The Hong Kong University of Science and Technology (Guangzhou)

AIAA 4051 Syllabus

Introduction to Natural Language Processing

Credits: 3

Prereq: UFUG 2601 OR UFUG 2602

Exclusive: DSAA 3051

Lecture location & time: E1-101, Every Monday and Wednesday 12:00 PM-13:20 PM

Instructor Name: Sihong Xie

Instructor Email: sihongxie@hkust-gz.edu.cn

TA team

- Yazheng Liu (yliu533@connect.hkust-gz.edu.cn)
- Rufeng Chen (rchen514@connect.hkust-gz.edu.cn)
- Yue Chang (ychang500@connect.hkust-gz.edu.cn)

Office hours

Monday	Tuesday	Wednesday	Thursday	Friday
Sihong Xie 13:30-14:30PM W4 303	Yazheng Liu 13:30-14:30PM W4 301	Sihong Xie 13:30-14:30PM W4 303	Rufeng Chen 13:30-14:30PM W4 301	Yue Chang 13:30-14:30PM W4 301

Course Description

This course provides students with a deep understanding of how computational systems process and generate human languages. The curriculum bridges foundational linguistic theories with cutting-edge deep learning techniques, moving from classic N-gram models and Hidden Markov Models to modern Transformer architectures and Large Language Models (LLMs). The course explores critical topics such as tokenization, embeddings, syntactic parsing, and machine translation, while also diving into advanced contemporary subjects like Reinforcement Learning from Human Feedback (RLHF), Retrieval-Augmented Generation (RAG), and model compression. Assessments include in-class quizzes, a closed-book final exam, one homework, one individual coding project, and a significant final team research project that includes a poster presentation and technical report. Students will gain the expertise necessary to build, evaluate, and analyze sophisticated NLP systems capable of addressing real-world challenges in the rapidly evolving AI landscape.

Prerequisites

1. Machine learning and deep learning: we will constructing loss functions, take derivatives, and use gradient descent/ascent algorithms to train NLP models. If you have not taken related courses, it is strongly recommended that you review some introductory machine learning materials (Chapter 5 of the recommended textbook 3 Deep Learning by Ian Goodfellow).
2. Multivariate calculus: we will formulate matrix-vector equations and take gradients with respect to vectors. You don't a comprehensive background in multi-variate calculus but can refer to Chapters 1-2 of Matrix Calculus (for Machine Learning and Beyond) by Alan Edelman, etc.

3. Linear algebra, probability, statistics: please refer to Chapter 2-4 of the recommended textbook 3 Deep Learning by Ian Goodfellow.
4. Programming: if you have been using Python, you will be fine; if you know other programming language, probably you can start review some Python basics. You need to have basic ideas about algorithms and data structures, as we will talk about lists, trees, graphs, search, iterations, etc..

Intended Learning Outcomes (ILOs)

1. **Mastery of NLP Foundations**: gain a comprehensive understanding of language processing, progressing from foundational mathematical concepts and N-grams to complex structures like Hidden Markov Models and syntactic parsing.
2. **Proficiency in Modern Neural Architectures**: develop the ability to design and implement advanced deep learning models, specifically focusing on Recurrent Neural Networks, Attention mechanisms, and Transformers.
3. **Practical Development and Optimization Skills**: become proficient in using PyTorch to build and optimize NLP systems, incorporating modern techniques such as LoRA for fine-tuning, model compression, and KV caching for efficient inference.
4. **Competence in Advanced Generative AI Techniques**: acquire the technical knowledge to work with state-of-the-art Large Language Model (LLM) workflows, including Reinforcement Learning from Human Feedback (RLHF) , Retrieval-Augmented Generation (RAG) , and Diffusion models.
5. **Research and Collaborative Communication**: demonstrate the ability to critically analyze existing literature , conduct original experiments, and communicate technical findings through professional poster presentations and formal reports.

Assessment 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.

Assessments:

Assessment Task	Contribution to Overall Course grade (%)	Due date
In-class Quiz	24%	At the end of each class
Homework	5%	See the schedule
Individual coding project	10%	
Final team research project	35%	
Final closed-book written exam	26%	

1. **In-class Quiz:**
    a. Description: at the end of each class, there will be a short closed-book and written quiz. These quizzes are designed to assess students' understanding of the material.
    b. Format:
        i. Number of Questions: 2 to 5 multiple-choice questions.
        ii. Duration: 5 minutes.
        iii. Bring your own paper (1 piece of A4 should be enough) and pens.
        iv. No electronic device is allowed.
    c. Content: covering topics in the same lecture.
2. **Homework**
    a. Description: one homework to evaluate students' deeper understanding of the topics.

b. Format:
    i. Number of Questions: 6 short-answer/calculation/proof questions.
    ii. Duration: two weeks.
c. Content: covering topics in the first 7 lectures.

3. Individual coding project:

a. Description: to evaluate students' ability to turn concepts lectured in class to codes.
b. Format:
    i. Codes are provided in PyTorch; datasets are given in raw texts.
c. Content: with blanks in the codes of a transformer, the students will complete the blanks, and then run pre-training and fine-tuning of the transformer on a small dataset. Plot the process of the optimization.

4. Final team research project:

a. Description: include team registration, poster presentations, codes, and final reports. Teams should consist of 2-3 members, and teams with one student are discouraged (due to lack of computational resources, i.e., GPUs for small teams). If you are having trouble finding a teammate, the instructor and TAs will assist in finding a team. Instructor and project managers (TA and graduates) will provide project advice through out all stages of the project. GPU hours for all teammembers can be aggregated for the team project. Excellent projects will be selected for additional support and can lead to a publication.
b. Format:
    i. Duration: about 2 months.
    ii. Need to submit proposal, milestone, poster, report, and codes.
    iii. Proposal: please register for your team and identify a seed paper and write about the resources needed (2%).
    iv. Milestone: report any codes/datasets found and processed (5%).
    v. Codes: to be submitted at the end of the semester (8%).
    vi. Poster presentation in the last lecture: in a conference poster format where students are encouraged to talk to peers about their projects. Necessary devices for demonstrating posters will be provided (10%).
    vii. TA will interview each team to ask questions regarding their codes and reports.
c. Content: papers on cutting edge NLP research topics will be provided as the starter and students are encouraged to talk to project managers (instructors and the teaching team) as the project progresses.

5. Final Report:

a. Description: A comprehensive written report about the final team project (10%).
b. Format:
    i. Length: limited to 8 pages, double-spaced.
    ii. Follow a given template in Latex.
c. Content:
    i. Introduction and background of the chosen topic.
    ii. Review of existing literature related to the topic.
    iii. Description of the research challenge and solution.
    iv. Experimental results demonstrated in figures and tables.
    v. Conclusion.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
In-class Quiz	ILO1, ILO4	This task assesses students' ability understand and apply fundamental NLP concepts and knowledge (ILO1) and modern transformer and diffusion model-based NLP techniques, including architecture, optimization algorithms, and evaluation methods (ILO4).
Homework	ILO1	Students develop in-depth understanding of fundamental NLP concepts via reading course slides and recommended additional materials and then completing the homework questions (ILO1).
Individual coding project	ILO2, ILO3, ILO4	The students complete codes of a transfomer architecture to learn about the details of modern NLP model (ILO4). Then the students will run pre-training and fine-tuning codes and then evaluate the resulting model (ILO2 and ILO3).
Final team research project	ILO1, ILO2, ILO3, ILO4, ILO5	Students will first read research papers on cutting edge NLP topics (ILO5), after communicating with the instructional team and peer students, they form teams to process text data using tokenization and embedding (ILO1), develop codes for models and fine-tuning (ILO2, ILO3, ILO4), and report their results in written reports, interviews, and oral presentation during poster sessions (ILO5).
Final exam	ILO1, ILO4	Comprehensively test students' understanding of both fundamental and more modern NLP techniques/algorithms/principles/theory (ILO1, ILO4).

Grading

1. For Homework, individual coding projects, and the Final closed-book written exam, we will grade them against grading rubrics.

2. Final team research project—Poster presentation (10% of total final grade)

Student	Introduction to the Topic	Organization and Logic	Clarity of Presentation	Fluency of Presentation	Overall

Score Rubrics: **1** (Poor) | **2** (Below Average) | **3** (Borderline) | **4** (Up Average) | **5** (Good) | **6** (Very Good) | **7** (Excellent)

3. Final team research project—Final report (10% of total final grade)

Content Quality (At the scale of 60 points)	Fundamental Concept Understanding (25 points)	Clearly explains key NLP concepts with well-structured definitions and real-world examples. (22-25 points)	Covers key NLP topics but lacks depth or fails to provide concrete examples. (17-21 points)	Mentions concepts but with vague explanations or missing critical details. (10-16 points)	Displays little to no understanding of core NLP concepts or applications. (0-9 points)
	Score				
	Review and Analyze Existing Research (20 points)	Provides a comprehensive literature review with relevant and up-to-date papers, critically analyzing methodologies and findings in NLP. (18-20 points)	Covers related works but lacks deep comparison or critical evaluation. (14-17 points)	Lists some related works without much discussion or relevance. (8-13 points)	Little to no research cited, or references are outdated/irrelevant. (0-7 points)
	Score				
	Practical Application & Critical Analysis (15 points)	Demonstrates strong analytical skills by applying NLP concepts to real-world scenarios with insightful evaluation and innovative solutions. (13-15 points)	Provides relevant applications and some critical analysis but lacks depth, originality, or a strong connection to practical challenges. (10-12 points)	Mentions basic applications but offers superficial analysis, with limited critical thinking or weak real-world relevance. (6-9 points)	Shows little to no application of concepts, with unclear, incorrect, or missing critical analysis. (0-5 points)
	Score				
Expression and Format (at the scale of 40 points)	Language Expression (15 points)	Clear and fluent language expression, easy to understand. (13-15 points)	Generally good expression, but some parts are not fluent. (10-12 points)	Unclear expression, affects understanding.(6-9 points)	Severely unclear expression. (0-5 points)
	Score				
	Format Adherence (15 points)	Strictly follows format requirements, well-structured. (13-15 points)	Basically follows format requirements, with minor deviations. (10-12 points)	Insufficient format adherence, noticeable deviations. (6-9 points)	Fails to follow format requirements. (0-5 points)
	Score				
	References (10 points)	Appropriately selected references, correct citation format. (9-10 points)	Generally appropriate references, minor issues in format or selection. (7-8 points)	Inadequate references or incorrect citation format. (4-6 points)	Severely lacking references or improper format. (0-3 points)
	Score				

4. Final research project—Code quality (15% graded by TAs and instructor)

Code Quality (at the scale of 100 points)	Specifications (50 points)	The code works and produces the correct results and displays them correctly. This code meets all of the specifications (40-50 points)	The code works and produces the correct results and displays them correctly. It also meets most of the other specifications. (30-39 points)	The program produces correct results but does not display them correctly. (15-29 points)	The code is producing incorrect results. (0-14 points)
	Score				
	Readability (30 points)	The code is exceptionally well organized and very easy to follow. (25-30 points)	The code is fairly easy to read. (20-24 points)	The code is readable only by someone who knows what it is supposed to be doing. (10-19 points)	The code is poorly organized and very difficult to read. (0-9 points)
	Score				
	Efficiency (20 points)	The code is extremely efficient without sacrificing readability and understanding. (15-20 points)	The code is fairly efficient without sacrificing readability and understanding. (10-14 points)	The code is brute force and unnecessarily long. (6-9 points)	The code is huge and appears to be patched together. (0-5 points)
	Score				

5. Regrading

If you feel your grades deserve a better one on an assignment, please submit a regrade request by emailing the TA/instructor within 3 days after the grades are released (release dates are marked on the schedule). Summarize the reason why your grade was unfair according to the grading rubrics (which are also released). If a TA cannot decide, the request will be upgraded to the instructor. Note that there is no guarantee that your grades will be better after re-grading, and we reserve the right to re-grade other parts of the same assignment beyond your request.

6. Teamwork contribution quantification

Given the nature of collaboration in the final research project, all teammembers will receive the same grade typically, but in extreme cases where contributions are highly uneven, discrepancy can exist in the project grades across teammembers. Each teammembers are required to make a statement about his/her contributions to the project in their final submitted report.

Final Grade Descriptors:

Grades	Description	The corresponding 100-point system
A+	Excellent Performance	95-100
A	Excellent Performance	90-94
A-	Excellent Performance	85-89
B+	Good Performance	80-84
B	Good Performance	75-79
B-	Good Performance	70-74
C+	Satisfactory Performance	67-69
C	Satisfactory Performance	64-66
C-	Satisfactory Performance	61-63
D	Marginal Pass	55-60
F	Fail	0-54

Course AI Policy

Acceptable use of any Artificial Intelligence (AI) generated or aided material in your work is governed by the same rules as using non-AI material: it must be in accordance with the class, department, and University policies on Integrity, and you must cite or otherwise provide attribution. For more details, please refer to the HKUST-GZ Academic Quality Policy Framework, which outlines the University's academic integrity guidelines. Regarding AI assistance, refer to the University guideline.

Three Always Principles for Using Generative AI in This Course:

1. Always Acknowledge AI Use: When using AI tools in coursework, clearly acknowledge and cite their contributions. Transparency about the use of AI ensures academic honesty and helps clarify the origin of ideas or content used in assignments.
2. Always Evaluate AI Output Critically: While AI can be a valuable aid, its responses must be critically assessed for accuracy and relevance. Students should verify AI-generated content against reliable sources and apply their own judgment to determine its appropriateness in academic work.
3. Always Uphold Academic Integrity: AI should complement—not replace—your own effort, understanding, and analysis. It must be used as a support tool, with all work ultimately reflecting your individual learning and intellectual engagement.

Communication and Feedback

Please use office hours as much as possible to guide your study on the topics. Be respectful of the instructor and TA's time by preparing your questions ahead of time. Outside office hours, emails and Canvas posts are used as the main communication venue.

In addition, the course actively collects anonymous student feedback through regular surveys throughout the semester. This feedback is reviewed in a timely manner and used to make responsive adjustments to course delivery and structure, ensuring an open and adaptive learning environment. Please feel free to let us know your opinions and concerns as soon as possible using either the surveys, emails, or in-person meetings.

Late and Resubmission Policy

1. Quiz: resubmission opportunities are not provided for in-class quiz missed due to absence, except in cases of special circumstances such as medical leave or official leave approved by the university.
2. Individual assignments: you will have a total quota of **7 days** to be used freely for individual assignments (homework and coding project). Due dates will be stated on all assignments. If we erroneously set conflicting dates across Canvas, and the assignment document, please inform us. Until any error is corrected the earliest date applies. Students are expected to be able to submit work correctly online and to back up their data. Therefore, "forgetting to click submit", "computer crashes", etc, are not acceptable lateness excuses. Note that online sites' clocks may not match yours perfectly, so don't wait until the last moment to submit. Note that other than these late days, we will not be making exceptions and extending deadlines except for health reasons, so please try to be frugal with your late days and use them only if necessary. Assignments that are late beyond the allowed late days will lose **25%** of the total score per late day (**any period less than 24 hours is counted as one late day**).
3. Team final research project written reports and codes: no late day quota can be used, and will lose **25%** of the total score per late day (**any period less than 24 hours is counted as one late day**).
4. Team final research project poster: no make up presentation.

5. Team final research project interview: the interview date will be communicated with the students at least two weeks before the interview. Date and time can be made flexible. ALL team members need to show up in-person on the interview to earn full credits.
6. Final exam: there is no makeup exam.

Textbooks and Materials

Recommended textbooks for classic NLP techniques:

1. SLP3=Speech and Language Processing. Dan Jurafsky and James H. Martin. 3rd editions (latest version can be found here https://web.stanford.edu/~jurafsky/slp3/).
2. FSNLP=Foundations of statistical natural language processing. C. Manning and H. Schutze. MIT Press, 2000.
3. DL=Deep Learning. Ian Goodfellow and Yoshua Bengio and Aaron Courville. MIT Press. 2016.
4. NNMNLP=Neural Network Methods for Natural Language Processing. Yoav Goldberg. Morgan & Claypool Publishers. 2017.
5. Natural Language Processing. Jacob Eisenstein. 2018 (free version: https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf)
6. Natural Language Processing with Transformers. Lewis Tunstall, Leandro von Werra, and Thomas Wolf. (https://transformersbook.com/).

More materials will be given on the course website.

Academic Integrity

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

Schedule of Topics

	Date	Topics	Assignments, dues, and important notes.
1	2026-01-26	Course logistics; basic math and AI review	
2	2026-01-28	Language modeling and N-Grams	
3	2026-02-02	Tokenization and embeddings	
4	2026-02-04	Hidden Markov Model I	
5	2026-02-09	Hidden Markov Model II	**Winter Break** (2026-02-11 to 2026-02-23)
6	2026-02-25	Syntactic trees	
7	2026-02-28	Syntactic parsing	Make up day HW release Final project release
8	2026-03-02	Recurrent neural networks	

9	2026-03-04	Machine translation	
10	2026-03-09	Attention and Transformers I	
11	2026-03-11	Attention and Transformers II	HW due Individual Project release
12	2026-03-16	Mixture of experts	Final research project proposal and team registration
13	2026-03-18	Learning: Pretraining	
14	2026-03-23	Supervised Fine-Tuning: basics	
15	2026-03-25	Supervised Fine-Tuning: Lora	HW grade release Research project p
16	2026-03-30	Reinforcement Learning: RLHF	Individual Project due
17	2026-04-01	Reinforcement Learning: PPO	**Qingming Festival** (2026-04-04 to 2026-04-06)
18	2026-04-08	Reinforcement Learning: DPO, GRPO	Research project milestone report
19	2026-04-13	Training data synthetization	
20	2026-04-15	RAG	Individual Project grade release
21	2026-04-20	Scaling laws	
22	2026-04-22	Inference: KV cache	
23	2026-04-27	Model compression	
24	2026-04-29	Diffusion model I	**Labor Day** (2026-05-01 to 2025-05-05) Project team interview date registration.
25	2026-05-06	Diffusion model II	
26	2026-05-11	Final project poster presentation	Final research project code and report due on 2026-05-18; Team interview afterwards. Final exam date TBD