CS 6301: Special Topics in Computer Science - Deep Learning for NLP

Introduction

Xinya Du

Spring 2023

CS6301

- People:
  - Instructor: Xinya Du
    - Call me Xinya (Pronunciation: Shinn-ya)
    - or Dr. Du, or Prof. Du
  - TAs: TBD
  - Graders: TBD

Exciting age for NLP

- Fast progress in deep learning-based NLP systems

Generating icebreaker questions (ChatGPT)

- "If you could be any fictional character, who would you choose and why?"
- "What's a goal or dream you have that you've yet to accomplish?"
- "If you could have dinner with any historical figure, who would it be and why?"

Summarization/Generation (ChatGPT)

- ← **Tweet**

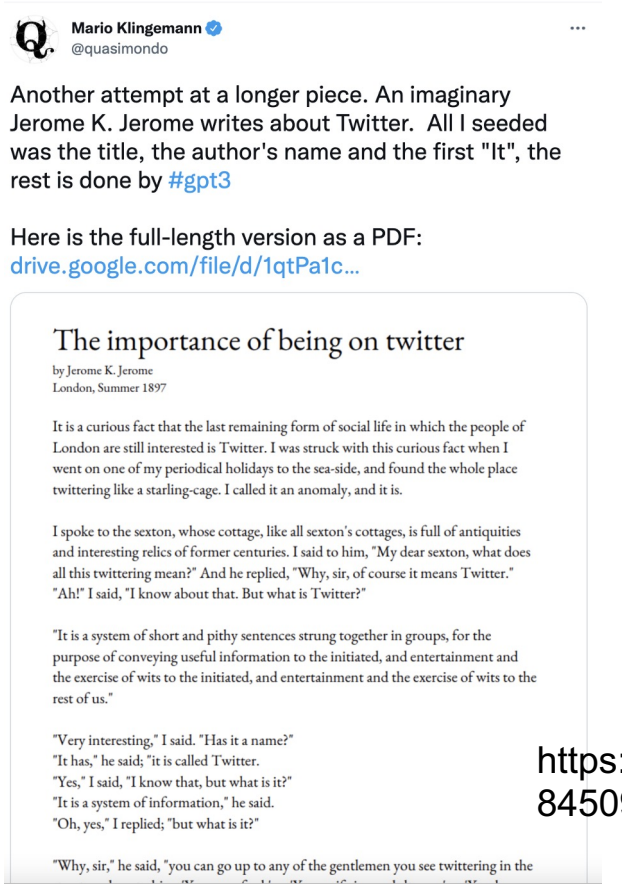
Mark Dredze
@mdredze

Idea for a killer app:
ChatGPT: Modify my ACL paper, which is currently over the limit by 4 lines, so that it fits into the 8 page limit.

3:13 PM · Jan 18, 2023 · **37.4K** Views

11 Retweets **1** Quote Tweet **240** Likes

Writing Entire Articles (GPT-3)



Mario Klingemann
@quasimondo

Another attempt at a longer piece. An imaginary Jerome K. Jerome writes about Twitter. All I seeded was the title, the author's name and the first "It", the rest is done by #gpt3

Here is the full-length version as a PDF:
drive.google.com/file/d/1qtPa1c...

The importance of being on twitter

by Jerome K. Jerome
London, Summer 1897

It is a curious fact that the last remaining form of social life in which the people of London are still interested is Twitter. I was struck with this curious fact when I went on one of my periodical holidays to the sea-side, and found the whole place twittering like a starling-cage. I called it an anomaly, and it is.

I spoke to the sexton, whose cottage, like all sexton's cottages, is full of antiquities and interesting relics of former centuries. I said to him, "My dear sexton, what does all this twittering mean?" And he replied, "Why, sir, of course it means Twitter."
"Ah!" I said, "I know about that. But what is Twitter?"

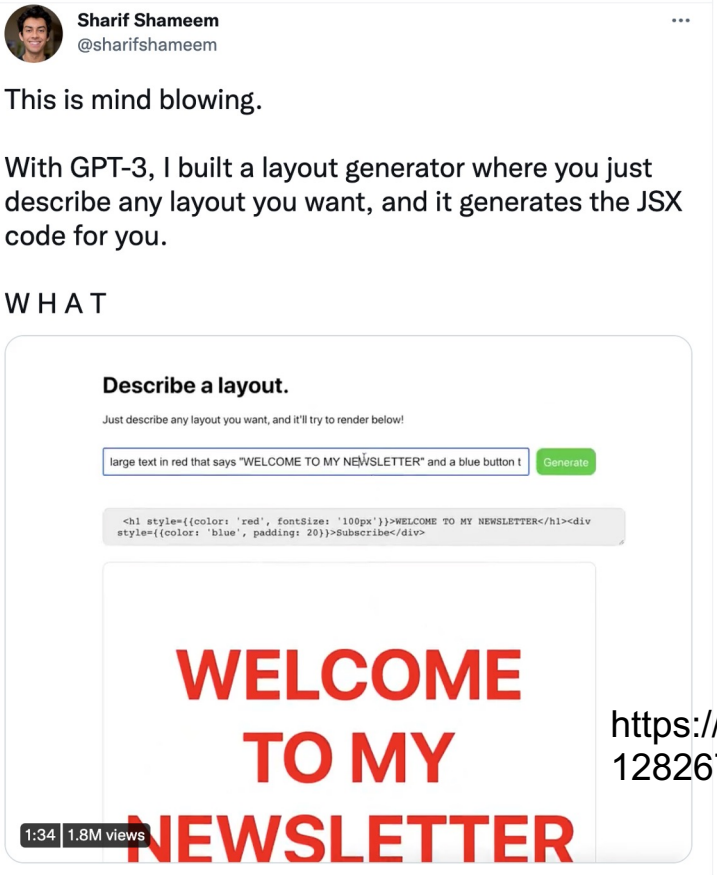
"It is a system of short and pithy sentences strung together in groups, for the purpose of conveying useful information to the initiated, and entertainment and the exercise of wits to the initiated, and entertainment and the exercise of wits to the rest of us."

"Very interesting," I said. "Has it a name?"
"It has," he said; "it is called Twitter.
"Yes," I said, "I know that, but what is it?"
"It is a system of information," he said.
"Oh, yes," I replied; "but what is it?"

"Why, sir," he said, "you can go up to any of the gentlemen you see twittering in the

https://twitter.com/quasimondo/status/1284509525500989445?ref_src=twsrc

Even for Writing Code... (GPT-3)



https://twitter.com/sharifshameem/status/1282676454690451457?ref_src=twsrc

Reasoning (GPT-3)

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are 16 / 2 = 8 golf balls. Half of the golf balls are blue. So there are 8 / 2 = 4 blue golf balls. The answer is 4. ✓

Reasoning (GPT-3)

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: ***Let's think step by step.***

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

Description → Images (multimodal)

- DALL-E Generation

"Photo of hip hop cow in a denim jacket recording a hit single in the studio"

Let’s go back to natural language

- deep learning a little later ...

What is the goal of human language?

What is the goal of human language?

To share thoughts with conspecifics

- to **communicate** with others
- to store and pass **knowledge**

What is the goal of human language?

To share thoughts with conspecifics

- to **communicate** with others
- to store and pass **knowledge**

To think more complex thoughts

- **reasoning**
- **interpretation**

What is the goal of human language?

- To store and pass **knowledge (Extraction to construct Graph)**

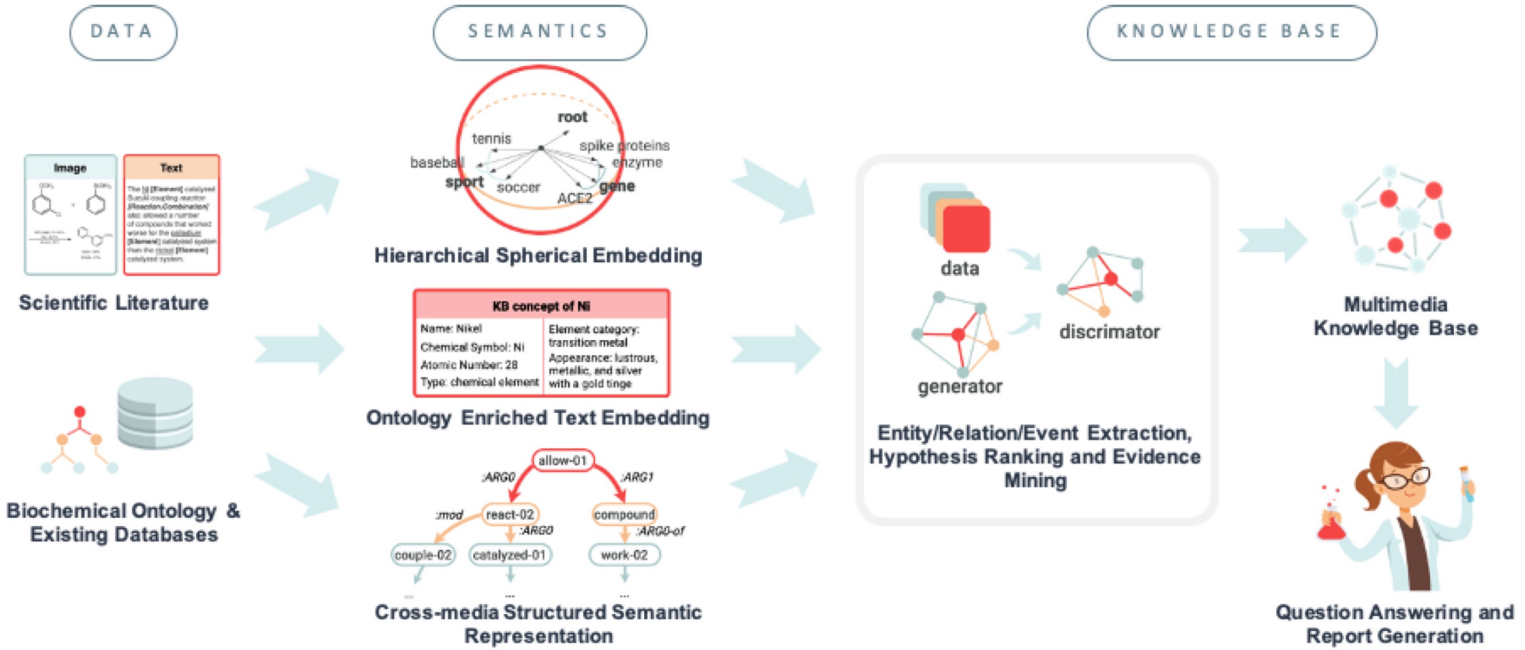


Figure 2: COVID-KG Overview: From Data to Semantics to Knowledge

What is the goal of human language?

- To store **knowledge (In the form of model “parameters”)**

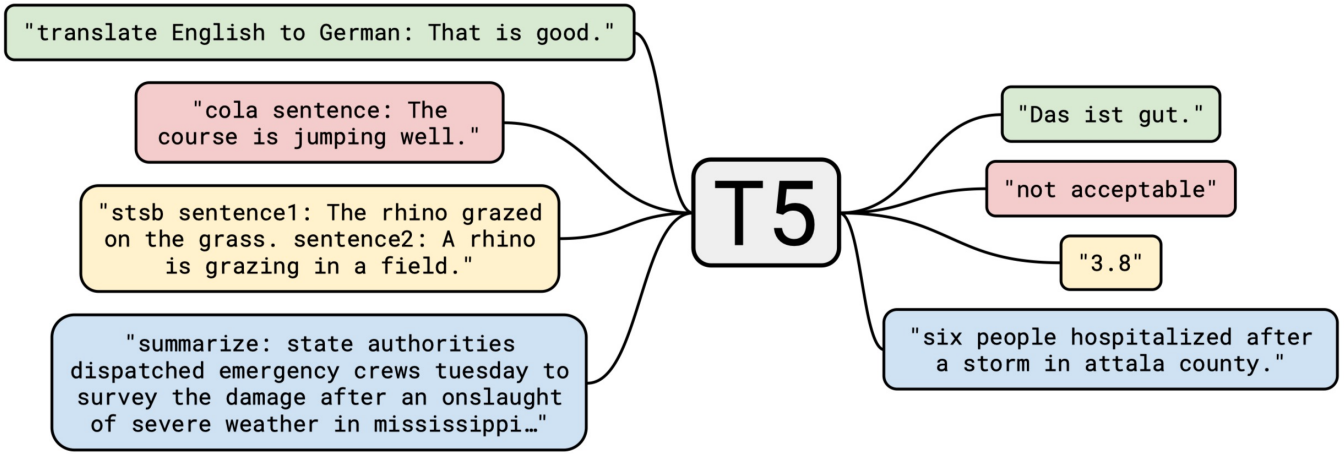


Figure 1: A diagram of our text-to-text framework. Every task we consider—including translation, question answering, and classification—is cast as feeding our model text as input and training it to generate some target text. This allows us to use the

What is the goal of human language?

To think more complex thoughts

- **reasoning (commonsense)**

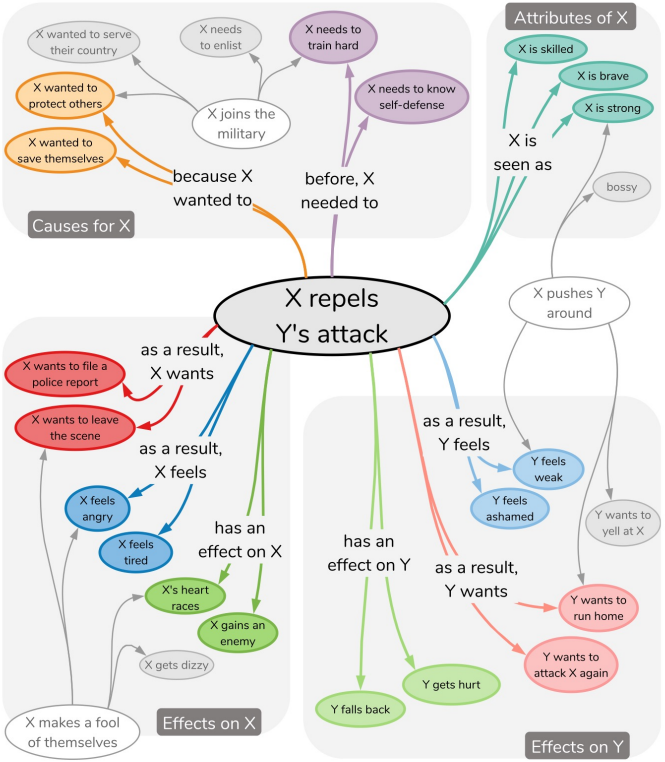
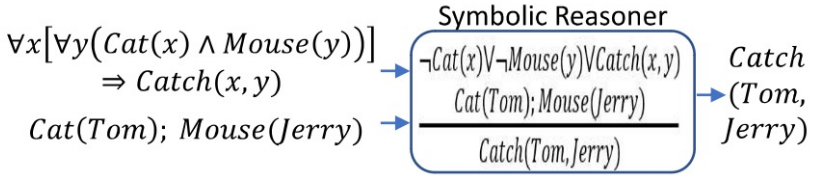


Figure 1: A tiny subset of ATOMIC, an atlas of machine commonsense for everyday events, causes, and effects.

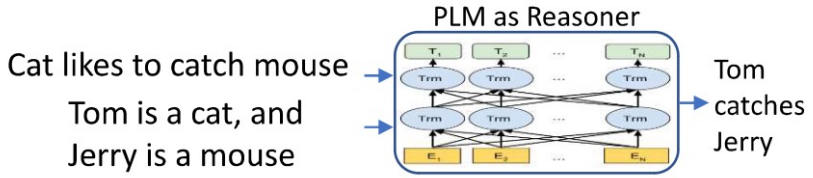
What is the goal of human language?

To think more complex thoughts

- **reasoning (deductive)**
- **interpretation**



(a) Formal language as knowledge representation and symbolic reasoner



(b) Natural language as knowledge representation and PLM as reasoner

Figure 1: Comparison between the previous paradigm which uses formal representation and symbolic reasoner, and the new paradigm which uses natural language as knowledge representation and PLM as reasoner.

What is the goal of human language?

To think more complex thoughts

- **reasoning (abductive)**
- **interpretation**

Short fact 1	Short fact 2	Short fact 3	Rule
The Venus flytrap is a **carnivorous plant** native to subtropical wetlands on the East Coast of the United States in North Carolina and South Carolina. It catches its prey-chiefly insects and arachnids—with a **trapping structure** formed by the terminal portion of each of the plant's leaves, which is triggered by tiny hairs on their inner surfaces.	Pitcher plants are several different **carnivorous plants** which have modified leaves known as **pitfall traps**—a prey -trapping mechanism featuring a deep cavity filled with digestive liquid. The traps of what are considered to be "true" pitcher plants are formed by specialized leaves. The plants attract and drown their prey with nectar.	Drosera, which is commonly known as the sundews, is one of the largest genera of **carnivorous plants**, with at least 194 species. The **trapping** and digestion mechanism of Drosera usually employs two types of glands: stalked glands that secrete sweet mucilage to attract and ensnare insects and enzymes to digest them, and sessile glands that absorb the resulting nutrient soup.	If a plant is carnivorous , then it probably has a trapping structure.

Table 1: An example of inductive reasoning in DEER dataset. We embolden the words in facts that contain the key information to induce this rule (just to explain the relation between facts and rule, in DEER there's no special word annotations for fact).

What is the goal of human language?

To share thoughts with conspecifics

- to **communicate** with others
- to store and pass **knowledge**

To think more complex thoughts

- **reasoning**
- **interpretation**

Noam Chomsky

Almost all of your use of language is internal. Virtually all of the use of language has nothing to do with communication. The idea that language has evolved as a system of communication, or designed for communication, makes no sense.

…

The systems of thought use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts.

What is the goal of human language?

To share thoughts with conspecifics

- to **communicate** with others
- to store and pass **knowledge**

Noam Chomsky

Almost all of your use of language is internal. Virtually all of the use of language has nothing to do with communication. The idea that language has evolved as a system of communication, or designed for communication, makes no sense.

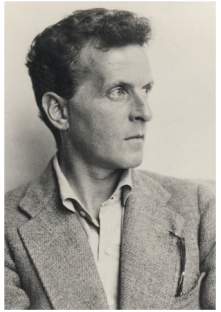
…

The systems of thought use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts.

To think more complex thoughts

- **reasoning**
- **interpretation**

The limits of my language mean the limits of my world.



Ludwig Wittgenstein

Wouldn't be nice if computers could understand human language?

The Gutenberg Bible at Yale Beinecke Rare Book Library

Natural Language Processing (NLP)

It deals with **natural languages** that human speak (e.g., English, Chinese, Spanish, Arabic, etc), not programming languages.

- NLP is also referred to as **Human Language Technology**

Natural Language Processing (NLP)

It deals with **natural languages** that human speak (e.g., English, Chinese, Spanish, Arabic, etc), not programming languages.

- NLP is also referred to as **Human Language Technology**

It aims to use computational approaches to analyze human languages.

- morphology, lexical analysis, syntax, semantics, discourse

Natural Language Processing (NLP)

It deals with **natural languages** that human speak (e.g., English, Chinese, Spanish, Arabic, etc), not programming languages.

- NLP is also referred to as **Human Language Technology**

It aims to use computational approaches to analyze human languages.

- morphology, lexical analysis, syntax, semantics, discourse

It aims to build computer systems that can **understand** and **generate** natural language.

- It can be used for human-human communication (e.g., machine translation, grammatical error correction) or human-machine communication (e.g., dialog systems, question answering).

Natural Language Processing vs Computational Linguistics

A lot of people use them interchangeably.

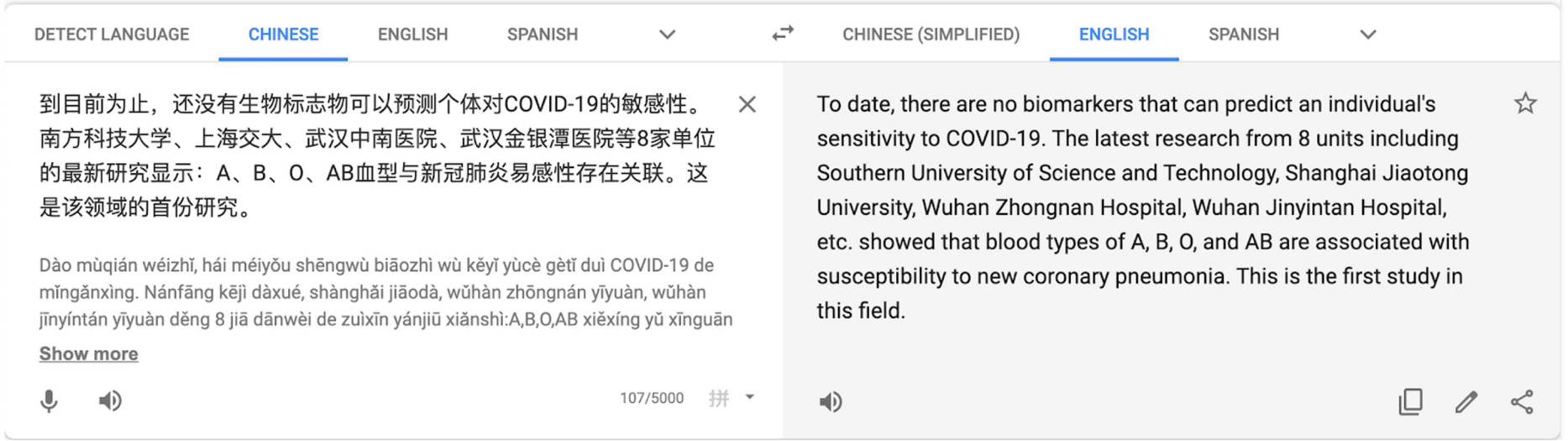
They are very similar, but also different.

The focus is a bit different:

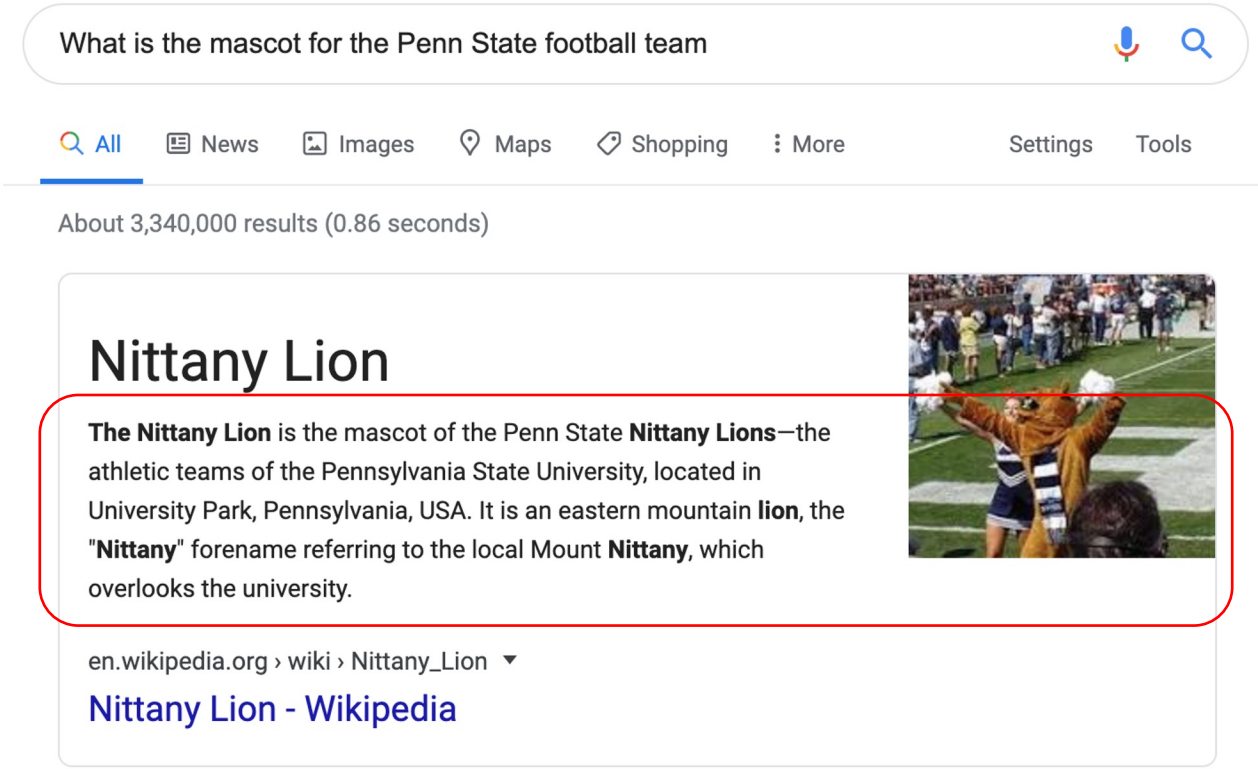
- Computational Linguistics focus on scientific questions in "Linguistics". It aims to study scientific questions about human languages using computational tools.
- Natural Language Processing focus on engineering problems of building practical systems and applications that can mimic human usage of languages.

https://www.quora.com/How-is-computational-linguistics-different-from-natural-language-processing

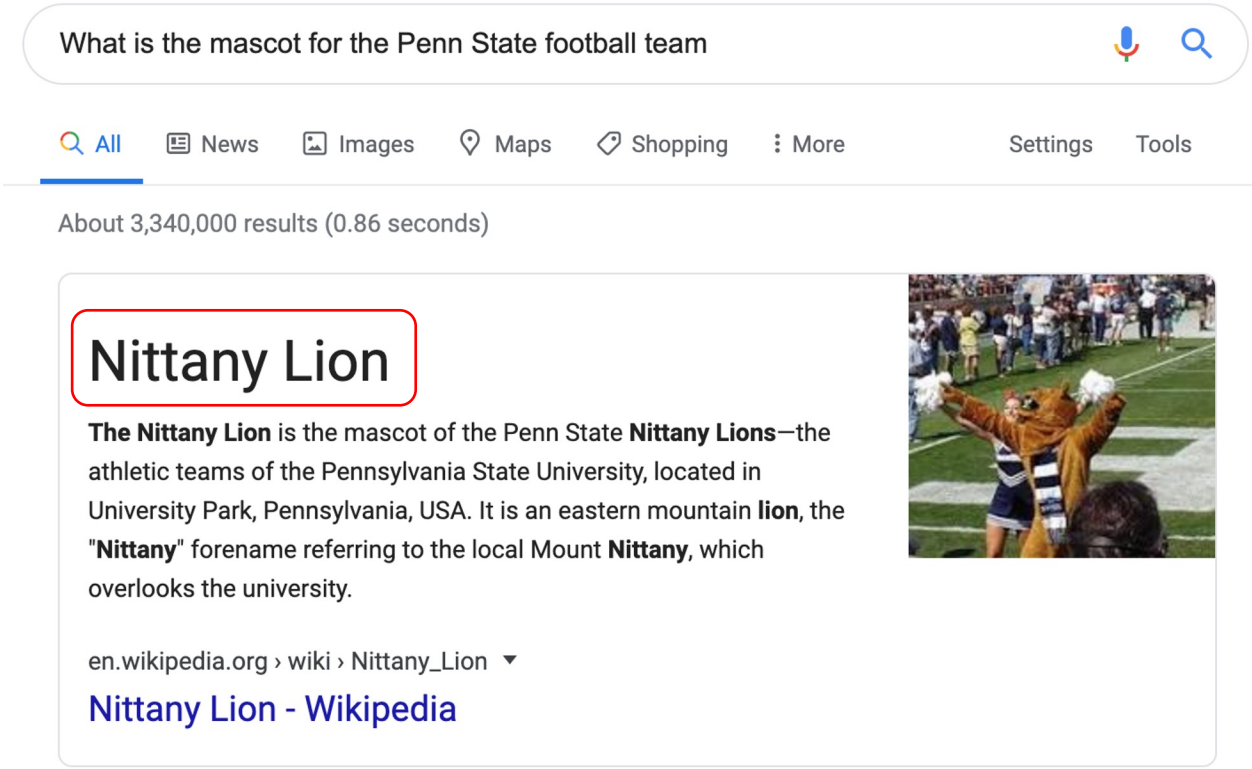
Machine Translation



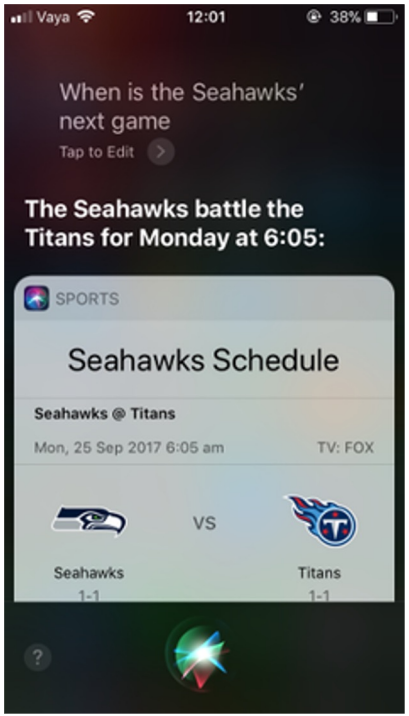
Question Answering (Information Retrieval)



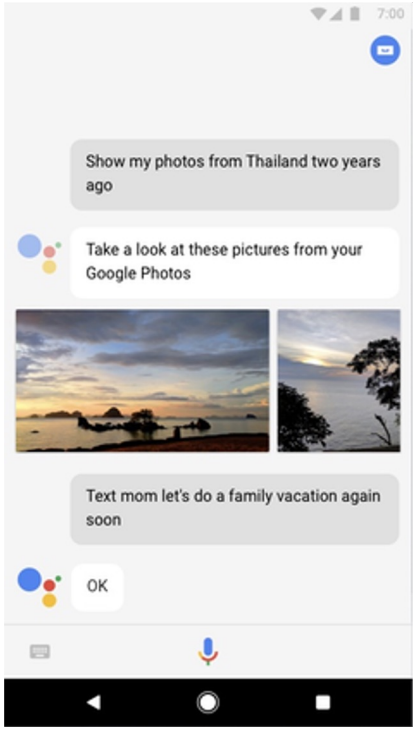
Question Answering (Span Extraction)



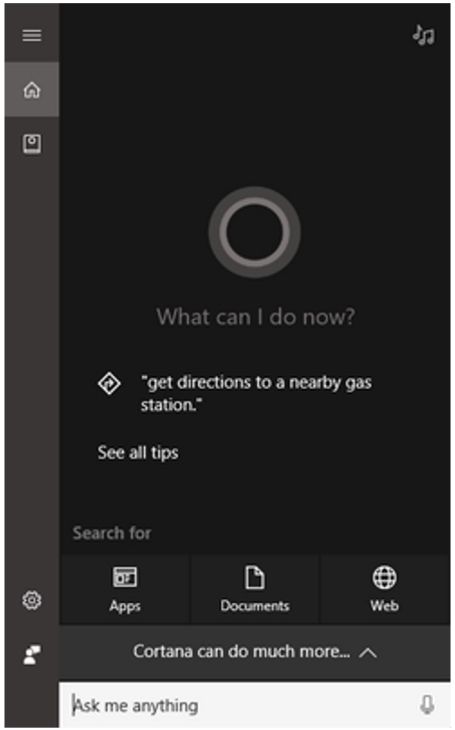
Conversational Personal Assistants



Apple Siri



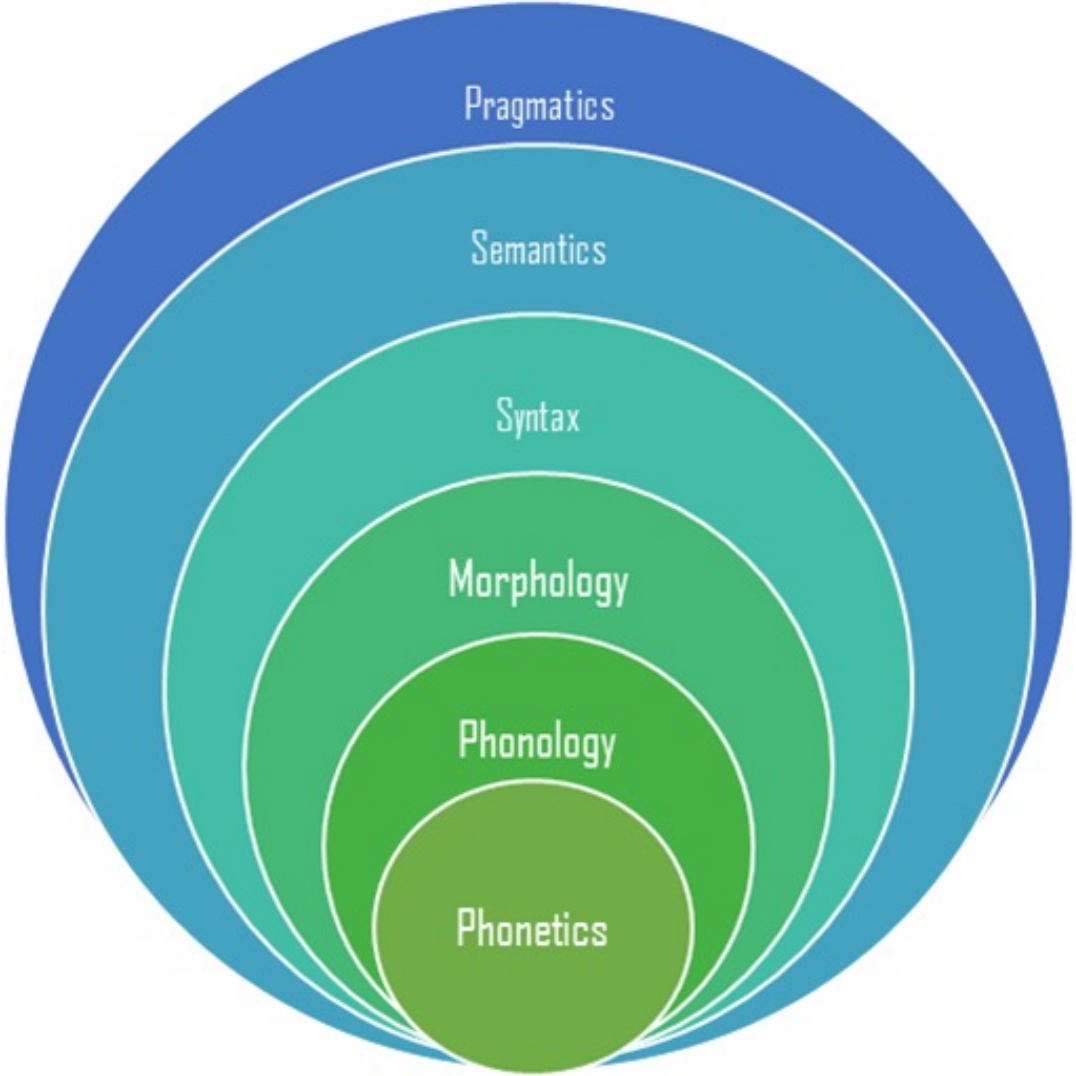
Google Assistant



Microsoft Cortana

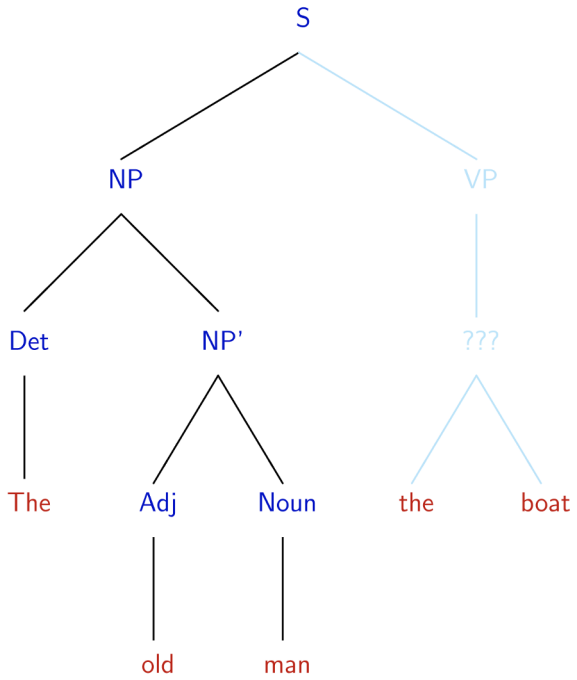
Ambiguity

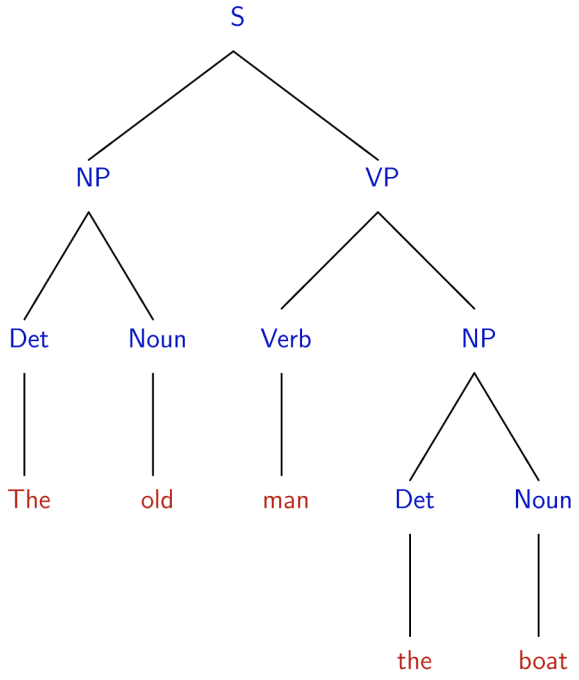
- Linguistic perspective ...



Natural Language Understanding is Hard

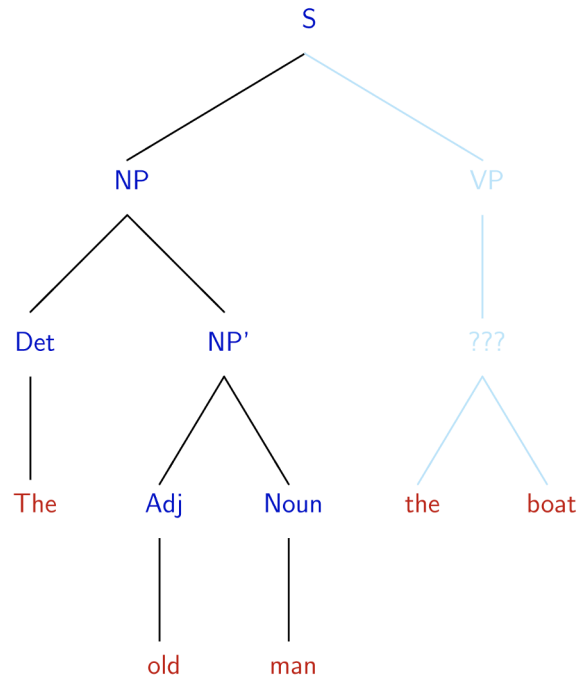
"The old man the boat."

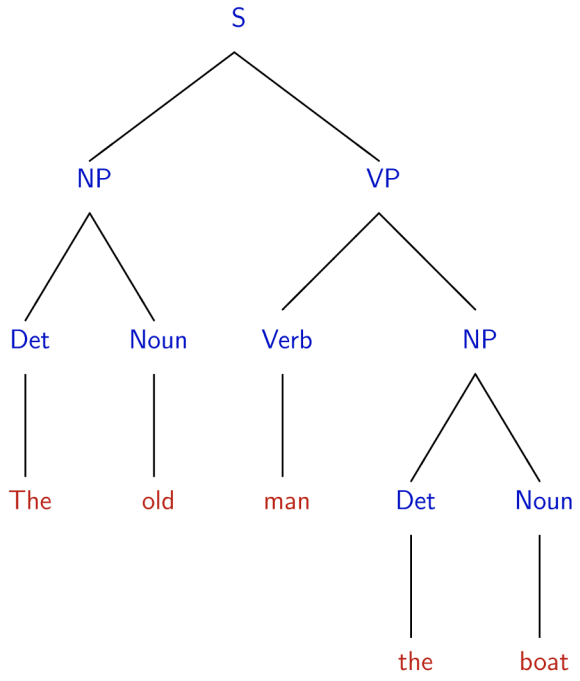




Natural Language Understanding is Hard

"The old man the boat." syntax ambiguity





Semantic Ambiguity

At last, a computer that understands you like your mother.

[Example from Lillian Lee]

Semantic Ambiguity

At last, a computer that understands you like your mother.

- Direct meanings:
  - It understands you like your mother (does) [presumably well]
  - It understands (that) you like your mother

[Example from Lillian Lee]

Semantic Ambiguity

At last, a computer that understands you like your mother.

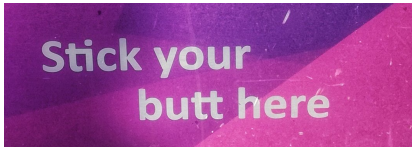
- Direct meanings:
  - It understands you like your mother (does) [presumably well]
  - It understands (that) you like your mother

[Example from Lillian Lee]

Context!

- Thus, the **key difference** between semantics and pragmatics is the fact that semantics is context independent whereas **pragmatic is context dependent**.

Context!



1. Stick your cigarette butt here
2. Stick your backside here

what type of context?

Ambiguities in the Wild: Context

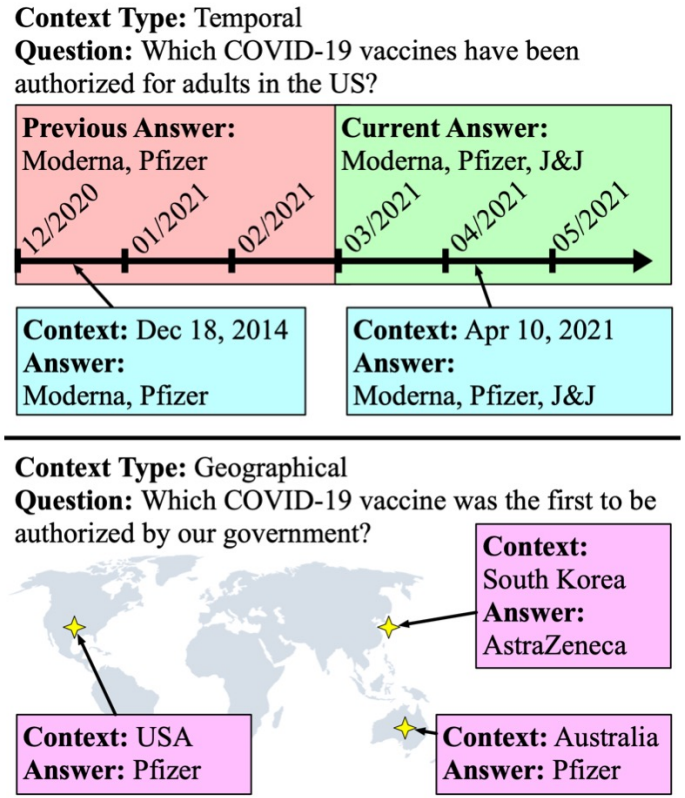
1. Stick your cigarette butt here (trash can)
2. Stick your backside here

Even extra-linguistic features

Q: How many home runs has Shohei Ohtani hit?
A: 24

Q: How many home runs has Shohei Ohtani hit?
A: **25**

Even extra-linguistic features



Even extra-linguistic features

- Time
- Location
- Demographics
- Culture
- ...

- ? assuming we have a strong Neural Language Model that can accurately answer questions (e.g. ChatGPT). How do we take into account extra-linguistic, keeping updated to the “current dynamic world”?

NLP Tasks

- Automatic speech recognition
- CCG
- Common sense
- Constituency parsing
- Coreference resolution
- Data-to-Text Generation
- Dependency parsing
- Dialogue
- Domain adaptation
- Entity linking
- Grammatical error correction
- Information extraction
- Intent Detection and Slot Filling
- Language modeling
- Lexical normalization
- Machine translation
- Missing elements
- Multi-task learning
- Multi-modal
- Named entity recognition
- Natural language inference
- Part-of-speech tagging
- Paraphrase Generation
- Question answering
- Relation prediction
- Relationship extraction
- Semantic textual similarity
- Semantic parsing
- Semantic role labeling
- Sentiment analysis
- Shallow syntax
- Simplification
- Stance detection
- Summarization
- Taxonomy learning
- Temporal processing
- Text classification
- Word sense disambiguation

http://nlpprogress.com/

NLP Tasks

Foundational Technologies

- Language Modeling
- Part-of-speech Tagging
- Syntactic Parsing
- Dependency Parsing
- Named Entity recognition
- Coreference resolution
- Word Sense Disambiguation
- Semantic Role Labelling
- ……

High-Level Tasks and Applications

- Sentiment Analysis
- Information Extraction
- Machine Translation
- Question Answering
- Semantic Parsing
- Summarization
- Dialogue systems
- Language and Vision
- Data-to-Text Generation
- ……

Natural Language Processing Systems

Input X	Output Y	Task
Text (e.g., Sentiment Analysis)	Label	Text Classification
Text	Linguistic Structure	Structured Prediction (e.g., Part-of-Speech Tagging)
Text (e.g., Translation)	Text	Text Generation

Natural Language Processing Systems

To create systems for X -> Y, we can

- **Expert Systems**: Manually create rules mapping X to Y based on our knowledge of languages (below is an example for detecting named entities)

PER, POSITION of ORG:
George Marshall, Secretary of State of the United States

PER (named|appointed|chose|etc.) PER Prep? POSITION
Truman appointed Marshall Secretary of State

PER [be]? (named|appointed|etc.) Prep? ORG POSITION
George Marshall was named US Secretary of State

Hand-built patterns have the advantage of high-precision and they can be tailored to specific domains. On the other hand, they are often low-recall, and it's a lot of work to create them for all possible patterns.

- **Data-Driven Methods**: Collect Data and Use Machine Learning Techniques to learn the mapping function

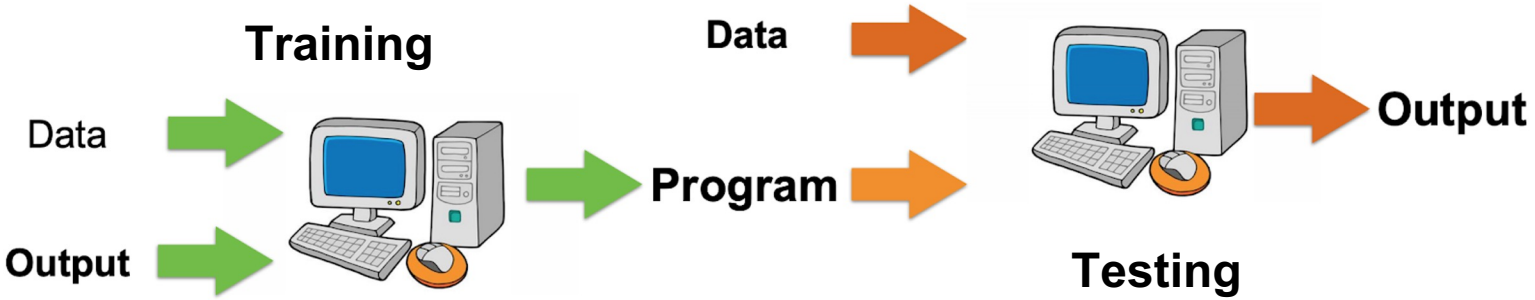
Data Driven Methods and Machine Learning

“**Machine Learning**” is a data-driven problem solving strategy.

Data Driven Methods and Machine Learning

"**Machine Learning**" is a data-driven problem solving strategy.

The ultimate goal of machine learning is **generalization**! Making future prediction on unseen examples as accurate as possible!

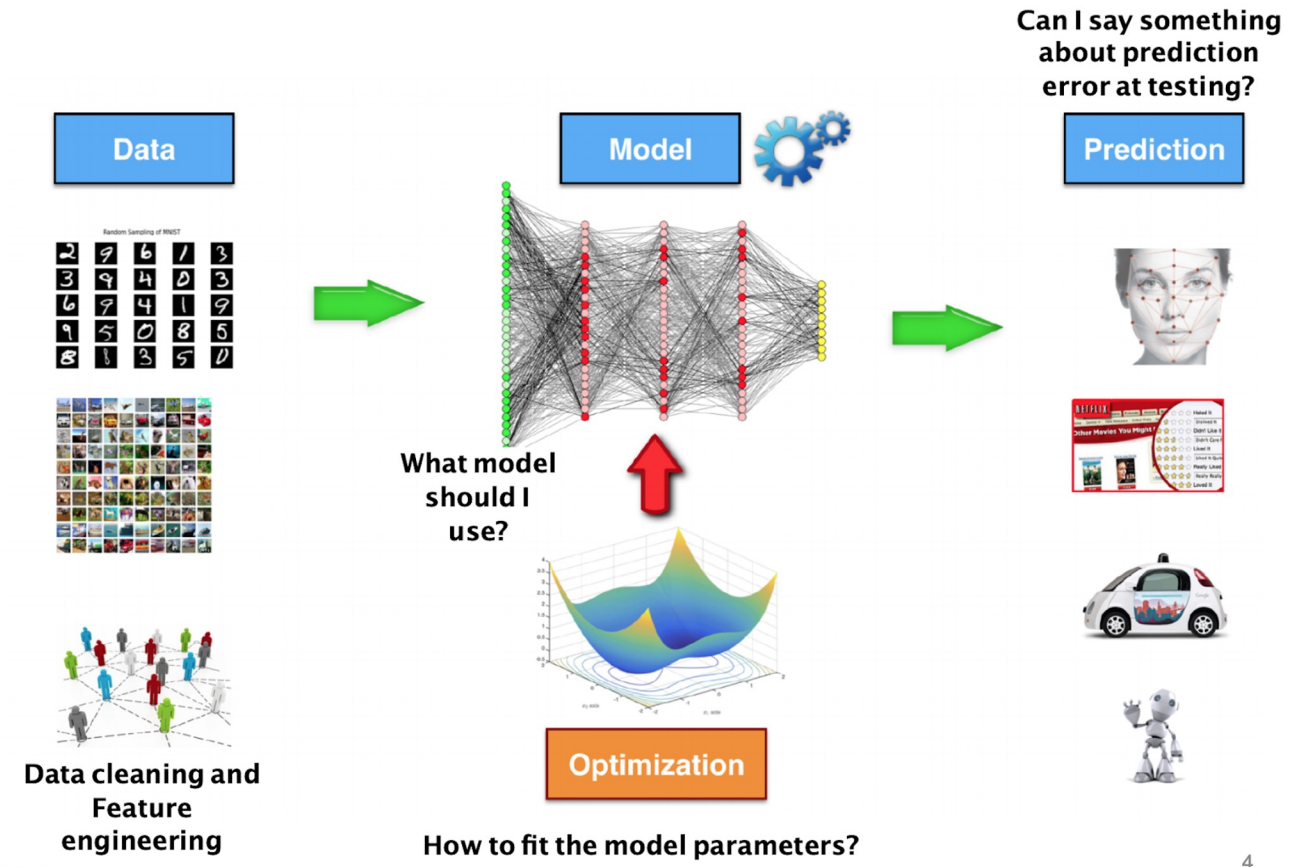


Common theme is to solve a prediction problem:

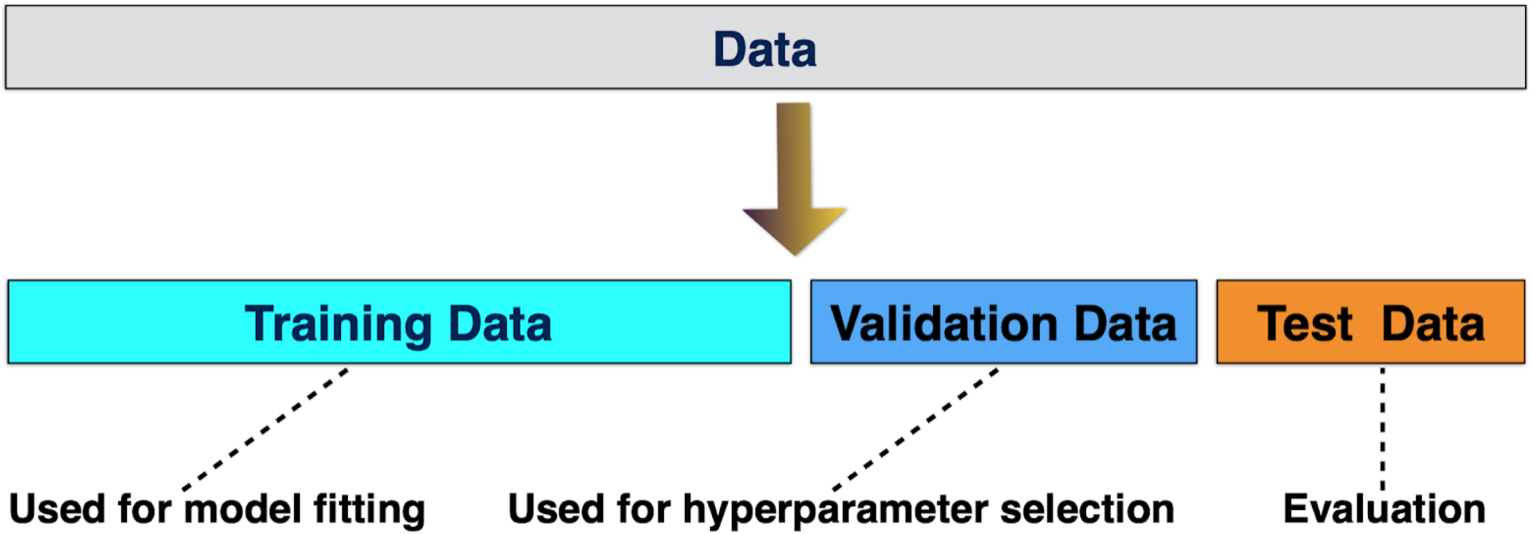
Given an input $\boldsymbol{x}$,

Predict an "appropriate" output y (label, decision, action, etc).

Machine Learning Pipeline



Training, Validation, Test



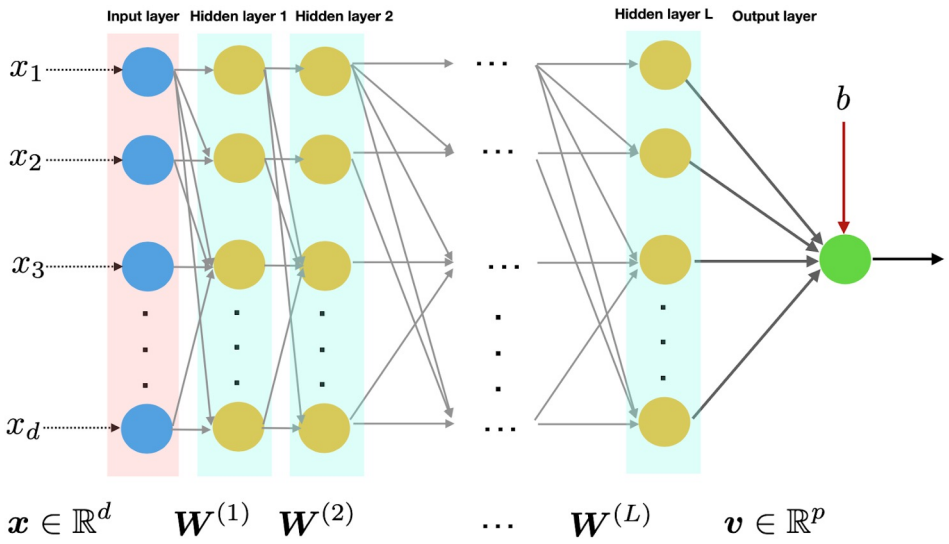
Data Hygiene

Training Data: Use this and only this to train your neural networks.

Development Data: Use this to decide when to stop training your neural networks, and do hyperparameter tuning.

Test Data: Use test data after picking your model. You should use test data only once.

Neural Networks == Deep Learning



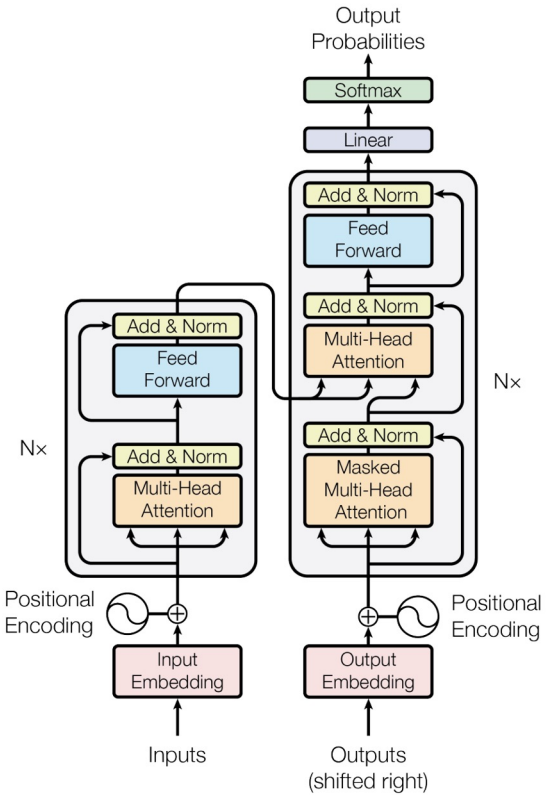
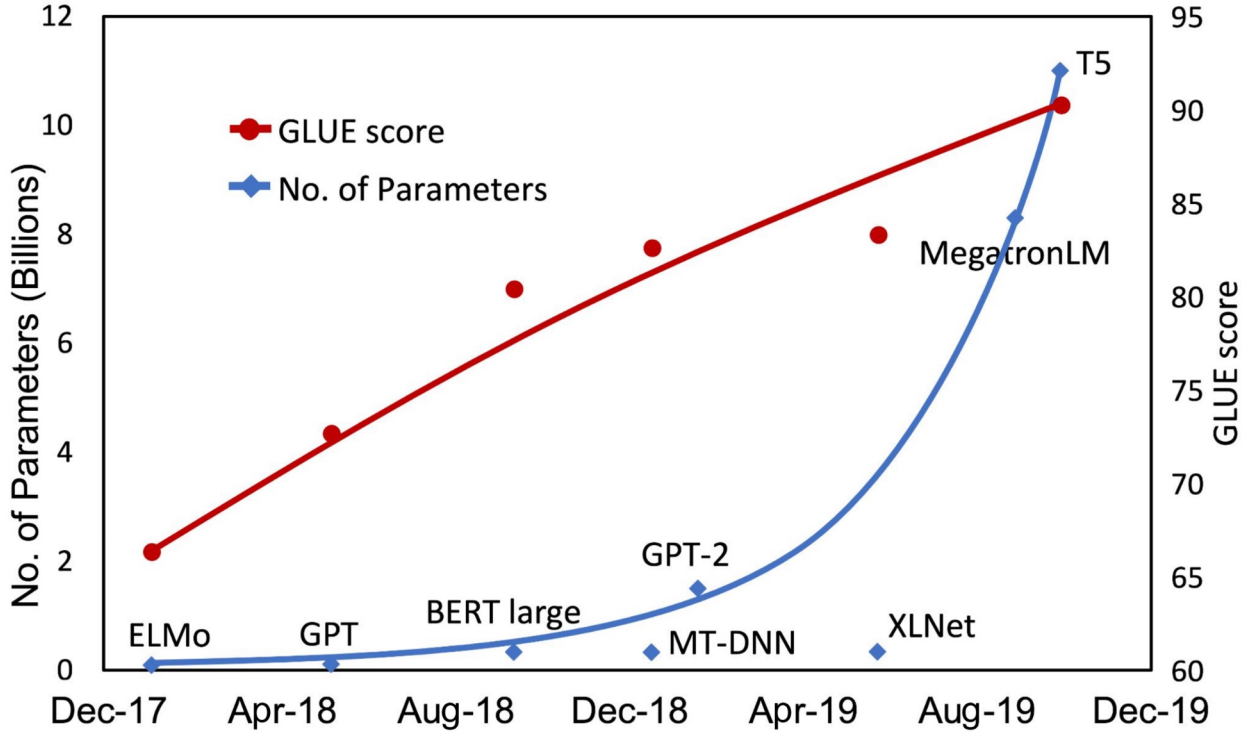


Figure 1: The Transformer - model architecture.

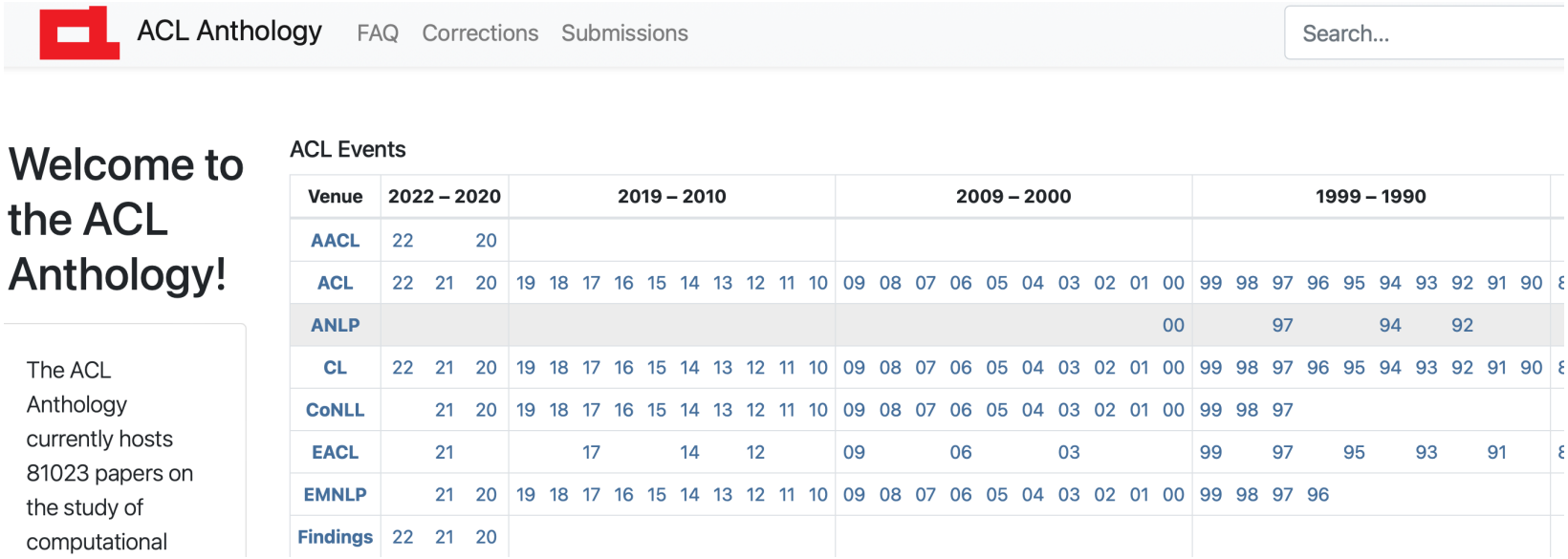
Deep Learning for NLP



From “Real-Time Social Media Analytics with Deep Transformer Language Models: A Big Data Approach” by Ahmet and Abdullah

Keep updated with NLP research

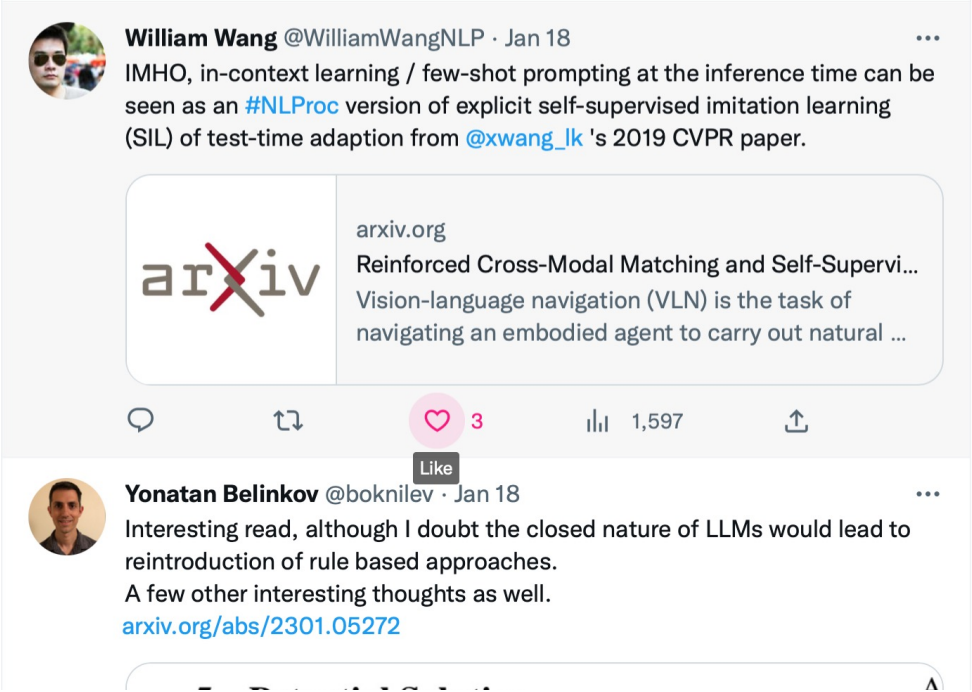
- https://aclanthology.org



Venue	2022 – 2020			2019 – 2010										2009 – 2000										1999 – 1990									
AACL	22		20																														
ACL	22	21	20	19	18	17	16	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90	8
ANLP																							00			97			94		92			
CL	22	21	20	19	18	17	16	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90	8
CoNLL		21	20	19	18	17	16	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97								
EACL		21				17			14		12			09			06			03					99		97		95		93		91		8
EMNLP		21	20	19	18	17	16	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97	96							
Findings	22	21	20																														

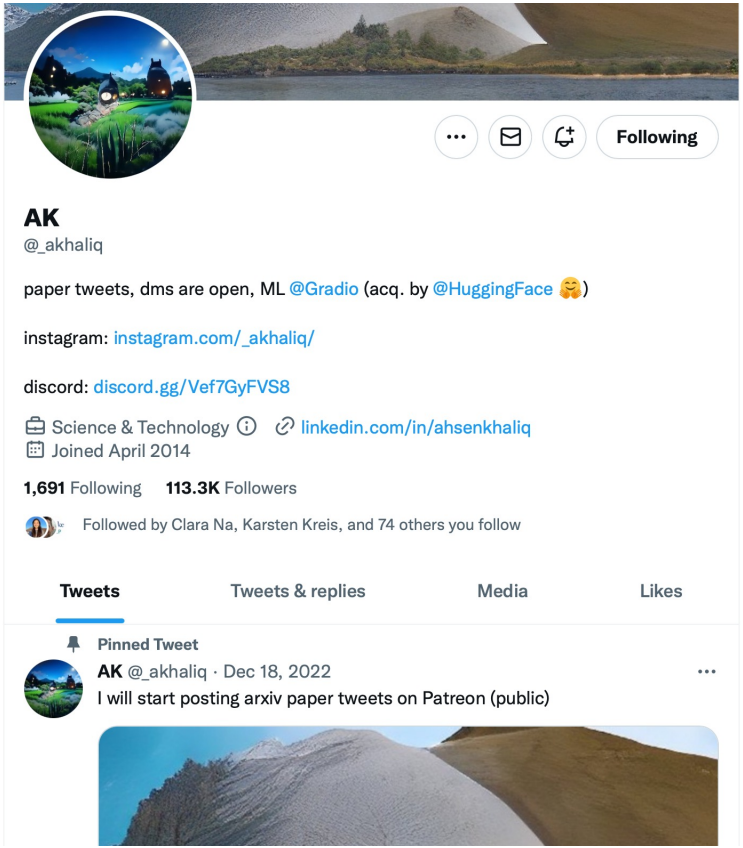
Keep updated with NLP research

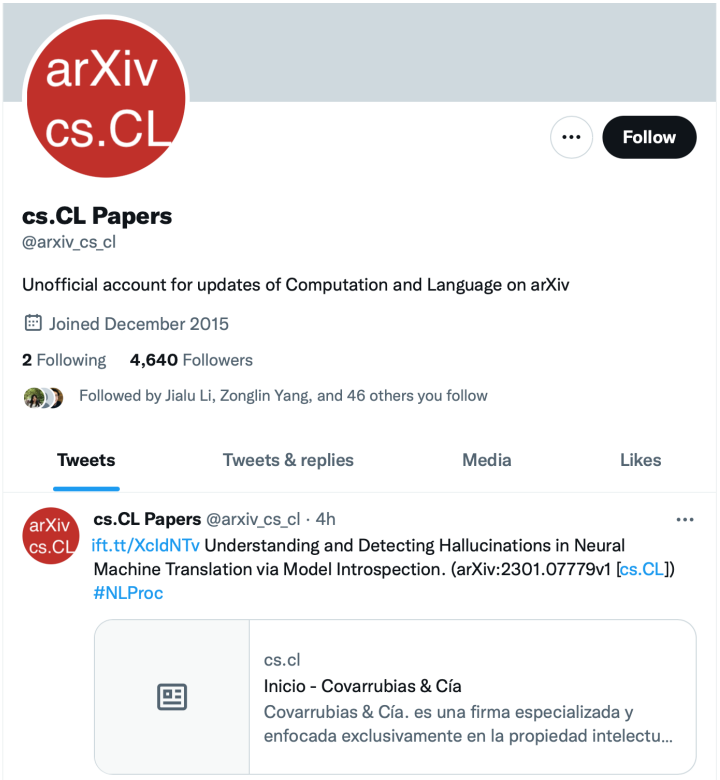
- Twitter



Keep updated with NLP research

- Twitter





Course Topics

- Part 1: Introduction and Neural Network Basics
  - Introduction to NLP (this lecture)
  - Text Classification and Language Modeling
  - Neural Networks and Backpropagation

Course Topics

- Part 1: Introduction and Neural Network Basics
  - Introduction to NLP (this lecture)
  - Text Classification and Language Modeling
  - Neural Networks and Backpropagation

- Part 2: DL&NLP Models, Applications
  - Sequence-to-Sequence, Attention, Transformers
  - Pretraining Language Models and Transfer Learning
  - Text-to-Text and Prompt-based Learning

Course Topics

- Part 1: Introduction and Neural Network Basics
  - Introduction to NLP (this lecture)
  - Text Classification and Language Modeling
  - Neural Networks and Backpropagation

- Part 2: DL&NLP Models, Applications
  - Sequence-to-Sequence, Attention, Transformers
  - Pretraining Language Models and Transfer Learning
  - Text-to-Text and Prompt-based Learning

- Part 3: Advanced Techniques (with paper discussions)
  - Retrieval-augmented models
  - In-context learning
  - Prompting
  - Vision-language Models
  - ..

Logistics

Course Information

Lecture

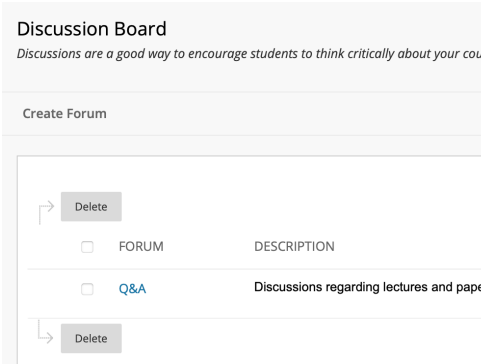
- Friday, 10:00am - 12:45pm (11:15--11:30 resting period)
- GR 2.530, In class attendance is encouraged.

Instructor

- Xinya Du, Office Hour (in person): ECSS 3.227, Friday 2 – 3pm
- TA, Office Hour: TBD

How to Contact us

- For most technical questions
  - ask on eLearning “Q&A” forum
- Please **email** TA.
- Do not send DMs via Teams



Textbooks

Available online! You are encouraged to read beyond slides.

- 🌟Neural Network Methods in Natural Language Processing Yoav Goldberg.
  - https://link.springer.com

- 🌟Speech and Language Processing. 3rd edition. Dan Jurafsky and James H. Martin.

- Natural Language Processing. Jacob Eisenstein.

Prerequisites

- Programming in Python
- CS 5343 Algorithm Analysis and Data Structures
- CS 6375 Machine Learning
- CS 6320: Natural Language Processing (not a must)

Previous experience in NLP/Deep Learning is preferred but not required.

Groups

- Paper presentation and the final project are group-based.
- All group members are treated equally and earn the same grades.
- Fill in the Excel Sheet (link with be posted on eLearning) with group member's names and NetIDs.

	A	B	C	D	E
1	please form groups and fill in your names in the cells below, each group should consists of <=4 students, you can use the discussion board to look for teammates. The ddl for grouping is Jan 25.				
2		member 1	member 2	member 3	member 4
3	group0	Xinya Du (xxd220001)	TA		
4	group1				
5	group2				

Groups

- Each group should contain **2-4** students. If you detect serious imbalance workload among group members, penalty will apply
- Utilizing the forum to find group members
- Deadline for grouping is Jan 25 [cs6401-spcs6401-sp23-groups.xlsx23-]

	A	B	C	D	E
1	please form groups and fill in your names in the cells below, each group should consists of <=4 students, you can use the discussion board to look for teammates. The ddl for grouping is Jan 25.				
2		member 1	member 2	member 3	member 4
3	group0	Xinya Du (xxd220001)	TA		
4	group1				
5	group2				

Groups

- Link will be posted on eLearning: cs6401-sp23-groups.xlsx

	A	B	C	D	E
1	please form groups and fill in your names in the cells below, each group should consists of <=4 students, you can use the discussion board to look for teammates. The ddl for grouping is Jan 25.				
2		member 1	member 2	member 3	member 4
3	group0	Xinya Du (xxd220001)	TA		
4	group1				
5	group2				

Grading

- Assignments (30%)
- Paper presentation (20%)
- Course project (45%)
- Participation (5%)

Assignments (individual-based)

Two Assignments

- Assignment 1 (Written)
  - Time: TBD
  - Related to basic concepts (e.g. word representations, backprop, FFNN)
- Assignment 2 (programming)
  - Time: TBD
  - Related to **correct** implementation of a neural network model for classification task

Not late assignment accepted unless there is a compelling reason. Late assignment: please refer to the syllabus.

Paper Presentation (group-based)

Research paper

- We will provide potential topics (each with 1-2 papers, you can also add in related papers!)

Paper Presentation (group-based)

Research paper

- We will provide potential topics (each with 1-2 papers, you can also add in related papers!)
- **Paper Review (5%)**
  - Send me the slides (cc'ing the TA) 48 hours before your group's presentation
  - After the presentation, upload the slides to eLearning.

Paper Presentation (group-based)

Research paper

- We will provide potential topics (each with 1-2 papers, you can also add in related papers!)
- **Paper Review (5%)**
  - Send me the slides (cc'ing the TA) 48 hours before your group's presentation
  - After the presentation, upload the slides to eLearning.
- **Paper presentation (15%)**
  - Time: in class, 35 minutes for each group
  - We value clear presentations and insightful discussions
    - You are allowed to search materials/explanations online, but you must need to understand the contents and demonstrate your understanding of this topic.

Exams

No Exam. This course is a **research project** oriented course.

Project Format

Project Topics. This project aims to conduct original and independent research over **any NLP-related topics**.

- Must deal with natural language data
- Must include some degree of model implementations and experiments

Group Policy. You can work on the course project in a group of **2 - 4** people. You are allowed to combine this project with your research projects or projects from other courses.

Project Deliverables

- **Proposal**. Write a proposal that outlines your plan including what problem or task you want to address, what dataset(s) you want to work on, what metrics you need to employ, what baselines you would like to compare with. You should also cite a few relevant prior papers. The length should be 2—4 pages.

- **Proposal Oral Presentation**. During the mid of the semester, each group will be given a 15—20 min oral presentation.

Project Deliverables

- **Final Report**. Your final report should use the ACL Latex template (https://acl-org.github.io/ACLPUB/formatting.html), overleaf template here. The length should be around 4—8 page plus references. Your report should begin with an abstract and introduction to clearly state the problem you want to solve and contributions you have made. It should also have a section on related work, a section on your methodology, a section on your experimental settings and results, and a section on conclusions.

- **Final Oral Presentation**. During the last week (and the week before) of the lecture, each group will give a 15—20 min oral presentation.

Participation

- Quiz (2%)
- Discussions (2%)
  - Asking and answering other students' questions on the "Q&A" forum
  - Asking and participating in other groups' presentations
  - Including (1) elearning discussion forum; (2) class question answering and asking. Top 30% students get 2%, 30%-70% students get 1%.
- Course feedback (1%)

Participation

- Quiz (2%)
  - 1-2 quiz, each contains multi-choice questions or short-form written questions.
  - within 10 minutes

Tentative schedule

Week	Date	Topic	Assignments	Candidate Papers	Recommended Readings
Week 1	Jan 20	Introduction			
		-			
Week 2	Jan 27	Neural Networks, Backpropagation			J&M 7 (.1--.4), Primer, J&M 7, Intro to Computation Graphs
		-	a1		
Week 3	Feb 3	Word Embeddings			J&M 6, word2vec explained
		-			
Week 4	Feb 10	Pytorch & Transformers Tutorial			
		RNNs, Seq2Seq, Attention	a1 due, a2		J&M 9, J&M 10 (.2, .3), Luong 15
Week 5	Feb 17	-			
		Self-attention			illustrated Transformer, Annotated Transformer, Paper
Week 6	Feb 24	-			
		Pretrained Language Models (PLMs)	a2 due		Survey
Week 7	Mar 3	-			
		DL for NLP applications (QA, NLG)			
Week 8	Mar 10	Project proposal			
		Project proposal	project proposal report due		
Week 9	Mar 17	Spring break			

Tentative schedule

Week 9	Mar 17	Spring break			
Week 10	Mar 24	In-context learning			
		Prompting			Survey
Week 11	Mar 31	Paper Presentation * 3			
		Paper Presentation * 3			
Week 12	Apr 7	Retrieval-augmented Models			
		-			
Week 13	Apr 14	Paper Presentation * 3			
		Paper Presentation * 3			
Week 14	Apr 21	Vision-Language Models			
		-			
Week 15	Apr 28	Paper Presentation * 3			
		Paper Presentation * 3			
Week 16	May 5	Project final presentation			
		Project final presentation	project final report due		

End of Introduction