



PHIL3400

Rights, Responsibilities, and AI

Session 2, In person-scheduled-weekday, North Ryde 2026

School of Humanities

Contents

<u>General Information</u>	2
<u>Learning Outcomes</u>	3
<u>General Assessment Information</u>	3
<u>Assessment Tasks</u>	4
<u>Delivery and Resources</u>	5
<u>Unit Schedule</u>	6
<u>Policies and Procedures</u>	9

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

Prof Paul Formosa

paul.formosa@mq.edu.au

Lecturer

Dr Regina Fabry

regina.fabry@mq.edu.au

Lecturer

Prof Niloufer Selvadurai

niloufer.selvadurai@mq.edu.au

Lecturer & Tutor

Dr Ines Hipolito

ines.hipolito@mq.edu.au

Credit points

10

Prerequisites

130cp at 1000 level or above

Corequisites

Co-badged status

PHIL8400

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 understanding of the ethical and other socioeconomic implications of AI.

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

ULO4: Critically reflect on the use of AI in relevant fields.

General Assessment Information

Late Penalty

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.

Generative AI (ChatGPT etc)

In this unit, unless notified otherwise in writing by the Unit Convenor, it is expected that the work you submit demonstrates your own understanding, analysis, research, reflection, critical thinking, and writing, and you will be marked on the basis of your demonstrated understanding of and engagement with the unit content and activities. This applies to all assessments, including online forums. If your teachers have concerns about whether a submission is your own work in a meaningful sense, you may be required to attend an interview with the Unit Convenor or other academic staff to discuss your work and demonstrate your attainment of unit and task learning outcomes. Further information and advice about policies and expectations around the use of Generative AI will be given in iLearn.

Special Considerations

For details around special consideration, including applying for short extensions, see <https://students.mq.edu.au/study/assessment-exams/special-consideration>

Assessment Tasks

Name	Weighting	Hurdle	Due	Groupwork/Individual	Short Extension	AI Approach
Media presentation	20%	No	11/10/2026	Individual	Yes	Open
Research Essay	45%	No	08/11/2026	Individual	Yes	Open
Reflective task	35%	No	30/08/2026	Individual	Yes	Open

Media presentation

Assessment Type ¹: Presentation task

Indicative Time on Task ²: 18 hours

Due: **11/10/2026**

Weighting: **20%**

Groupwork/Individual: Individual

Short extension ³: Yes

AI Approach: Open

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 understanding of the ethical and other socioeconomic implications of AI.
- Demonstrate an understanding of what Responsible AI means, or will mean, in our current as well future society.

Research Essay

Assessment Type ¹: Written Submission

Indicative Time on Task ²: 35 hours

Due: **08/11/2026**

Weighting: **45%**

Groupwork/Individual: Individual

Short extension ³: Yes

AI Approach: Open

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 understanding of the ethical and other socioeconomic implications of AI.
- Demonstrate an understanding of what Responsible AI means, or will mean, in our current as well future society.
- Critically reflect on the use of AI in relevant fields.

Reflective task

Assessment Type ¹: Written Submission

Indicative Time on Task ²: 30 hours

Due: **30/08/2026**

Weighting: **35%**

Groupwork/Individual: Individual

Short extension ³: Yes

AI Approach: Open

Present arguments and defend your own view 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 understanding of the ethical and other socioeconomic implications of AI.
- Demonstrate an understanding of what Responsible AI means, or will mean, in our current as well future society.
- Critically reflect on the use of AI in relevant fields.

¹ 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
- [Academic Success](#) for academic skills support.

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

³ An automatic short extension is available for some assessments. Apply through the [Service Connect Portal](#).

Delivery and Resources

Delivery: All lectures are delivered live; echo recordings are available via iLearn. Online forums are available in iLearn. Tutorials for in-person students.

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

Unit Schedule

Unit Schedule

W1 – Introduction: Rights, Responsibilities, and AI (Prof Paul Formosa) – 28 July 2026

- No readings
- In person tutorial 1 OR Online forum 1

W2 – Ethical AI Frameworks (Prof Paul Formosa) – 4 August 2026

- 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., Cowls, 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>
- In person tutorial 2 OR Online forum 2

W3 – Robot Rights & Moral Status (Prof Paul Formosa) – 11 August 2026

- Reading 1: Schwitzgebel, E., & Garza, M. (2015). A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy*, 39(1), 98–119. <https://doi.org/10.1111/misp.12032>
- Reading 2: 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>
- In person tutorial 3 OR Online forum 3

W4 – Explainable AI (Dr Regina Fabry) – 18 August 2026

- 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. (2024). Connecting ethics and epistemology of AI. *AI & Society*, 39(4), 1585–1603. <https://doi.org/10.1007/s00146-022-01617-6>
- In person tutorial 4 OR Online forum 4

W5 – Equitable AI (Dr Regina Fabry) – 25 August 2026

- Reading 1: Cossette-Lefebvre, H., & Maclure, J. (2023). AI's fairness problem: Understanding wrongful discrimination in the context of automated decision-making. *AI and Ethics*, 3(4), 1255–1269. <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>

- In person tutorial 5 OR Online forum 5
- Assignment 1 (Reflective task)

W6 – Trustworthy AI: Chatbots (Dr Regina Fabry) – 1 September 2026

- 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>
- In person tutorial 6 OR Online forum 6

W7 – The Regulation of AI (Prof Niloufer Selvadurai) – 8 September 2026

- 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>
- In person tutorial 7 OR Online forum 7

W8 – AI, Automation & the Future of Work (Prof Paul Formosa) – 15 September 2026

- Reading 1: Bankins, S., & Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics*, 185(4), 725–740. <https://doi.org/10.1007/s10551-023-05339-7>
- Reading 2: Kim, T. W., & Scheller-Wolf, A. (2019). Technological unemployment, meaning in life, purpose of business, and the future of stakeholders. *Journal of Business Ethics*, 160(2), 319–337. <https://doi.org/10.1007/s10551-019-04205-9>
- In person tutorial 8 OR Online forum 8

RECESS FROM 21 SEPTEMBER TO 2 OCTOBER 2026

W9 – Power, Politics & AI: Censorship & Content Moderation (Dr Ines Hipolito) – 6 October 2026

- Reading 1: Lazar, S. (2022). Power and AI: Nature and justification. In J. B. Bullock et al. (Eds.), *The Oxford Handbook of AI Governance*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780197579329.013.12>

- Reading 2: Kashyap, B., & Formosa, P. (2026). From tweets to tyranny: Exploring the symmetry between state and social media censorship of speech. *Journal of Social Philosophy*. <https://doi.org/10.1111/josp.70052>

- In person tutorial 9 OR Online forum 9

- Assignment 2 (Media presentation)

W10 – Moral Responsibility & the Responsibility Gap (Dr Ines Hipolito) – 13 October 2026

- Reading 1: Gunkel, D. J. (2020). Mind the gap: Responsible robotics and the problem of responsibility. *Ethics and Information Technology*, 22(4), 307–320. <https://doi.org/10.1007/s10676-017-9428-2>

- Reading 2: Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, 26(4), 2051–2068. <https://doi.org/10.1007/s11948-019-00146-8>

- In person tutorial 10 OR Online forum 10

W11 – The Singularity & AI Safety (Dr Ines Hipolito) – 20 October 2026

- Reading 1: Chalmers, D. J. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies*, 17(9–10), 7–65. [Pages 7–40 ONLY]

- Reading 2: Millière, R. (2025). Normative conflicts and shallow AI alignment. *Philosophical Studies*, 182(7), 2035–2078. <https://doi.org/10.1007/s11098-025-02347-3> [Pages 2035–2062 ONLY]

- In person tutorial 11 OR Online forum 11

W12 – What Is AI? Intelligence & the Turing Test (Dr Ines Hipolito) – 27 October 2026

- 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>

- In person tutorial 12 OR Online forum 12

W13 – AI, Virtue & Human Flourishing (Dr Ines Hipolito) – 3 November 2026

- Reading 1: Vallor, S. (2015). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. *Philosophy & Technology*, 28(1), 107–124. <https://doi.org/10.1007/s13347-014-0156-9>

- Reading 2: Smith, N., & Vickers, D. (2024). Living well with AI: Virtue, education, and artificial intelligence. *Theory and Research in Education*, 22(1), 19–44. <https://doi.org/10.1177/14778785241231561>

- Assignment 3 (Research 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 connect.mq.edu.au or if you are a Global MBA student contact 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/>

Academic Success

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

- [Ask a Study Coach](#)
- [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 the [Service Connect Portal](#), or contact [Service Connect](#).

IT Help

For help with University computer systems and technology, visit <https://students.mq.edu.au/support/technology/service-desk>.

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 2026.06 of the [Handbook](#)