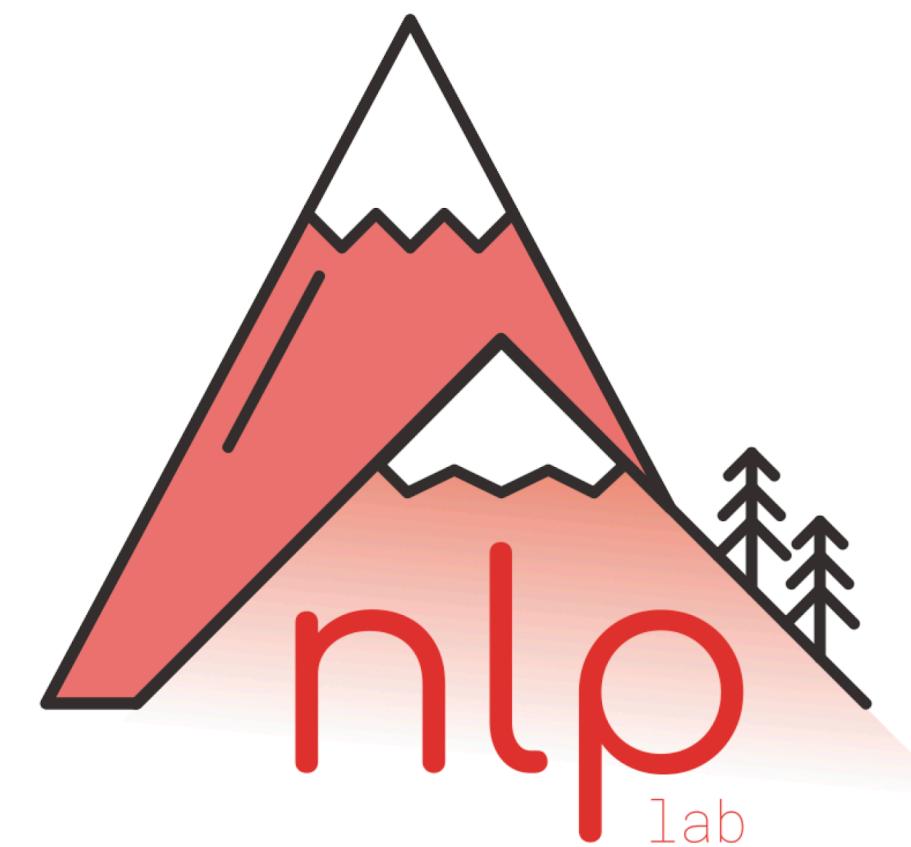


CS-552: Modern NLP

Introduction

Antoine Bosselut



Natural Language Processing

**Enabling
Human-Machine
Collaboration**

Search Engines

Dialogue Agents

Text Generation

**Accelerating
Human-Human
Communication**

Machine Translation

Text Summarization

Information Extraction

**Mining
Human Insights**

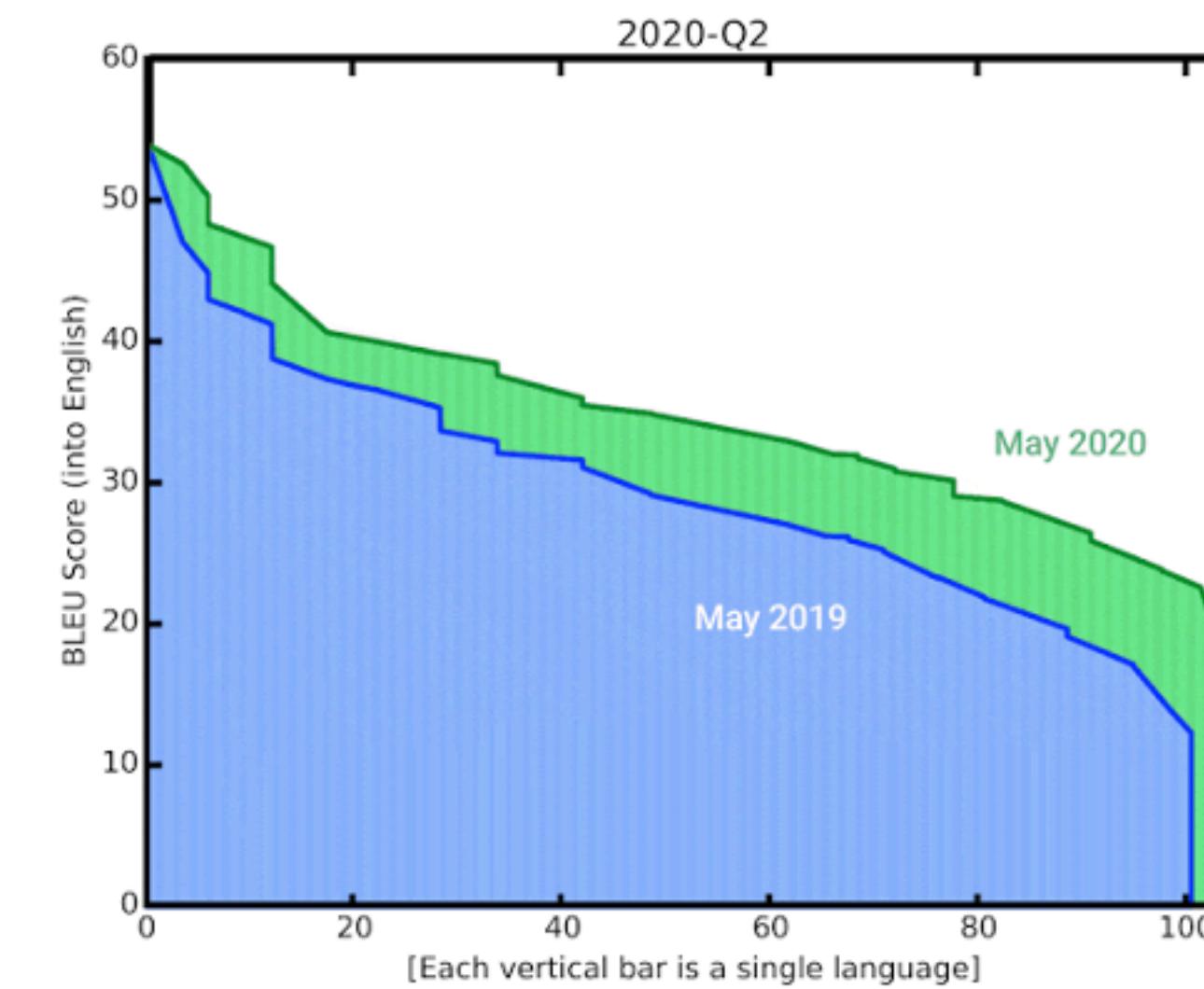
Sentiment Analysis

Motivation Analysis

Emotion Detection



Machine Translation



DETECT LANGUAGE FRENCH ENGLIS ▾ ↗ ENGLISH FRENCH SPANISH ▾

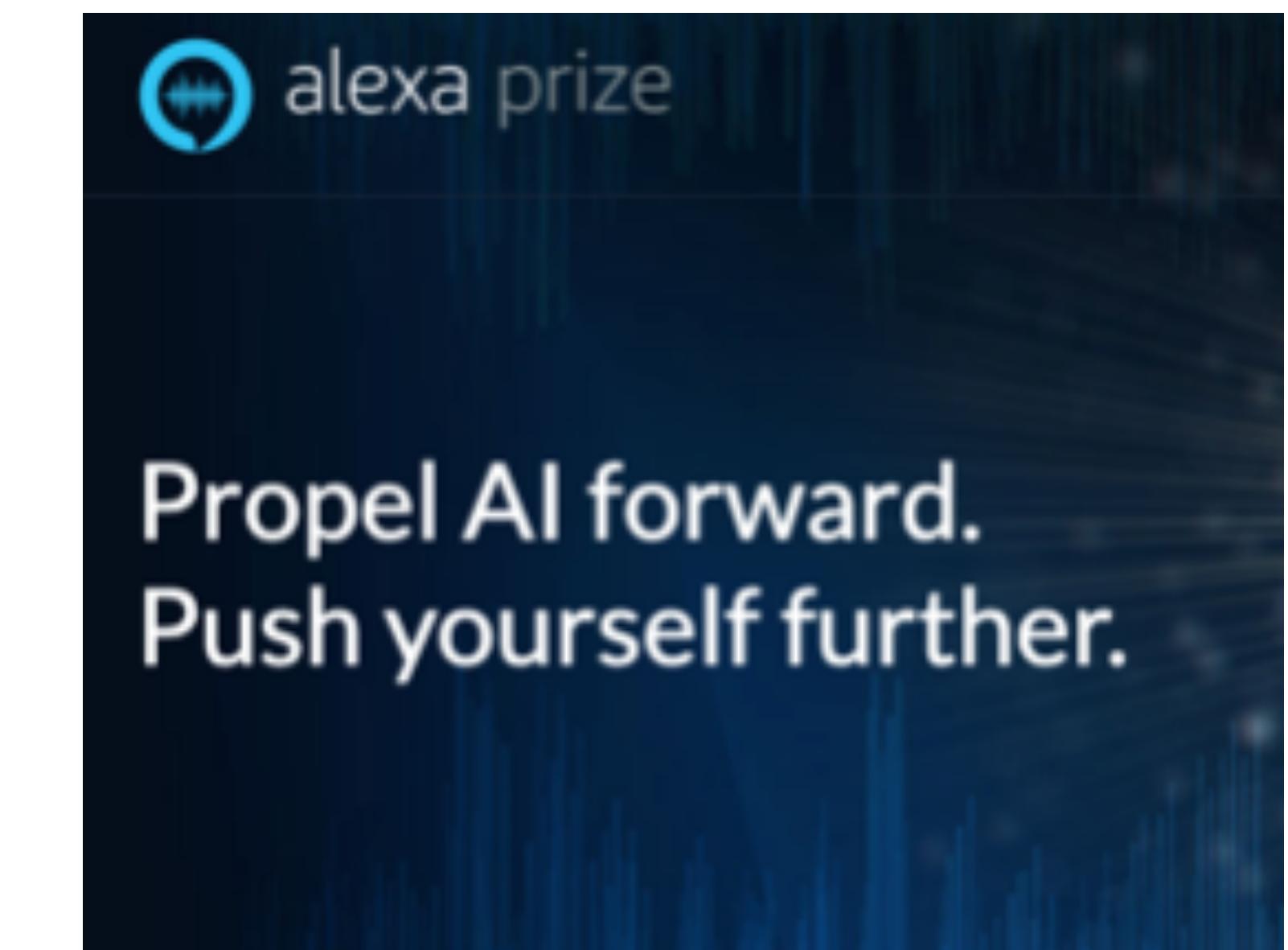
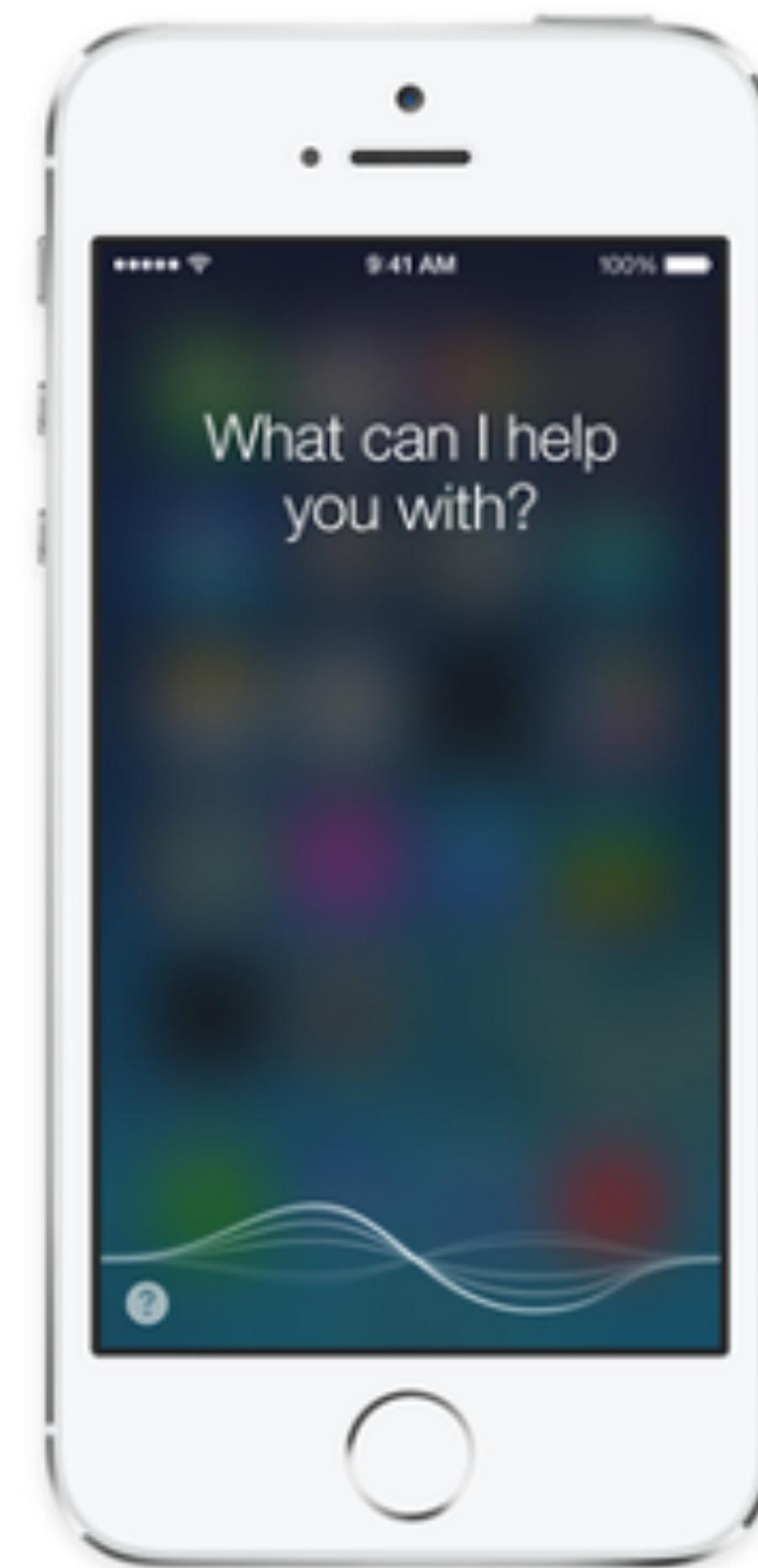
J'ai mangé avec mon avocat × I ate with my lawyer today ☆

aujourd'hui

38 / 5000

3

Conversational Systems



Question Answering

what is the tallest mountain in europe ?

All Images Maps News Videos More Tools

About 12'400'000 results (1.10 seconds)

Europe / Mountains / Maximal / Elevation

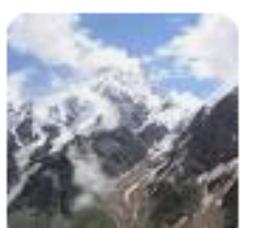
Mount Elbrus

5,642 m

Mountains



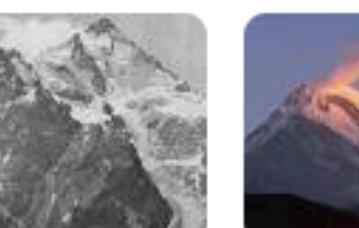
Mount
Elbrus
5,642 m



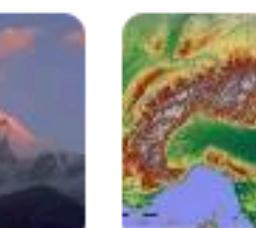
Dykh-Tau
5,205 m



Shkhara
5,193 m



Koshtan-Tau
5,151 m
Mount
Kazbek
5,033 m



Mount
Kazbek
5,033 m



Alps



Mont Blanc
4,810 m

who is the president of switzerland ?

All Images News Maps Videos More Tools

About 415'000'000 results (0.82 seconds)

Guy Parmelin

The President of the Swiss Confederation in 2021 is **Guy Parmelin** from the canton of Vaud. He was elected on 9 December 2020. The President's department in 2021 is the Federal Department of Economic Affairs, Education and Research EAER.

<https://www.admin.ch> › gov › start › federal-presidency

Presidential year 2021

[About featured snippets](#) • [Feedback](#)

Text Generation

AN

Can you give me a two-day itinerary for a weekend trip to Paris ? Please don't include more than three activities per day.



Day 1:

1. Visit the Eiffel Tower - Start your day with the iconic symbol of Paris, the Eiffel Tower. Take a stroll around the Champ de Mars and enjoy the views from the tower.
2. Explore the Louvre Museum - After lunch, head to the Louvre Museum and admire the world-famous artwork such as the Mona Lisa and the Venus de Milo.
3. Stroll along the Seine River - In the evening, take a romantic stroll along the Seine River and enjoy the Parisian atmosphere.

Day 2:

1. Visit Notre-Dame Cathedral - Start your day by visiting one of the most famous Gothic cathedrals in the world, Notre-Dame Cathedral.
2. Explore the Marais neighborhood - After lunch, explore the Marais neighborhood, known for its historical architecture, trendy boutiques, and restaurants.
3. Enjoy a Cabaret Show - In the evening, enjoy a cabaret show at the famous Parisian venue, the Moulin Rouge, and end your trip with a memorable experience.

Why is natural language processing difficult?

Ambiguity

Kids make nutritious snacks

Stolen painting found by tree

Miners refuse to work after death

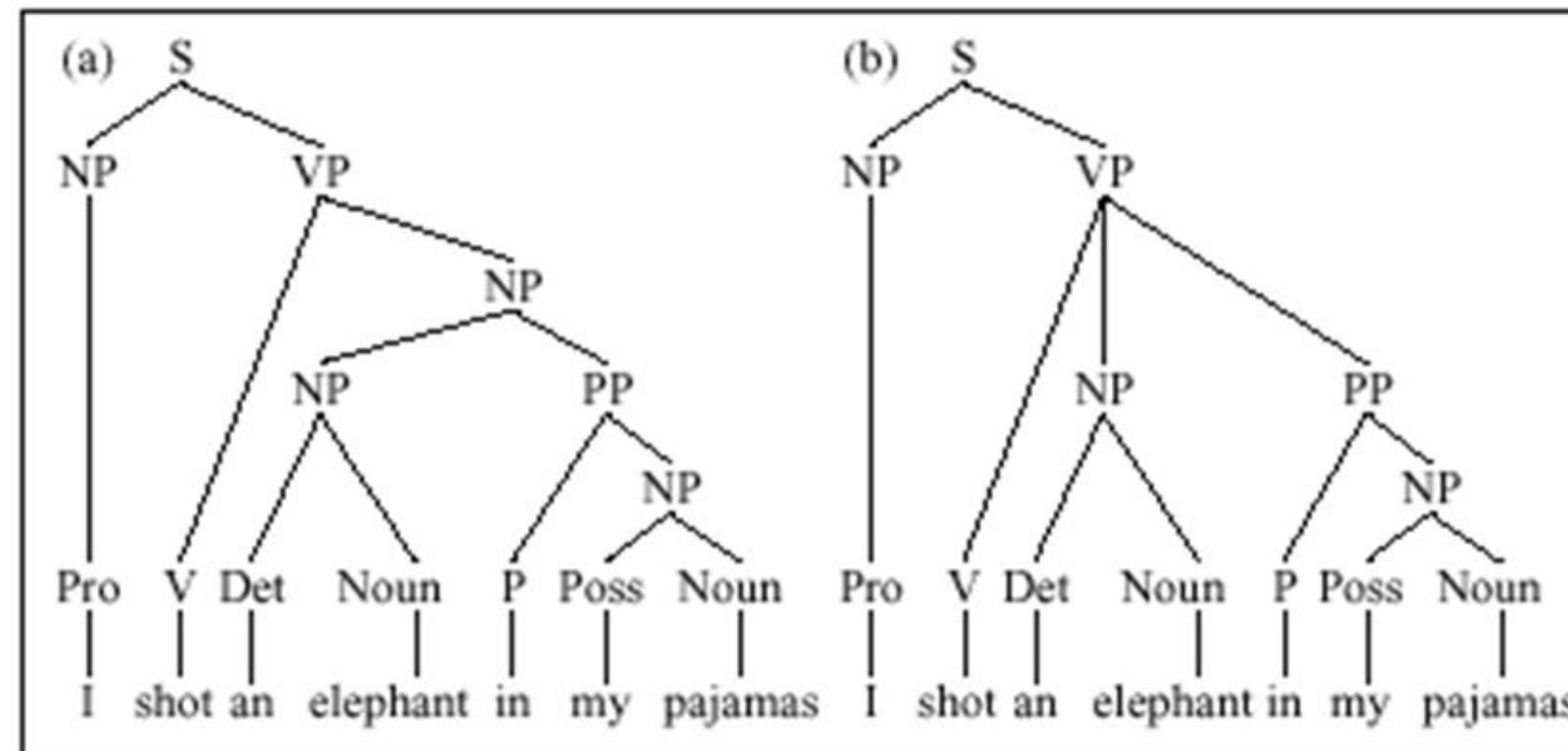
Squad helps dog bite victim

Killer sentenced to die for second time in 10 years

Lack of brains hinders research

Ambiguity: Syntax

I shot an elephant in my pajamas



Ambiguity: Semantics

The city council refused the demonstrators a permit because they advocated violence

The city council refused the demonstrators a permit because they feared violence

The city council refused the demonstrators a permit because they _____ violence

Ambiguity?

- Often, there aren't only a few possibilities that can be resolved pragmatically

il fait vraiment beau



It is really nice out

It's really nice

The weather is beautiful

It is really beautiful outside

~~He makes truly beautiful~~

~~It fact actually handsome~~

- There could be combinatorially many possibilities
- Many won't even register as ambiguities, but systems must resolve them

AN

In the sentence, "The city council refused the demonstrators a permit because they advocated violence" who does "they" refer to ?



In this sentence, the pronoun "they" refers to the demonstrators. The sentence states that



How did we get so good at natural language processing over the last few years?



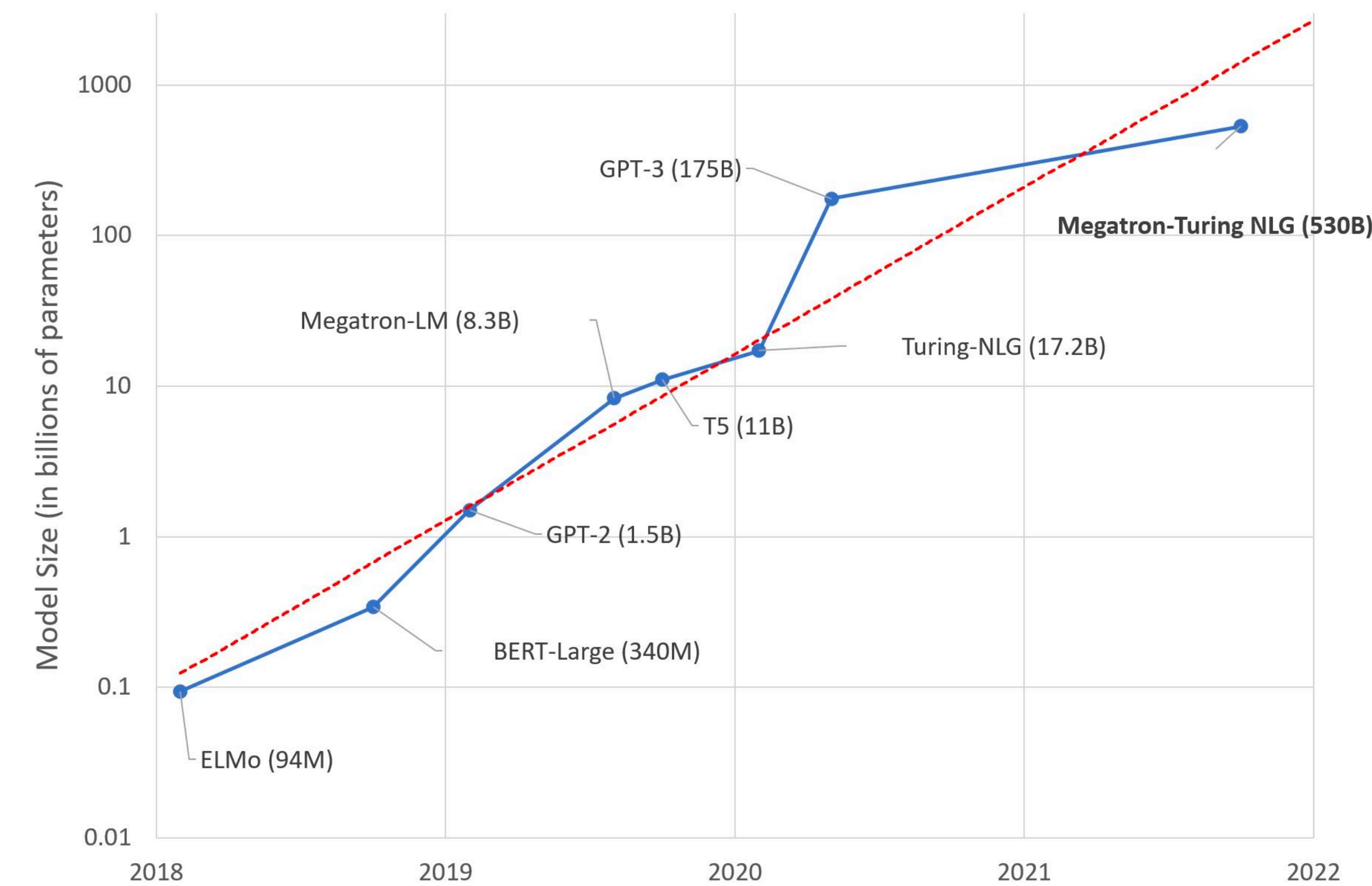
In this sentence, the pronoun "they" refers to the city council. The sentence states that



the city council refused the demonstrators a permit because they (the city council) feared violence.

Language models got a lot bigger!

- Advances in hardware and algorithms allow researchers to train larger language models
- **Growth:** ~5000x increases in model size from 2018 to 2022
- **Comparison:** human cerebral cortex contains ~1000x more neurons than that of mice



AN

In the sentence, "The city council refused the demonstrators a permit because they advocated violence" who does "they" refer to ? 



In this sentence, the pronoun "they" refers to the demonstrators. The sentence states that   the city council refused the demonstrators a permit because they (the demonstrators) advocated violence.

Neural modeling at scale makes things better!

Still a lot of work to be done!



I apologize for the confusion in my previous response. In the sentence, "The city council refused the demonstrators a permit because they feared violence," the pronoun "they" actually refers to the demonstrators.  

The sentence means that the city council denied the permit requested by the demonstrators, and the reason for that denial was that the city council was afraid of violence that might occur if the permit was granted to the demonstrators. So, in this sentence, "they" refers to the demonstrators.

Today's Outline

- **Section 1 - Modern NLP**
- **Course Introduction**
- **Section 2 - An simple introduction to Neural NLP**

What will we cover in this class?

Course Goals

- Define basic problems and tasks in natural language processing (e.g., text classification; generation)
- Implement common modern approaches for tackling NLP problems and tasks (e.g., embeddings, neural models, transformers)
- Understand failure modes of these models and learning algorithms (e.g., robustness, ethics, evaluation)
- Deploy your understanding of these concepts in applied scenarios (assignments, projects)



Building Blocks: Models

Week	Date	Topic
Week 1	2/21/2024	Introduction + Building a simple neural classifier
	2/22/2024	Neural word embeddings
Week 2	2/28/2024	Classical and Fixed-context Language Models
	2/29/2024	Recurrent Neural Networks + Gated RNN Variants
Week 3	3/6/2024	Attention + Transformers
	3/7/2024	Transformers continued
Week 4	3/13/2024	Pretraining: ELMo, BERT
	3/14/2024	Transfer Learning: Introduction

Building Blocks: Learning

Week 5	3/20/2024	Transfer Learning: Dataset Biases
	3/21/2024	Transfer Learning: Prompts
Week 6	3/27/2024	Text Generation
	3/28/2024	Text Generation
Week 7	4/3/2024	NO CLASS
	4/4/2024	NO CLASS
Week 8	4/10/2024	In-context Learning + RLHF
	4/11/2024	Project Description

Building Blocks: Capabilities

Week 9 4/17/2024 Ethics in NLP

4/18/2024 **No Class - Work on your project**

Week 10 4/24/2024 Scaling Laws + Model Compression

4/25/2024 **No Class - Work on your project**

Week 11 5/1/2024 Tokenization + Multilingual LMs

5/2/2024 **Guest Lecture:** Signed Language Processing

Week 12 5/8/2024 Interpretability & Analysis of Language Models

5/9/2024 **No Class - Work on your project**

Applications

Week 13 5/15/2024 Reading Comprehension & Retrieval-Augmented LMs

5/16/2024 **No Class - Work on your project**

Week 14 5/22/2024 Language & Vision

5/23/2024 Looking Forward + Wrap-up

Week 15 5/29/2024 **No Class - Work on your project**

5/30/2024 **No Class - Work on your project**

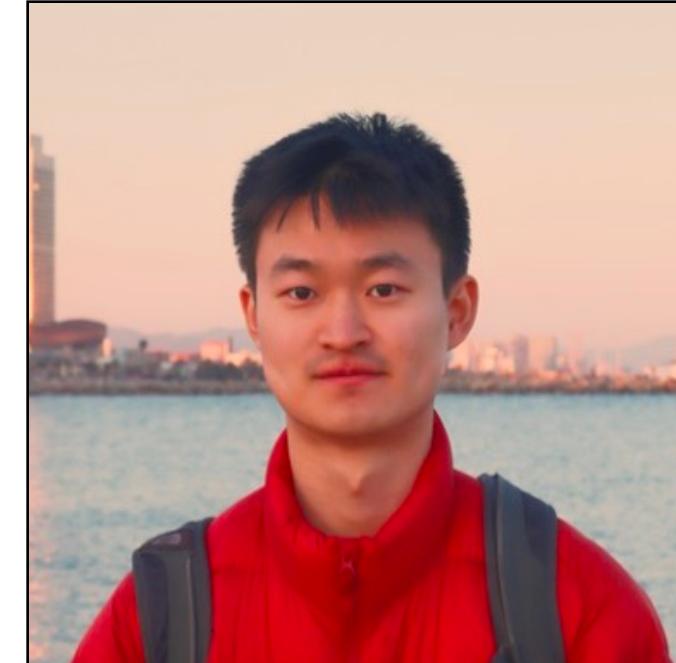
Course Team



Antoine Bosselut
Instructor



Negar Foroutan
TA



Silin Gao
TA



Beatriz Borges
TA



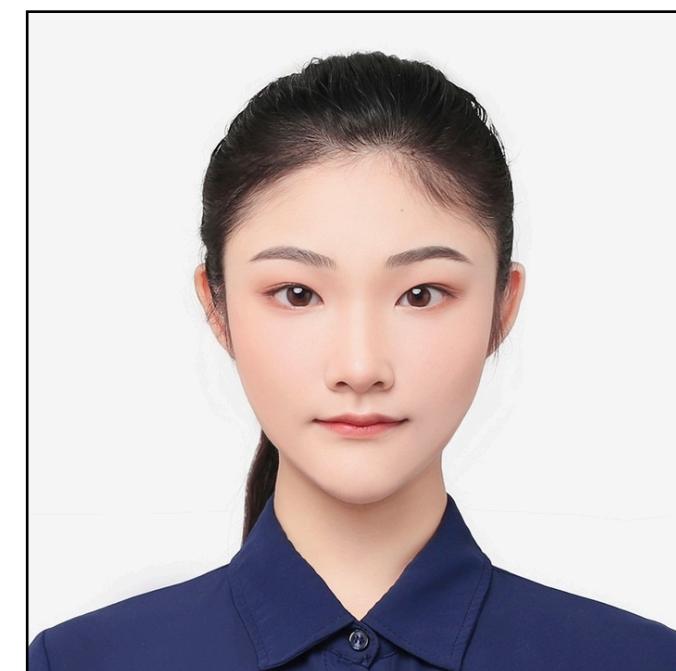
Mete Ismayil
TA



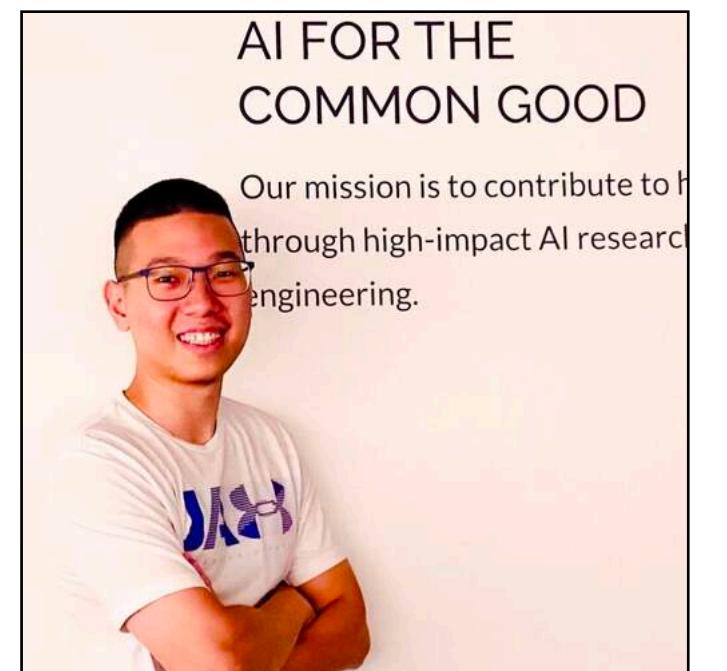
Deniz Bayazit
TA



Badr AlKhamissi
TA



Simin (Olivia) Fan
TA



Zeming (Eric) Chen
TA

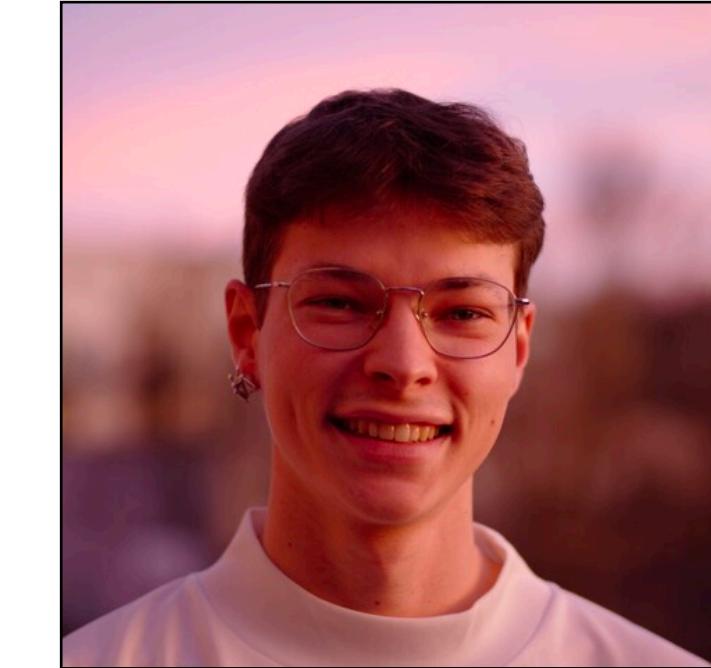
Course Team



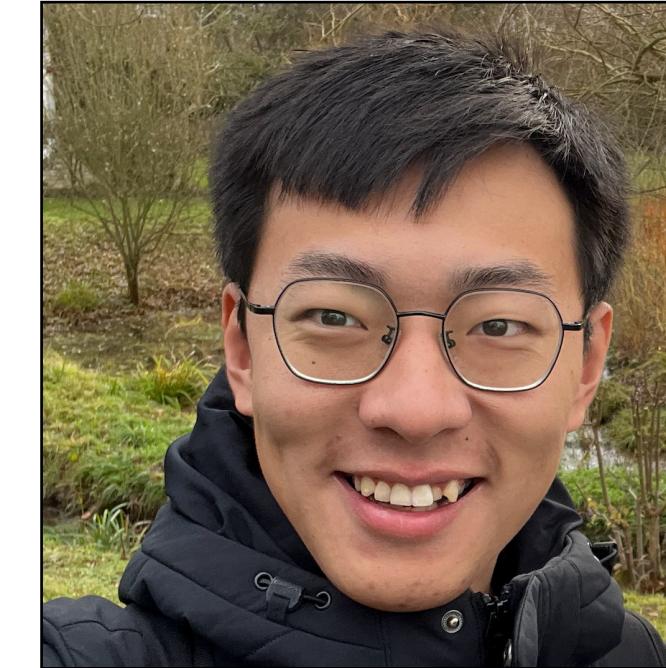
Arina Rak
AE



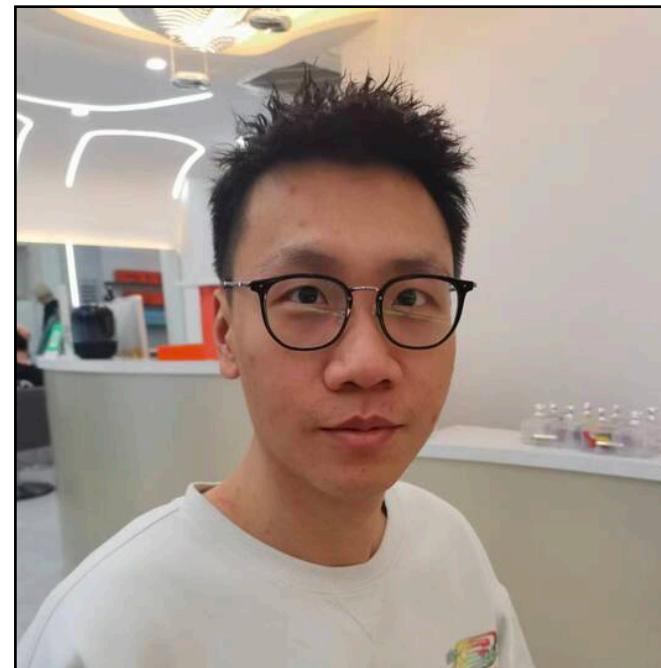
Paul Boulenger
AE



Max Conti
AE



Haotian Wu
AE



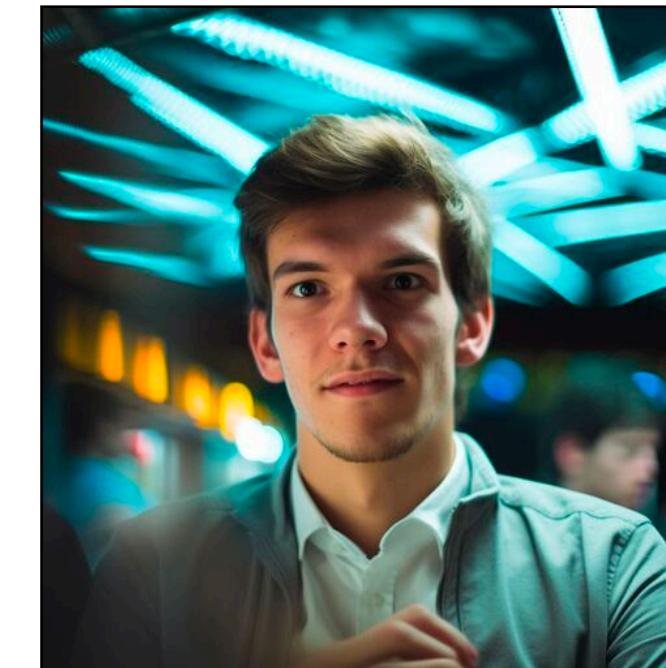
Yiyang Feng
AE



Chau Kot
AE



Ahmet Sencan
AE



Antonin Faure
AE

Prerequisites

- This course assumes some prior exposure to the following topics:
 - Comfort with Python Programming
 - Comfort with mathematical concepts: linear algebra, probability and statistics
 - Machine Learning (CS-233; CS-433 or some equivalent)
- If you're missing background in these topics, brush up on them, though I'll try keep the course as self-contained!

Logistics

- Course Webpage: <https://nlp.epfl.ch/cs-552-modern-nlp/>
 - Contains all detailed information about the course (slides, reading lists, assignments, etc.)
- Moodle: <https://moodle.epfl.ch/course/view.php?id=17143>
 - Used to point you to links to other course resources (Webpage, Ed, etc.)
- Ed: <https://edstem.org/eu/courses/1159/discussion/>
 - Will be used for class-related questions, discussion, and communication with course staff
 - Queries of a more personal or private nature can be sent to:
nlp-cs552-spring2024-ta-team@groupes.epfl.ch

Logistics

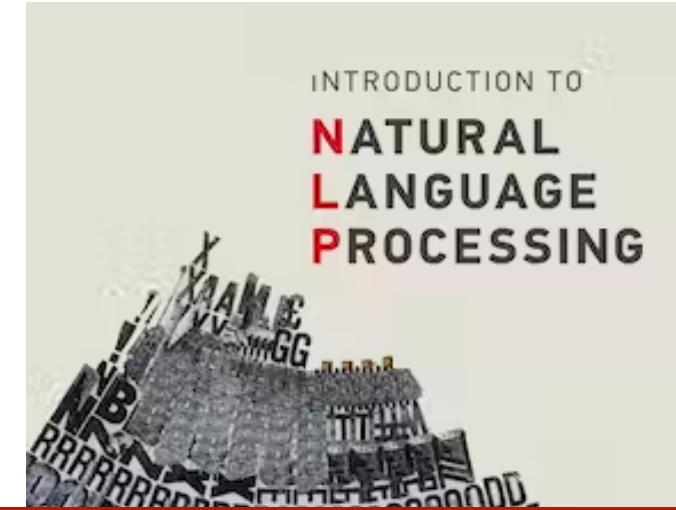
- Course Github: <https://nlp.epfl.ch/cs-552-modern-nlp/>
 - Contains all detailed information about submitting assignments and project milestones
 - Please fill out the following form to register your GitHub username & SCIPER:
<https://forms.gle/ustWbWHAXMXiLG8X9>

Lectures

- Two sessions / week
- **Wednesdays:** 11:15 AM - 1:00 PM in **STCC - Cloud C**
- **Thursdays:** 1:15 - 2:00 PM in **CE16**
- Lecture and topic schedule found on website
- Lecture slides should be posted before each lecture

Textbooks (optional!)

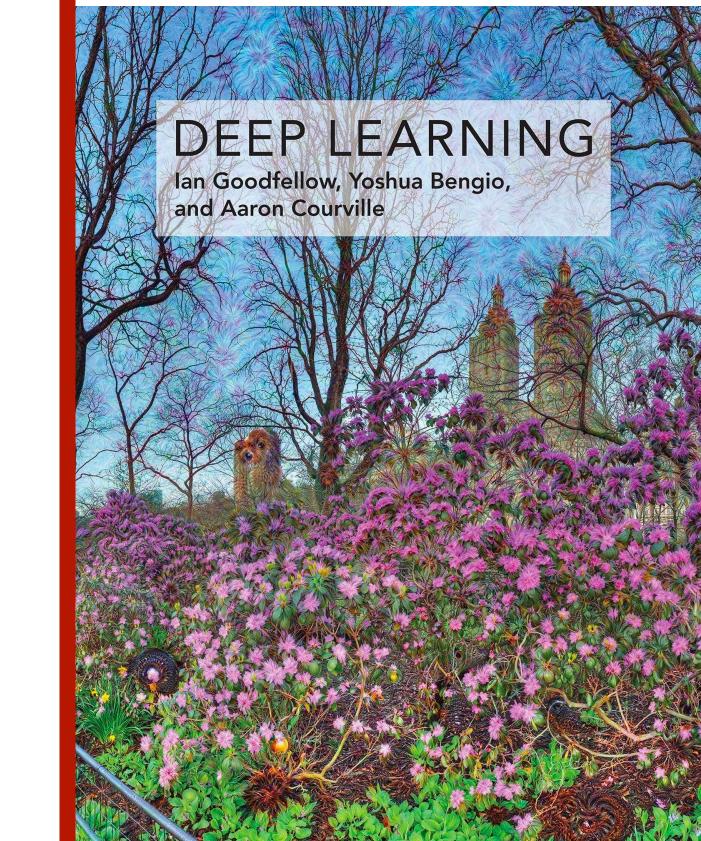
- Jacob Eisenstein. *Natural Language Processing*.



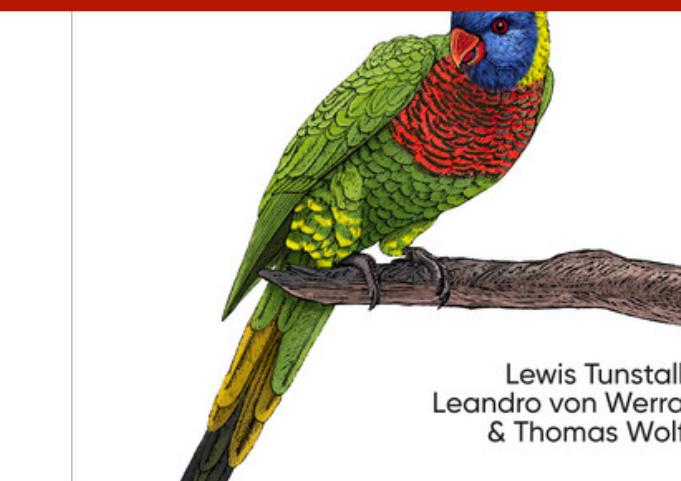
- Yoav Goldberg. *Models for Language Processing*.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*.

Caution: The NLP field moves very fast!

Books may be a bit out of date!



- Lewis Tunstall, Leandro von Werra, and Thomas Wolf. *Natural Language Processing with Transformers*.



Exercise Sessions

- **Thursdays:** 2:15 - 4:00 pm in **CE11**
- **First half of semester:**
 - **First part:** Review of previous week's exercises
 - **Second part:** New week's exercises
- **Second half of semester:**
 - Optional meetings with Project Supervisors if you have questions about the project
- 1-2 TAs + multiple AEs will be present to answer questions at each exercise session
- Schedule on website

How will you be graded in this course?

Assignments (40%)

- **Assignment 1:** Language models (10%)
- **Assignment 2:** Fine-tuning Pretrained Language Models (15%)
- **Assignment 3:** Text Generation and Scaling (15%)
- Three weeks for each assignment
- Release dates, due dates, grade release dates, etc. on course website
- Assignments will require GPU / TPU use — Google Colaboratory is a good tool

Assignment Review Sessions

- **Wednesdays:** 1:15 - 2:00 pm in **CE11**
 - See website for exact timeline
 - **First half of semester:**
 - TAs answer questions you have about assignments
 - **Second half of semester:**
 - TAs that graded your assignments will be available to answer questions about your grades
- 1-2 TAs + AEs will be present to give guidance about the assignment
 - Please post questions to the Ed discussion board ahead of time. You may get an answer more quickly and all students can benefit from the response

Assignment Schedule (at a Glance)

	Assignment Release Date	Assignment Due Date	Assignment Grades Released	Grade Review Session 1	Grade Review Session 2
A1	2/28/2024	3/17/2024	4/14/2024	4/18/2024	4/25/2024
A2	3/20/2024	4/7/2024	5/5/2024	5/9/2024	5/16/2024
A3	4/3/2024	4/21/2024	5/19/2024	5/22/2024	5/30/2024

- **Grade Review Requests** can also be made by
- **Grade Review is a two-way street:** If you ask for your grade to be reviewed and you were given extra points, they will be removed.
- **No additional grade reviews will be granted after the review sessions.**

Final Project (60%)

- Complete in teams of 3
- Deliverables (due dates on website):
 - 2 Milestones (30%)
 - Final report, code, and data (30%)
- Topic: **Create your own ChatGPT !**
- **More details in following weeks!**

Late Deliverable Policy

- All assignments and milestones are due at **11:59 PM** on their due date
- You get 6 late days throughout the semester to turn in your assignments and project milestones late.
 - Late day is used as soon as the clock strikes 12:00 AM
 - Project is done in teams. **The entire team is penalised by three late days.**
- **Final project must be handed in on June 14th. No exceptions.**

Course Integrity Policy

- For the assignments, you **should not use outside codebases** unless explicitly allowed by the course staff in the assignment description. **You should avoid using ChatGPT or other AI-based tools for any assignment or part of your project. Any use of ChatGPT and other AI-based tools must be cited and mentioned.**
- For the project, you may build your work upon existing open-source codebases, but are **required to write new code** to perform your experiments. In the project, clearly specify your team's contributions and how they differ from the pre-existing codebase in your milestone reports and final report.
- You are free to discuss ideas and implementation details with other project teams. However, you **should not look at another team's code**, or incorporate their code into your project (unless explicitly allowed by the course staff).

- **Natural Language Processing** Lab
 - Master's Theses, Semester Projects available every term
- Other **NLP** courses
 - **Fall 2024:** Introduction to Natural Language Processing (6 credits)
 - ▶ Lectures, Quizzes, Final Exam
 - **Fall 2024:** Topics in Natural Language Processing (2 credits)
 - ▶ Paper reading, paper reviewing, discussion

Let's have a great semester!