

# Course Information

## Contents

- Learning outcomes
- What is Natural Language Processing (NLP)?
- Example: Lexical ambiguity
- Example: Referential ambiguity
- Course roadmap



# DSCI 575 Advanced Machine Learning

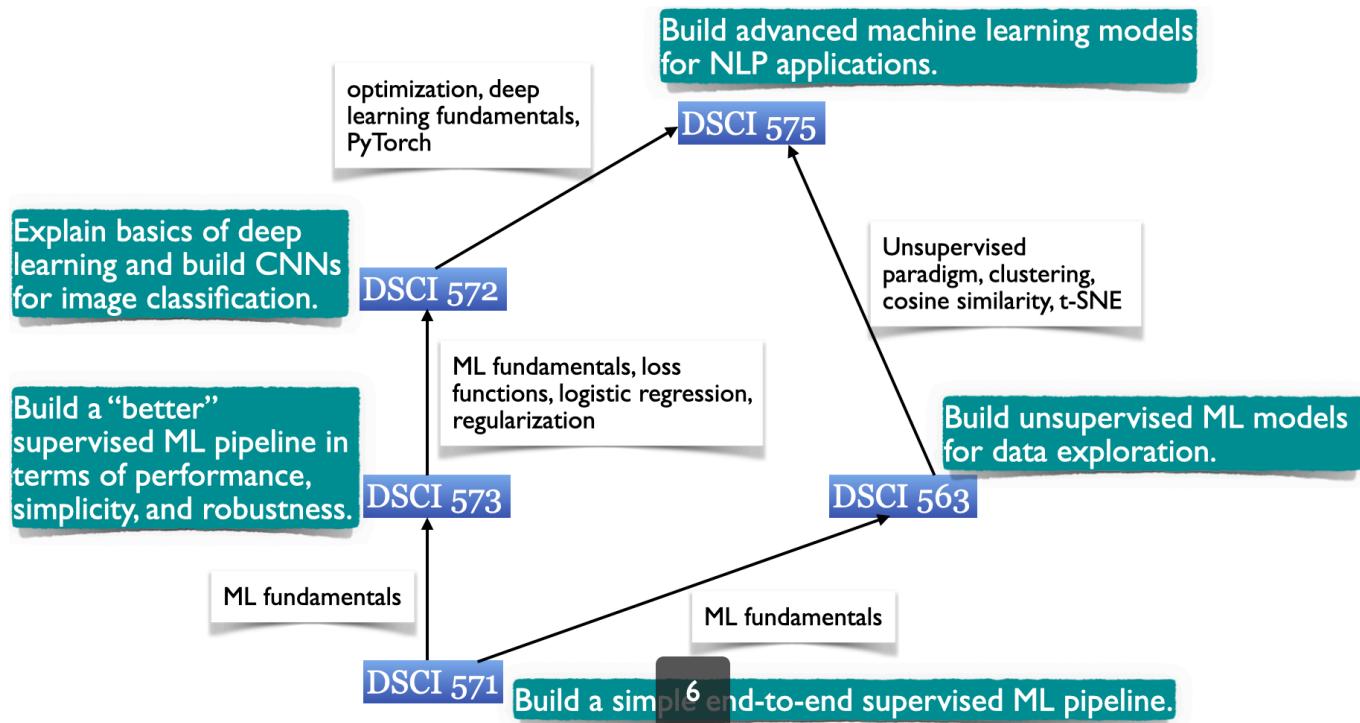
UBC Master of Data Science program, 2024-25

## Learning outcomes

From this lecture, students are expected to be able to:

- Explain relevance of NLP and some goals of natural language processing (NLP).
- Explain difficulties associated with NLP.
- Name a few applications of natural language processing.

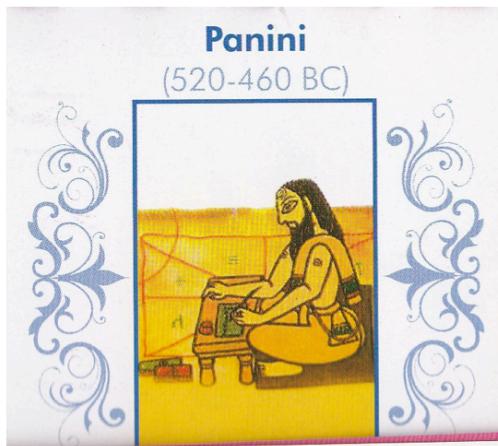
# Map of ML courses in MDS



## What is Natural Language Processing (NLP)?

- What should a search engine return when asked the following question?

Who is Panini?



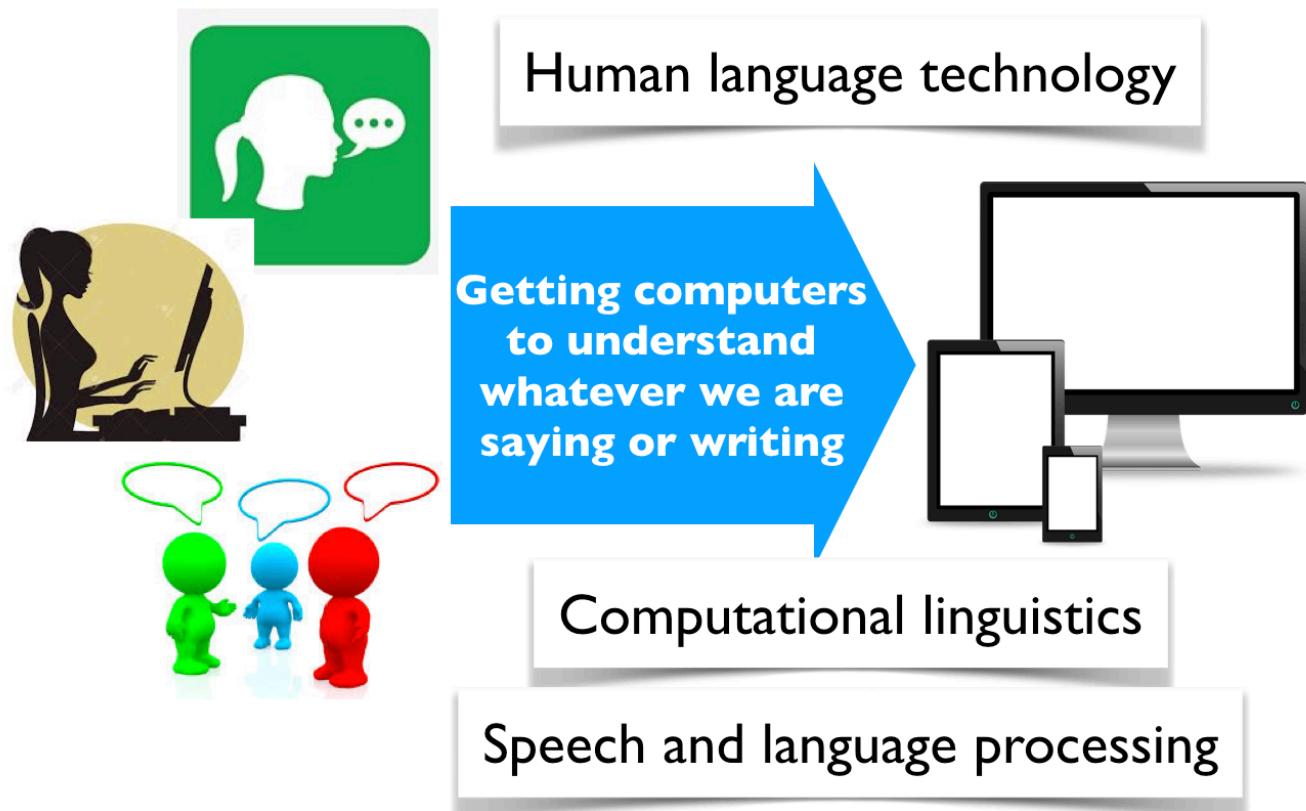
# What is Natural Language Processing (NLP)?

How often do you search everyday?

Google processes > **4.5 billion** queries per day!



# What is Natural Language Processing (NLP)?



## Everyday NLP applications

**Voice assistants**

amazon Google Microsoft SAMSUNG

amazon alexa Google Assistant Cortana Siri Bixby

**Smart compose**

Dinner next week

Evan Brown, Maalika Patel

Dinner next week

Hey Evan and Maalika,

Haven't seen you in a while! Wanna grab dinner next week at my place? My address is 34 Smith Street, Somers, CT 06071.

Looking forward to seeing you!

**Translation**

Text Documents

ENGLISH - DETECTED ENGLISH SPANISH FRENCH GERMAN ENGLISH SPANISH

I find it hard to talk about myself because I am work in progress.

Es fällt mir schwer, über mich selbst zu sprechen, weil ich in Arbeit bin.

# NLP in news

Often you'll see NLP in news. Some examples:

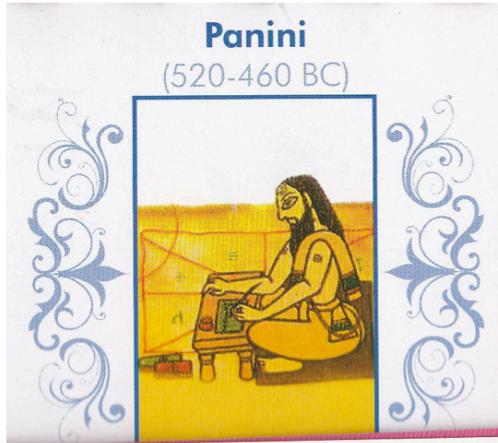
- [How suicide prevention is getting a boost from artificial intelligence](#)
- [Meet GPT-3. It Has Learned to Code \(and Blog and Argue\).](#)
- [How Do You Know a Human Wrote This?](#)
- ...

## Why is NLP hard?

- Language is complex and subtle.
- Language is ambiguous at different levels.
- Language understanding involves common-sense knowledge and real-world reasoning.
- All the problems related to representation and reasoning in artificial intelligence arise in this domain.

## Example: Lexical ambiguity

### Who is Panini?



# Example: Referential ambiguity

If the **baby** does not thrive on raw **milk**, boil **it**.



**it = ?**



## Ambiguous news headlines

PROSTITUTES APPEAL TO POPE

- **appeal to** means make a serious or urgent request or be attractive or interesting?

KICKING BABY CONSIDERED TO BE HEALTHY

- **kicking** is used as an adjective or a verb?

MILK DRINKERS ARE TURNING TO POWDER

- **turning** means becoming or take up?

# High-level goals of this course

- Learn and apply new ML algorithms and methods with the theme of NLP applications.
- Prepare you a bit for employment in the NLP area.
- Have fun!



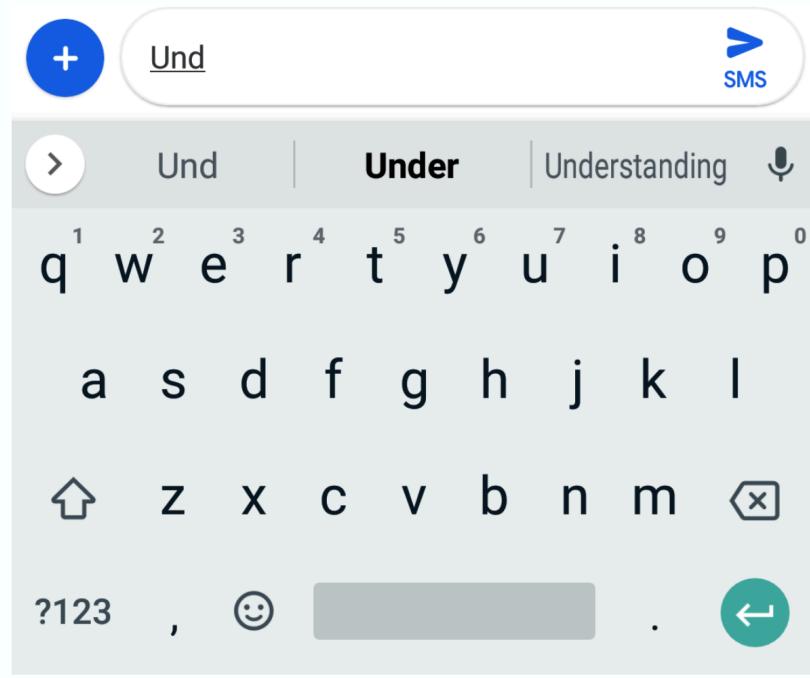
## Course roadmap

### Week 1

- Markov models

# Everyday example of Markov chains

Autocomplete in text messaging



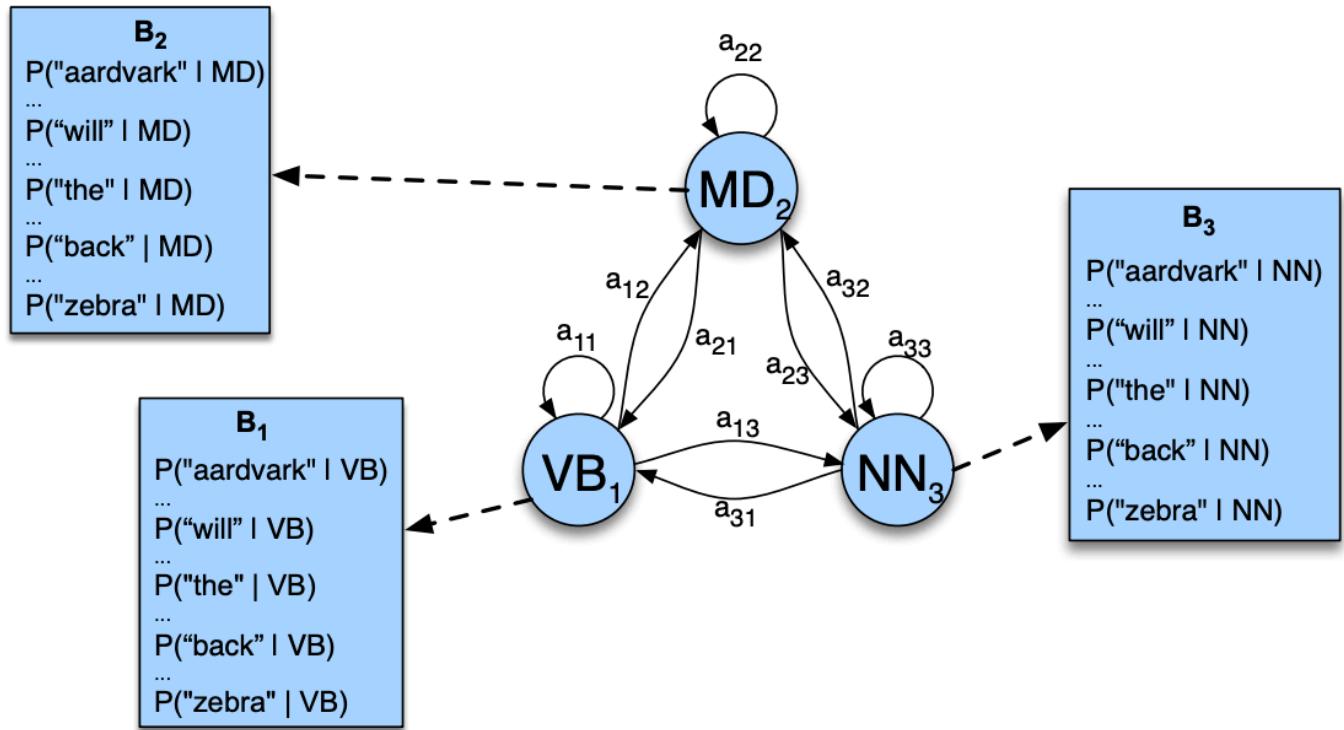
# Applications of Markov models

## 18 Applications

- 18.1 Physics
- 18.2 Chemistry
- 18.3 Testing
- 18.4 Speech recognition
- 18.5 Information and computer science
- 18.6 Queueing theory
- 18.7 Internet applications
- 18.8 Statistics
- 18.9 Economics and finance
- 18.10 Social sciences
- 18.11 Mathematical biology
- 18.12 Genetics
- 18.13 Games
- 18.14 Music
- 18.15 Baseball
- 18.16 Markov text generators
- 18.17 Bioinformatics

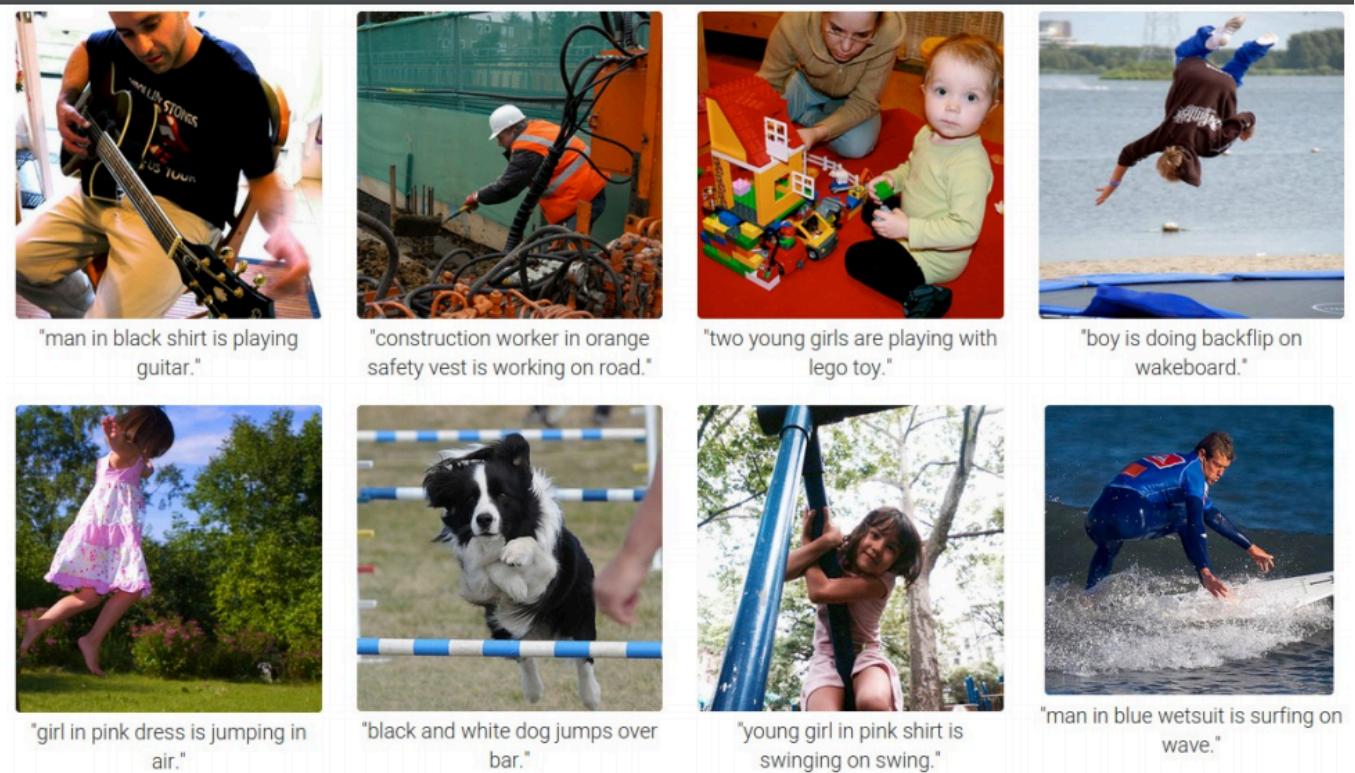
## Week 2

- Hidden Markov models



## Week 3 and Week 4

- Introduction to Recurrent Neural Networks (RNNs)
- Introduction to self attention and transformers
- Applications of transformers



[Source](#)

## ASIDE: [Neural Storyteller](#)



We were barely able to catch the breeze at the beach , and it felt as if someone stepped out of my mind . She was in love with him for the first time in months , so she had no intention of escaping . The sun had risen from the ocean , making her feel more alive than normal . She 's beautiful , but the truth is that I do n't know what to do ...

## Source

# Tools we will be using in this class

- [sklearn](#)
- [Gensim](#)
- [spaCy](#)
- [nltk](#)
- [PyTorch](#)
- [Kaggle Notebooks](#) or [Google Colab](#)

Make sure to set up the environment using [the course environment file](#).

That's all about course information!