The Hong Kong University of Science and Technology

UG Course Syllabus (Spring 25-26)

[Course Title] Introduction to Natural Language Processing

[Course Code] COMP4221

[No. of Credits] 3 credits

[Any pre-/co-requisites]

ELEC 2600 OR IEDA 2520 OR IEDA 2540 OR MATH 2411 OR MATH 2421 OR MATH 2431

- Computer science: object-oriented programming and data structures, design, and analysis of algorithms.
- Mathematics: multivariable calculus, linear algebra, probability, stochastic process, and statistics.
- Students are expected to have probability, linear algebra, algorithm design and machine learning background. It is suggested to take an introductory algorithm and an introductory machine learning course before taking this course.

Name: Yi R. (May) Fung

Email: yrfung@ust.hk

Office Hours: Room 3539

Course Description

[Briefly describe the course content, key topics or themes, objectives, methods of instruction, e.g., lectures, discussions, projects].

This course provides an introduction to the tasks, models and algorithms of natural language processing (NLP) including the fundamental NLP tasks and models (such as language model, text classification, word vectors, sequence labelling, etc.), and several typical NLP applications (such as question answering, dialogue systems and machine translation, etc.), and the linguistic structure annotation. The course also introduces some advanced topics such as pre-trained language models and their prompt tuning and instruct tuning methods to give the students more perspectives of the NLP industry.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

1. Explain the principles behind fundamental NLP algorithms

2. Design, implement, and evaluate NLP approaches to solve practical problems, applying appropriate models and metrics.

3. Get a basic understanding of the techniques in the frontline of NLP industries

Assessments and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below, outlining the criteria used for evaluation.

Assessment

[List specific assessed tasks, exams, quizzes, their weightage, and due dates; perhaps, add a summary table as below, to precede the details for each assessment.]

- Assignments: (30%); due around week 4, 8, 12
- Course Project: (20%); due around week 13 (last week of class)
- Attendance: (10%); due throughout the course semester
- Mid-term Examination: (20%); due Oct 31st latest
- Final Examination: (20%); due May pending date finalization from registrar

Mapping of Course ILOs to Assessment Tasks

[add to/delete table as appropriate]

Assessed Task	Mapped ILOs	Explanation
Assignments	ILO1, ILO2	The assignments assess students' ability to explain the core principles and mechanics of fundamental NLP algorithms (ILO1) and to demonstrate practical competence by designing, implementing, and evaluating these models on defined tasks using appropriate metrics (ILO2).
Course Project	ILO2, ILO3	The project comprehensively assesses the ability to design, implement, and evaluate a complete NLP pipeline to solve a practical problem (ILO2). It also requires research and application of contemporary techniques, providing a basic understanding of the tools and methods at the forefront of the NLP industry (ILO3).
Attendance	ILO1, ILO3	Consistent attendance ensures ongoing engagement with the course material, facilitating the understanding of fundamental principles (ILO1) and exposure to advanced, cutting-edge topics and industry perspectives discussed in lectures (ILO3).
Midterm Examination	ILO1	The midterm examination primarily assesses students' ability to explain, compare, and contrast the principles behind fundamental NLP algorithms and models covered in the first half of the course (ILO1).
Final Examination	ILO1, ILO3	The final examination assesses the comprehensive understanding of principles behind both fundamental and advanced algorithms (ILO1). It also tests the basic understanding of modern techniques (e.g., pre-trained models,

		prompt tuning) that represent the frontline of NLP industries (ILO3).

Grading Rubrics

[Detailed rubrics for each assignment will be provided. These rubrics clearly outline the criteria used for evaluation. Students can refer to these rubrics to understand how their work will be assessed.]

- **Assignments (30%):** Graded on a 100-point scale per assignment.
    - **Correctness & Performance:** Code runs correctly and produces accurate results on test cases. Models achieve expected performance metrics.
    - **Code Quality & Implementation:** Code is well-structured, documented, efficient, and follows instructions.
    - **Analysis & Report:** Answers to theoretical questions and analysis of results are clear, concise, and correct.
- **Course Project (20%):** Graded on a 100-point scale.
    - **Proposal:** Clarity, feasibility, and motivation of the proposed project.
    - **Technical Soundness & Implementation:** Appropriateness of the chosen models, quality of the code, and complexity of the task.
    - **Evaluation & Analysis:** Thoroughness of experimentation, quality of result analysis, and understanding of limitations.
    - **Final Report & Presentation:** Clarity, organization, and professionalism of the written report and oral presentation.
- **Examinations:** Graded based on the correctness and completeness of answers to questions covering theoretical concepts, model derivations, and algorithmic comparisons.

Final Grade Descriptors:

[As appropriate to the course and aligned with university standards]

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	[Example: Demonstrates a comprehensive grasp of subject matter, expertise in problem-solving, and significant creativity in thinking. Exhibits a high capacity for scholarship and collaboration, going beyond core requirements to achieve learning goals.]
B	Good Performance	[Example: Shows good knowledge and understanding of the main subject matter, competence in problem-solving, and the ability to analyze and evaluate issues. Displays high motivation to learn and the ability to work effectively with others.]

C	Satisfactory Performance	[Example: Possesses adequate knowledge of core subject matter, competence in dealing with familiar problems, and some capacity for analysis and critical thinking. Shows persistence and effort to achieve broadly defined learning goals.]
D	Marginal Pass	[Example: Has threshold knowledge of core subject matter, potential to achieve key professional skills, and the ability to make basic judgments. Benefits from the course and has the potential to develop in the discipline.]
F	Fail	[Example: Demonstrates insufficient understanding of the subject matter and lacks the necessary problem-solving skills. Shows limited ability to think critically or analytically and exhibits minimal effort towards achieving learning goals. Does not meet the threshold requirements for professional practice or development in the discipline.]

Course AI Policy

[State the course policy on the use of generative artificial intelligence tools to complete assessment tasks.]

The use of generative AI tools (e.g., ChatGPT, Claude, Copilot) in this course is permitted nder the following conditions, with strict transparency:

1. Allowed for **b**rainstorming ideas, debugging code, explaining fundamental concepts, and improving the writing clarity of your *own* already-written work.
2. Not Allowed for directly generating complete solutions to assignment or project problems. Your submissions must be your own original work. Using AI to circumvent the learning process (e.g., generating code or answers to be submitted as your own) is considered a violation of academic integrity.
3. Any use of generative AI in the process of completing your work **must** be explicitly declared in a "AI Use Statement" at the beginning of your submission, detailing the tool used and the purpose it served. Failure to declare will be treated as an academic integrity violation.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include comments on the correctness of solutions, code quality, and depth of analysis. For the project, feedback will cover the proposal, milestone reports, and the final submission. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission Policy

[If applicable, explain the policy for resubmitting work or reassessment opportunities, including conditions and deadlines.]

N/A

Required Texts and Materials

[List required textbooks, readings, and any other materials]

Dan Jurafsky and James Martin, Speech and Language Processing

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Academic Integrity | HKUST – Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.