



## UNSW Course Outline

# COMP6713 Natural Language Processing - 2024

Published on the 28 Jan 2024

## General Course Information

Course Code : COMP6713

Year : 2024

Term : Term 1

Teaching Period : T1

Is a multi-term course? : No

Faculty : Faculty of Engineering

Academic Unit : School of Computer Science and Engineering

Delivery Mode : In Person

Delivery Format : Standard

Delivery Location : Kensington

Campus : Sydney

Study Level : Postgraduate, Undergraduate

Units of Credit : 6

### Useful Links

[Handbook Class Timetable](#)

## Course Details & Outcomes

### Course Description

Natural language processing (NLP) is a branch of artificial intelligence that deals with computational approaches used to process text.

Human language (i.e., natural language) is inherently ambiguous. Ambiguity resolution using computational techniques is at the heart of NLP. As a result, the advancements in NLP can be visualized as three generations: rule-based, statistical and neural. The course introduces the three generations of NLP through the philosophy of ambiguity resolution being the core task of NLP. The content covers different NLP sub-problems (such as POS tagging, sentiment classification, named entity recognition, machine translation and summarisation), and typical approaches in the three generations to tackle these sub-problems.

With recent advancements in large language models, there has been a renewed interest in NLP from industry and research alike. However, NLP precedes large language models. The exposition of NLP centered around ambiguity resolution helps to develop an understanding of the past and the present of NLP.

## Course Aims

This course aims to introduce students to the main topics and methods in the field of natural language processing, ranging from rule-based and statistical models to the recent innovation in neural models for various NLP sub-problems.

## Relationship to Other Courses

COMP6714 (Information Retrieval and Web Search) : The course covers n-grams and vector representations of words. COMP6713 will describe the concepts in the context of NLP, specifically: N-grams as features for classification, and dense representations of words: word2vec, GloVe etc.

COMP9444 (Neural Networks and Deep Learning): The course covers fundamentals of neural networks (backpropagation, etc.) and foundational neural models (such as RNNs, ConvNets, and autoencoders). Language processing is one of the modules in the course. COMP6713 will dig deeper into the NLP fundamentals. This will include a deeper investigation into attention, introduction to Transformers and the development of encoder-only (BERT) and decoder-only (GPT) models based on Transformers. The proposed course will cover NLP approaches beyond neural networks as well.

COMP3411/COMP9814 (Artificial intelligence): The course covers fundamental topics in artificial intelligence. The modules on neural networks, supervised learning, and natural language understanding will serve as a basis for students taking COMP6713.

# Course Learning Outcomes

Course Learning Outcomes
CLO1 : Describe NLP problems such as POS tagging, sentiment analysis, information extraction and machine translation along with their challenges in terms of ambiguity resolution
CLO2 : Explain typical NLP strategies based on statistical and neural approaches.
CLO3 : Train NLP models using NLP libraries (e.g. NLTK, scikit-learn, Transformers).
CLO4 : Design an NLP solution by selecting the NLP problem formulation, approach and evaluation strategy, by analysing the requirements of a specific application.

Course Learning Outcomes	Assessment Item
CLO1 : Describe NLP problems such as POS tagging, sentiment analysis, information extraction and machine translation along with their challenges in terms of ambiguity resolution	<ul style="list-style-type: none"><li>• Quizzes</li><li>• Final Exam</li></ul>
CLO2 : Explain typical NLP strategies based on statistical and neural approaches.	<ul style="list-style-type: none"><li>• Group Project</li><li>• Quizzes</li><li>• Final Exam</li></ul>
CLO3 : Train NLP models using NLP libraries (e.g. NLTK, scikit-learn, Transformers).	<ul style="list-style-type: none"><li>• Programming Assignment</li><li>• Group Project</li></ul>
CLO4 : Design an NLP solution by selecting the NLP problem formulation, approach and evaluation strategy, by analysing the requirements of a specific application.	<ul style="list-style-type: none"><li>• Programming Assignment</li><li>• Group Project</li><li>• Final Exam</li></ul>

## Learning and Teaching Technologies

Moodle - Learning Management System | Echo 360

## Learning and Teaching in this course

The lectures will cover the 'what', 'how' and 'why' of NLP through an exposition to linguistic phenomena, model fundamentals and coding examples. The tutorials will provide an opportunity to clarify the concepts using problem sets.

The group project will be an excellent opportunity for students to put the learned concepts to practice. Submission of papers based on the group project is encouraged and will be supported via mentorship.

# Other Professional Outcomes

## Learning Outcomes

CL01: Describe NLP problems such as POS tagging, sentiment analysis, information extraction and machine translation along with their challenges in terms of ambiguity resolution.

CL02: Explain typical NLP approaches based on statistical and neural approaches.

CL03: Use NLP libraries (e.g. NLTK, scikit-learn, Transformers) to implement the training of models for NLP problems and use them for inference.

CL04: Design an NLP solution by selecting the NLP problem formulation, approach and evaluation strategy, by analysing the requirements of a specific application.

## Additional Course Information

Natural language processing (NLP) is a branch of artificial intelligence that deals with computational approaches used to process text. Human language (i.e., natural language) is inherently ambiguous. Ambiguity resolution using computational techniques is at the heart of NLP. As a result, the advancements in NLP can be visualized as three generations: rule-based, statistical and neural. The course introduces the three generations of NLP through the philosophy of ambiguity resolution being the core task of NLP. The content covers different NLP sub-problems (such as POS tagging, sentiment classification, named entity recognition, machine translation and summarisation), and typical approaches in the three generations to tackle these sub-problems. With recent advancements in large language models, there has been a renewed interest in NLP from industry and research alike. However, NLP precedes large language models. The exposition of NLP centered around ambiguity resolution helps to develop an understanding of the past and the present of NLP.

# Assessments

## Assessment Structure

Assessment Item	Weight	Relevant Dates
Quizzes Assessment Format: Individual	20%	Start Date: Weekly Due Date: Weeks 2-5, 7-9
Programming Assignment Assessment Format: Individual	10%	Due Date: Weeks 3-5
Group Project Assessment Format: Group	25%	Start Date: Week 5 Due Date: Week 10
Final Exam Assessment Format: Individual	45%	Due Date: Exam Period

## Assessment Details

### Quizzes

#### Assessment Overview

Weekly online quizzes in Weeks 2-5, 7-9. The quizzes will be auto marked on the Learning Management System.

Overall feedback will be provided in the class.

#### Course Learning Outcomes

- CL01 : Describe NLP problems such as POS tagging, sentiment analysis, information extraction and machine translation along with their challenges in terms of ambiguity resolution
- CL02 : Explain typical NLP strategies based on statistical and neural approaches.

#### Assignment submission Turnitin type

This is not a Turnitin assignment

## Programming Assignment

#### Assessment Overview

This is a programming assignment to be completed individually and will be held between weeks 3-5. The dataset, NLP task and recommended model will be provided. They must (a) submit their codebase in the form of a Jupyter Notebook, (b) submit a project report describing and analysing their solution.

The assignment will be evaluated based on a rubric provided to the students. The evaluation will

be semi-automatic: using test scripts that generate automatic metrics (5 marks), and manual inspection of code (5 marks).

### **Course Learning Outcomes**

- CLO3 : Train NLP models using NLP libraries (e.g. NLTK, scikit-learn, Transformers).
- CLO4 : Design an NLP solution by selecting the NLP problem formulation, approach and evaluation strategy, by analysing the requirements of a specific application.

### **Assignment submission Turnitin type**

This is not a Turnitin assignment

## **Group Project**

### **Assessment Overview**

The project will be conducted between weeks 7-10. Students are expected to form themselves into groups of 3-5 (to be decided based on class strength) for the Group Project, by the end of Week 4. Each group will be assigned a Mentor.

Groups will choose a specific NLP problem to be implemented using a neural approach and one among rule-based/statistical approaches. They must (a) submit their codebase in the form of a Jupyter Notebook, (b) submit a project report describing and analysing their solution, and (c) give a presentation on their project work.

The Mentor assesses the group's work and provides written feedback throughout the term.

The allocation of 25 marks includes 20 marks for group work plus 5 marks for the individual component (based on each student's contribution to the final presentation).

### **Course Learning Outcomes**

- CLO2 : Explain typical NLP strategies based on statistical and neural approaches.
- CLO3 : Train NLP models using NLP libraries (e.g. NLTK, scikit-learn, Transformers).
- CLO4 : Design an NLP solution by selecting the NLP problem formulation, approach and evaluation strategy, by analysing the requirements of a specific application.

### **Assignment submission Turnitin type**

This is not a Turnitin assignment

## **Final Exam**

### **Assessment Overview**

The exam is 2-hours long and will be held in the UNSW exam period. It involves multiple-choice

questions, short-answers and code analysis.

Marking will be against specific criteria in a marking guide and no formal feedback will be provided.

To pass this course, students must score more than 40% on the final exam. Note that the hurdle will be enforced after any required scaling.

### Course Learning Outcomes

- CLO1 : Describe NLP problems such as POS tagging, sentiment analysis, information extraction and machine translation along with their challenges in terms of ambiguity resolution
- CLO2 : Explain typical NLP strategies based on statistical and neural approaches.
- CLO4 : Design an NLP solution by selecting the NLP problem formulation, approach and evaluation strategy, by analysing the requirements of a specific application.

### Assignment submission Turnitin type

This is not a Turnitin assignment

### Hurdle rules

To pass this course, students must score more than 40% on the final exam. Note that the hurdle will be enforced after any required scaling.

## **General Assessment Information**

This course aims to introduce students to the main topics and methods in the field of natural language processing, ranging from rule-based and statistical models to the recent innovation in neural models for various NLP sub-problems. The assessments will include individual and group assessments, along with online quizzes.

Quizzes (20%): These are multiple quizzes over the term. Each quiz will consist of multiple-choice questions that will be completed on Moodle. Overall feedback will be provided in the class.

Assignment (10%): This is a programming assignment to be completed individually and will be held between weeks 3-5. The dataset, NLP task and recommended model will be provided. They must (a) submit their codebase in the form of a Jupyter Notebook, (b) submit a project report describing and analysing their solution. The assignment will be evaluated based on a rubric provided to the students. The evaluation will be semiautomatic: using test scripts that generate automatic metrics (5 marks), and manual inspection of code (5 marks).

Group Project (25%): The project will be conducted between weeks 7-10. Students are expected to form themselves into groups of 3-5 (to be decided based on class strength) for the Group Project, by the end of Week 4. Each group will be assigned a Mentor. Groups will choose a specific NLP problem to be implemented using a neural approach and one among rule-based/statistical approaches. They must (a) submit their codebase in the form of a Jupyter Notebook, (b) submit a project report describing and analysing their solution, and (c) give a presentation on their project work. The Mentor assesses the group's work and provides written feedback throughout the term. The allocation of 25 marks includes 20 marks for group work plus 5 marks for the individual component (based on each student's contribution to the final presentation).

Final Exam (45%): The exam is 2-hours long and will be held in the UNSW exam period. It involves multiple-choice questions, short-answers and code analysis. Marking will be against specific criteria in a marking guide and no formal feedback will be provided.

### **Grading Basis**

Standard

### **Requirements to pass course**

To pass this course, students must score more than 40% on the final exam, along with the University rules regarding the overall marks. Note that the hurdle will be enforced after any required scaling.

# Course Schedule

Teaching Week/Module	Activity Type	Content
Week 1 : 12 February - 18 February	Lecture	<p>Topic: Introduction to NLP</p> <ul style="list-style-type: none"> <li>• Mon12:00 - 14:00; Old Main Building 150 (K-K15-150)</li> <li>• Tue11:00 - 13:00; E19 Patricia O'Shane G03 (K-E19-G03)</li> </ul>
Week 2 : 19 February - 25 February	Lecture	<p>Topic: Representation learning (&amp; Potential guest lecture)</p> <ul style="list-style-type: none"> <li>• Mon12:00 - 14:00; Old Main Building 150 (K-K15-150)</li> <li>• Tue11:00 - 13:00; E19 Patricia O'Shane G03 (K-E19-G03)</li> </ul>
	Tutorial	<ul style="list-style-type: none"> <li>• Mon17:00 - 18:00Mathews 106 (K-F23-106)</li> <li>• Tue14:00 - 15:00Mathews 230 (K-F23-230)</li> </ul> <p>Tutorial days to be advised.</p>
Week 3 : 26 February - 3 March	Lecture	<p>Topic: Transformers - 1</p> <ul style="list-style-type: none"> <li>• Mon12:00 - 14:00; Old Main Building 150 (K-K15-150)</li> <li>• Tue11:00 - 13:00; E19 Patricia O'Shane G03 (K-E19-G03)</li> </ul>
	Tutorial	<ul style="list-style-type: none"> <li>• Mon17:00 - 18:00Mathews 106 (K-F23-106)</li> <li>• Tue14:00 - 15:00Mathews 230 (K-F23-230)</li> </ul> <p>Tutorial days to be advised.</p>
Week 4 : 4 March - 10 March	Lecture	<p>Topic: Transformers - 2</p> <ul style="list-style-type: none"> <li>• Mon12:00 - 14:00; Old Main Building 150 (K-K15-150)</li> <li>• Tue11:00 - 13:00; E19 Patricia O'Shane G03 (K-E19-G03)</li> </ul>
	Tutorial	<ul style="list-style-type: none"> <li>• Mon17:00 - 18:00Mathews 106 (K-F23-106)</li> <li>• Tue14:00 - 15:00Mathews 230 (K-F23-230)</li> </ul> <p>Tutorial days to be advised.</p>
Week 5 : 11 March - 17 March	Lecture	<p>Topic: Sentiment Analysis</p> <ul style="list-style-type: none"> <li>• Mon12:00 - 14:00; Old Main Building 150 (K-K15-150)</li> <li>• Tue11:00 - 13:00; E19 Patricia O'Shane G03 (K-E19-G03)</li> </ul>
	Tutorial	<ul style="list-style-type: none"> <li>• Mon17:00 - 18:00Mathews 106 (K-F23-106)</li> <li>• Tue14:00 - 15:00Mathews 230 (K-F23-230)</li> </ul> <p>Tutorial days to be advised. This will be a lab Mentoring Session to assist with the Group Project.</p>
Week 7 : 25 March - 31 March	Lecture	<p>Topic: POS Tagging &amp; Named Entity Recognition (NER)</p> <ul style="list-style-type: none"> <li>• Mon12:00 - 14:00; Old Main Building 150 (K-K15-150)</li> <li>• Tue11:00 - 13:00; E19 Patricia O'Shane G03 (K-E19-G03)</li> </ul>
	Tutorial	<ul style="list-style-type: none"> <li>• Mon17:00 - 18:00Mathews 106 (K-F23-106)</li> <li>• Tue14:00 - 15:00Mathews 230 (K-F23-230)</li> </ul> <p>Tutorial days to be advised. This will be a lab Mentoring Session to assist with the Group Project.</p>
Week 8 : 1 April - 7 April	Lecture	<p>Topic: Machine Translation</p> <ul style="list-style-type: none"> <li>• Mon12:00 - 14:00; Old Main Building 150 (K-K15-150)</li> <li>• Tue11:00 - 13:00; E19 Patricia O'Shane G03 (K-E19-G03)</li> </ul>
	Tutorial	<ul style="list-style-type: none"> <li>• Mon17:00 - 18:00Mathews 106 (K-F23-106)</li> <li>• Tue14:00 - 15:00Mathews 230 (K-F23-230)</li> </ul> <p>Tutorial days to be advised. This will be a lab Mentoring Session to assist with the Group Project.</p>
Week 9 : 8 April - 14 April	Lecture	<p>Topic: Summarisation</p> <ul style="list-style-type: none"> <li>• Mon12:00 - 14:00; Old Main Building 150 (K-K15-150)</li> <li>• Tue11:00 - 13:00; E19 Patricia O'Shane G03 (K-E19-G03)</li> </ul>
	Tutorial	<ul style="list-style-type: none"> <li>• Mon17:00 - 18:00Mathews 106 (K-F23-106)</li> <li>• Tue14:00 - 15:00Mathews 230 (K-F23-230)</li> </ul> <p>Tutorial days to be advised. This will be a lab Mentoring Session to assist with the Group Project.</p>
Week 10 : 15 April - 21 April	Lecture	<p>Topic: Deploying NLP: Benchmarks, shared tasks, and applications</p> <ul style="list-style-type: none"> <li>• Mon12:00 - 14:00; Old Main Building 150 (K-K15-150)</li> <li>• Tue11:00 - 13:00; E19 Patricia O'Shane G03 (K-E19-G03)</li> </ul>
	Tutorial	<ul style="list-style-type: none"> <li>• Mon17:00 - 18:00Mathews 106 (K-F23-106)</li> <li>• Tue14:00 - 15:00Mathews 230 (K-F23-230)</li> </ul> <p>Tutorial days to be advised. This will be a lab Mentoring Session to assist with the Group Project.</p>

# Attendance Requirements

Students are strongly encouraged to attend all classes and review lecture recordings.

## General Schedule Information

The course will include Lectures (Weeks 1-5, 7-10); Tutorials (Weeks 2-5) and Lab Mentoring sessions (Weeks 6-10). Tutorials will be used in Weeks 2 to 6, to discuss worked examples and develop a deeper understanding of fundamental topics. Lab Mentoring Sessions will be used in Weeks 7 to 10, to assist with the Group Project.

Lectures may be conducted online. The students will be informed as such in advance.

# Course Resources

## Prescribed Resources

The course materials include content in the form of slides, suggested reading, online discussion forums, quizzes, coding exercises using Jupyter notebooks, and tutorial questions. Students are encouraged to read through the materials before each lecture. The lecture time will be used to summarize the material, discuss recent developments, and answer any questions about the topic.

## Recommended Resources

- Speech and Language Processing, Daniel Jurafsky and James Martin, Prentice Hall, 2023.
- (Optional due to uncertain availability) Natural Language Processing, Pushpak Bhattacharyya and Aditya Joshi, Wiley, 2023.

## Additional Costs

Enrolled students will be provided a coupon for Google Cloud Credits, closer to the project start date.

## Course Evaluation and Development

This is the first offering of the course. The course feedback will influence future offerings of the course.

# Staff Details

Position	Name	Email	Location	Phone	Availability	Equitable Learning Services Contact	Primary Contact
Head tutor	Saurav Jha					No	No
Convenor	Aditya Joshi				In-person consultations upon email request	Yes	Yes

## Other Useful Information

### Academic Information

#### I. Special consideration and supplementary assessment

If you have experienced an illness or misadventure beyond your control that will interfere with your assessment performance, you are eligible to apply for Special Consideration prior to, or within 3 working days of, submitting an assessment or sitting an exam.

Please note that UNSW has a Fit to Sit rule, which means that if you sit an exam, you are declaring yourself fit enough to do so and cannot later apply for Special Consideration.

For details of applying for Special Consideration and conditions for the award of supplementary assessment, please see the information on UNSW's [Special Consideration page](#).

#### II. Administrative matters and links

All students are expected to read and be familiar with UNSW guidelines and policies. In particular, students should be familiar with the following:

- [Attendance](#)
- [UNSW Email Address](#)
- [Special Consideration](#)
- [Exams](#)
- [Approved Calculators](#)
- [Academic Honesty and Plagiarism](#)
- [Equitable Learning Services](#)

#### III. Equity and diversity

Those students who have a disability that requires some adjustment in their teaching or learning environment are encouraged to discuss their study needs with the course convenor prior to, or at

the commencement of, their course, or with the Equity Officer (Disability) in the Equitable Learning Services. Issues to be discussed may include access to materials, signers or note-takers, the provision of services and additional exam and assessment arrangements. Early notification is essential to enable any necessary adjustments to be made.

#### IV. Professional Outcomes and Program Design

Students are able to review the relevant professional outcomes and program designs for their streams by going to the following link: <https://www.unsw.edu.au/engineering/student-life/student-resources/program-design>.

*Note: This course outline sets out the description of classes at the date the Course Outline is published. The nature of classes may change during the Term after the Course Outline is published. Moodle or your primary learning management system (LMS) should be consulted for the up-to-date class descriptions. If there is any inconsistency in the description of activities between the University timetable and the Course Outline/Moodle/LMS, the description in the Course Outline/Moodle/LMS applies.*

#### Academic Honesty and Plagiarism

UNSW has an ongoing commitment to fostering a culture of learning informed by academic integrity. All UNSW students have a responsibility to adhere to this principle of academic integrity. Plagiarism undermines academic integrity and is not tolerated at UNSW. *Plagiarism at UNSW is defined as using the words or ideas of others and passing them off as your own.*

Plagiarism is a type of intellectual theft. It can take many forms, from deliberate cheating to accidentally copying from a source without acknowledgement. UNSW has produced a website with a wealth of resources to support students to understand and avoid plagiarism, visit: [student.unsw.edu.au/plagiarism](https://student.unsw.edu.au/plagiarism). The Learning Centre assists students with understanding academic integrity and how not to plagiarise. They also hold workshops and can help students one-on-one.

You are also reminded that careful time management is an important part of study and one of the identified causes of plagiarism is poor time management. Students should allow sufficient time for research, drafting and the proper referencing of sources in preparing all assessment tasks.

Repeated plagiarism (even in first year), plagiarism after first year, or serious instances, may also

be investigated under the Student Misconduct Procedures. The penalties under the procedures can include a reduction in marks, failing a course or for the most serious matters (like plagiarism in an honours thesis or contract cheating) even suspension from the university. The Student Misconduct Procedures are available here:

[www.gs.unsw.edu.au/policy/documents/studentmisconductprocedures.pdf](http://www.gs.unsw.edu.au/policy/documents/studentmisconductprocedures.pdf)

## Submission of Assessment Tasks

Work submitted late without an approved extension by the course coordinator or delegated authority is subject to a late penalty of five percent (5%) of the maximum mark possible for that assessment item, per calendar day.

The late penalty is applied per calendar day (including weekends and public holidays) that the assessment is overdue. There is no pro-rata of the late penalty for submissions made part way through a day. This is for all assessments where a penalty applies.

Work submitted after five days (120 hours) will not be accepted and a mark of zero will be awarded for that assessment item.

For some assessment items, a late penalty may not be appropriate. These will be clearly indicated in the course outline, and such assessments will receive a mark of zero if not completed by the specified date. Examples include:

- Weekly online tests or laboratory work worth a small proportion of the subject mark;
- Exams, peer feedback and team evaluation surveys;
- Online quizzes where answers are released to students on completion;
- Professional assessment tasks, where the intention is to create an authentic assessment that has an absolute submission date; and,
- Pass/Fail assessment tasks.

## Faculty-specific Information

[Engineering Student Support Services](#) – The Nucleus - enrolment, progression checks, clash requests, course issues or program-related queries

[Engineering Industrial Training](#) – Industrial training questions

[UNSW Study Abroad](#) – study abroad student enquiries (for inbound students)

[UNSW Exchange](#) – student exchange enquiries (for inbound students)

[UNSW Future Students](#) – potential student enquiries e.g. admissions, fees, programs, credit transfer

## Phone

(+61 2) 9385 8500 – Nucleus Student Hub

(+61 2) 9385 7661 – Engineering Industrial Training

(+61 2) 9385 3179 – UNSW Study Abroad and UNSW Exchange (for inbound students)

## School Contact Information

**CSE Help! - on the Ground Floor of K17**

- For assistance with coursework assessments.

**The Nucleus Student Hub** - <https://nucleus.unsw.edu.au/en/contact-us>

- Course enrolment queries.

**Grievance Officer** - [grievance-officer@cse.unsw.edu.au](mailto:grievance-officer@cse.unsw.edu.au)

- If the course convenor gives an inadequate response to a query or when the course convenor does not respond to a query about assessment.

**Student Reps** - [stureps@cse.unsw.edu.au](mailto:stureps@cse.unsw.edu.au)

- If some aspect of a course needs urgent improvement. (e.g. Nobody responding to forum queries, cannot understand the lecturer)