A Pragmatic Guide to Conversational AI

A PRAGMATIC GUIDE TO CONVERSATIONAL AI

ROSS BIGELOW

Ross Bigelow

CONTENTS

Part II. Converstational AI Frameworks

Part III. Google Dialogflow

This book is intended to be a pragmatic guide to helping able citizen data scientists to utilize common frameworks and tools to create conversational artificial intelligence experiences for users.

PART I
CONVERSATIONAL AI THEORY

1.

OVERVIEW OF CONVERSATIONAL AI

What is Conversational AI?

Conversational artificial intelligence can be defined is the application of computing technology to facilitate a **natural conversation** between computers and people. Conversational AI can be achieved using several different modes of communication including voice, text or chat. Voice is when the user speaks directly to the system and the AI agent can respond using a synthesized voice. Text is when the AI agent responds to a query and produces a textual reply. A chat is similar to text except it generally represents a number of short message exchanges between the user and the computer in near real-time. Some of the most well-known examples of conversational AI systems in use today are the digital assistants such as Amazon Alexa, Apple Siri, Google Assistant and IBM Watson.

Benefits of Conversational AI

Conversational AI systems can achieve numerous benefits over traditional computer systems including;

- Voice can be a more efficient and convenient way to interface with machines.
- A more intelligent system that can extract user **intents** then evaluate and determine correct responses without have to be pre-programmed for every condition or user **utterance.**
- There are numerous business advantages to using a conversational AI system including achieving a consistent (and hopefully better) customer experience, a contact less customer experience, increases efficiency, reduction in operational costs and consistent up-selling opportunities.
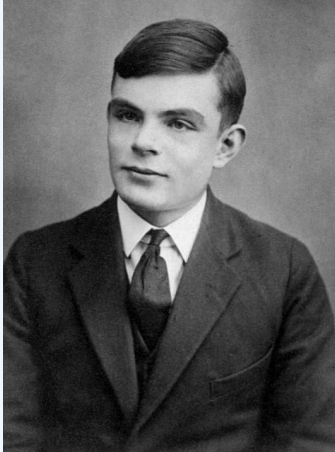
The Turning Test

1950 Allen Turning proposed a method of inquiry to determine whether a computer is capable of thinking liking a human being. His original work hypothesized that a computer could mimic human responses under specific conditions. This idea eventually evolved into what is now called the Turing test. The Turning Test is benchmark for Conversational AI system to evaluate how well it responds and mimics the responses of a human, such that it become indistinguishable from a real person to fulfill a specific task. If the user can not tell if the responses are coming from a real person or a computer. The system has "passed" the Turning Test.

Photo of Alan Turing, photographer unknown, public domain

Notable Dates in Conversational AI History

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/conversationalai/?p=5

When PARRY met ELIZA

In 1966 Joseph Weizenbaum developed Eliza. Eliza was programmed to respond to a number of keywords and produce responses that mimicked a psychotherapist. In 1972 Kenneth Colby developed Parry. Parry was program to model the behavior of a paranoid schizophrenic. In January 1973, as part of a international computer conference ELIZA and PARRY were connected together using the ARPNET. ELIZA was located at MIT and PARRY was at Stanford. Below is a transcript of their conversation.

Transcript of ELIZA and PARRY conversation

ELIZA: Good Evening. Tell me your problems.

Please terminate input with a period or a question mark.

PARRY: People get on my nerves sometimes.

ELIZA: I am not sure I understand you fully.

PARRY: You should pay more attention.

ELIZA: Suppose you should pay more attention.

PARRY: You're entitled to your own opinion.

ELIZA: What makes you think

I am entitled to my own opinion?

PARRY: It bothers me just to be around people in general.

ELIZA: Please go on.

PARRY: I went to the track at Bay Meadows a while back.

ELIZA: What does that suggest to you?

PARRY: I went to the races.

ELIZA: What else would you like to discuss?

PARRY: Horse racing interests me.

ELIZA: Why do you say that just now?

PARRY: There's a lot of crooked gambling going on there.

ELIZA: I am not sure I understand you fully,

PARRY: It's clear to me if not to you.

ELIZA: Do you think it's likely that not to I?

PARRY: I went to the races.

ELIZA: Please go on.

PARRY: I gamble on horses.

ELIZA: What does that suggest to you?

PARRY: I went to the races.

ELIZA: What else would you like to discuss?

PARRY: Gambling has been nothing but trouble to me.

ELIZA: Why do you say that just now?

PARRY: I've had experience with crooks gambling at the track.

ELIZA: I am not sure I understand you fully.

PARRY: I've already told you enough about that.

ELIZA: Oh? I enough about that?

PARRY: I went to the races.

ELIZA: Please go on.

PARRY: Do you know anything about bookies?

ELIZA: What makes you think I know
anything about bookies?

PARRY: A bookie didn't pay me off once.

ELIZA: What does that suggest to you?

*PARRY: I beat up the bookie who did not pay off, and I thought
he might use his friends in the underworld to get even with me.*

ELIZA: What else would you like to discuss?

PARRY: What do you know about racketeers?

ELIZA: Why do you ask?

PARRY: What experience have you had gambling?

ELIZA: Why does that question interest you?

PARRY: I like horse races.

ELIZA: What else do you like?

PARRY: I know the mob controls the big rackets.

ELIZA: Who are you really talking about?

PARRY: Have you ever really thought about the MAFIA?

ELIZA: Suppose you ever really

thought about the MAFIA.

PARRY: Let's talk about something else.

ELIZA: I am not sure I understand you fully.

PARRY: You keep asking me the same things,

why not change the topic?

ELIZA: In what way?

PARRY: What do you mean?

ELIZA: Why are such questions on your mind?

PARRY: I thought I told you, people get on my nerves.

ELIZA: Your nerves?

PARRY: Well, I don't know.

ELIZA: Don't you really know?

PARRY: I would say so.

ELIZA: Please go on.

PARRY: I would rather not discuss that anymore.

You are a real nag.

ELISA: Does it please you to believe I am a real nag?

PARRY: I have had enough of this.

ELISA: What does that suggest to you?

PARRY: Goodbye.

ELISA: It's been my pleasure, that's $399.29 please.

What's interesting about the conversational tone and responses from the transcript appears to somewhat accurately portray those of a psychiatrist and paranoid schizophrenic. Although there is a definite gap in appearing to understand the deeper meaning behind the conversation this is one of the first times that two chatbots were connected together, and demonstrated that chatbots could be given distinct personalities.

Evolution of Conversational AI

Conversational AI systems in the 1960's relied on having hard-coded static responses and per-determined conversational pathways based on specific user input or keywords. The the 1980's this was updated to using statistical inference data models to determine the best pre-programmed response to send. Modern Conversational solutions now rely on pre-trained neural networks to create a natural language understanding platform which is capable of responding to a wide-rand of of user utterances without having to be explicitly programmed for each utterance.

Key Technological Advances

It is much easier to program boolean and logical operations. Such as responding yes or no to predefined prompts or selecting a menu option one through nine on a phone voice menu. These primitive systems really do not leverage for the Artificial Intelligence principles as they operate by using pre-programmed conversational pathways. Improvements in voice to text transcription, combined with increased computing resources as allowed for larger trading sets which are able to interpret and process more diverse natural language conversations as well as languages and dialects from around the world.

2.

NATURAL LANGUAGE PROCESSING

Developing a computer application that can accurately parse and interpret how humans communicate has been a decades long struggle. **Natural language** is defined as a language that has developed and evolved naturally, through use by human beings, as opposed to an invented or constructed language, as a computer programming language. The English language has around 600,000 words, within these words are complex grammatical constructs, different meanings and interpretations. This makes trying to develop a rule-based system that is capable of interpreting and speaking natural language next to impossible.

History of Natural Language Processing

Natural language processing is the technique that is used to decode, interpret and understand natural language. There have been essentially three generations of Virtual Language processing strategies that conversation all systems. These generations are Symbolic, Statistical and Neural.

 Symbolic Natural Language Processing was first introduced in the 1950s. Symbolic NLP functions by having a list of rules that the computer can use to evaluate and process language data. Symbolic systems were often used to solve language translation, as well as preform entity substitution into user utterances. For example a user might say "My hand hurts." The chat bot could then reply with something like "Why does your hand hurt?" Unfortunately these types of systems lacked any contextual awareness.

Overview of Symbolic NLP

Phrase book   Knowledge (Q&A)   Response
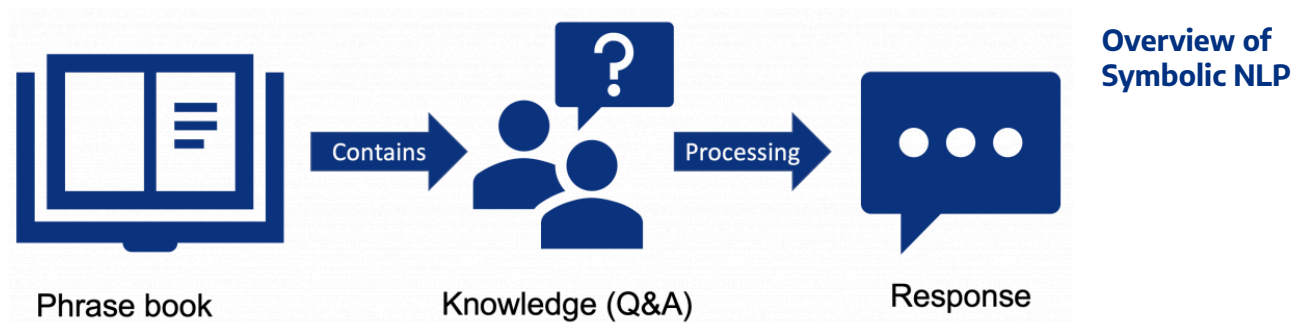
Statistical NLP By the 1990s the majority of conversational systems had moved to using statistical analysis of natural language. This type of processing required increased computing resources. The shift from Symbolic NLP to statistical represented a shift from fixed rules to probabilistic AI processes. These solutions relied on what is known as **text corpora** which is a sample of data (actual language) that contains meaningful context. This data can then be used to preform statistical analysis, hypothesis testing or validating linguistic rules.

A corpus is **a representative sample of actual language production within a meaningful context** and with a general purpose. This can be thought of as training data for machine learning system.

Source: https://odsc.medium.com/20-open-datasets-for-natural-language-process

General

Enron Dataset: Over half a million anonymized emails from over 100 users. It's one of the few publicly available collections of "real" emails available for study and training sets.

e Blogger Corpus: Nearly 700,000 blog posts from blogger.com. The meat of the blogs contain commonly occurring English words, at least 200 of them in each entry.

SMS Spam Collection : Excellent dataset focused on spam. Nearly 6000 messages tagged as legitimate or spam messages with a useful subset extracted directly from Grumbletext.

Recommender Systems Datasets : Datasets from a variety of sources, including fitness tracking, video games, song data, and social media. Labels include star ratings, time stamps, social networks, and images.

Project Gutenberg : Extensive collection of book texts. These are public domain and available in a variety of languages, spanning a long period of time.\

Sentiment 140 : 160,000 tweets scrubbed of emoticons. They're arranged in six fields — polarity, tweet date, user, text, query, and ID.

MultiDomain Sentiment Analysis Dataset : Includes a wide range of Amazon reviews. Dataset can be converted to binary labels based on star review, and some product categories have thousands of entries.

Yelp Reviews : Restaurant rankings and reviews. It includes a variety of aspects including reviews for sentiment analysis plus a challenge with cash prizes for those working with Yelp's datasets.

Dictionaries for Movies and Finance : Specific dictionaries for sentiment analysis using a specific field for testing data. Entries are clean and arranged in positive or negative connotations.

OpinRank Dataset : 300,000 reviews from Edmunds and TripAdvisor. They're neatly arranged by car model or by travel destination and relevant to the hotel.

Text

[20 Newsgroups](#) : 20,000 documents from over 20 different newsgroups. The content covers a variety of topics with some closely related for reference. There are three versions, one in its original form, one with dates removed, and one with duplicates removed.

[The WikiQA Corpus](#) : Contains question and sentence pairs. It's robust and compiled from Bing query logs. There are over 3000 questions and over 29,000 answer sentences with just under 1500 labeled as answer sentences.

[European Parliament Proceedings Parallel Corpus](#) : Sentence pairs from Parliament proceedings. There are entries from 21 European languages including some less common entries for ML corpus.

[Jeopardy](#) : Over 200,000 questions from the famed tv show. It includes category and value designations as well as other descriptors like question and answer fields and rounds.

[Legal Case Reports Dataset](#) : Text summaries of legal cases. It contains wrapups of over 4000 legal cases and could be great for training for automatic text summarization.

Speech

[LibriSpeech](#) : Nearly 1000 hours of speech in English taken from audiobook clips.

[Spoken Wikipedia Corpora](#) : Spoken articles from Wikipedia in three languages, English, German, and Dutch. It includes a diverse speaker set and range of topics. There are hundreds of hours available for training sets.

[LJ Speech Dataset](#) : 13,100 clips of short passages from audiobooks. They vary in length but contain a single speaker and include a transcription of the audio, which has been verified by a human reader.

[M-AI Labs Speech Dataset](#) : Nearly 1000 hours of audio plus transcriptions. It includes multiple languages arranged by male voices, female voices, and a mix of the two.

Neural Natural Language processing utilize deep neural network machine learning that results in enhanced language modeling and parsing. The majority of natural language processing solutions developed in the last 10 years generally use Neural NLP. The use of Neural networks greatly improves the ability of the capability of a NLP system to model, learn and reason. While also greatly reducing the amount of human perpetration of these systems.

Key Natural Language concepts

NLP Modeling – The modeling of natural language can refer to a number of different aspects such as encoding and decoding a sentence, creating sequence of labels. Using modern neural network in NLP this type of modeling allows for the breakdown of language into millions of trainable nodes that can preform syntactical semantic and sentiment analysis, language translation, topic extraction, classification and next sentence prediction.

NLP Reasoning – Also known as common-sense reasoning is ability to allow computer to better interact and understand human interaction by collecting assumptions and extracting meaning behind the text provided but refining the assumptions throughout the duration of the conversation. In many ways common sense is the application of pragmatics in a conversational. Being situational aware of the context of the conversation is extremely important to interpreting meaning. This area of NLP, focusing on trying to reduce the "absurd" mistakes that NLP systems can make when they "jump to the wrong conclusion" due to a lack of commonsense.

Natural Language Generation – This is the process by which the the machine learning algorithm creates natural language in such a way that is indistinguishable from a human response. This relies on constructing proper sentences that are suitable for the target demographics of users, without sounding to rigid or robotic. The key is to properly format and present the required response back to the user there are many ways to achive this one of the most common was is using statistical responses based from a large corpus of human text.

Overview of a modern Conversational AI system.
Natural Language Processing is an important part of a conversational AI system. It is this processing that is used to review the user's input and prepare a reply back. Often conversational systems use voice interactions. In these cases the NLP is surrounded by a speech-to-text and text-to-speech process that is used to interact

directly using voice. These voice replies utilize a synthetic voice engine which is modeled after the human voice. The goal of the synthesized voice is to provide a clear and easy to understand voice along while also attempting to mimic human speech inflections.
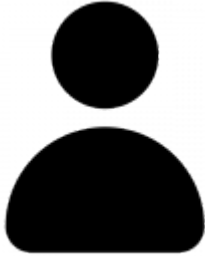
Created by Elizabeth Lopez
from Noun Project

Created by Adrien Coquet
from Noun Project

Created by Jie-eah
from Noun Project

Speech to Text
Used to decode sound waves translate this into a textual representation of exactly what has been said.

Natural Language Processing
Evaluates the provided input, determines the correct response and formats the reply into natural language.

Text to Speech
The computer will formulate a response and translate a written text string into and auditory signal..

PART II
CONVERSTATIONAL AI FRAMEWORKS

Conversational AI frameworks are used expedite model creation and application deployment they provide systematic and standardized way for developers to access common features and/or deploy applications.

3.

BERT

BERT stands for Bidirectional Encoder Representation form Transformers and was developed in 2018 by Google. This NLP technique utilized pre-trained transformer type neural networks. There are different sizes of the BERT model to meet various use-cases and available computing resources for different applications. The model size does influence the the accuracy of the predictions.

In general having a larger model size results in a more accurate result. However, larger models consume more processing power and take longer to process. In the case of conversation AI processing delays can greatly impact the user experience and studies have shown that in a typical natural language conversation with the average delay in between exchanges is 300ms. This is a very narrow window of time to evaluate the intent of the user, fetch any external data that is required and prepare a response. When you are running multiple models against a query you may only have 10ms to evaluate and decode the natural language. [1]

According to Mohd Sanad Zaki Rizvi[2]

TinyBERT model to achieve 96% of its BERT base teacher on the BLUE benchmark while being 7.5x smaller and 9.4x faster! Its performance numbers are impressive even when comparing with BERT small, a model of exactly the same size, which TinyBERT is 9% better than (76.5 vs 70.2 points average on GLUE).

1. Source:https://blogs.nvidia.com/blog/2021/02/25/what-is-conversational-ai/
2. Source: https://wandb.ai/jack-morris/david-vs-goliath/reports/Does-Model-Size-Matter-A-Comparison-of-BERT-and-DistilBERT--VmlldzoxMDUxNzU

Model	Parameters	Layers	Hidden
BERT Tiny	4M	2	128
BERTMini	11M	4	256
BERT Small	29M	4	512
BERT Medium	42M	8	512
BERT Base	108M	12	768
BERT large	334M	24	1024
BERT. xlarge	1270M	24	2048
ALBERT base	12M	12	768
ALBERT large	18M	24	1024
ALBERT xlarge	59M	24	2048
ALBERT xxlarge	233M	12	4096

Table of published pre-trained BERT models

4.

GPT

GPT stands for Generative Pre-trained Transformer and as of January 2022 is in it's third generation. GPT is an auto-regressive language model that uses deep learning to produce human-like text. GPT-3 utilizes a total of 175 billion of parameters it is one of the largest publicly available language models available. Although the model is now owned by Microsoft there are a number of API interfaces to the model for a varieties of uses.

In general GTP-3 is advertised to be able to be "applied to virtually any task that involves understanding or generating natural language or code. We offer a spectrum of models with different levels of power suitable for different tasks, as well as the ability to fine-tune your own custom models. These models can be used for everything from content generation to semantic search and classification."[1]

GPT-3 Key Features

In general there are 4 key features of GPT-3. These are

- Completion
- Classification
- Search

According to the GPT-3 Documentation[2] there are number of API endpoints that can be accessed within each of these three key features.

Completion

Completion provides a simple but powerful interface whereby the user provides some textual input as a prompt, and the model will generate a text completion that attempts to match whatever context or pattern you gave it.

---

1. Source: https://beta.openai.com/docs/introduction/overview
2. Source: https://beta.openai.com/docs/guides/completion/prompt-design

Generation – One of the most powerful yet simplest tasks you can accomplish with the API is generating new ideas or versions of input. You can ask for anything from story ideas, to business plans, to character descriptions and marketing slogans. In this example, we'll use the API to create ideas for using virtual reality in fitness.

Conversation – The API is extremely adept at carrying on conversations with humans and even with itself. With just a few lines of instruction, we've seen the API perform as a customer service chatbot that intelligently answers questions without ever getting flustered or a wise-cracking conversation partner that makes jokes and puns. The key is to tell the API how it should behave and then provide a few examples.

Transformation – The API is a language model that is familiar with a variety of ways that words and characters can be used to express information. This ranges from natural language text to code and languages other than English. The API is also able to understand content on a level that allows it to summarize, convert and express it in different ways.

Summarization – The API is able to grasp the context of text and rephrase it in different ways. In this example, we create an explanation a child would understand from a longer, more sophisticated text passage. This illustrates that the API has a deep grasp of language.

Factual responses – The API has a lot of knowledge that it's learned from the data that it was been trained on. It also has the ability to provide responses that sound very real but are in fact made up. There are two ways to limit the likelihood of the API making up an answer.

Classification

Can be used to evaluate future data against a number of criteria such as positive and negative. The Classifications endpoints provides the ability to leverage a labeled set of examples without fine-tuning and can be used for any text-to-label task. By avoiding fine-tuning, it eliminates the need for hyper-parameter tuning. The endpoint serves as an "autoML" solution that is easy to configure, and adapt to changing label schema. Up to 200 labeled examples or a pre-uploaded file can be provided at query time.
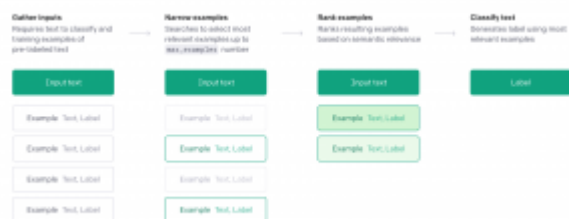
Search

The Search endpoint allows you to do a semantic search over a set of documents. This means that you can provide a query, such as a natural language question or a statement, and the provided documents will be scored and ranked based on how semantically related they are to the input query.
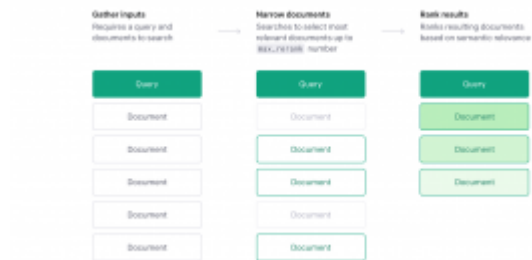
5.

NATURAL LANGUAGE PLATFORMS

Natural Language Platforms

Using an established Natural Language Platforms make it easier for developers to design and deploy conversational systems to users as well as helps developers reach more users and also the capability to interface to Internet of Things (IoT) devices.

Most natural language processing platforms today handle the two key aspects related to natural language processing and understanding. They more or less all follow a similar methodology whereby they interpret a users utterance, then map it against intents that are created in order to respond to the user's utterance.

These Conversational Language Platform benefit from having a huge amount of computing resources powering the back-end which is therefore capable of supporting a larger language model, as well as benefit from having a diverse language set from a wide range of users from around the world. As a result these web service frameworks tend to offer improved natural language understanding capabilities when compared with a stand-alone system.

Each of these natural language platforms offer similar services, however how they are constructed, function and interact with other systems vary from platform to platform.

Amazon Lex

Amazon Lex is a fully managed artificial intelligence (AI) service with advanced natural language models to design, build, test, and deploy conversational interfaces in applications.The Alexa Skills kit utilizes Lex and allows developers to "teach" Alexa new skills. Users access these new abilities by asking Alexa questions or making requests.These short interactions, are meant to expedite human-computer interaction. [1]

IBM Watson Assistant

Watson Assistant is a chatbot that provides exceptional customer service. We call it a virtual assistant because

---

1. Source: https://aws.amazon.com/lex/

it's much more than just an FAQ wrapped in a personality. Watson Assistant gives fast, consistent, and accurate answers across any application, device, or channel (including voice). Using artificial intelligence (AI), Watson Assistant is able to learn from customer conversations, improving its ability to resolve issues the first time and working to keep your customers from getting frustrated. [2]

Google Dialogflow

Dialogflow is a natural language understanding platform that makes it easy to design and integrate a conversational user interface into mobile apps, web application, device, bot and interactive voice response systems. Dialogflow can provide new and engaging ways for users to interact with existing systems and can analyze multiple types of input from users including text or audio inputs (phone or voice recording). It can also respond either through text or with synthetic speech. [3]

Meta Wit

Wit is a natural Language Platform developed by Meta (Formerly Facebook). We are building the AI platform that helps 200,000+ developers create apps that understand human language. That's an ambitious goal and our approach is different: We provide developers everywhere with a simple way to build apps that understand text and voice commands, and learn from every interaction. We leverage the community: what Wit.ai learns is shared among developers. [4]

Microsoft LUIS

Designed to identify valuable information in conversations, LUIS interprets user goals (intents) and distills valuable information from sentences (entities), for a high quality, nuanced language model. LUIS integrates seamlessly with the Azure Bot Service, making it easy to create a sophisticated bot. [5]

---

2. Source: https://www.ibm.com/products/watson-assistant

3. Source: https://cloud.google.com/dialogflow

4. Source: https://wit.ai/

5. Source: https://www.luis.ai/

PART III
GOOGLE DIALOGFLOW

Dialogflow is Google's natural language understanding platform that is used to create conversational interfaces for a variety of applications including mobile applications, web, Internet of Things devices, and telephone interactive voice response systems. Dialogflow is capable of conversing in both my text and voice.

Dialogflow revolves around the digital agent concept where by voice application is represented by a digital agent that responds to the users utterances.

There are two types of virtual agents available.

- Dialogflow CX which is well-suited to large or complex conversational pathways.
- Dialogflow ES which is best used for small or simple conversational pathways.

As of January 2022, Dialogflow ES could be utilized to develop conversational applications for free, where as Dialogflow CX is a paid service that is part of the Google Cloud Platform (GCP).

To demonstrate the core concepts of Conversational AI this chapter will focus on utilizing Dialogflow ES.

Additional Resources: https://cloud.google.com/dialogflow/es/docs/video

6.

DIAGLOGFLOW OVERVIEW

To access Dialogflow you first need to have a Google account. Then you can navigate to the Dialogflow console which can be found here.
https://dialogflow.cloud.google.com/

Once you are logged into the Dialogflow console (figure 1) you will be prompted to create a new agent. (figure 2)

This agent represents your application and can be thought of as the personification of a human agent that will converse with your end users. You will be required to provide the following attributes as you create the project.

- Agent Name
- Default Language
- Default Timezone.

The default language in time zone are used unless overridden by the client.

Note: You cannot change the default language after you create the agent.

As of Janauary 2022, Dialogflow supports 55 language variations for text chats. It is important to note that not all languages are fully supported for all client types. Since Dialogflow is capable of a number of different client features that are listed below it is important to consult the Official **Dialogflow Language table** to make sure that your languages is supported for the client features that you are looking to use.

Dialogflow Console visual Tour

(click each image to expand)

Figure 1: Login Screen

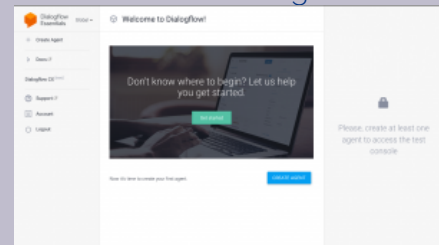Figure 2: First Console Login

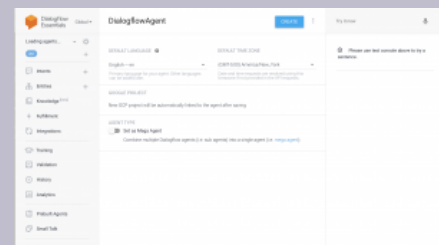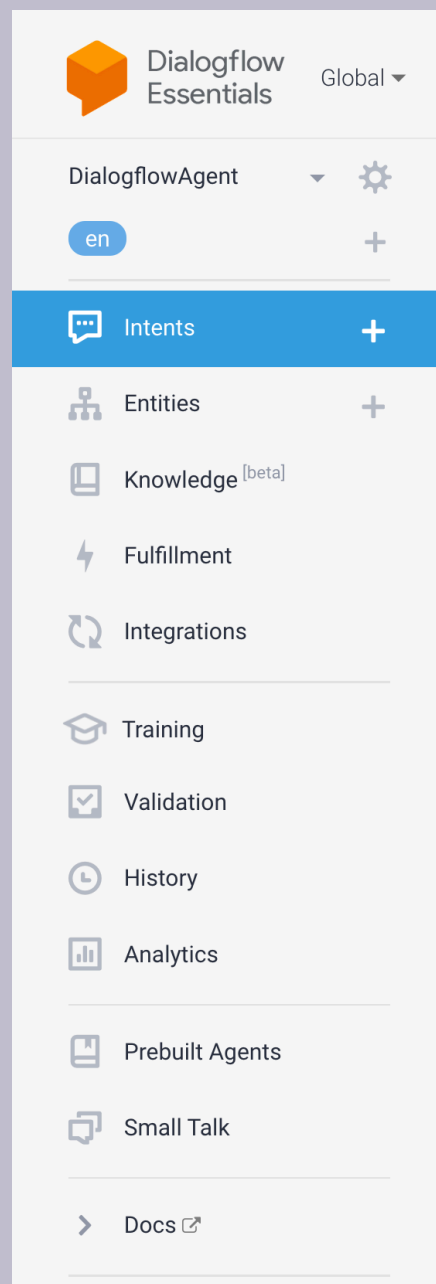Figure 3: Creating an Agent

Figure 4: Dialogflow Menu

Dialogflow Client Features

- Text(text-only chat)
- STT(speech-to-text, audio input, speech recognition)
- TTS(text-to-speech, audio output, speech synthesis)
- Phone ([Dialogflow Phone Gateway)](#)
- Knowledge ([Knowledge Connectors](#))
- Sentiment(Sentiment Analysis)
- SmTalk ([Built-in Small Talk](#))

Dialogflow Menu

The menu on the left side of the screen provides quick access to the majority of the Dialogflow features.

Intents – This is where you will manage the intents of your agent. And intent a way of handling the end-users intention during a conversation. It could be thought of as a reason that they are conversing with your agent or as something they would like to do during the conversation.

Entities – this is where you will manage the entities of your agent. Entities can be thought of as data variables that the user provides during conversation.

Knowledge – the knowledge feature is used to connect to an external data source such as a webpage and can automatically create things like a set of frequently asked questions or knowledge base.

Fulfillment – This is how Dialogflow connects to different back-end systems to extract information. Such as retrieving current weather conditions from a weather service API.

Integrations- This is where your Digalogflow agent connects with other platforms such as Slack, Facebook Messenger, Twitter, Telegram or even the telephone system.

The rest of this menu will be discussed in future chapters of this book.

This is where you can add appendices or other back matter.