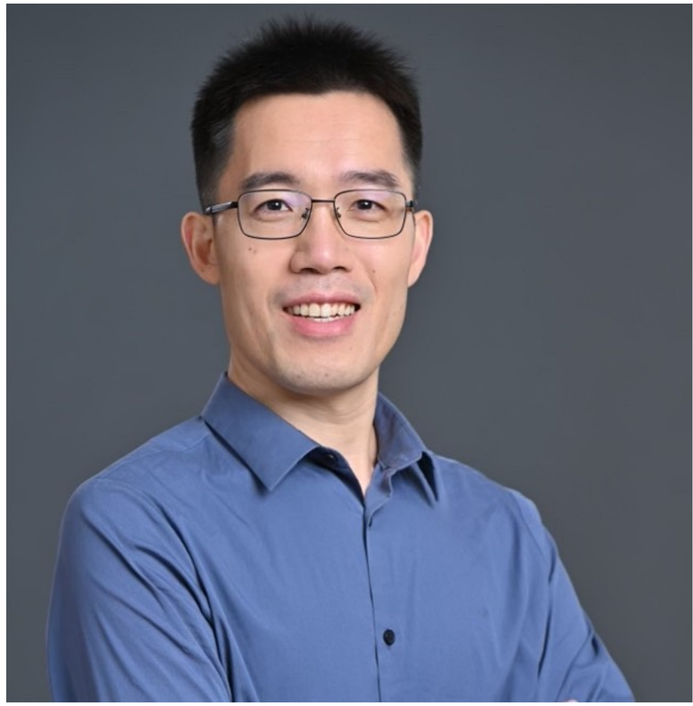
Fundamentals of Speech and Language Processing

CSC3160/MDS6002

Zhizheng Wu

Lecture 1: Course introduction

Myself



- Associate professor joined in Aug 2022
  - https://drwuz.com/
  - Email: wuzhizheng@cuhk.edu.cn
  - CTE: 5.9 of 6.0

- Ex-Meta, ex-Apple, ex-Microsoft, ex-JD.COM

- Associate editor of IEEE/ACM Transactions on Audio Speech and Language Processing
- Member of the IEEE Speech and Language Processing Technical Committee Co-founder of ASVspoof challenge, Voice conversion challenge
- Organizer of Blizzard challenge 2019

Course logistics

- Instructor: Zhizheng Wu
- TAs: Xueyao Zhang, Xi Chen

- Course website: https://drwuz.com/CSC3160/

- Lecture time and location
  - Tuesday/Thursday 4:00PM - 5:20PM in TB103 (Online for the first two weeks)
- Tutorials
  - There will be five tutorials. The specifics are available on the course website.

- Office hours
  - Zhizheng Wu: **Thu 2-3PM**. Daoyuan Building 321b
  - Xueyao Zhang: **Wed 7-9PM**. SDS Research Lab (4th Floor, Zhi Xin Building) Seat No.78
  - Xi Chen: **Wed 7-9PM**. SDS Research Lab (4th Floor, Zhi Xin Building) Seat No.100

Communication and feedback

- We will send out two course feedback surveys during the semester (0.5% credit each)

- Feel free to send me or TAs any feedback regarding the course
  - Both the instructor and TAs can **possibly make mistakes**! Communication will help

- Email is the preferred way for communication. BB is encouraged.

> Hands down top five instructors I've ever had. Prof. Wu taught really clearly & simply and he's also very enthusiastic, thus keeping me motivated throughout the semester. He's also very open to questions and feedback unlike any other instructor I had before. Prof. Wu as an academic and professional has inspired me since the start of this semester. As for the course, I think this course is well-built. Workload is still okay, and the exams too. No suggestions from me for now.

Presuming prior knowledge

- **Solid background** of python programming
- PyTorch is suggested, but you can learn during the course
- Knowledge of machine learning is preferred. If you don’t have machine learning background, we will have one lecture on machine learning, you can learning during the semester (basic classification or regression)

- LaTex + GitHub

Grading (details are available on course website)

- Assignment (30%)
- Midterm exam (25%)
- Final project (40%)
- Participation (5%)
  - Guest lecture attendance
  - Course evaluation

- LaTex + GitHub (provide a tutorial on GitHub)
  - Code needs to be tracked in GitHub
  - All reports need to write in LaTex. Template will be provided.
    - Template: https://www.overleaf.com/read/fvhykwvngdwz

Assignment (30%)

- Assignment 1 (12%): first assignment due Feb 14
  - Speech alignment and audio synthesis
- Assignment 2 (8%): second assignment due Feb 23
  - Text processing
- Assignment 3 (10%): due Mar 21
  - Word embedding and classification

  - **All assignments will be available by Jan 17th**

Final project (40%)

- Project proposal (2%): Mar 2nd due
- Project milestone (2%)
  - Mar 30th: Project milestone 1 due
  - Apr 18th: Project milestone 2 due
- Project poster (5%)
  - Peer review (1%): evaluating other students' poster, each evaluate 4 posters
  - Expert grading (4%)
    - Average of at least three expert gradings (professors or research scientists from companies)
- Reproducibility (1%)
    - Visual inspection of code quality and readability

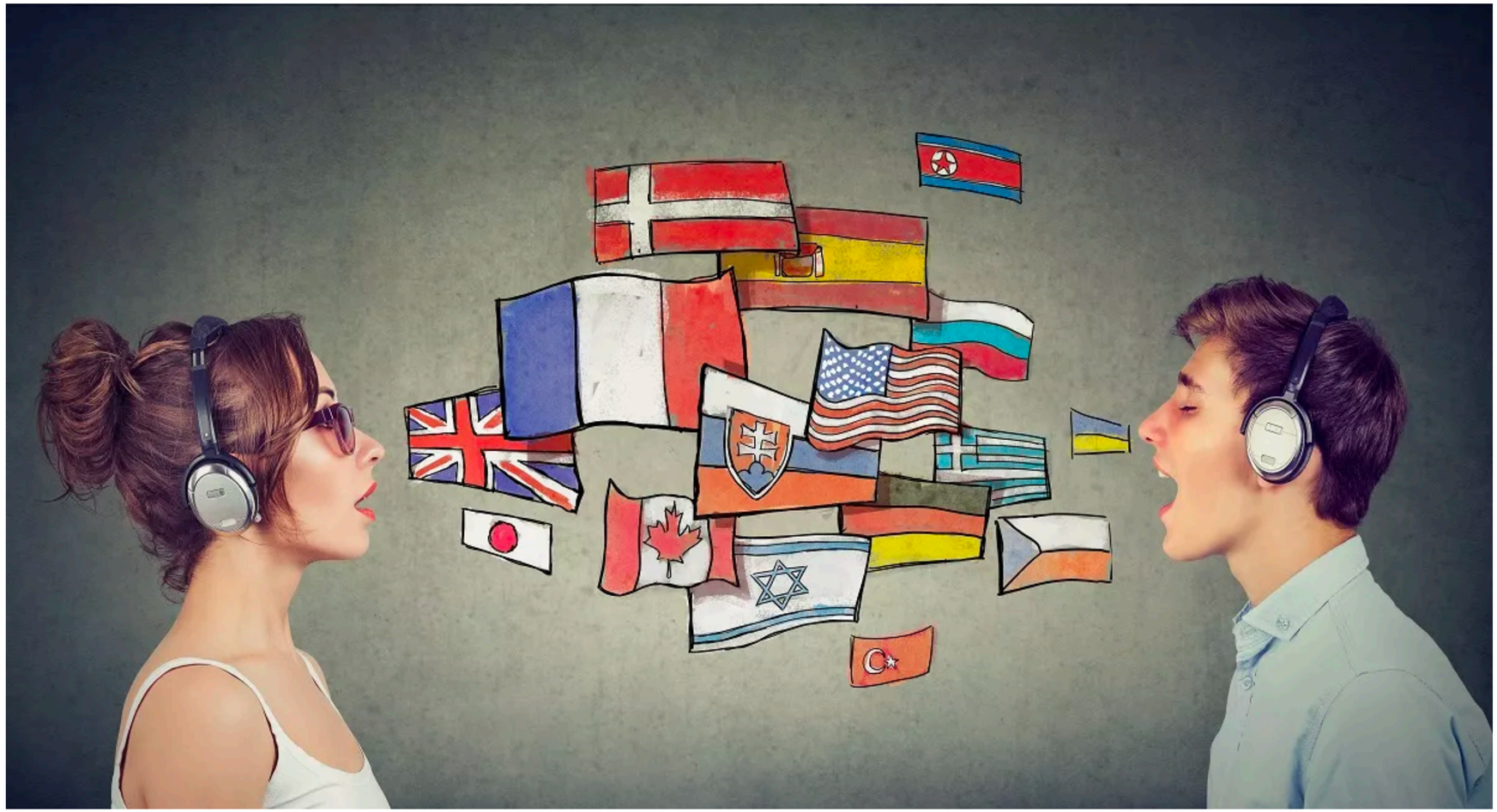
Final project (40%)

- Project report (30%)
  - Overall quality of report (5%)
  - Originality and creativity (1%)
  - Reference to Prior Work (3%)
  - Technical Correctness (4%)
  - Experimental validation (5%)
  - Analysis quality (6%)
  - Clarity of writing (3%)
  - Peer review (3%)
    - Constructive review to other reports. Each review 6 reports. The reviews are NOT used to rate other reports.

Final project

- Individual project: A few projects will be provided for you to choose.
- Customized project: Subject to approval from the teaching team
- Group project is OK, but need approval from the teaching team. The scope of a group project will be larger than an individual project
- MDS6002
  - Group project is encouraged! Team size: 2-3 students.

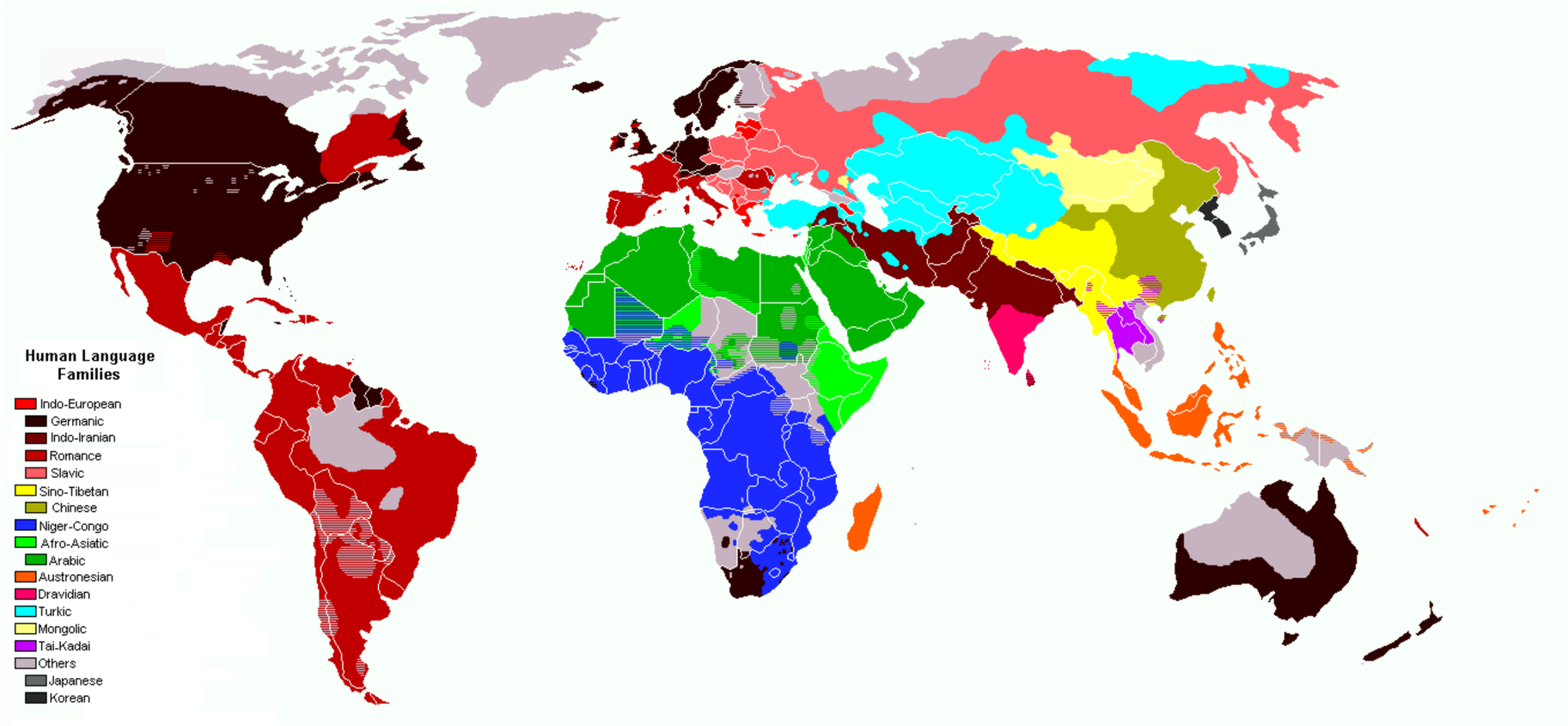
Human language



https://news.yahoo.com/did-humans-first-start-speak-152613750.html

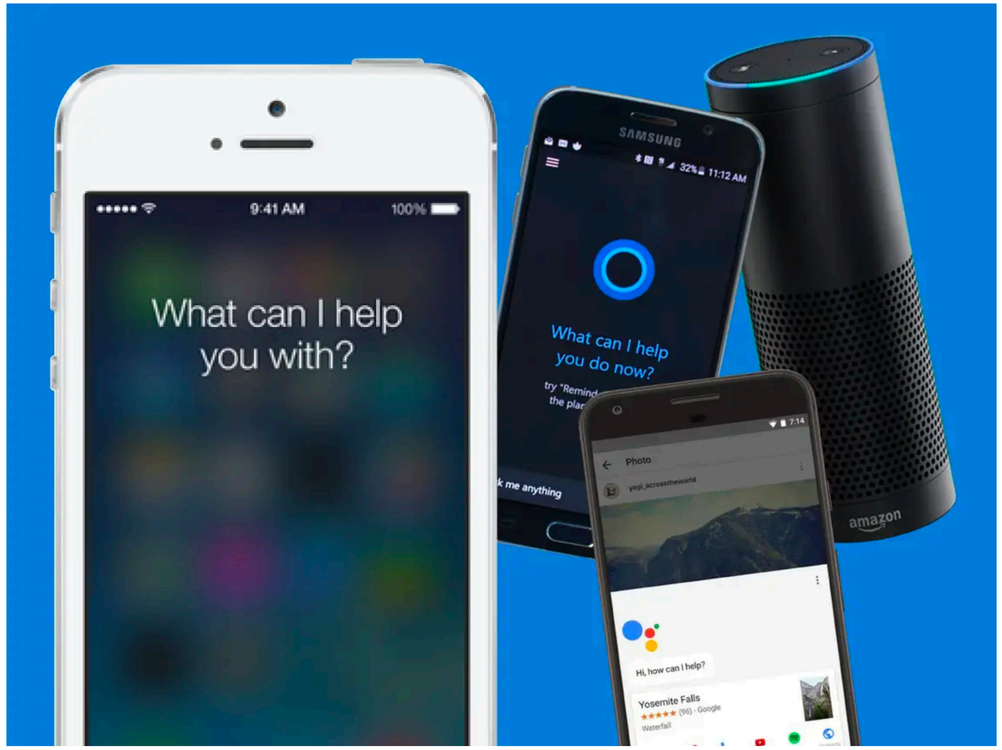
Languages in the world

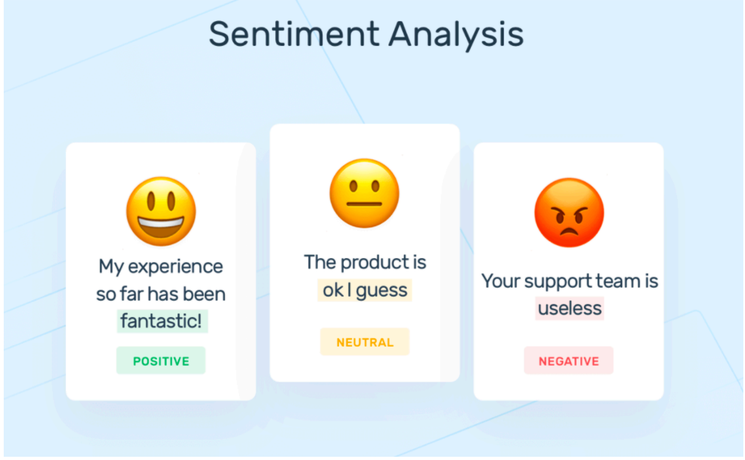
- About 7,000 languages spoken as of 2010. More than half of them have no written form



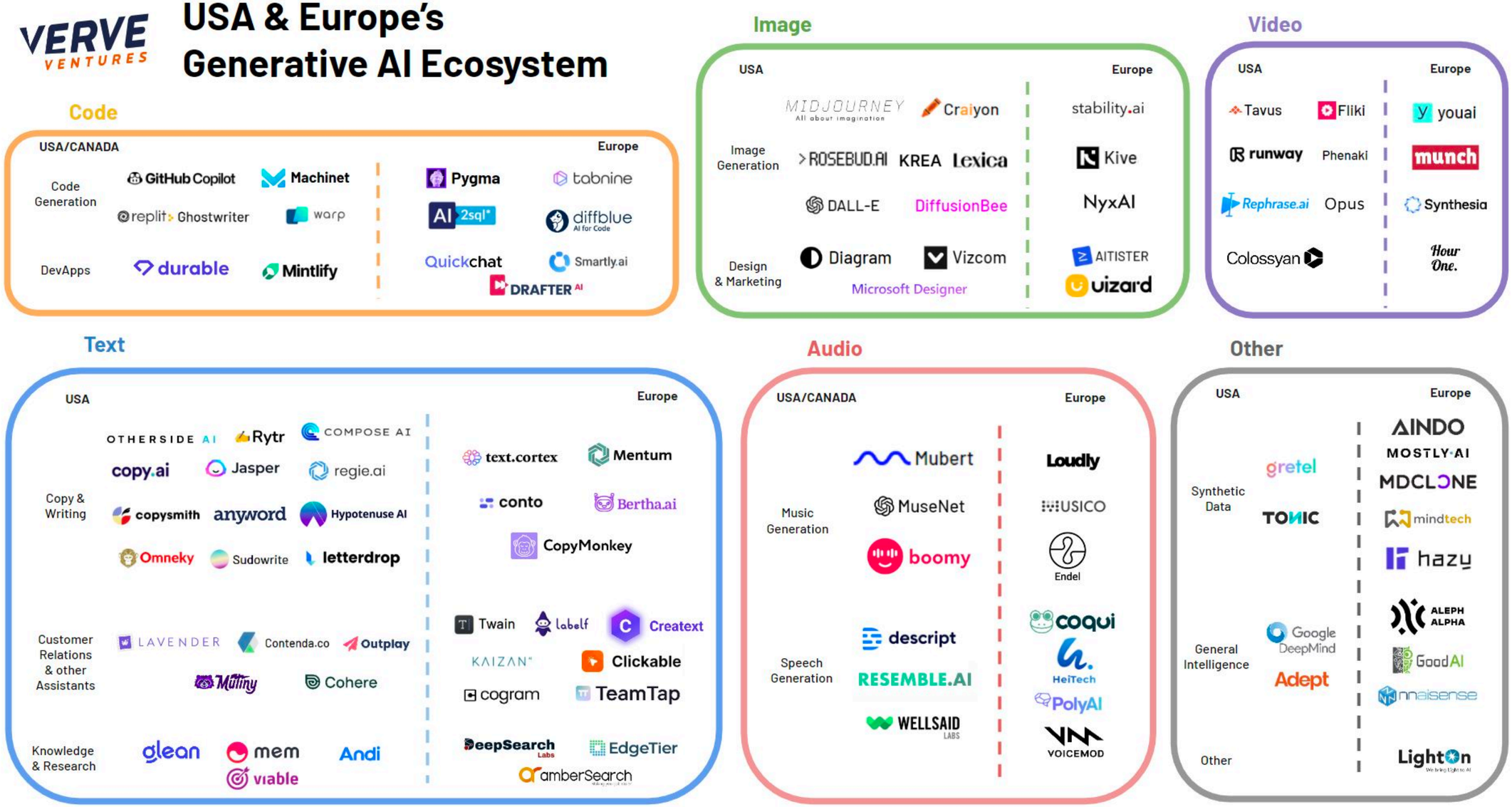
https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers

Applications of speech and language processing





Applications: Generative AI startups



What is this course about?

- Natural language can be ***speech*** or ***text,*** in other words, in ***spoken*** form or ***written*** form (NLP != Text processing)

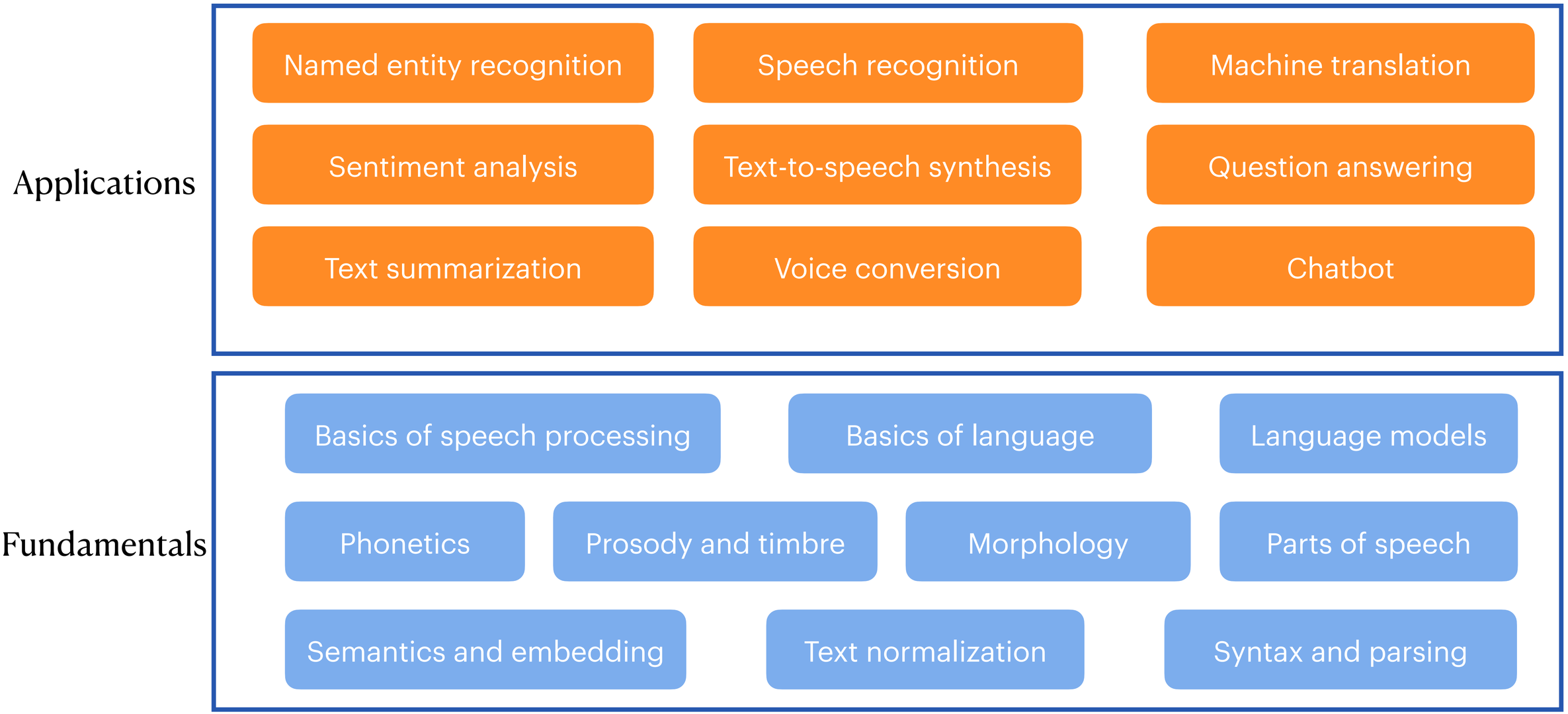
- First half: Fundamentals knowledge of speech signals and language elements
  - Fundamentals of speech processing
    - Spectrogram, prosody, pronunciation, etc
  - Fundamentals of text processing
    - Language models, word embedding, syntax, tokenization, etc

- Second half: Applications of speech and language processing
  - Speech recognition, synthesis, question answering, chatbot, etc

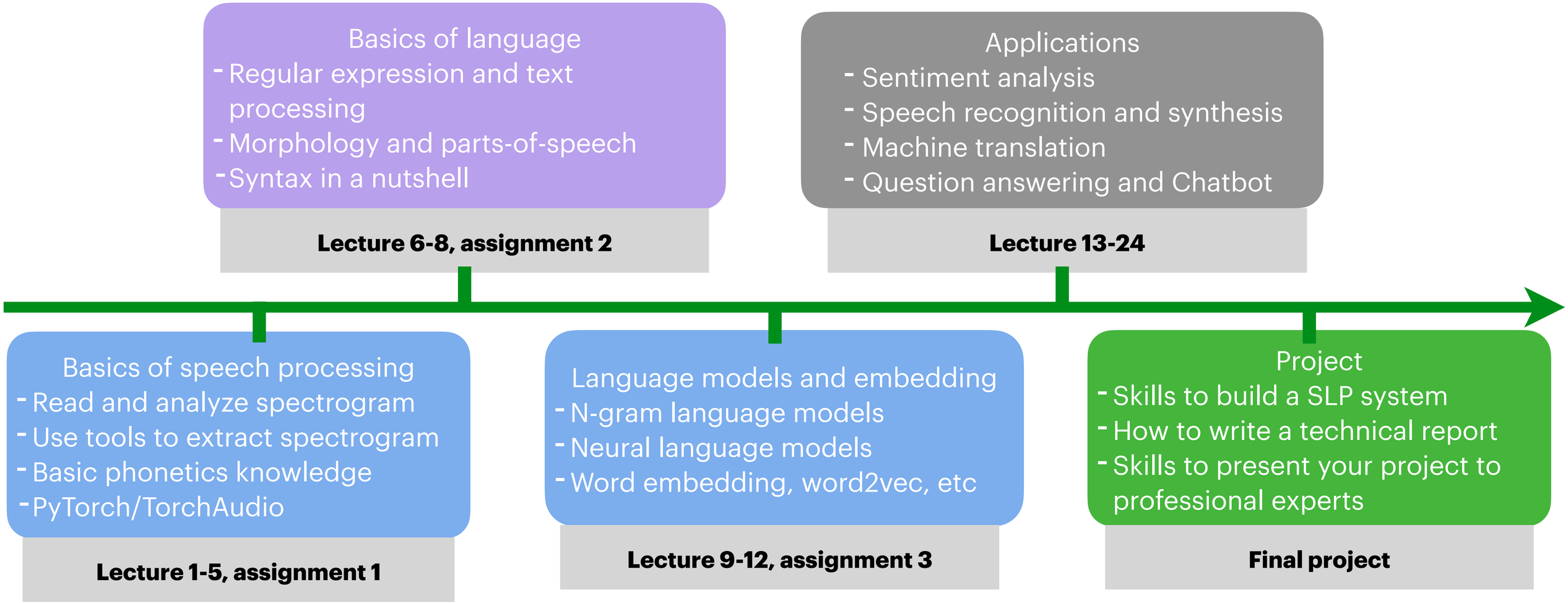
Learning objectives

- Upon successful completion, students should be able to
  - Explain fundamental knowledge of speech and language processing
  - Understand and build systems for some projects in SLP
  - Organize and implement a speech and language project in a business environment
  - Interpret the results of a speech and language project

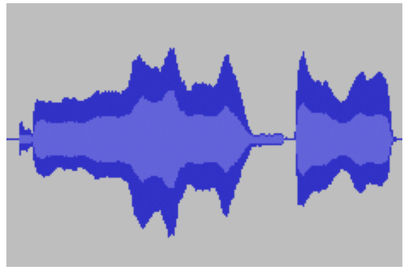
One slide overview

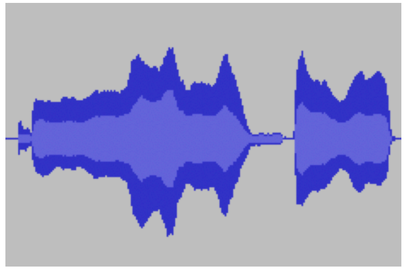


Learning progression



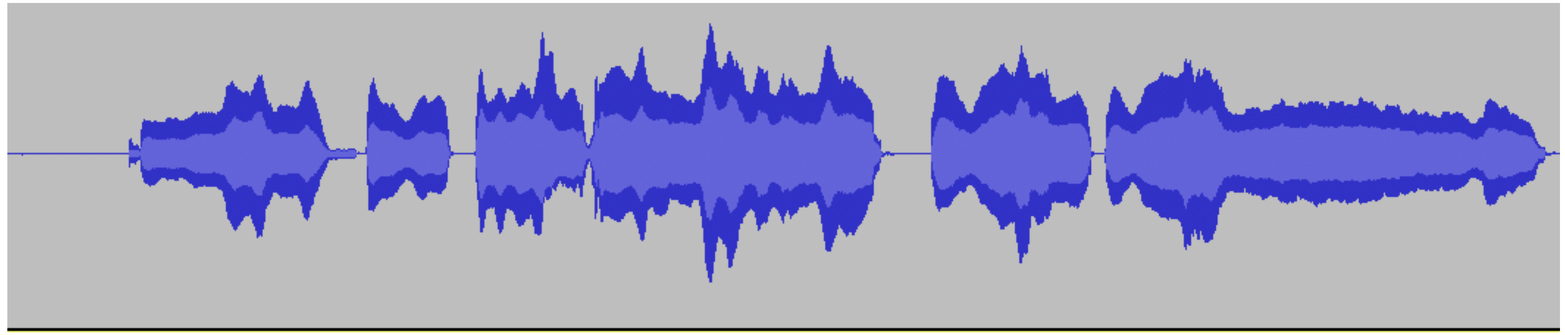
Lecture 2: Machine learning in a nutshell

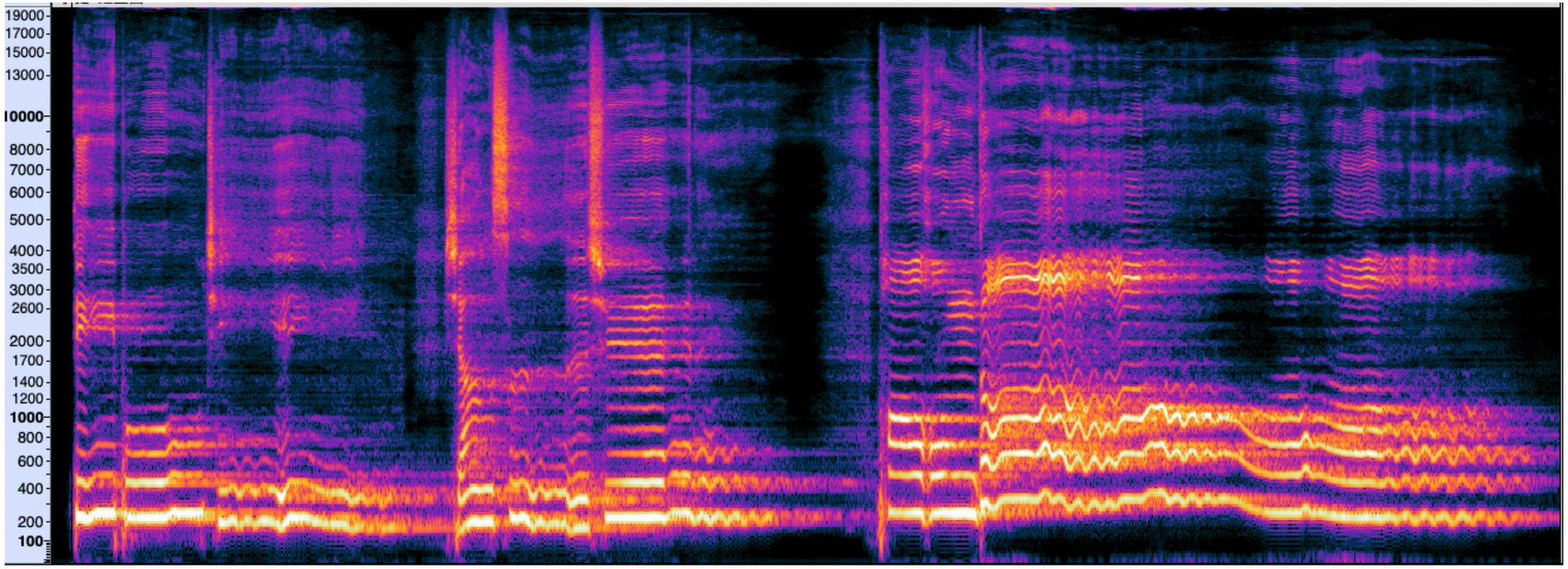
$f($  $) =$ “Hello world”

$f($“Hello world”$) =$ 

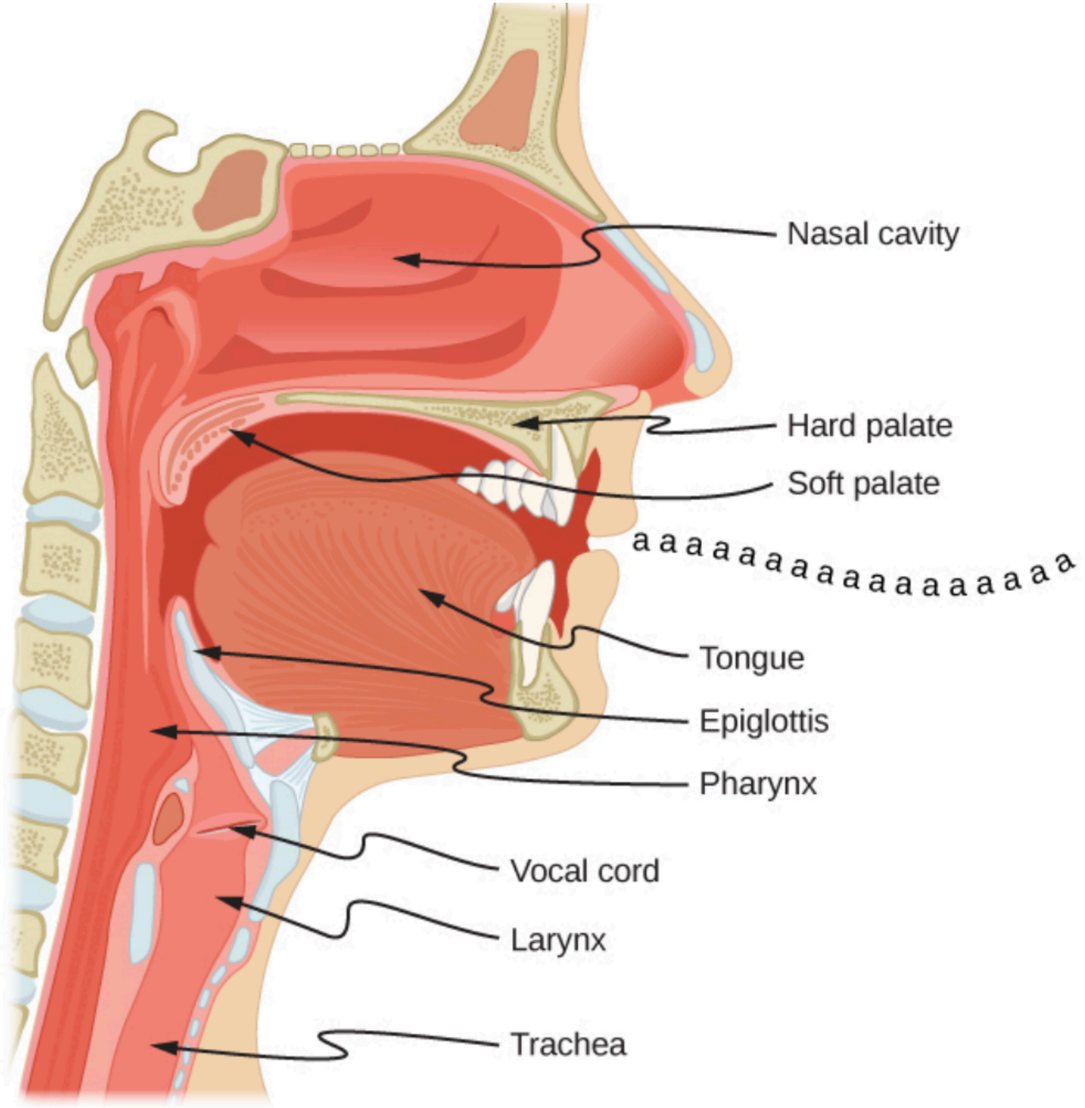
$f($ “你好” $) =$ “Hello world”

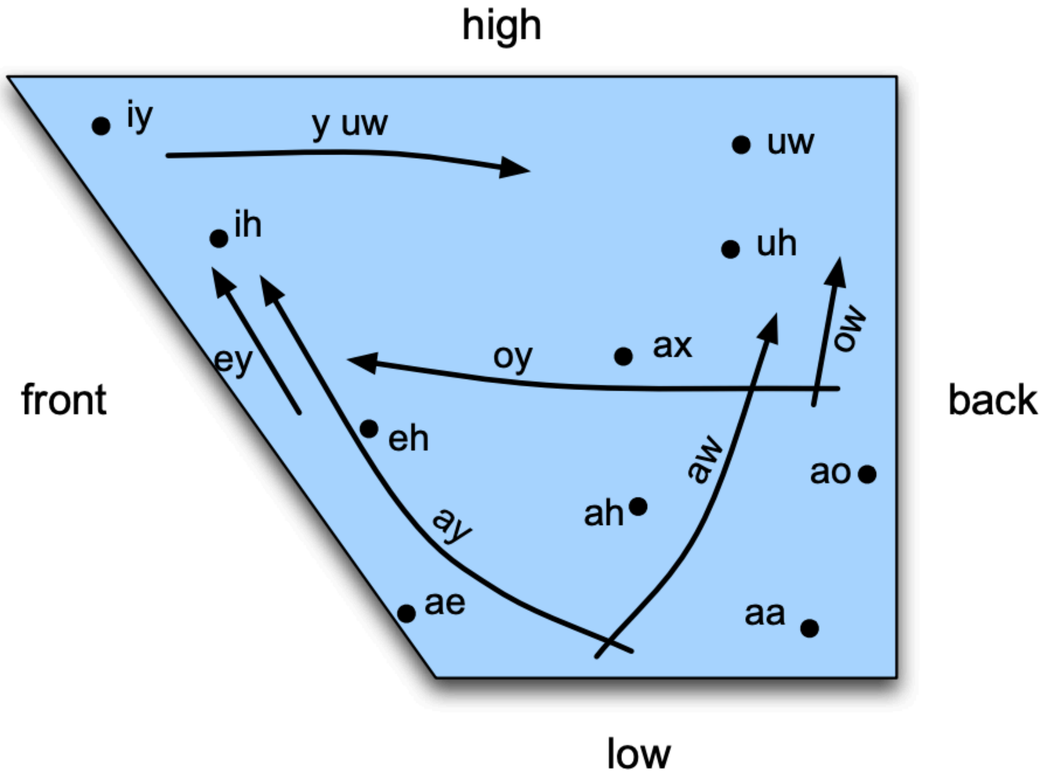
Lecture 3-4 Understanding human speech





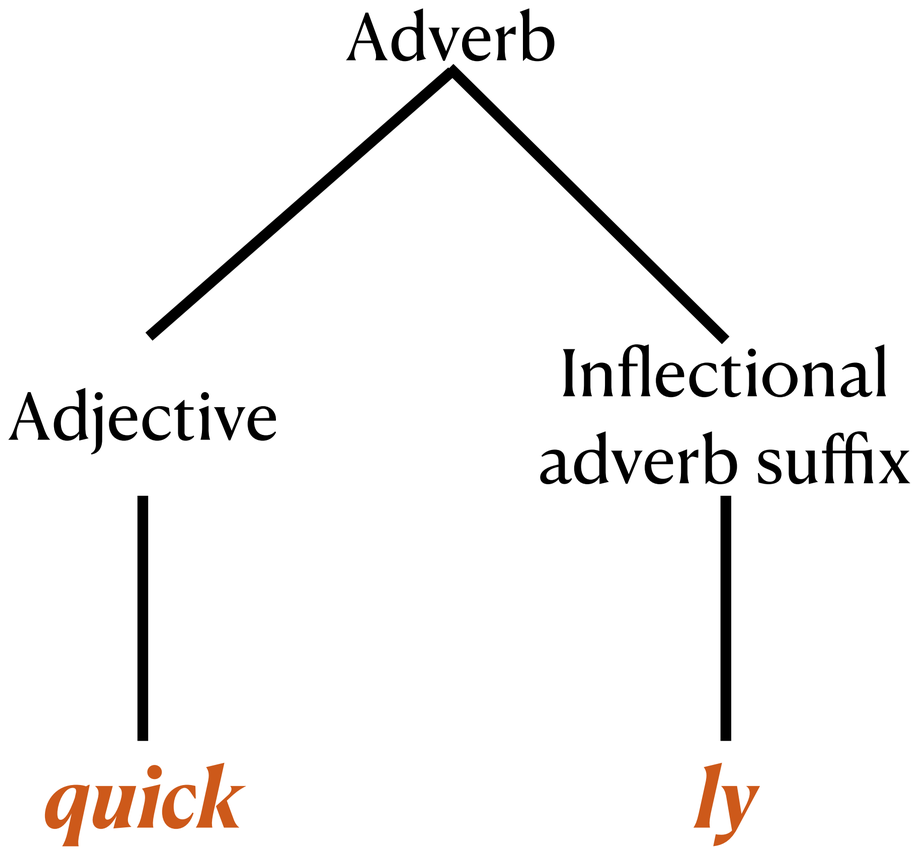
Lecture 5: Human sounds and their organization



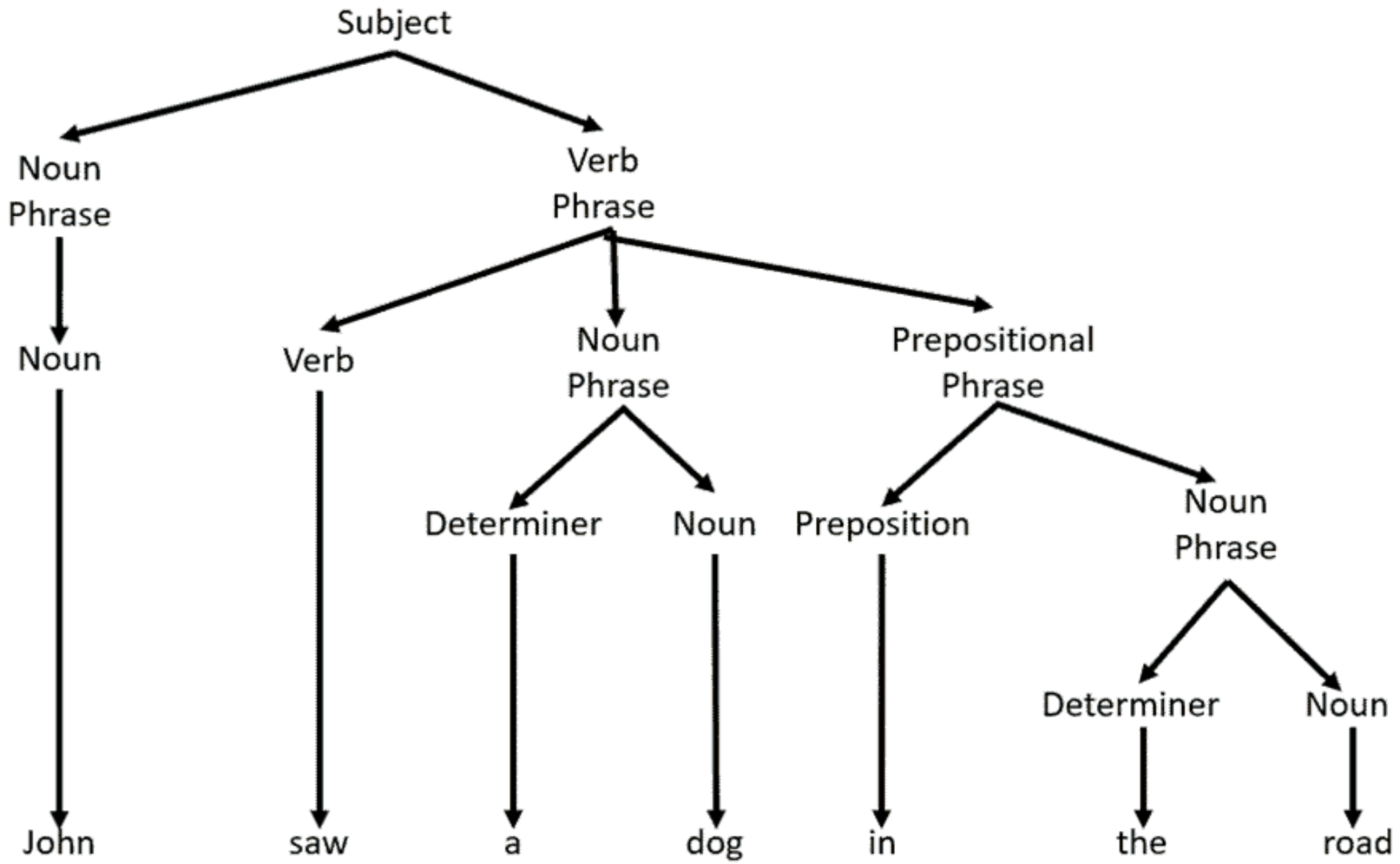


Lecture 6: Text processing and regular expressions

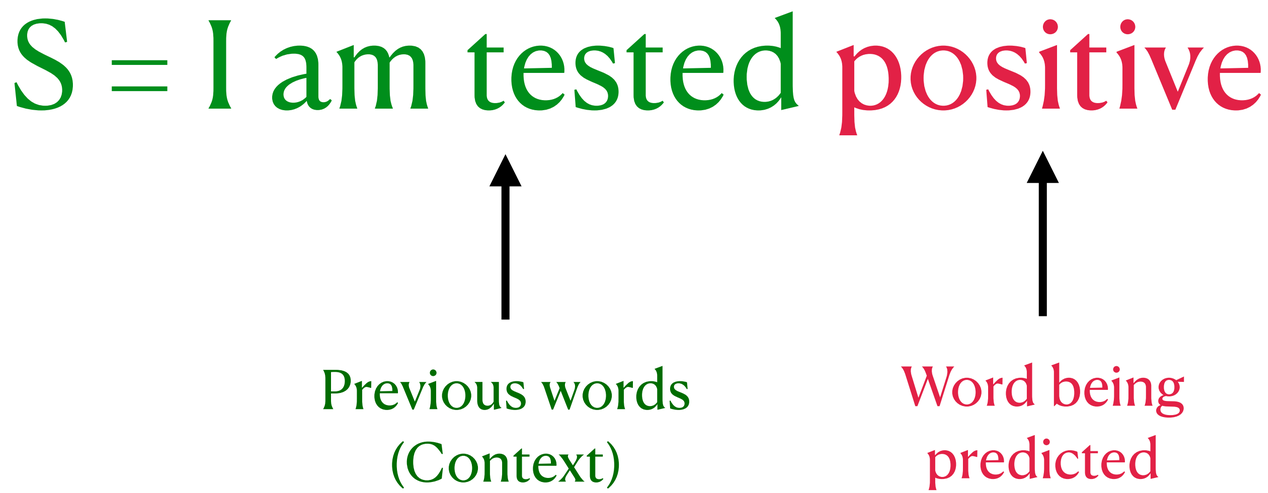
Lecture 7: Words, parts of speech and morphology



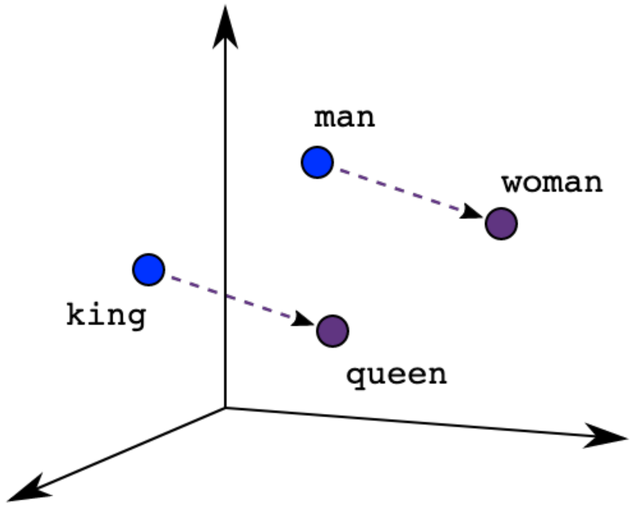
Lecture 8: Syntax - Structure of sentences



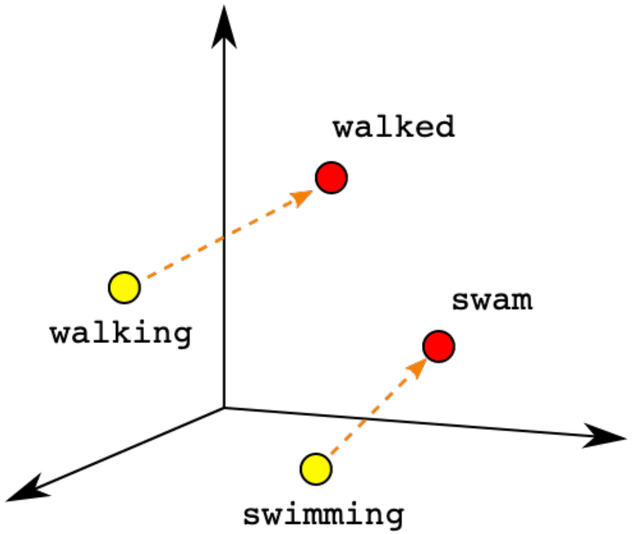
Lecture 9&10: Language models



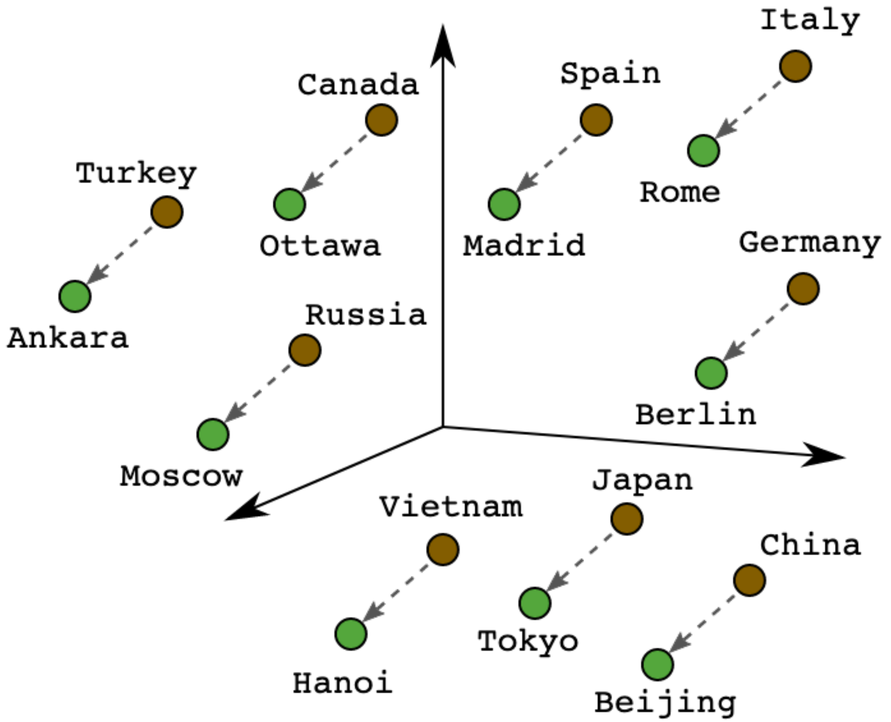
Lecture 11&12: Embedding



Male-Female



Verb Tense



Country-Capital

https://developers.google.com/machine-learning/crash-course/embeddings/translating-to-a-lower-dimensional-space

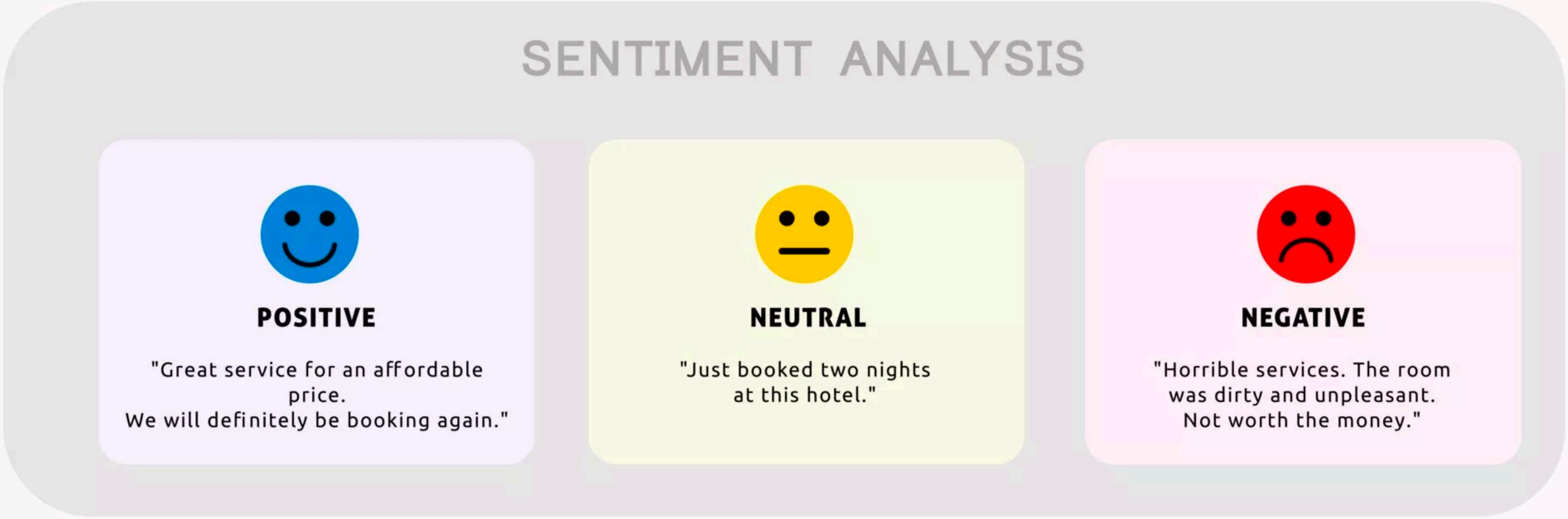
Lecture 13: Named entity recognition

In fact, the Chinese NORP market has the three CARDINAL most influential names of the retail and tech space – Alibaba GPE , Baidu ORG , and Tencent PERSON (collectively touted as BAT ORG), and is betting big in the global AI GPE in retail industry space . The three CARDINAL giants which are claimed to have a cut-throat competition with the U.S. GPE (in terms of resources and capital) are positioning themselves to become the 'future AI PERSON platforms'. The trio is also expanding in other Asian NORP countries and investing heavily in the U.S. GPE based AI GPE startups to leverage the power of AI GPE . Backed by such powerful initiatives and presence of these conglomerates, the market in APAC AI is forecast to be the fastest-growing one CARDINAL , with an anticipated CAGR PERSON of 45% PERCENT over 2018 - 2024 DATE .

To further elaborate on the geographical trends, North America LOC has procured more than 50% PERCENT of the global share in 2017 DATE and has been leading the regional landscape of AI GPE in the retail market. The U.S. GPE has a significant credit in the regional trends with over 65% PERCENT of investments (including M&As, private equity, and venture capital) in artificial intelligence technology. Additionally, the region is a huge hub for startups in tandem with the presence of tech titans, such as Google ORG , IBM ORG , and Microsoft ORG .

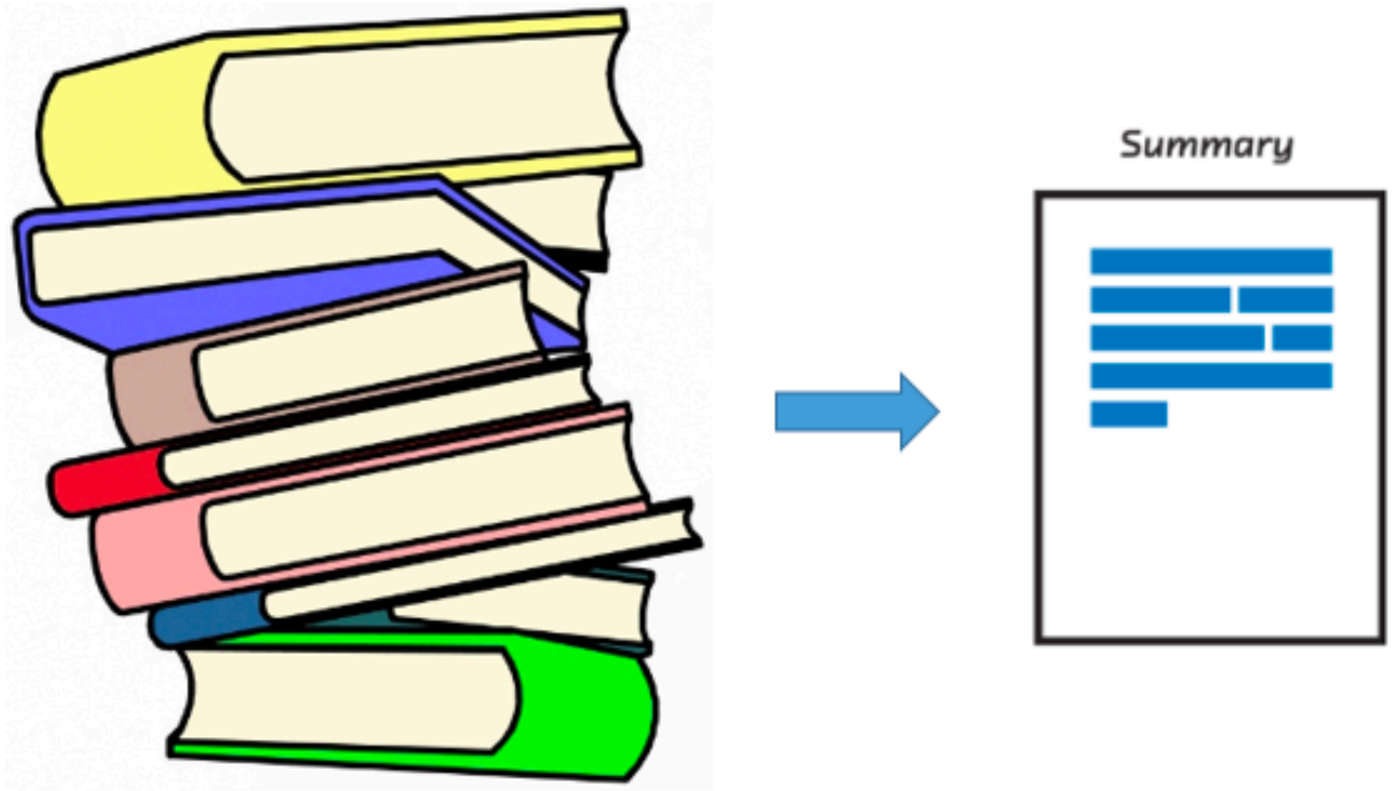
https://umagunturi789.medium.com/everything-you-need-to-know-about-named-entity-recognition-2a136f38c08f

Lecture 14: SLP Application - Sentiment analysis



https://www.expressanalytics.com/blog/social-media-sentiment-analysis/

Lecture 15: SLP Application - Text summarization

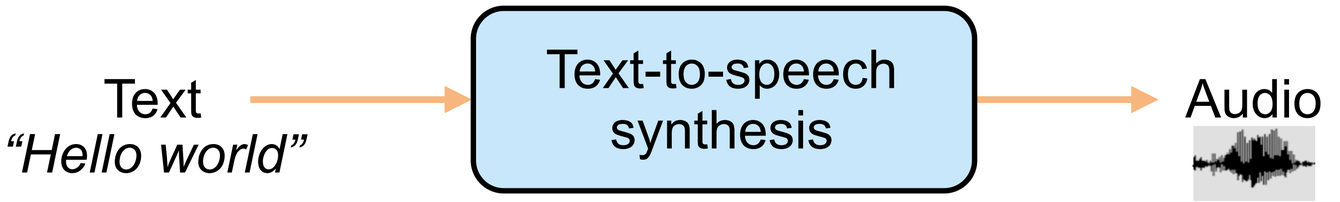


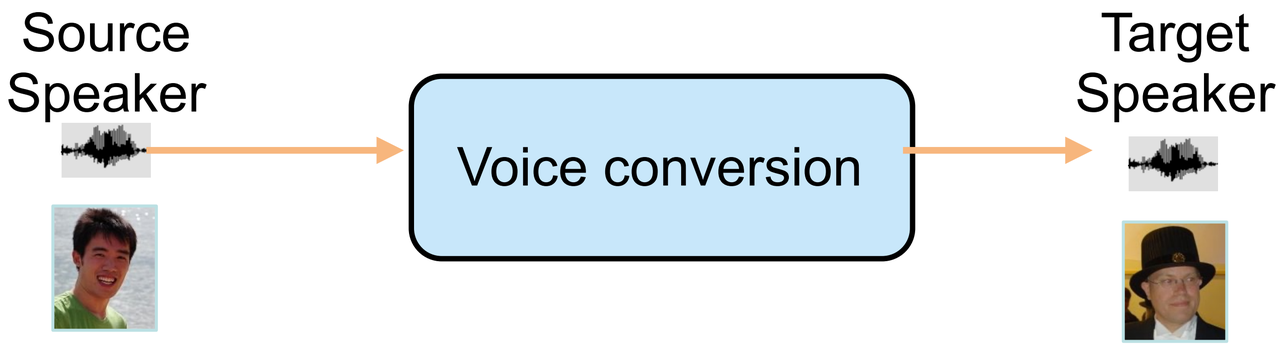
https://www.analyticsvidhya.com/blog/2018/11/introduction-text-summarization-textrank-python/

Lecture 16-17: Speech recognition

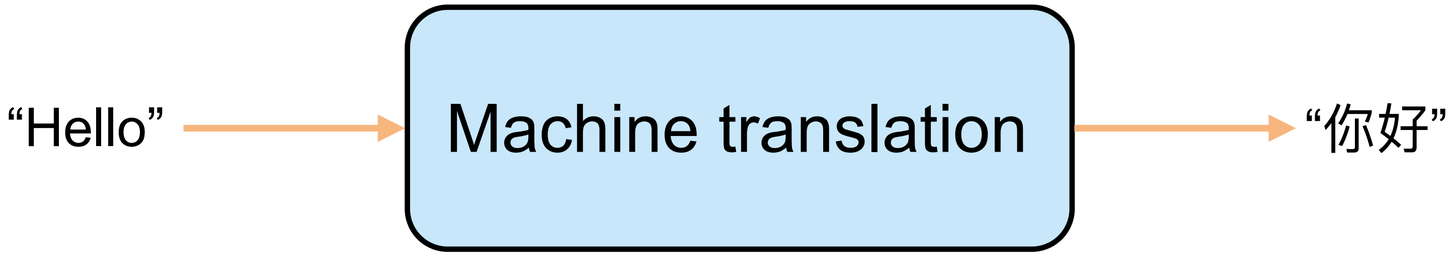
https://developer.nvidia.com/blog/solving-automatic-speech-recognition-deployment-challenges/

Lecture 18-19: Speech synthesis and voice conversion

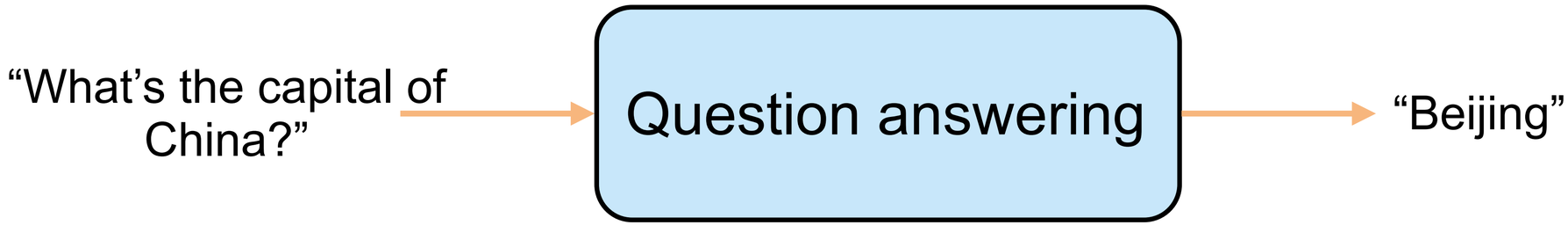




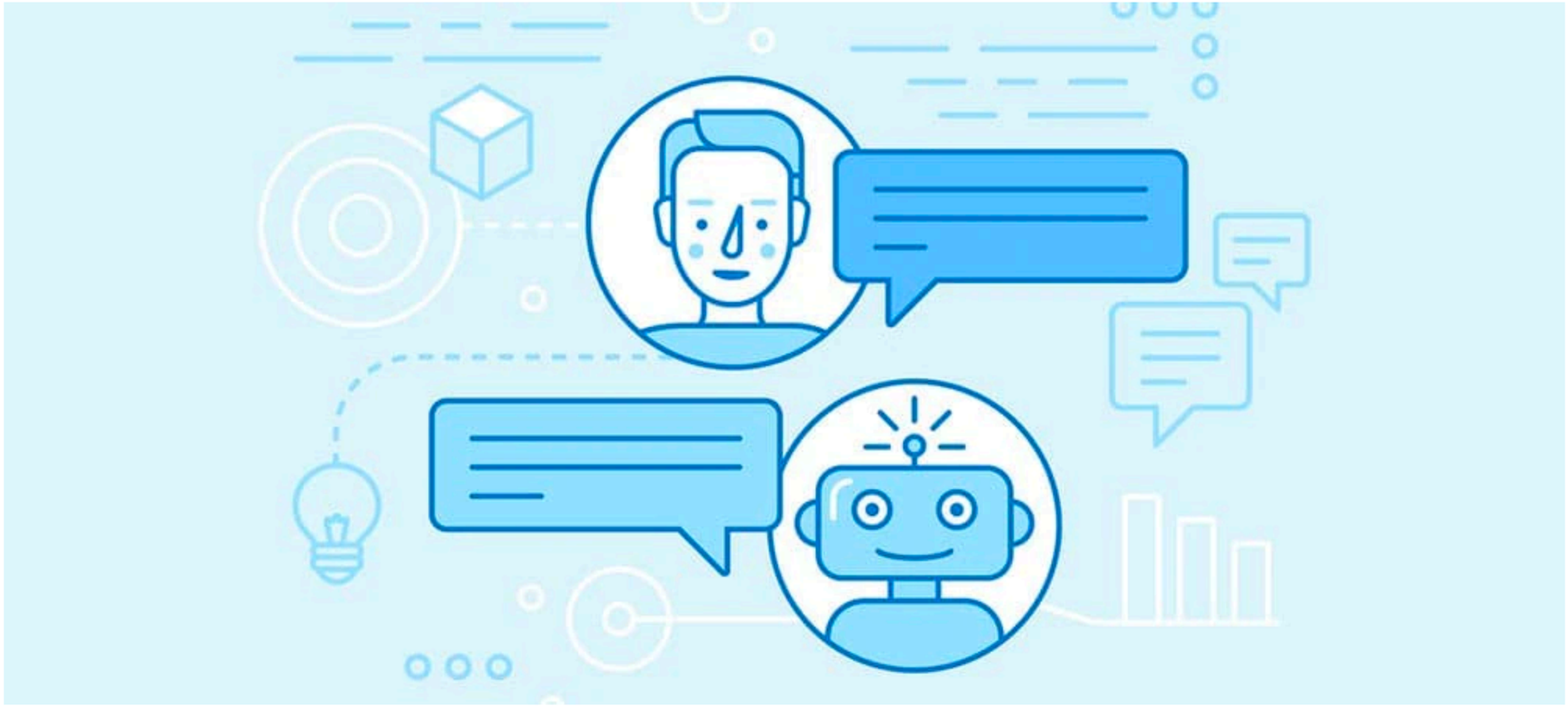
Lecture 20: Machine translation



Lecture 21: Question answering

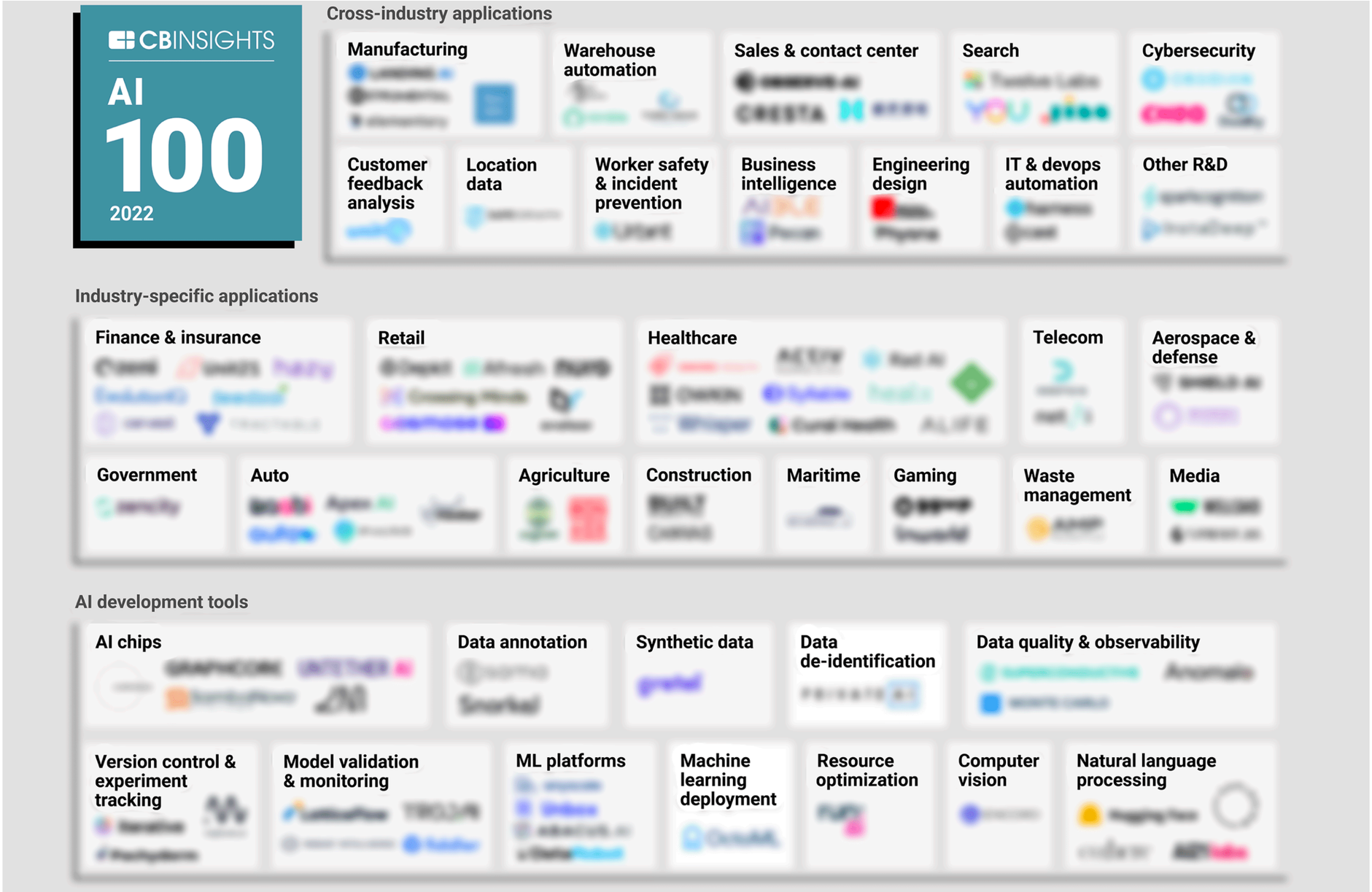


Lecture 22: Chatbot

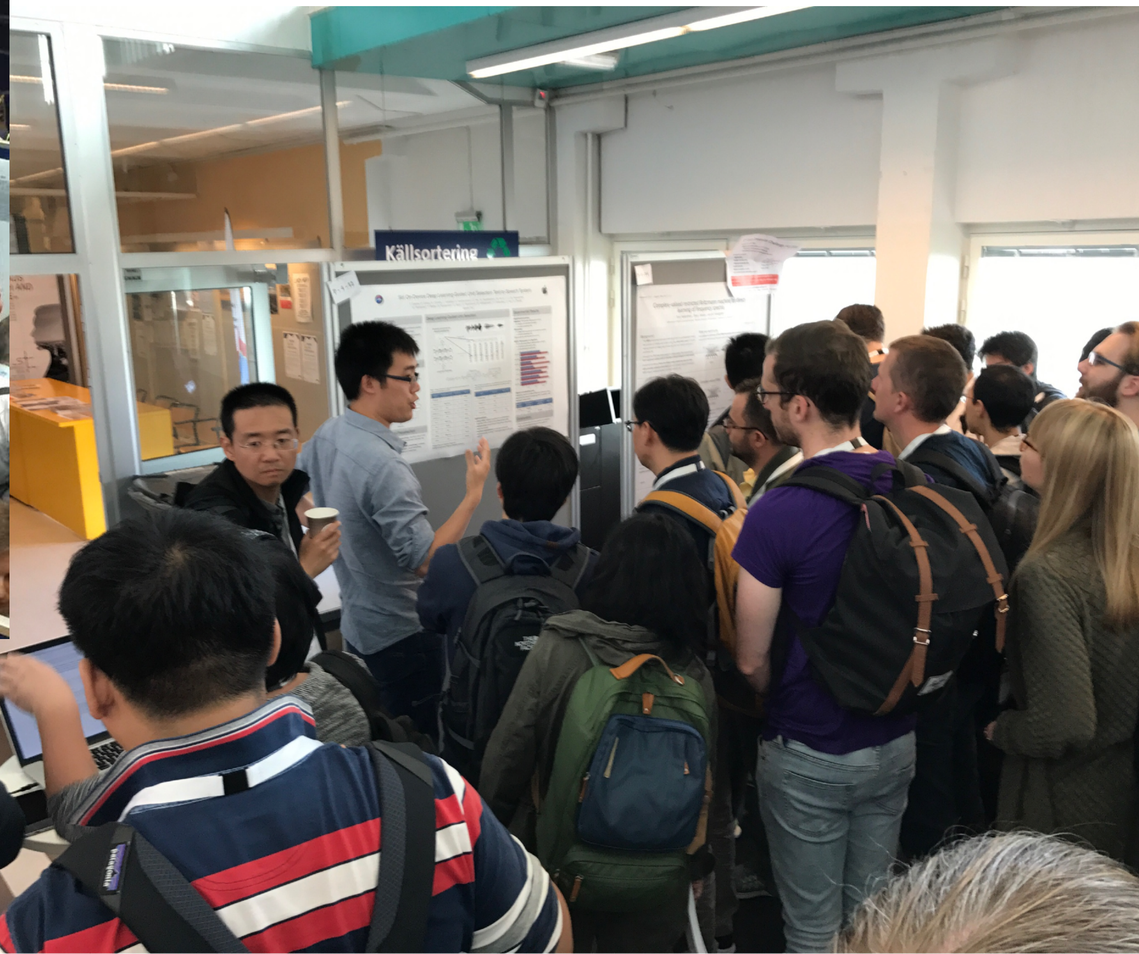


https://www.hp.com/us-en/shop/tech-takes/what-is-a-chatbot

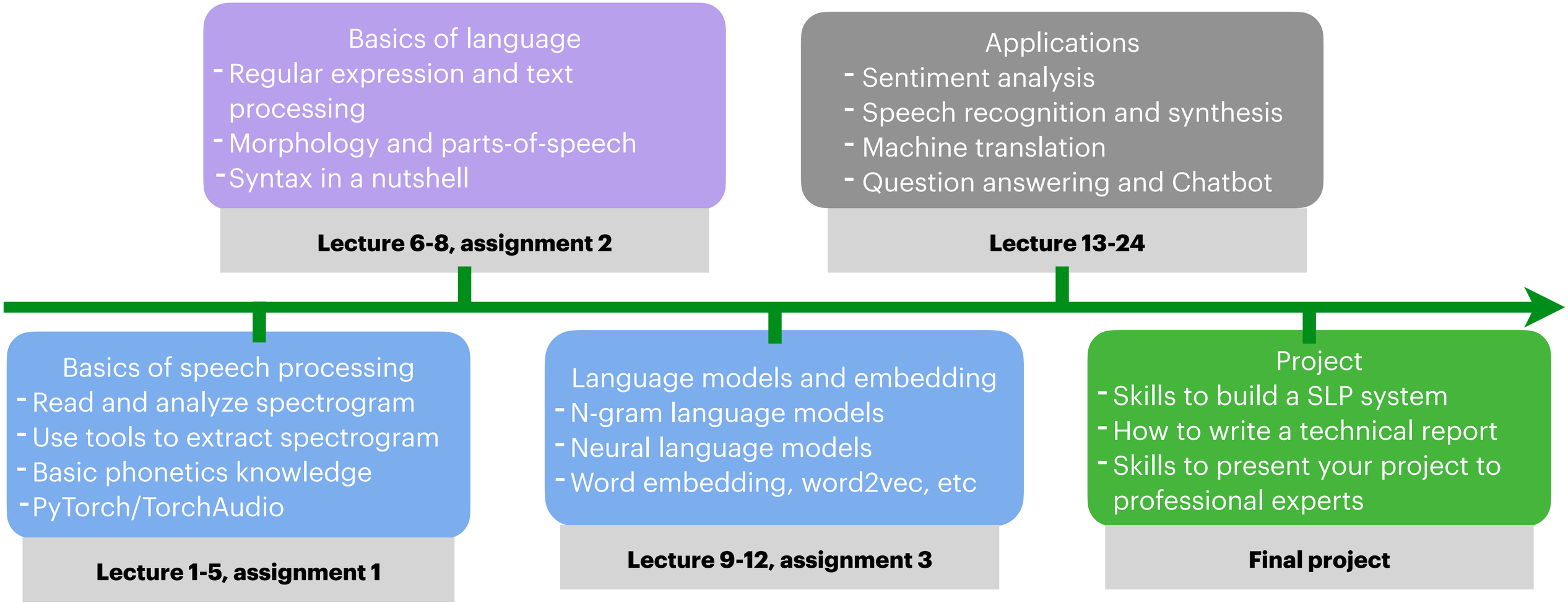
Lecture 23-24: Industry applications of SLP (Invited Lectures)



Poster session: May 20th (Tentatively)



Learning progression




no rain, no flowers

香港中文大學(深圳) 数据科学学院
The Chinese University of Hong Kong, Shenzhen | School of Data Science

Thanks!

Zhizheng Wu
Associate professor
https://drwuz.com/

Course website: https://drwuz.com/CSC3160/