

Natural Language Processing

Session 1: Motivation and Course Overview

Sascha Göbel
goebel@hertie-school.org

September 10, 2025

What is Natural Language Processing (NLP)?

- A field of Artificial Intelligence
- Make machines process, understand, and put language to use

What is Natural Language Processing (NLP)?

- A field of Artificial Intelligence
- Make machines process, understand, and put language to use

Natural Language

- Primary medium of communication between humans
- Complex, hierarchical structure
- Diversity and Ambiguity
- In its raw textual state unstructured data to a machine

What is Natural Language Processing (NLP)?

- A field of Artificial Intelligence
- Make machines process, understand, and put language to use

Natural Language

- Primary medium of communication between humans
- Complex, hierarchical structure
- Diversity and Ambiguity
- In its raw textual state unstructured data to a machine

Processing

- Turning unstructured data into actionable knowledge
- Using Algorithmic tools in several stages
- Fast developing field

Why do we want machines to understand human language?

Why do we want machines to understand human language?

Search engines



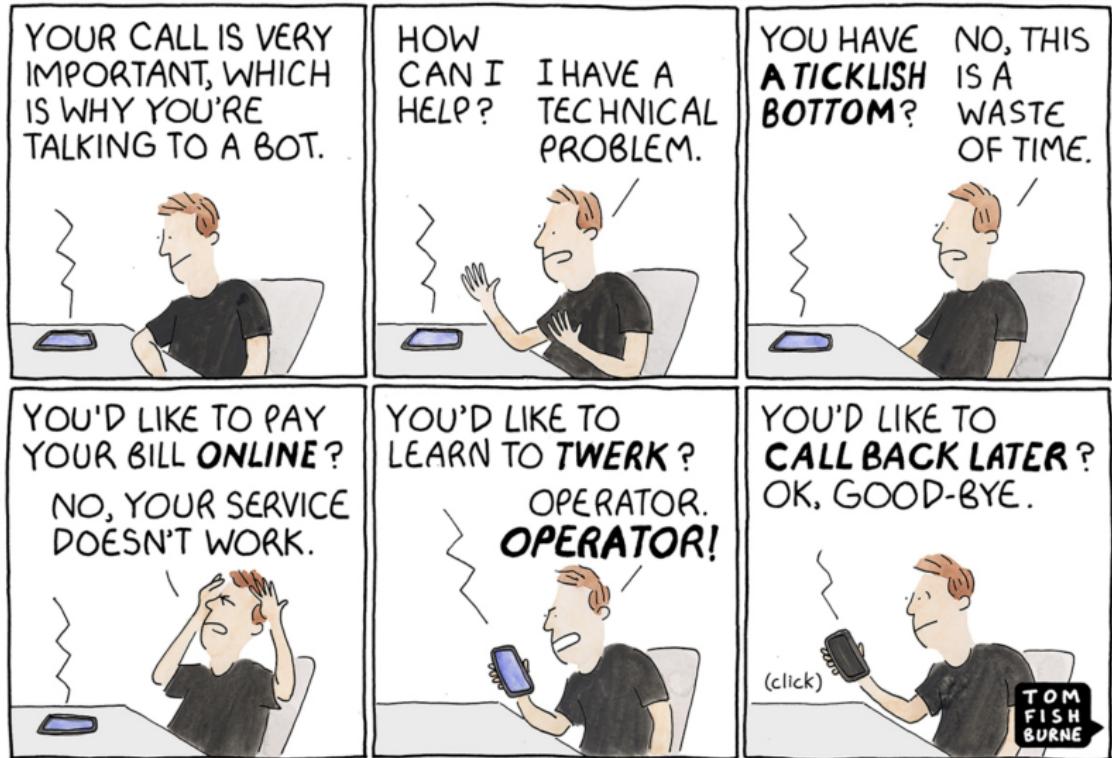
Human resources



Voice-activated assistants



Customer support



© marketoonist.com

Machine translation

The screenshot shows a machine translation interface. At the top, there are two buttons: "Translate text" (33 languages) and "Translate files" (.pdf, .docx, .pptx). Below these, the source language is set to "Czech" and the target language is "English". A glossary button is also present.

The main area displays a conversation pair:

- Czech input: "Jak mohu obejít chatbota a mluvit se skutečným člověkem..."
- English output: "How can I bypass the chatbot and talk to a real person..."

At the bottom of the interface, there are several small icons: a microphone, a double arrow, a left arrow, a right arrow, a reply icon, a like icon, a dislike icon, a share icon, and a more options icon.

Autocomplete

```
1  from __future__ import print_function
2  import argparse
3  import torch
4  import torch.nn as nn
5  import torch.optim as optim
6  import numpy as np
7  import matplotlib
8  matplotlib.use('Agg')
9  import matplotlib.pyplot as plt
10
11 class Sequence(nn.Module):
12     def __init__(self):
13         super(Sequence, self).__init__()
14         self.lstm1 = nn.LSTMCell(1, 51)
15         self.lstm2 = nn.LSTMCell(51, 51)
16         self.linear = nn.Linear(51, 1)
17
18     def forward(self, input, future = 0):
```

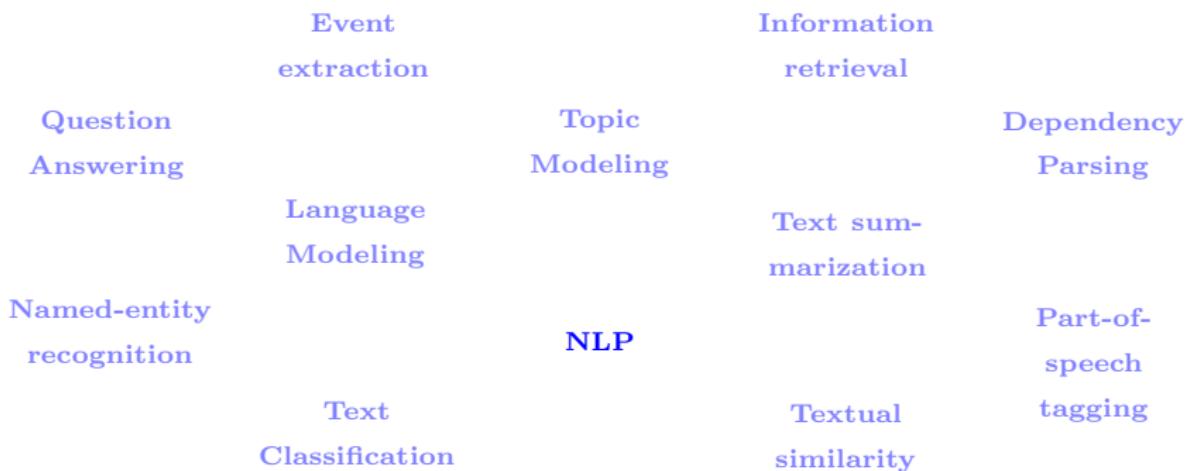
Content moderation

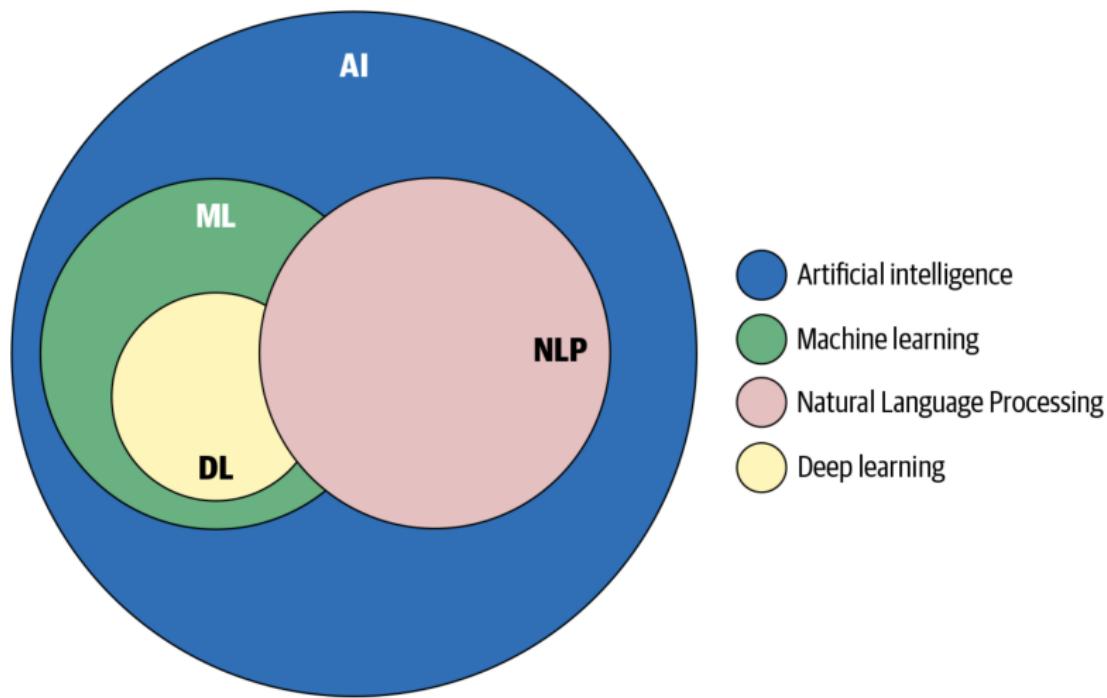


And more ...

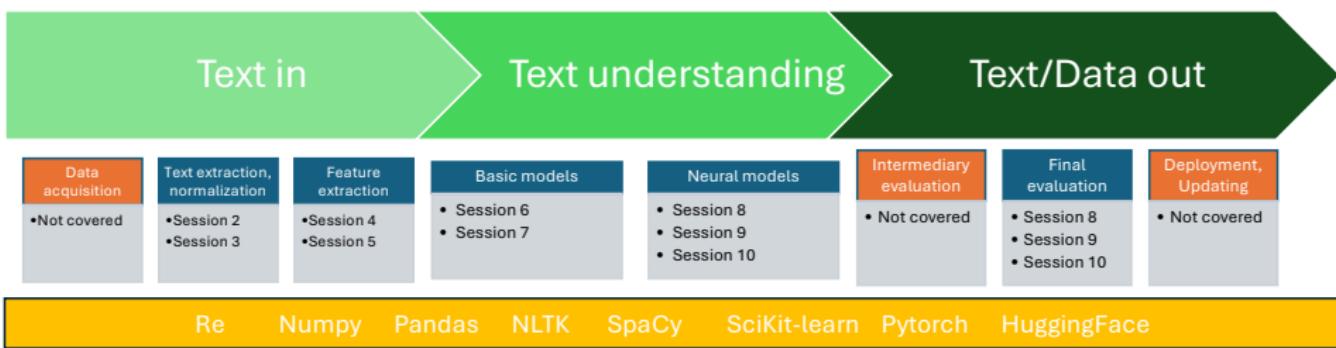
- Email spam filters
- Grammar checking
- Virtual meeting transcription
- Product recommendations
- Symptom checkers
- Financial risk analysis
- Dating apps
- In-car voice commands
- Research
- Policy understanding and policy-making
- ...

Some common NLP tasks

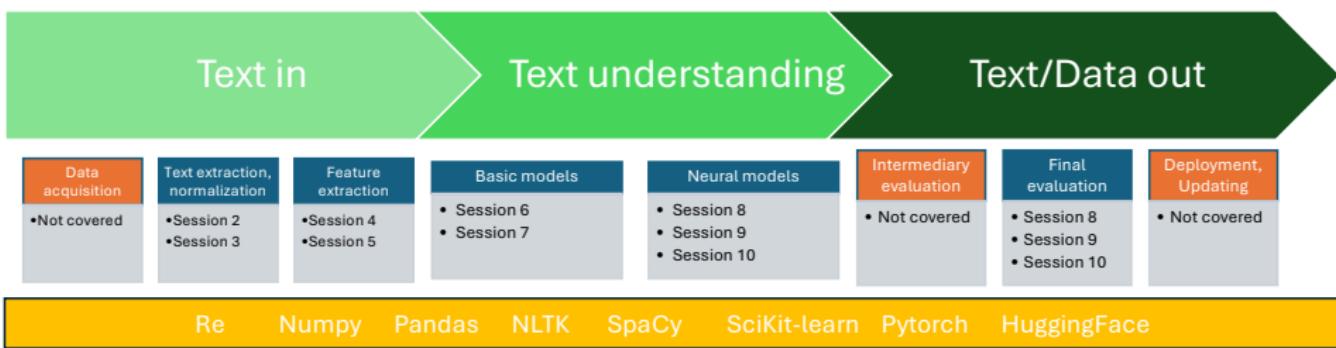




NLP Pipeline



NLP Pipeline



Conceptual and applied perspective

- Build an understanding of NLP (concepts, concepts, concepts)
- Actionable code (recipes for direct implementation)
- Starting point for deeper engagement (build your own pipelines)

Rules of the game

→ Sessions and literature

- Required literature is actually required, please come prepared!
- Please ask questions!
- But note that we cannot address all questions in all detail

→ Attendance

- If you cannot attend, send your sick note to the exam office
- If you miss more than two classes, you have to do a make-up assignment
- If you miss more than four classes, you lose access to course materials

→ Assignments

- Participation in class (10%)
- 5 quizzes (45%)
- 1 research note (45%), start early!

→ Office hour

- By appointment only
- Write an email with a suggested time