Course Proposal	
New/Revise	New
Course Name	Natural Language Processing
Course Name – SiMs	Natural Language Processing
Owning Faculty	Faculty of Engineering
Owning Academic Unit	School of Computer Science & Engineering
Administrative Campus	Sydney
Units of Credit	6
Grading Basis	Standard UNSW Grades
Academic Calendar Type	3+
Career	Undergraduate Postgraduate
Academic Details	
Course Description for Handbook	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 via the lens of ambiguity resolution. 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.
Field of Education (Broad)	020000 Information Technology
Field of Education (Narrow)	020100 Computer Science
Field of Education (Detailed)	020119 Artificial Intelligence
Level	Level 3
Author	Aditya Joshi
Proposal Sponsor	John Shepherd
Teaching Strategies and Rationale	The course will include Lectures (Weeks 1-5, 7-10); Tutorials (Weeks 2-5) and Lab Mentoring sessions (Weeks 6-10).

	The course materials include content in the form of slides, suggested reading, online discussion forums, quizzes, coding exercises using Jupyter notebooks, as well as links to the 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. Tutorials will be used in Weeks 2 to 5, to discuss worked examples and develop a deeper understanding of fundamental topics. Lab Mentoring Sessions will be used in Weeks 6 to 10, to assist with the Group Project.
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 research in neural models for various NLP sub-problems.
Features, Fit & Market Demand	NLP has been effectively applied to automate the processing of textual data in several sectors including industry and government. Businesses are looking for ways to integrate Transformer-based models such as BERT and GPT into their use cases. Tech companies in Australia and around the world use NLP in their products, and require professionals who have NLP skills. A Forbes article from May 2023 describes how NLP is "revolutionising business operations". A 2021 report from IBM states that the use of NLP technology can reduce information gathering time by 50%. Similar reports from Accenture and McKinsey highlight the potential of Generative NLP/AI. Given the utility of NLP, the course aligns with the University's strategic priority of Academic Excellence. The NLP skills learned in the course can be applied to any discipline where automated text processing is required. The students can apply NLP to research problems for social good. As a result, the course aligns with the University's strategic priority of social engagement.
Course Properties	
Course Type	Award Course
Deliver Mode	In-person/Multimodal

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: Devise an NLP solution by selecting the NLP problem formulation, approach and evaluation strategy, by analysing the requirements of a specific application. 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: Explain typical NLP approaches based on Hidden Markov Models, Support Vector Machines, Transformers (BERT as well decoder-only models). CL05: Appreciate the past, present and future of NLP.
Assessments	

Assessment Type	Quiz
Assessment Name	Quiz
Weighting (%)	10
Group or Individual	Individual
Assessment Overview	This is a quiz of multiple-choice questions that will be completed on Moodle. Overall feedback will be provided in the class.
Mapping to Learning Outcomes	CL01, CL04, CL05
Applies to All Delivery Variations?	
Assessment Number	

Assessment Type	Assignment
Assessment Name	Assignment 1
Weighting (%)	10
Group or Individual	Individual
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).
Mapping to Learning Outcomes	CL02, CL03, CL05
Applies to All Delivery Variations?	
Assessment Number	

Assessment Type	Assignment
Assessment Name	Assignment 2
Weighting (%)	10
Group or Individual	Individual
Assessment Overview	This is a non-programming assignment and will be conducted between weeks 7-10. Students are expected to read an NLP paper among a repository of papers, and submit a short report (500 words), followed by randomly generated follow-up question(s). The assignment will be submitted via LMS (Learning Management System) in the form of a Moodle quiz. The Mentor will assess the report and provide feedback.
Mapping to Learning Outcomes	CL02, CL03, CL04, CL05
Applies to All Delivery Variations?	
Assessment Number	

Assessment Type	Group Project
Assessment Name	Group Project
Weighting (%)	25
Group or Individual	Group
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 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).
Mapping to Learning Outcomes	CL02, CL03, CL04, CL05
Applies to All Delivery Variations?	
Assessment Number	

Assessment Type	Examination
Assessment Name	Final Exam
Weighting (%)	45
Group or Individual	Individual
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.
Mapping to Learning Outcomes	CL01, CL02, CL04, CL05
Applies to All Delivery Variations?	
Assessment Number	

Enrolment Requirements and Relationships	
Enrolment Requirements	Postgraduate: COMP9020 and COMP9024 AND (COMP9444 or COMP9414)

	Undergraduate: COMP2521 AND (COMP3411 or COMP9444)
Course Relationships	(See below)
Other Information for Handbook	
Key Search Terms	Natural language processing, NLP, computational linguistics, LLM, GPT, language models, ChatGPT

Syllabus

Module 1: Fundamentals

Introduction: Layers of NLP, introduction to NLP tasks, the role of ambiguity and ambiguity resolution. Ethical considerations for NLP. Using Spacy, NLTK for tokenization/lemmatization, POS tagging, parsing and entity extraction.

Representation learning: Logic-based representation of grammar, probabilistic language models, word vectors (word2vec, GloVe).

Transformers: Attention. Architecture of Transformer. Positional encodings. BERT. XLNet as an alternative to BERT. Pre-training, fine-tuning. GPT. Role of masked attention and Rotary embeddings. Prompting methods namely zero-shot, few-shot, chain-of-thought. Augmented language models, LangChain.

Module 2: NLP as a classification task

Sentiment analysis: Challenges. Sentiment lexicons. Statistical classifiers for sentiment analysis. Introduction to HuggingFace. Fine-tuning HuggingFace models for sentiment analysis.

Figurative language processing: Sarcasm and metaphor detection. Statistical classifiers with lexicon-derived features. LSTM-LSTM stacks for sarcasm detection.

Module 3: NLP as sequence labeling

POS tagging - Challenges. POS tagset. HMM-based POS tagging. POS tagging as a seq2seq model.

Named entity recognition (NER): Challenges. BIO tagset. CRF based NER. BiLSTM+CRF for NER. BERT-based fine-tuning. Applications to medical documents.

Module 4: NLP as seq2seq generation

Machine translation- Challenges. Parallel corpora. Rule-based machine translation. Statistical language models: alignment and language model. Neural MT: revisiting Transformers.

Summarisation: Extractive and abstractive summarisation. Extractive summarisation using text classifiers. Abstractive summarisation using seq2seq models, pointer-generator networks.

Module 5: Other NLP tasks

Natural language inference, question-answering, conversational AI. Challenges, datasets and analogies with NLP problems covered in previous modules of the course.

NLP applications to health, productivity, education and so on.

Relationship with other courses

COMP6714 (Information Retrieval and Web Search) : The course covers n-grams and vector representations of words. The proposed course will include them 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. The proposed course 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/COMP9414 (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 good basis for students taking the proposed NLP course.