# Case - Natural Language Processing

INTRODUCTION  
Assume that your organization is developing a product with a chat bot functionality, i.e. the end user talks to a device, the speech gets translated into text and now it is your task to map the given text to an assumed intent of the end user.

The data comes partitioned into training, validation and test, but whether you need to train a model or approach the task with pre-trained models is up to you. But the solution should no matter be evaluated on the test data set, that is, evaluated on whether for each text snippet in the test data set, your algorithm maps that text into the right intent.

DATA SET  
In this folder you will a json file named json\_data.json, which contains text-based questions/commands and the intent of these questions/commands. In the json file the first layer of key values will be separated into training, validation and test. Ensure you have a good understanding of the data set.

PERFORMANCE METRICS  
Your solution does not need to perform the task perfectly, but the performance needs to be reported in some statistics of your choosing, and the task will be used as an offset for further discussion.

PERSPECTIVE  
Consider the following situations and questions for discussion in the presentation

* How should we handle text input from a user which has an intent not currently present in our data?
* How do we handle text inputs which could have multiple intents?
* How will dialect or slang impact the solution?
* What do you expect your solutions weaker sides to be?

LIBRARIES  
You are free to choose any solution approach. Some libraries which may or may not be helpful are:

* Spacy: <https://github.com/explosion/spaCy>
* Haystack: <https://github.com/deepset-ai/haystack>
* NeuralCoref: <https://github.com/huggingface/neuralcoref>
* BERT: <https://huggingface.co/docs/transformers/model_doc/bert>
* USE: <https://spacy.io/universe/project/spacy-universal-sentence-encoder>
* Faiss: https://github.com/facebookresearch/faiss