

# Dataset

[https://drive.google.com/drive/folders/11sglwm6-cY7gjeqIZaMxL\\_MDKDMLdhym](https://drive.google.com/drive/folders/11sglwm6-cY7gjeqIZaMxL_MDKDMLdhym)  
([https://drive.google.com/drive/folders/11sglwm6-cY7gjeqIZaMxL\\_MDKDMLdhym](https://drive.google.com/drive/folders/11sglwm6-cY7gjeqIZaMxL_MDKDMLdhym)).

The dataset includes about 600 conversation records. Each record include a description of symptom and a single round conversation between patient and doctor. In the future, the dataset needs to be enlarged. For now, we can use this for our template-based chatbot.

(Thanks for Kavi found a paper which provides this conversation dataset) The group has reviewed this dataset together and we all agree that it can be adopted for now.

## Existing Work

### 1. NER

<https://cloud.ibm.com/catalog?category=ai#services> (<https://cloud.ibm.com/catalog?category=ai#services>).

The IBM provide a cloud service called "Annotator for Clinical Data" which can help us to do NER. Below is a result of entities recognition using one of our conversation data

This system is for demonstration purposes only and is not intended to process Personal Data. No Personal Data is to be entered into this system as it may not have the necessary controls in place to meet the requirement of the General Data Protection Regulation (EU) 2016/679.

### IBM Watson Annotator for Clinical Data I Try it out

The Annotator for Clinical Data can be used to surface insights in a variety of healthcare scenarios. Select a use case to see relevant insights extracted from unstructured text.

[Get Started on IBM Cloud](#)

#### Clinical Insights

Derives insights from clinical text, such as medications, diagnoses, procedures, labs, and abnormal findings.

#### COVID Medical Literature

Extracts information about the novel coronavirus from medical literature.

#### Preview

JSON Output

↺ Edit Text ↗

"patient: my 19 year old son was diagnosed with pneumonia 5 months ago and was given antibiotics but with minimal results. unfortunately he still suffers from the same symptoms but since he is in the military he refuses to go back to the doctor for

#### Entities

Show/hide all

- ☒ Attribute Values
- ☐ Concepts
- ☐ SymptomDiseaseInd 3
- ☐ ConceptValue 5
- ☐ NegatedSpan 1

### 2. Healthcare chatbot demo

[https://github.com/RasaHQ/medicare\\_locator](https://github.com/RasaHQ/medicare_locator) ([https://github.com/RasaHQ/medicare\\_locator](https://github.com/RasaHQ/medicare_locator)).

Medicare Locator is an open source starter pack based on Rasa framework. It is a rule-based chatbot framework.

The dataset used in this project comes from: <https://data.cms.gov/provider-data/> (<https://data.cms.gov/provider-data/>) which unfortunately not include doctor-patient conversation.

The work helps people search for a hospital, nursing home or home health agency in a US city. It also hands basic chitchat like greeting and saying goodbye.

Following is a demo of hospital-locator chatbot.

```

Bot loaded. Type a message and press enter (use '/stop' to exit):
Your input -> Hey!
I'm good, thanks!
127.0.0.1 - - [2021-04-11 15:28:24] "POST /webhooks/rest/webhook?stream=true&token= HTTP/1.1" 200 185 0.155521
Your input -> I'm in Sitka, I need a hospital
What is your current city?
127.0.0.1 - - [2021-04-11 15:28:36] "POST /webhooks/rest/webhook?stream=true&token= HTTP/1.1" 200 194 0.027149
Your input -> Sitka
Here is a hospital near you:
Buttons:
1: Mt Edgecumbe Hospital (/inform{"facility_id":"021314"})
127.0.0.1 - - [2021-04-11 15:28:44] "POST /webhooks/rest/webhook?stream=true&token= HTTP/1.1" 200 295 0.699193
Your input -> Bye
Have a good day.
127.0.0.1 - - [2021-04-11 15:28:51] "POST /webhooks/rest/webhook?stream=true&token= HTTP/1.1" 200 184 0.019094
Your input -> /stop

```

## Conclusion

Next week I will replace some diagnosis data on this framework and see the results, continue to study deep learning basic knowledge and do literature review. Here's my plan for the first version of the chatbot:

