

# LING3401 Linguistics and Information Technology

## Tutorial: Demonstrating NLP tasks and applications

Yige Chen

The Chinese University of Hong Kong

January 8, 2025





# Before We Start...

- This is the tutorial session of **LING3401 Linguistics and Information Technology**
- You are now at **Lee Shau Kee Building 302**
- Make sure you are in the right tutorial session



# About Me

- Yige Chen
- Third-year PhD student in Linguistics at CUHK
- Research interests: Natural language processing
  - Information extraction in economics, finance and business
  - Incorporating linguistic knowledge into NLP models and tasks
  - Effects of languages and NLP on economic decision-making
  - The interplay of conversational agents (multi-agent systems)





# About Me

- Education

- M.S., Computational Linguistics, University of Washington
- M.Phil., Theoretical and Applied Linguistics, University of Cambridge
- B.S., Economics, Linguistics, Mathematics, Asian Studies & Cert., Computer Sciences, University of Wisconsin–Madison

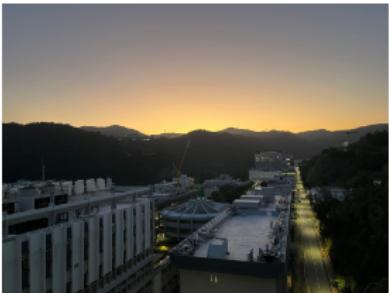
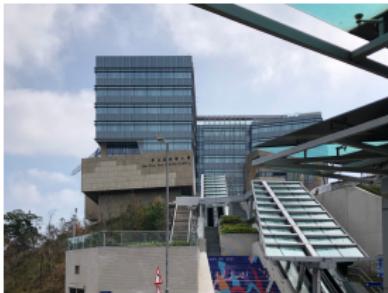


UW-Madison, Cambridge, and UW Seattle; I took those photos



# About Me

- Experience
  - Teaching Assistant, at Chinese University of Hong Kong
  - Research Assistant, at HKUST Business School
  - Financial NLP Intern, at China Merchants Bank
- Languages: Wu Chinese, English, Mandarin Chinese, Korean
  - My Cantonese is very poor. If you have to talk to me in the language, please speak as slowly as possible. Thanks for your understanding.



CMB Shanghai, HKUST Business School, and CUHK; I took those photos



# Contact

- Email: [yigechen@link.cuhk.edu.hk](mailto:yigechen@link.cuhk.edu.hk)
- Office hours
  - Time: Tuesday 12:00 – 14:00
  - Location: G19, Leung Kau Kui Bldg. or via Zoom
    - Meeting on Zoom is feasible if you let me know beforehand, i.e., send me an email and I'll set up the Zoom meeting
  - If you visit G19 and do not find me there during my office hours, please shoot me an email, and I'll be back in 5 minutes
  - If the time slot above does not fit in your schedule...
    - ① Send me an email and (hopefully) we can schedule a meeting then
    - ② You may also want to attend the office hours of Prof. Feng
- Website: <https://yigechen.com>



# Course Materials

- Blackboard: <https://blackboard.cuhk.edu.hk/>
- Tutorial materials will also be posted at  
<https://yigechen.com/teaching/ling3401/sp25>
- Textbook I would recommend:
  - Dan Jurafsky and James H. Martin. *Speech and Language Processing* (3rd ed. draft). 2024.
  - <https://web.stanford.edu/~jurafsky/slp3/>





# Academic Honesty

- Any kind of plagiarism is prohibited
- But we will allow the use of LLMs, AI tools, etc. to some extent due to the nature of this course
  - With explicit acknowledgment and proper citation
  - Prohibited for midterm and final exams
- Please refer to  
<https://www.cuhk.edu.hk/policy/academichonesty/> or ask us if you are unsure



# Academic Accommodations

- Please contact Special Educational Needs (SEN) Service if you need special accommodations, facilities or arrangements due to physical or mental illnesses
- Please refer to <https://www2.osa.cuhk.edu.hk/sens/en-GB/>



# About Google Colaboratory (Colab)

- Google Colaboratory is a research tool for machine learning education and research. It's a Jupyter notebook environment that requires no setup to use.
- ...which means that you don't have to worry about setting up the Python environment on your device (at least for now)
- Today's Colab notebook has been posted
- What do you need to do today?
  - Simply watching me running all the codes and demonstrating these NLP tasks
  - And if you'd like to have a try, simply change the example sentence and run all the codes as I do.
  - (That's why we said that no programming background is needed!)



# NLP Tasks: Linguistic Analysis

- Tokenization
- Part-of-speech tagging
- Syntactic parsing
- ...
- Check Universal Dependencies and its visualization tools!
  - [https://universaldependencies.org/conllu\\_viewer.html](https://universaldependencies.org/conllu_viewer.html)



# NLP Tasks: Applications

- Information extraction
  - e.g. Named entity recognition
- Sentiment analysis
- Machine translation
- Dialogue systems
- ...
- Can be performed with or without LLMs



# Miscellaneous

- Midterm: February 26, in class
- Final Exam: April 16, in class
- Please do not hesitate to ask questions
- We enjoy feedback from you, so please let us know if you feel there's anything we could have done better
- It would be great if you'd bring your laptop to the class every week