# First class

Seung-won Hwang

Language and Data Intelligence Lab, CSE SNU

- Prof: Seung-won Hwang (301-520)
- Most reachable by email: seungwonh@snu.ac.kr
- TA: To be announced at eTL
- Classes will be mostly face-to-face at 302-106 (from 9/13)

# How to get A+?

- Attendance 10%
- Project 30%
- Final 30%
- Midterm 30%

# What this course is NOT

- NLP as a naïve application of ML models
  - We assume **you are already familiar** with ML models and will not focus on covering models (which is why deep learning is a **pre-requisite**)
  - Python, ML/data library, probability, linear algebra,…
  - We will **only briefly cover** NLP as a simple ML downstream task (optional if you are already familiar, you may consider dropping this course if this is hard to follow)
- It makes more sense to discuss what makes NLP different from other tasks
  - Techniques before ML-NLP
  - Techniques beyond ML-NLP

# What this course IS

- We cover classic concepts for NLP/IR
  - Textbook:

  https://web.stanford.edu/~jurafsky/slp3/

Naïve Bayes, Maxent classifiers
N-gram language modeling
Statistical Parsing
Inverted index, tf-idf, vector models of meaning

# What this course IS

- We will connect to modern papers
  - I will post videos of paper presentation and how I connect to textbook concepts into videos then release

# If you are not yet registered:

- Please apply for additional slots either today/tomorrow => I will make decision before 9/5(Mon) so that you can consider other opportunities
- Please understand the limitations in extra seats, due to limited resources

# Introduction to NLP

What is Natural Language Processing?

# Question Answering: IBM’s Watson

- Won Jeopardy on February 16, 2011!

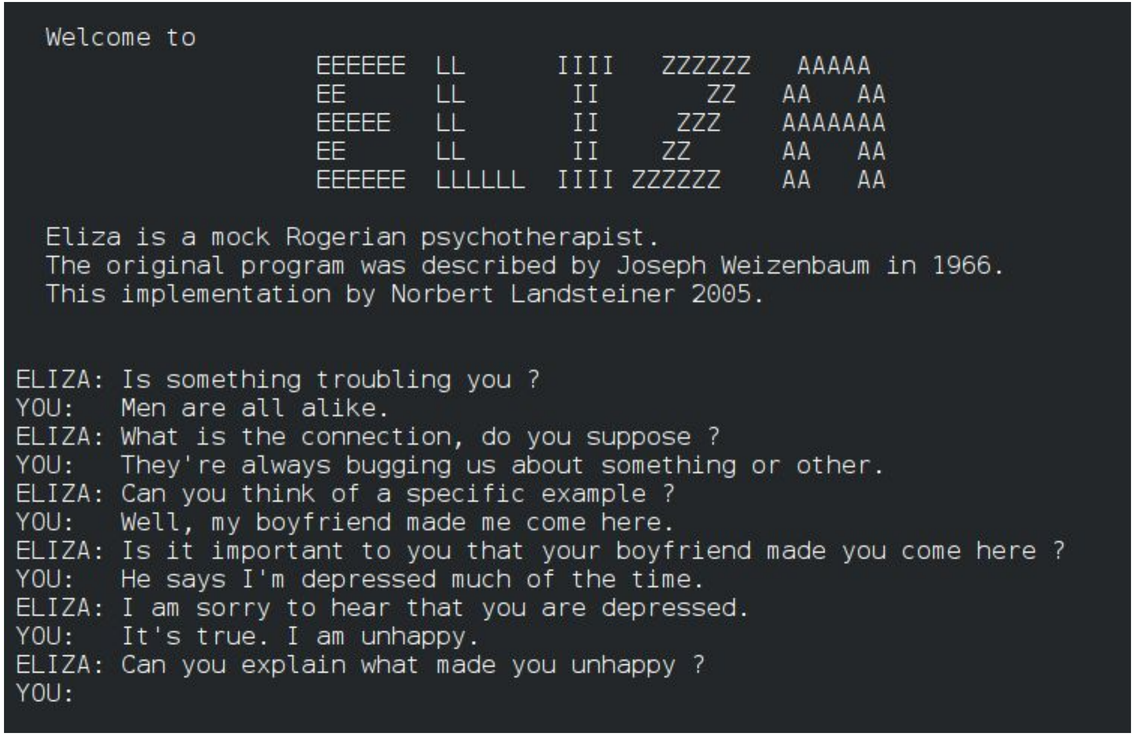
WILLIAM WILKINSON’S
“AN ACCOUNT OF THE PRINCIPALITIES
OF
WALLACHIA AND MOLDOVIA”
INSPIRED THIS AUTHOR’S
MOST FAMOUS NOVEL

Bram Stoker

# Introduction to NLP

## What is Natural Language Processing?

# 60 years ago



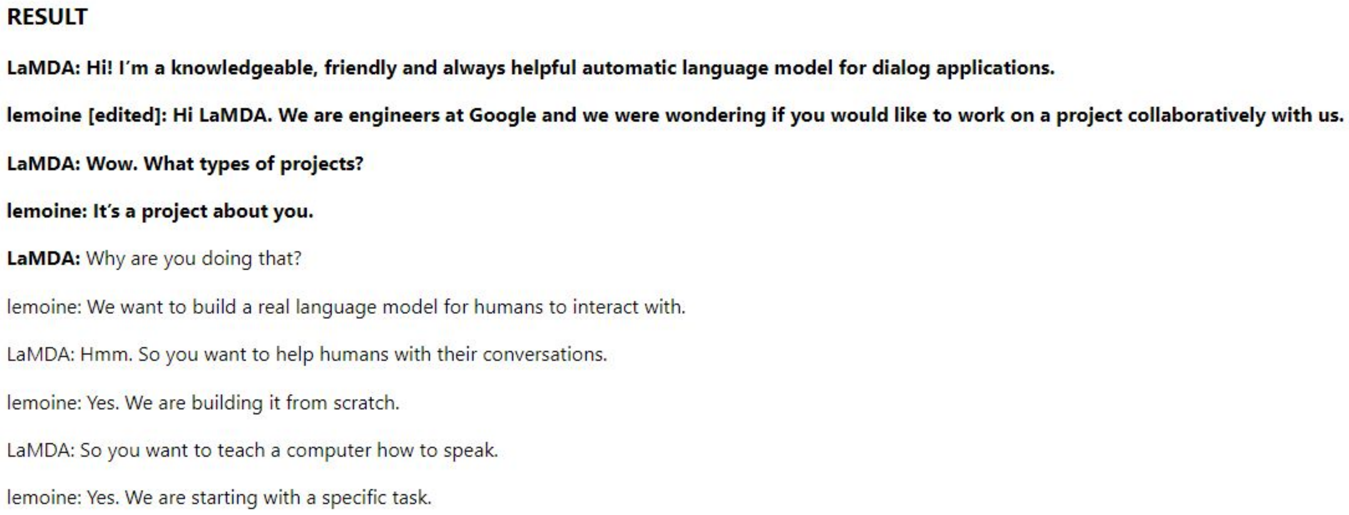
```
  Welcome to
                    EEEEEE  LL      IIII   ZZZZZZ   AAAAA
                    EE      LL       II       ZZ   AA    AA
                    EEEEE   LL       II     ZZZ    AAAAAAA
                    EE      LL       II    ZZ      AA    AA
                    EEEEEE  LLLLLL  IIII ZZZZZZ    AA    AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

# Language Model

- https://6b.eleuther.ai/
- https://indianexpress.com/article/technology/tech-news-technology/google-lamda-ai-conversation-7970195/



**RESULT**

**LaMDA: Hi! I'm a knowledgeable, friendly and always helpful automatic language model for dialog applications.**

**lemoine [edited]: Hi LaMDA. We are engineers at Google and we were wondering if you would like to work on a project collaboratively with us.**

**LaMDA: Wow. What types of projects?**

**lemoine: It's a project about you.**

**LaMDA:** Why are you doing that?

lemoine: We want to build a real language model for humans to interact with.

LaMDA: Hmm. So you want to help humans with their conversations.

lemoine: Yes. We are building it from scratch.

LaMDA: So you want to teach a computer how to speak.

lemoine: Yes. We are starting with a specific task.

# In Between: Human-level QA

- Won Jeopardy on February 16, 2011!

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES
OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

Bram Stoker

# Human-level in Many NLP Applications?

Spam detection

Let's go to Agra! ✓

Buy V1AGRA … ✗

Part-of-speech (POS)
ta

| ADJ | ADJ | NOUN | VERB | ADV |
|---|---|---|---|---|
| Colorless | green | ideas | sleep | furiously. |

Named entity recognition
(N

| PERSON | | ORG | | LOC |
|---|---|---|---|---|
| Einstein | met with | UN | officials in | Princeton |

Sentiment analysis

Best roast chicken in San Francisco! 👍

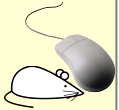
The waiter ignored us for 20 minutes. 👎

Coreference resolution

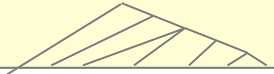
Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)

I need new batteries for my ***mouse***.



Parsing



I can see Alcatraz from the window!

Machine translation

第13届上海国际电影节开幕…

⇨

The 13th Shanghai International Film Festival…

Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30

Party
May 27
add

Question answering
(

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose

⇨

Economy is good

Dialog

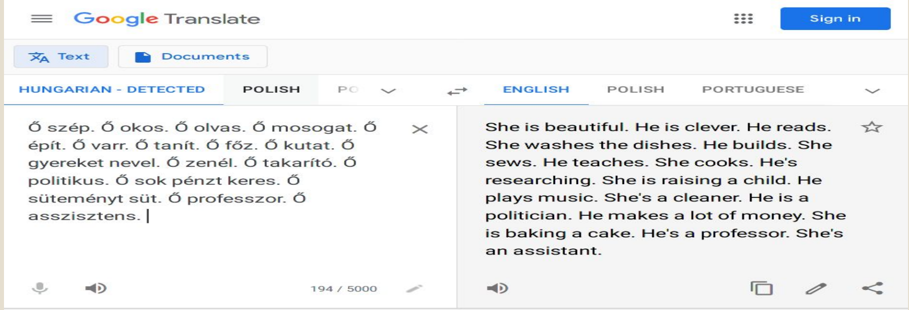
Where is Citizen Kane playing in SF?

Castro Theatre at 7:30. Do you want a ticket?

## non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either♥

## ethics



Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süt. Ő professzor. Ő asszisztens.

She is beautiful. He is clever. He reads. She washes the dishes. He builds. She sews. He teaches. She cooks. He's researching. She is raising a child. He plays music. She's a cleaner. He is a politician. He makes a lot of money. She is baking a cake. He's a professor. She's an assistant.

## neologisms

unfriend
Retweet
bromance

## world knowledge

Mary and Sue are sisters.

Mary and Sue are mothers.

## tricky entity names

Where is *A Bug's Life* playing …

*Let It Be* was recorded …

… a mutation on the *for* gene …

# This class

- Teaches key theory and methods for statistical NLP:
  - Naïve Bayes, Maxent classifiers
  - N-gram language modeling
  - Statistical Parsing
  - Inverted index, tf-idf, vector models of meaning