



# PHIL8400

## Rights, Responsibilities, and AI

Session 2, In person-scheduled-weekday, North Ryde 2024

*Department of Philosophy*

### Contents

---

<a href="#"><u>General Information</u></a>	2
<a href="#"><u>Learning Outcomes</u></a>	3
<a href="#"><u>General Assessment Information</u></a>	3
<a href="#"><u>Assessment Tasks</u></a>	3
<a href="#"><u>Delivery and Resources</u></a>	6
<a href="#"><u>Unit Schedule</u></a>	6
<a href="#"><u>Policies and Procedures</u></a>	10

---

#### **Disclaimer**

Macquarie University has taken all reasonable measures to ensure the information in this publication is accurate and up-to-date. However, the information may change or become out-dated as a result of change in University policies, procedures or rules. The University reserves the right to make changes to any information in this publication without notice. Users of this publication are advised to check the website version of this publication [or the relevant faculty or department] before acting on any information in this publication.

## General Information

Unit convenor and teaching staff

Unit Convenor

Dr Regina Fabry

[regina.fabry@mq.edu.au](mailto:regina.fabry@mq.edu.au)

Lecturer

A/Prof Abhaya Nayak

[abhaya.nayak@mq.edu.au](mailto:abhaya.nayak@mq.edu.au)

Lecturer

Prof Niloufer Selvadurai

[niloufer.selvadurai@mq.edu.au](mailto:niloufer.selvadurai@mq.edu.au)

Tutor

Siavosh Sahebi

[siavosh.sahebi@mq.edu.au](mailto:siavosh.sahebi@mq.edu.au)

Credit points

10

Prerequisites

COMP6400 or COMP2400

Corequisites

Co-badged status

Unit description

With increasing entrenchment of AI in human affairs, its scientific, moral, political, economic, and other social aspects are becoming a significant issue. For instance, there is a significant concern that machine learning algorithms contribute to the discrimination against members of oppressed groups (e.g., women, people of colour). This unit, co-designed and co-taught by relevant experts in Computing, Philosophy, and cognate disciplines, will present and discuss key theoretical, ethical, and empirical questions about the conditions of explainable, safe, fair, and responsible AI. Furthermore, it will explore scientific, ethical, political, economic, and other social implications of topical issues such as algorithmic decision making, applications of deep learning models, and robot rights. Students will be exposed to ideas such as balancing risks and responsibilities, both in the scientific and moral sense, in the context of the evolving AI technologies.

## Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are available at <https://www.mq.edu.au/study/calendar-of-dates>

## Learning Outcomes

On successful completion of this unit, you will be able to:

**ULO1:** Explain the fundamental principles underlying AI, and the normative constraints that it needs to satisfy.

**ULO2:** Demonstrate an advanced understanding of the ethical and other socioeconomic implications of AI.

**ULO3:** Demonstrate an advanced understanding of what Responsible AI means, or will mean, in our current as well as future society.

**ULO4:** Effectively communicate your findings to different stakeholders

## General Assessment Information

*Unless a Special Consideration request has been submitted and approved, a 5% penalty (of the total possible mark) will be applied each day a written assessment is not submitted, up until the 7th day (including weekends). After the 7th day, a mark of '0' (zero) will be awarded even if the assessment is submitted. Submission time for all written assessments is set at 11.55pm. A 1-hour grace period is provided to students who experience a technical issue. This late penalty will apply to written reports and recordings only. **Late submission of time sensitive tasks** (such as tests/exams, performance assessments/presentations, scheduled practical assessments/labs **will be addressed by the unit convenor in a Special consideration application.***

*Information about this unit's policy on the use of AI will be made available in the Assessment block in iLearn. Please check that information and contact the convenor if you have any questions.*

## Assessment Tasks

Name	Weighting	Hurdle	Due
<a href="#">Essay</a>	40%	No	03/11/2024 at 11:55 PM
<a href="#">Participation</a>	15%	No	Weeks 2 to 11

Name	Weighting	Hurdle	Due
<a href="#">Media presentation</a>	15%	No	09/10/2024 at 11:55 PM
<a href="#">Case Study</a>	30%	No	28/08/2024 at 11:55 PM

## Essay

Assessment Type <sup>1</sup>: Essay

Indicative Time on Task <sup>2</sup>: 33 hours

Due: **03/11/2024 at 11:55 PM**

Weighting: **40%**

Research essay on a topic from the unit

On successful completion you will be able to:

- Explain the fundamental principles underlying AI, and the normative constraints that it needs to satisfy.
- Demonstrate an advanced understanding of the ethical and other socioeconomic implications of AI.
- Demonstrate an advanced understanding of what Responsible AI means, or will mean, in our current as well as future society.

## Participation

Assessment Type <sup>1</sup>: Participatory task

Indicative Time on Task <sup>2</sup>: 10 hours

Due: **Weeks 2 to 11**

Weighting: **15%**

Active engagement in class discussions and associated activities

On successful completion you will be able to:

- Explain the fundamental principles underlying AI, and the normative constraints that it needs to satisfy.
- Demonstrate an advanced understanding of the ethical and other socioeconomic implications of AI.

- Demonstrate an advanced understanding of what Responsible AI means, or will mean, in our current as well as future society.
- Effectively communicate your findings to different stakeholders

## Media presentation

Assessment Type <sup>1</sup>: Media presentation

Indicative Time on Task <sup>2</sup>: 15 hours

Due: **09/10/2024 at 11:55 PM**

Weighting: **15%**

Media presentation

On successful completion you will be able to:

- Explain the fundamental principles underlying AI, and the normative constraints that it needs to satisfy.
- Demonstrate an advanced understanding of the ethical and other socioeconomic implications of AI.
- Demonstrate an advanced understanding of what Responsible AI means, or will mean, in our current as well as future society.
- Effectively communicate your findings to different stakeholders

## Case Study

Assessment Type <sup>1</sup>: Case study/analysis

Indicative Time on Task <sup>2</sup>: 25 hours

Due: **28/08/2024 at 11:55 PM**

Weighting: **30%**

Case study involving application of theoretical concepts to a practical context

On successful completion you will be able to:

- Explain the fundamental principles underlying AI, and the normative constraints that it needs to satisfy.
- Demonstrate an advanced understanding of the ethical and other socioeconomic implications of AI.
- Demonstrate an advanced understanding of what Responsible AI means, or will mean, in

our current as well as future society.

- Effectively communicate your findings to different stakeholders

---

<sup>1</sup> If you need help with your assignment, please contact:

- the academic teaching staff in your unit for guidance in understanding or completing this type of assessment
- the [Writing Centre](#) for academic skills support.

<sup>2</sup> Indicative time-on-task is an estimate of the time required for completion of the assessment task and is subject to individual variation

## Delivery and Resources

Delivery: All lectures are delivered live. The tutorial is held in person.

Resources: All required readings are provided in iLearn and Leganto. You must read the required readings before class.

## Unit Schedule

### W1 – Introduction (Dr Regina Fabry) – 25 July 2024

- No readings
- No tutorial

### W2 – Ethics and Robotics (A/Prof Abhaya Nayak) – 1 August 2024

- Reading 1: Birhane, A., & van Dijk, J. (2020). Robot rights? Let's talk about human welfare instead. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 207–213. <https://doi.org/10.1145/3375627.3375855>
- Reading 2: Vanderelst, D., & Winfield, A. (2018). An architecture for ethical robots inspired by the simulation theory of cognition. *Cognitive Architectures for Artificial Minds*, 48, 56–66. <https://doi.org/10.1016/j.cogsys.2017.04.002>
- Tutorial 1

### W3 – Algorithmic Decision Making (A/Prof Abhaya Nayak) – 8 August 2024

- Reading 1: Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 2053951719897945. <https://doi.org/10.1177/2053951719897945>

5

- Reading 2: Lim, H. S., & Taeihagh, A. (2019). Algorithmic decision-making in AVs: Understanding ethical and technical concerns for smart cities. *Sustainability*, 11(20). <https://doi.org/10.3390/su11205791>
- Tutorial 2

**W4 – AI and Safety (A/Prof Abhaya Nayak) – 15 August 2024**

- Reading 1: Limarga, R., Song, Y., Nayak, A., Rajaratnam, D., & Pagnucco, M. (forthcoming). Formalisation and evaluation of properties for consequentialist machine ethics. *Proceedings of the 33rd International Joint Conference in Artificial Intelligence*.
- Reading 2: Burton, S., Habli, I., Lawton, T., McDermid, J., Morgan, P., & Porter, Z. (2020). Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. *Artificial Intelligence*, 279, 103201. <https://doi.org/10.1016/j.artint.2019.103201>
- Tutorial 3

**W5 – Moral Responsibility of AI Researchers (Dr Qiongkai Xu / Dr Xiaohan Yu) – 22 August 2024**

- Reading 1: Freedman, R., Borg, J. S., Sinnott-Armstrong, W., Dickerson, J. P., & Conitzer, V. (2020). Adapting a kidney exchange algorithm to align with human values. *Artificial Intelligence*, 283, 103261. <https://doi.org/10.1016/j.artint.2020.103261>
- Reading 2: Schaich Borg, J. (2022). The AI field needs translational Ethical AI research. *AI Magazine*, 43(3), 294–307. <https://doi.org/10.1002/aaai.12062>
- Tutorial 4

**W6 – Ethical/Social AI Frameworks (Dr Regina Fabry) – 29 August 2024**

- Reading 1: Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Reading 2: Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks,

principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>

- Tutorial 5
- Assessment 1 (Case Study)

### **W7 – The Regulation of AI (Prof Niloufer Selvadurai) – 5 September 2024**

- Reading 1: Gacutan, J., & Selvadurai, N. (2020). A statutory right to explanation for decisions generated using artificial intelligence. *International Journal of Law and Information Technology*, 28(3), 193–216. <https://doi.org/10.1093/ijlit/eaad016>
- Reading 2: Smuha, N. A. (2021). From a ‘race to AI’ to a ‘race to AI regulation’: Regulatory competition for artificial intelligence. *Law, Innovation and Technology*, 13(1), 57–84. <https://doi.org/10.1080/17579961.2021.1898300>
- Tutorial 6

### **W8 – Power, Politics, and AI (Dr Regina Fabry) – 12 September 2024**

- Reading 1: Lazar, S. (2022). Power and AI: Nature and justification. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (p. 0). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780197579329.013.12>
- Reading 2: Campolo, A., & Crawford, K. (2020). Enchanted determinism: Power without responsibility in artificial intelligence. *Engaging Science, Technology, and Society*, 6, 1–19. <https://doi.org/10.17351/ests2020.277>
- Tutorial 7

### **W9 – What Is AI After All? The Turing Test Revisited (Dr Regina Fabry) – 3 October 2024**

- Reading 1: Proudfoot, D. (2013). Rethinking Turing’s Test. *The Journal of Philosophy*, 110(7), 391–411. <https://doi.org/10.5840/jphil2013110722>
- Reading 2: Wheeler, M. (2020). Deceptive Appearances: The Turing Test, Response-Dependence, and Intelligence as an Emotional Concept. *Minds and Machines*, 30(4), 513–532. <https://doi.org/10.1007/s11023-020-09533-8>



- Tutorial 8

### W10 – Explainable AI (Dr Regina Fabry) – 10 October 2024

- Reading 1: Zednik, C. (2021). Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy & Technology*, 34(2), 265–288. <https://doi.org/10.1007/s13347-019-00382-7>
- Reading 2: Russo, F., Schliesser, E., & Wagemans, J. (2023). Connecting ethics and epistemology of AI. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-022-01617-6>
- Tutorial 9
- Assessment 2 (Media presentation)

### W11 – Equitable AI (Dr Regina Fabry) – 17 October 2024

- Reading 1: Cossette-Lefebvre, H., & Maclure, J. (2022). AI's fairness problem: Understanding wrongful discrimination in the context of automated decision-making. *AI and Ethics*. <https://doi.org/10.1007/s43681-022-00233-w>
- Reading 2: Kasirzadeh, A. (2022). Algorithmic fairness and structural injustice: Insights from feminist political philosophy. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 349–356. <https://doi.org/10.1145/3514094.3534188>
- Tutorial 10

### W12 – Trustworthy AI? The Case of Chatbots (Dr Regina Fabry) – 24 October 2023

- Reading 1: Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Reading 2: Heersmink, R., de Rooij, B., Clavel Vázquez, M. J., & Colombo, M. (2024). A phenomenology and epistemology of large language models: Transparency, trust, and trustworthiness. *Ethics and Information Technology*, 26(3), 41. <https://doi.org/10.1007/s10676-024-09777-3>
- No Tutorial

## W13 – Writing and Review

- No Readings
- No Lecture
- No Tutorial
- Assessment 3 (Essay)

## Policies and Procedures

Macquarie University policies and procedures are accessible from [Policy Central \(https://policies.mq.edu.au\)](https://policies.mq.edu.au). Students should be aware of the following policies in particular with regard to Learning and Teaching:

- [Academic Appeals Policy](#)
- [Academic Integrity Policy](#)
- [Academic Progression Policy](#)
- [Assessment Policy](#)
- [Fitness to Practice Procedure](#)
- [Assessment Procedure](#)
- [Complaints Resolution Procedure for Students and Members of the Public](#)
- [Special Consideration Policy](#)

Students seeking more policy resources can visit [Student Policies \(https://students.mq.edu.au/support/study/policies\)](https://students.mq.edu.au/support/study/policies). It is your one-stop-shop for the key policies you need to know about throughout your undergraduate student journey.

To find other policies relating to Teaching and Learning, visit [Policy Central \(https://policies.mq.edu.au\)](https://policies.mq.edu.au) and use the [search tool](#).

## Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: <https://students.mq.edu.au/admin/other-resources/student-conduct>

## Results

Results published on platform other than [eStudent](#), (eg. iLearn, Coursera etc.) or released directly by your Unit Convenor, are not confirmed as they are subject to final approval by the University. Once approved, final results will be sent to your student email address and will be made available in [eStudent](#). For more information visit [ask.mq.edu.au](https://ask.mq.edu.au) or if you are a Global MBA student contact [globalmba.support@mq.edu.au](mailto:globalmba.support@mq.edu.au)

## Academic Integrity

At Macquarie, we believe [academic integrity](#) – honesty, respect, trust, responsibility, fairness and courage – is at the core of learning, teaching and research. We recognise that meeting the

expectations required to complete your assessments can be challenging. So, we offer you a range of resources and services to help you reach your potential, including free [online writing and maths support](#), [academic skills development](#) and [wellbeing consultations](#).

## Student Support

Macquarie University provides a range of support services for students. For details, visit <http://students.mq.edu.au/support/>

### The Writing Centre

[The Writing Centre](#) provides resources to develop your English language proficiency, academic writing, and communication skills.

- [Workshops](#)
- [Chat with a WriteWISE peer writing leader](#)
- [Access StudyWISE](#)
- [Upload an assignment to Studiosity](#)
- [Complete the Academic Integrity Module](#)

The Library provides online and face to face support to help you find and use relevant information resources.

- [Subject and Research Guides](#)
- [Ask a Librarian](#)

## Student Services and Support

Macquarie University offers a range of [Student Support Services](#) including:

- [IT Support](#)
- [Accessibility and disability support](#) with study
- Mental health [support](#)
- [Safety support](#) to respond to bullying, harassment, sexual harassment and sexual assault
- [Social support including information about finances, tenancy and legal issues](#)
- [Student Advocacy](#) provides independent advice on MQ policies, procedures, and processes

## Student Enquiries

Got a question? Ask us via [AskMQ](#), or contact [Service Connect](#).

## IT Help

For help with University computer systems and technology, visit [http://www.mq.edu.au/about\\_us/offices\\_and\\_units/information\\_technology/help/](http://www.mq.edu.au/about_us/offices_and_units/information_technology/help/).

When using the University's IT, you must adhere to the [Acceptable Use of IT Resources Policy](#). The policy applies to all who connect to the MQ network including students.

---

Unit information based on version 2024.04 of the [Handbook](#)