UNSW Course Outline

ARTS2115 Philosophy of Artificial Intelligence - 2024

Published on the 28 Jan 2024

General Course Information

Course Code : ARTS2115
Year : 2024
Term : Term 1
Teaching Period : T1
Is a multi-term course? : No
Faculty : Faculty of Arts, Design and Architecture
Academic Unit : School of Humanities and Languages
Delivery Mode : In Person
Delivery Format : Standard
Delivery Location : Kensington
Campus : Sydney
Study Level : Undergraduate
Units of Credit : 6

Useful Links

Handbook Class Timetable

Course Details & Outcomes

Course Description

The possibility of intelligent machines has long been a favourite subject of speculation for philosophers, computer scientists, and science-fiction writers. But things are different today: technology has advanced to the point where the fascinating and troubling questions raised by

Artificial Intelligence must be addressed head-on. Some of those we consider in this course are foundational questions in the philosophy of mind. What is intelligence, and can machines really possess it? Could it be that—as many have argued—we ourselves are no more than intelligent machines (designed by evolution rather than engineers)? How do technologies such as artificial neural networks and machine learning change our understanding of the mind? Others are ethical, social, and political. What are the risks associated with these technologies, and how can we minimise them? What are their benefits, and how can we ensure that they are equitably shared? Conversely, assuming that true A.I. is possible, what are our own moral obligations towards our non-human but intelligent creations?

No background in computer science is required for this course.

Course Learning Outcomes

Course Learning Outcomes
CLO1 : Understand and explain key philosophical arguments around the possibility of artificial intelligence
CLO2 : Understand and evaluate arguments concerning the ethical, social, and political implications of existing and projected forms of artificial intelligence
CLO3 : Analyse and critically respond to philosophical texts
CLO4 : Communicate complex ideas clearly and concisely

Course Learning Outcomes	Assessment Item
CLO1 : Understand and explain key philosophical arguments around the possibility of artificial intelligence	• In-class Test • Major Essay
CLO2 : Understand and evaluate arguments concerning the ethical, social, and political implications of existing and projected forms of artificial intelligence	• Critique of AI Reasoning • In-class Test • Major Essay
CLO3 : Analyse and critically respond to philosophical texts	• Critique of AI Reasoning • Major Essay
CLO4 : Communicate complex ideas clearly and concisely	• Critique of AI Reasoning • Major Essay

Learning and Teaching Technologies

Moodle - Learning Management System | Echo 360

Assessments

Assessment Structure

Assessment Item	Weight	Relevant Dates
Critique of AI Reasoning Assessment Format: Individual	25%	Due Date: 10/03/2024 11:59 PM Post Date: 25/03/2024 12:00 AM
In-class Test Assessment Format: Individual	25%	Start Date: Not Applicable Due Date: Test 1: Week 5, Session 2; Test 2: Week 10, Session 2.
Major Essay Assessment Format: Individual	50%	Due Date: 26/04/2024 11:59 PM

Assessment Details

Critique of AI Reasoning

Assessment Overview

Students will generate a response to a philosophical question, using one of the publicly available Generative tools such as Bard or ChatGPT. They are then asked to evaluate the response and correct its mistakes (if any).

Length: 750 words

Feedback via written comments.

Course Learning Outcomes

- CLO2 : Understand and evaluate arguments concerning the ethical, social, and political implications of existing and projected forms of artificial intelligence
- CLO3 : Analyse and critically respond to philosophical texts
- CLO4 : Communicate complex ideas clearly and concisely

Detailed Assessment Description

Please see Moodle for Moodle for more details.

Assignment submission Turnitin type

This assignment is submitted through Turnitin and students can see Turnitin similarity reports.

In-class Test

Assessment Overview

Students will take two in-class tests (short-answer questions), each covering approximately 4-5 weeks of material.

Duration: 45 mins per test.

Feedback via numerical score.

Course Learning Outcomes

- CLO1 : Understand and explain key philosophical arguments around the possibility of artificial intelligence
- CLO2 : Understand and evaluate arguments concerning the ethical, social, and political implications of existing and projected forms of artificial intelligence

Detailed Assessment Description

Please see Moodle for more details.

Major Essay

Assessment Overview

Students will write a major essay on a topic in the philosophy of AI.

Length: 2,500 words

Feedback via rubric and written comments.

Course Learning Outcomes

- CLO1 : Understand and explain key philosophical arguments around the possibility of artificial intelligence
- CLO2 : Understand and evaluate arguments concerning the ethical, social, and political implications of existing and projected forms of artificial intelligence
- CLO3 : Analyse and critically respond to philosophical texts
- CLO4 : Communicate complex ideas clearly and concisely

Detailed Assessment Description

Please see Moodle for more details.

Assignment submission Turnitin type

This assignment is submitted through Turnitin and students can see Turnitin similarity reports.

Hurdle rules

Students who have failed to meet the attendance requirement without justification will be barred from submitting the essay.

General Assessment Information

Grading Basis

Standard

Course Schedule

Teaching Week/Module	Activity Type	Content
Week 1 : 12 February - 18 February	Seminar	Session 1 (Monday): Introduction to course; What is intelligence? What is artificial intelligence?
	Seminar	Session 2 (Thursday): Ontology I: Materialism in the philosophy of mind Lecturer: Markos Valaris
Week 2 : 19 February - 25 February	Seminar	Session 1 (Monday): Ontology II: Multiple realisability and functionalism Lecturer: Markos Valaris
	Seminar	Session 2 (Thursday): Minds and machines Lecturer: Markos Valaris
Week 3 : 26 February - 3 March	Seminar	Session 1 (Monday): The problem of intentionality Lecturer: Markos Valaris
	Seminar	Session 2 (Thursday): Classical AI and Neural networks Lecturer: Markos Valaris
Week 4 : 4 March - 10 March	Seminar	Session 1 (Monday): Brains in a vat? Lecturer: Markos Valaris
	Seminar	Session 2 (Thursday): Understanding and embodiment Lecturer: Markos Valaris
Week 5 : 11 March - 17 March	Seminar	Session 1 (Monday): mid-term wrap-up
	Assessment	Session 2 (Thursday): IN-CLASS TEST
Week 7 : 25 March - 31 March	Seminar	Session 1 (Monday): Understanding and personhood Lecturer: Melissa Merritt
	Seminar	Session 2 (Thursday): Machine ethical learning Lecturer: Melissa Merritt
Week 8 : 1 April - 7 April	Seminar	Session 1 (Monday): Are they coming to kill us? Autonomous weapons, learned bias, and other moral dangers of AI. Lecturer: Melissa Merritt Since this falls on a public holiday, there is no face-to-face lecture on this day. A lecture will be recorded and made available via Moodle.
	Seminar	Session 2 (Thursday): Ethics of human-machine interaction I Lecturer: Melissa Merritt
Week 9 : 8 April - 14 April	Seminar	Session 1 (Monday): Ethics of human-machine interaction II Lecturer: Melissa Merritt
	Seminar	Session 1 (Monday): AI and creativity Lecturer: Melissa Merritt
Week 10 : 15 April - 21 April	Seminar	Session 1 (Monday): wrap-up
	Assessment	Session 2 (Thursday): IN-CLASS TEST

Attendance Requirements

This is a seminar-style course, based on face-to-face interaction. Attendance is **mandatory**. Failing to attend at least 12 out of 18 seminar sessions without justification will be barred from participating in the final assessment in the course (final essay). Note that simply watching the seminar recording on Echo360 does **not** count as attendance.

Attendance and participation in the course is essential to meet the Course Learning Outcomes 1, 2, and 4. Attendance will be recorded at each session by the lecturer.

Course Resources

Recommended Resources

Useful internet resources for this course include:

- Stanford Encyclopedia of Philosophy (plato.stanford.edu)
- Internet Encyclopedia of Philosophy (iep.utm.edu)

Staff Details

Position	Name	Email	Location	Phone	Availability	Equitable Learning Services Contact	Primary Contact
Convenor	Melissa Merritt		MB341		By email and by appointment	Yes	Yes
Lecturer	Markos Valaris		MB339		By email and by appointment.	No	No

Other Useful Information

Academic Information

Due to evolving advice by NSW Health, students must check for updated information regarding online learning for all Arts, Design and Architecture courses this term (via Moodle or course information provided).

Please see: https://www.unsw.edu.au/arts-design-architecture/student-life/resources-support/protocols-guidelines for essential student information relating to:

- UNSW and Faculty policies and procedures;
- Student Support Services;
- Dean's List;
- review of results;
- credit transfer;
- cross-institutional study and exchange;
- examination information;
- enrolment information;
- Special Consideration in the event of illness or misadventure;
- student equity and disability;

And other essential academic information.

Academic Honesty and Plagarism

Plagiarism is using the words or ideas of others and presenting them as your own. It can take many forms, from deliberate cheating to accidentally copying from a source without acknowledgement.

UNSW groups plagiarism into the following categories:

- Copying: Using the same or very similar words to the original text or idea without acknowledging the source or using quotation marks. This includes copying materials, ideas or concepts from a book, article, report or other written document, presentation, composition, artwork, design, drawing, circuitry, computer program or software, website, internet, other electronic resource, or another person's assignment without appropriate acknowledgement.
- Inappropriate paraphrasing: Changing a few words and phrases while mostly retaining the original information, structure and/or progression of ideas of the original without acknowledgement. This also applies in presentations where someone paraphrases another's ideas or words without credit and to piecing together quotes and paraphrases into a new whole, without appropriate referencing.
- Collusion: Working with others but passing off the work as a person's individual work. Collusion also includes providing your work to another student for the purpose of them plagiarising, paying another person to perform an academic task, stealing or acquiring another person's academic work and copying it, offering to complete another person's work or seeking payment for completing academic work.
- Inappropriate citation: Citing sources which have not been read, without acknowledging the "secondary" source from which knowledge of them has been obtained.
- Duplication ("self-plagiarism"): Submitting your own work, in whole or in part, where it has previously been prepared or submitted for another assessment or course at UNSW or another university.

The UNSW Academic Skills support offers resources and individual consultations. Students are also reminded that careful time management is an important part of study. One of the identified causes of plagiarism is poor time management. Students should allow sufficient time for research, drafting and proper referencing of sources in preparing all assessment items. UNSW Library has the ELISE tool available to assist you with your study at UNSW. ELISE is designed to introduce new students to studying at UNSW, but it can also be a great refresher during your study.

Completing the ELISE tutorial and quiz will enable you to:

- analyse topics, plan responses and organise research for academic writing and other assessment tasks
- effectively and efficiently find appropriate information sources and evaluate relevance to your

needs
- use and manage information effectively to accomplish a specific purpose
- better manage your time
- understand your rights and responsibilities as a student at UNSW
- be aware of plagiarism, copyright, UNSW Student Code of Conduct and Acceptable Use of UNSW ICT Resources Policy
- be aware of the standards of behaviour expected of everyone in the UNSW community
- locate services and information about UNSW and UNSW Library

Use of AI for assessments

As AI applications continue to develop, and technology rapidly progresses around us, we remain committed to our values around academic integrity at UNSW. Where the use of AI tools, such as ChatGPT, has been permitted by your course convener, they must be properly credited and your submissions must be substantially your own work.

In cases where the use of AI has been prohibited, please respect this and be aware that where unauthorised use is detected, penalties will apply.

Use of AI for assessments | UNSW Current Students

Submission of Assessment Tasks

Turnitin Submission

If you encounter a problem when attempting to submit your assignment through Turnitin, please telephone External Support on 9385 3331 or email them on externalteltsupport@unsw.edu.au

Support hours are 8:00am – 10:00pm on weekdays and 9:00am – 5:00pm on weekends (365 days a year). If you are unable to submit your assignment due to a fault with Turnitin, you may apply for an extension, but you must retain your ticket number from External Support (along with any other relevant documents) to include as evidence to support your extension application. If you email External Support, you will automatically receive a ticket number, but if you telephone, you will need to specifically ask for one. Turnitin also provides updates on their system status on Twitter.

Generally, assessment tasks must be submitted electronically via either Turnitin or a Moodle assignment. In instances where this is not possible, alternative submission details will be stated on your course's Moodle site. For information on how to submit assignments online via Moodle: https://student.unsw.edu.au/how-submit-assignment-moodle

Late Submission Penalty

UNSW has a standard late submission penalty of:

- 5% per calendar day,
- for all assessments where a penalty applies,
- capped at five calendar days (120 hours) from the assessment deadline, after which a student cannot submit an assessment, and
- no permitted variation.

Students are expected to manage their time to meet deadlines and to request Special Consideration as early as possible before the deadline. Support with Time Management is available here.

School Contact Information

School of Humanities & Languages

Email: hal@unsw.edu.au

Location: School Office, Morven Brown Building, Level 2, Room 258

Opening Hours: Monday - Friday, 9am - 5pm