | In Azure, Azure functions are used. | In google cloud, Cloud functions are used. |
| API management | Amazon API Gateway | Azure API gateway | Cloud endpoints |
| Media services | Amazon Elastic Transcoder | Azure media services | Cloud video intelligence API |
| Website | Aws.amazon.com | Azure.microsoft.com | Cloud.google.com |

## AZURE vs AWS

* Competitive pricing

AWS is 5 times more expensive than Azure for Windows Server and SQL Server. Azure matches AWS pricing for comparable services.

* Achieve more with open source on Azure

Use any open-source OS, languages, and tools on Azure. Azure made the most contributions to GitHub in 2017 and it’s the only cloud with integrated support for Red Hat.

* Enhanced proactive security and compliance

Compare AWS and Azure and you’ll find that Azure’s compliance offerings—including 70+ compliance certifications—and are more comprehensive.

* Get more value from your existing Microsoft investment

Keep using your organization’s existing tools and knowledge: get a consistent experience across your on-premises and cloud technologies by integrating them with Azure Active Directory (AD).

## Microsoft AZURE

What is Microsoft AZURE?

It is a cloud service provider owned and managed by Microsoft. It offers various services on the cloud such as compute, storage & database. It has a pay-as you go model & is the second largest cloud provider in the market right now.

* AZURE has a lot of free services but the most that are relevant to our project are
  + AI and Cognitive Services
    - Machine Learning
    - Cognitive Services
  + AZURE App services
  + AZURE chatbot which is built mainly on Microsoft bot framework
* Find below the documentation for AZURE: <https://docs.microsoft.com/en-us/azure/guides/developer/azure-developer-guide>

## As per our project

Microsoft AZURE can be used as cloud service provider where:

1. The application should be accessible to the public
2. Deploying the application on the cloud will provide us with the opportunity to scale up as the demand increases on the chatbot.
3. It may be used in training if the training data is very big.
4. The models will be deployed on the server which is provided by the cloud

## IBM Watson

IBM Watson is rule-based AI chatbot developed by IBM's DeepQA project. It is designed for information retrieval and question-answering system that incorporates natural language processing and hierarchical machine-learning method. Watson uses a broad range of mechanisms to identify and assign feature values such as names, dates, geographic locations or other entities to generated response. The machine learning system then learns how to combine the values of these features into a final score for each response. Based on that score, it ranks all possible answers and selects one as its top answer. Watson incorporates a variety of technologies including Hadoop, Apache Unstructured Information Management Architecture (UIMA) framework to examines the phrase structure and the grammar of the question to better gauge what's being asked. Applications for the Watson's underlying cognitive computing technology are almost endless. Because it can process text mining and complex analytics on huge volumes of unstructured data and handle enormous quantities of data. As the application gains experience with more input, it can find enough patterns to make accurate predictions. Besides the advantages of Watson, it has some major drawback such as it does not process structure data directly, no relational databases, higher maintenance cost, targeting towards bigger organizations and take longer time and effort to teach Watson in order to use its full potential.

## Data Mining

The enormous usage of computers has provided a huge amount of data for one’s disposal. Because of the spiraling amount of data, experts have been facing challenges in extracting useful and meaningful information from it. This has led to data mining.

* Data mining is a non-trivial process of extraction of information which is hidden, previously unknown and is potentially useful, from large databases. Data mining can also be explained as finding the correlations in a large relational database based on the different depth of angles, we analyze it. It is a powerful tool with high potential that helps the organizations or companies to increase their sales and gain more profit from the information about the dealings of their customers. Data mining provides us with the useful information that queries, and reports are not able to provide us efficiently. The information that is extracted by the data mining etiquette is not explicitly available in the database, whereas database application only projects the information that is available in the info bank with a restricted manipulation capacity. So, data mining is best described as knowledge unearthing in databases
* There are different process and techniques used to carry out data mining successfully and can be broadly classified in two categories those are descriptive and predictive. Descriptive is the technique that tells us about all the properties of the input data. Predictive is the type of technique that performs inferences in the input data to generate or predict the information which is hidden. The different techniques are listed below:
  + Data classification
  + Prediction data mining technique
  + Data clustering technique
  + Outlier analysis technique
  + Association rule mining (ARM)
  + Characterization – cum –discrimination

**Cognitive Search**

The stages of development of searching   
The ideal searching process would be able to match the search queries to the exact context and return results within that context. To achieve this, searching went through several stages of development

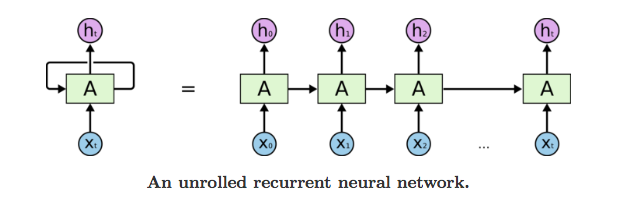
**1- Keyword Search:**

It uses dictionaries that have many words and indexes that refers to results that have this word then start to search these keywords into them and returns results by the indexes of these keywords in dictionaries in what is known as indexing dictionaries. This type of searching has existed almost since the creation of internal office networks (which pre-date the Internet)

**2- semantic Search:**

Semantic Search is look like the keyword search, but it uses natural language processing and neural networks to define which words have more importance in context than others then returns its values first.

**3- Cognitive Search:**

Search is no longer just about unstructured text contained in documents and web pages. Cognitive search solutions can also accommodate structured data contained in databases and even nontraditional enterprise data like images, video, audio, and machine data such as from internet-of-things (IoT) devices. Cognitive search uses Recurrent neural networks and attention method to get data from its context not just ordinary searching. RNN depends on networks with loops in them, allowing information to persist.

So, we recommend using Cognitive Search as it has more natural human interact

## References

AI Components Part

1. <https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence>
2. <https://www.researchworld.com/what-are-the-opportunities-and-practical-applications-of-ai-in-research-and-insights/>
3. <https://medium.com/@chethankumargn/artificial-intelligence-definition-types-examples-technologies-962ea75c7b9b>
4. <https://towardsdatascience.com/role-of-data-science-in-artificial-intelligence-950efedd2579>

NLP

1. <https://www.researchgate.net/publication/319164243_Natural_Language_Processing_State_of_The_Art_Current_Trends_and_Challenges>

Microsoft AZURE documentation

1. <https://docs.microsoft.com/en-us/azure/guides/developer/azure-developer-guide>

Building with Bot Framework: Publishing to Azure

1. <https://www.youtube.com/watch?v=ox455b98qKE>

**Important papers:**

1. https://arxiv.org/pdf/1706.03762.pdf
2. [**https://www2.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf**](https://www2.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf)
3. [**https://ieeexplore.ieee.org/document/6385297**](https://ieeexplore.ieee.org/document/6385297)
4. [**https://ieeexplore.ieee.org/document/8740963**](https://ieeexplore.ieee.org/document/8740963)
5. [**https://ieeexplore.ieee.org/abstract/document/6918822**](https://ieeexplore.ieee.org/abstract/document/6918822)
6. **Paper\_12 Survey\_on\_Chatbot\_Design\_Techniques\_in\_Speech\_Conversation\_Systems.pdf**
7. **Surveyofchatbotsystems.pdf**
8. **Chatbots\_Are\_they\_Really\_Useful.pdf**
9. **10.1109@ICEBE.2018.00019.pdf(A Survey on Chatbot Implementation in Customer Service Industry through Deep Neural Networks)**
10. **https://pdfs.semanticscholar.org/8e60/5c49d4a7cba9bf077d97b401ba78aafe693f.pdf (A SURVEY OF VARIOUS CHATBOT IMPLEMENTATION TECHNIQUES)**
11. **https://www.aclweb.org/anthology/W18-5215 DocChat: An Information Retrieval Approach for Chatbot Engines Using Unstructured Documents)**
12. **https://www.businessnewsdaily.com/10315-siri-cortana-google-assistant-amazon-alexa-face-off.html**
13. [**https://discover.bot/bot-talk/guide-to-bot-frameworks/**](https://discover.bot/bot-talk/guide-to-bot-frameworks/)
14. **A Comparative Study of Neural Network Models for Sentence Classification**
15. **Deep Learning for Natural Language Processing and Language Modelling**
16. **Review on Natural Language Processing (NLP) and Its Toolkits for Opinion Mining and Sentiment Analysis**
17. [**https://www.mdpi.com/2079-9292/8/3/292**](https://www.mdpi.com/2079-9292/8/3/292)
18. [**https://www.researchgate.net/publication/312432980\_A\_survey\_of\_deep\_neural\_network\_architectures\_and\_their\_applications**](https://www.researchgate.net/publication/312432980_A_survey_of_deep_neural_network_architectures_and_their_applications)