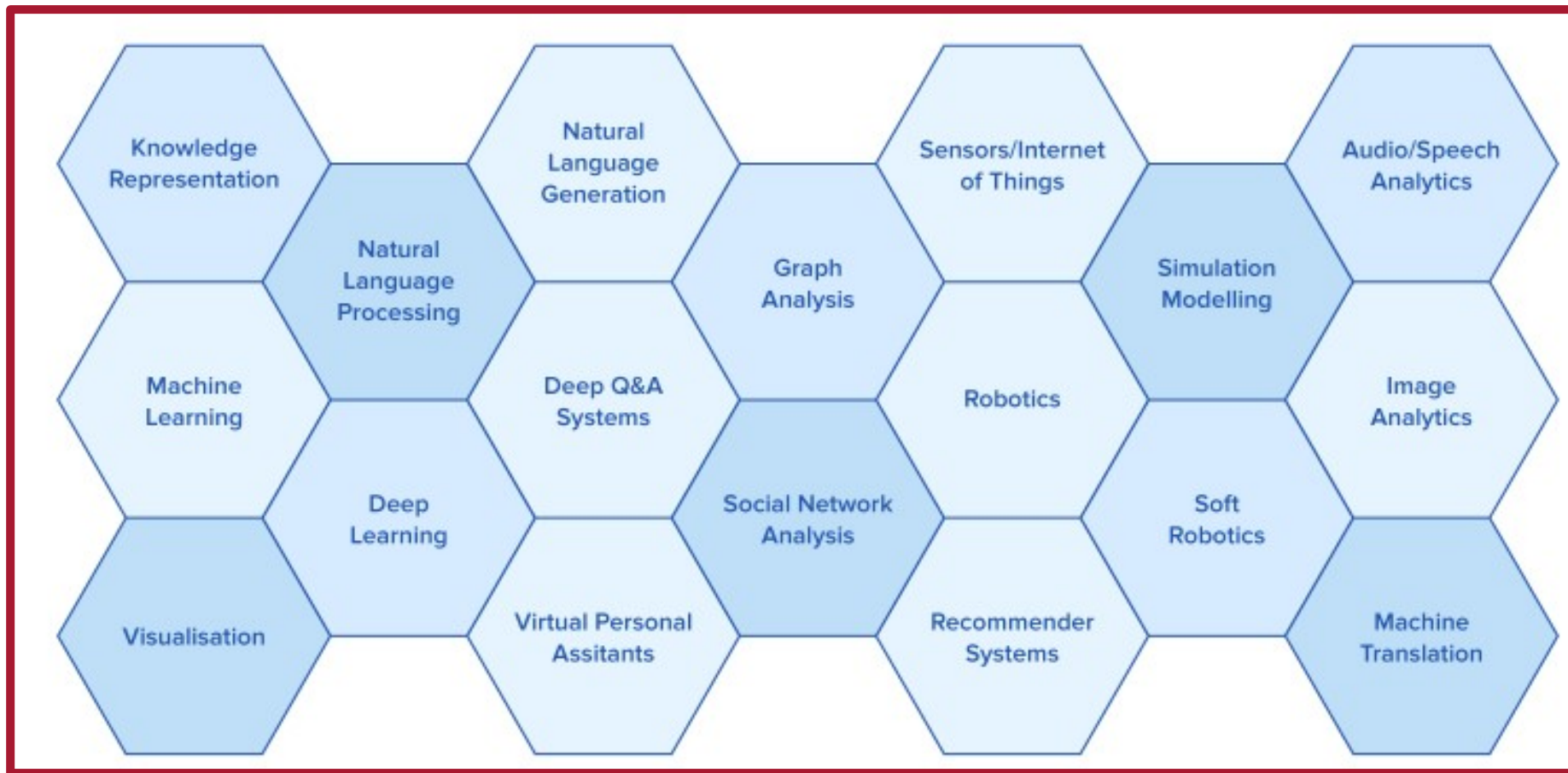


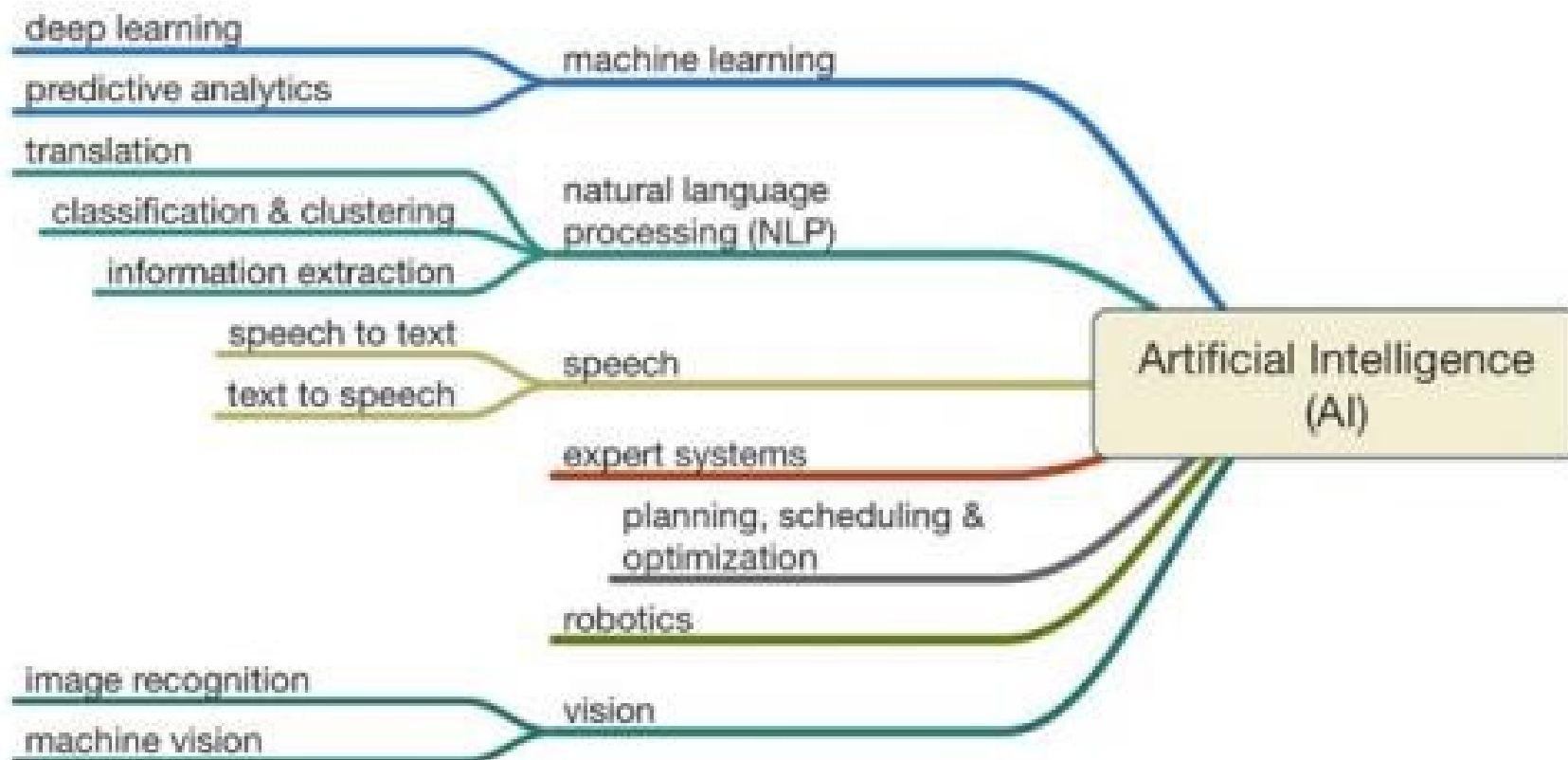
Deep Learning and Computer Vision: Overview

AI is driving business value worth several billions

AI Landscape (Source: PWC)



AI Landscape (Source: Oracle)



Walkthrough: Problem Descriptions

Problem#1: Quora Questions

Can we **distinguish** sincere questions from the insincere ones given a dataset that has a mix of both?

If so, **which specific words** make a question either sincere or insincere?

Notes on the dataset

Important: The dataset in the Kaggle site is a zip of 6 GB. This contains many other datasets besides the core Quora dataset. You can download just the core dataset from: <http://quora-insincere.s3-website-us-east-1.amazonaws.com/> (About 57 MB)

Warning: The dataset contains sensitive content such as sentences that have foul language, abusive, explicitly sexual content and so on. We ask you to use this dataset only for developing the technology to identify and eliminate such offensive content. We don't endorse or promote the content of the dataset beyond using it to develop the technology. We have not curated the data.

What is an insincere question? (Ref Kaggle)

An insincere question is defined as a question intended to make a statement rather than look for helpful answers.

Some characteristics:

Has a non-neutral tone

- Has an exaggerated tone to underscore a point about a group of people
- Is rhetorical and meant to imply a statement about a group of people

Is disparaging or inflammatory

- Suggests a discriminatory idea against a protected class of people, or seeks confirmation of a stereotype
- Makes disparaging attacks/insults against a specific person or group of people
- Based on an outlandish premise about a group of people
- Disparages against a characteristic that is not fixable and not measurable

Isn't grounded in reality

- Based on false information, or contains absurd assumptions

Uses sexual content (incest, bestiality, pedophilia) for shock value, and not to seek genuine answers

Samples

How did Quebec nationalists see their province as a nation in the 1960s?
Do you have an adopted dog, how would you encourage people to adopt and not shop?
Why does velocity affect time? Does velocity affect space geometry?
How did Otto von Guericke used the Magdeburg hemispheres?
Can I convert montra helicon D to a mountain bike by just changing the tyres?
Is Gaza slowly becoming Auschwitz, Dachau or Treblinka for Palestinians?
Why does Quora automatically ban conservative opinions when reported, but does not do the same for liberal views?
Is it crazy if I wash or wipe my groceries off? Germs are everywhere.
Is there such a thing as dressing moderately, and if so, how is that different than dressing modestly?

Has the United States become the largest dictatorship in the world?
Which babies are more sweeter to their parents? Dark skin babies or light skin babies?
If blacks support school choice and mandatory sentencing for criminals why don't they vote Republican?

What is expected?

Basic level: Build a model that can classify sincere versus insincere questions as per the dataset and definitions given and achieve a high accuracy. You should measure the F1 score and report it as a metric.

Advanced Level: Develop an attention based mechanism that allows determining the relative weights of words in a text that contribute to the classification decision. You should visualize these and demonstrate.

General approach to solving NLP problems

Data analysis (Use packages like Pandas)

Text preprocessing: Tokenization, stop word removal, stemming, lemmatization, TF-IDF etc

Text representation: Word2Vec, Glove, BERT, FastText, etc

Architecture design and development: TensorFlow, Keras, PyTorch, etc

Application development: Browser, Mobile, Desktop app

Technologies


Data Analysis: Pandas, Jupyter Notebook

Text preprocessing libraries such as NLTK

Text representation

LSTM, Recurrent Neural Networks

Attention models

Visualization Techniques
 Mindtree