

Lingnan University
Department of Philosophy

Course Title	:	AI Ethics and Social Policy
Course Code	:	PAI4001
Recommended Study Term	:	Term 2, Fourth Year
No. of Credits/Term	:	3
Mode of Tuition	:	Sectional approach with optional Service-Learning component
Class Contact Hours	:	3 hours per week (plus 30 hours over the term for SL students)
Category	:	Required course
Prerequisite(s)	:	NIL
Co-requisite(s)	:	NIL
Exclusion(s)	:	NIL
Exemption Requirement(s)	:	NIL

Brief Course Description

The rapid advancement of AI technology poses novel and serious ethical challenges to society. These challenges must be addressed in part through rules and regulations laid down by governments and other policy-makers. But what sorts of policies are appropriate to deal with the challenges posed by AI? Which members of society should be empowered to make these decisions? What values should AI regulation aim to promote? In this course, we draw on resources from ethical and political philosophy and value theory to provide students with the philosophical tools necessary to evaluate and formulate policies to address the risks posed by emerging AI systems. Depending on the availability of partner agencies, Service-Learning will be implemented in the course. When implemented, students will have the option to participate in the Service-Learning activity or to complete the course through traditional teaching and learning methods.

Aims

The primary aim of this course is to educate students about the ethical risks posed by AI technology and to provide them with philosophical tools to evaluate various approaches that governments, businesses, or other policy-makers might take to address them. The course will impart students an understanding of both the philosophical underpinnings of the ethical challenges posed by AI systems as well as of contemporary discussions about how to address these challenges. Equipped with this knowledge, students will be able to meaningfully contribute to discussions with stakeholders in industry, government, and academia about AI risks and strategies for addressing them. For those students who opt for Service-Learning, a further aim of the course is that students develop the skills to lead workshops in which they

share the knowledge gained in the course with audiences in the partner agencies, particularly in relation to AI literacy, ethical awareness, and understanding of risks.

Learning Outcomes (LOs)

Upon completion of this course, students will be able to:

1. Characterize some main uses of AI tools as well as the potential risks deriving from them.
2. Evaluate and explain the impact of various AI tools from a variety of ethical, political, legal, and policy perspectives.
3. Anticipate future applications of AI and plan for their potential impacts.
4. Interpret and assess existing regulatory proposals put forth by governments, corporations, and other actors for addressing the social risks posed by AI systems.
5. Formulate and articulate one's own perspective on the correct social policies to implement regarding the ethical and social challenges posed by AI.
6. SL students only: Apply philosophical analysis of AI ethics to real-world contexts through engagement with community stakeholders.

Indicative Content

(Actual content may vary from term to term.)

1. The uses of AI technology in business, government, and academia.
2. Ethical frameworks for assessing AI's impact on society
3. Ethical and social AI risks including catastrophic risks
4. Data privacy and protection
5. Regulating 'frontier models'
6. Regulation for AI ethics (e.g. AI-caused misinformation, bias, job loss, etc.)
7. Business and financial use of AI
8. Aligning AI with human values
9. Existing and future efforts to regulate AI
10. International collaboration on AI regulation

Teaching Method

A sectional approach. The lecture component of the class will be used to introduce course materials (LO1, LO2, LO3, LO4). The sectional component of the course will focus on enabling students to formulate their own views regarding the course topics (LO5) and to reinforce the material covered in lectures (LO1, LO2, LO3, LO4). Sections will include open discussions as well as in-class activities including assessments such as the quizzes (LO1, LO2, LO4) and in-class presentations (LO1, LO2, LO3, LO4, LO5). Midterm and final examinations will consolidate students' understanding of key issues (LO1, LO2, LO4). And completing a research paper and presentation will tie together all of the themes of the class (LO1, LO2, LO3, LO4, LO5, LO6).

Service-Learning Project

Under the guidance of the course instructor and a project officer from the Office of Service-Learning, students who opt for Service-Learning will conduct educational workshop sessions with partner agencies, enabling them to apply and share the knowledge gained in the course. Working in small groups, students will develop a series of workshops covering different ethical and social issues raised by AI tools. The precise topics and modes of delivery will be determined by the students in collaboration with the project officer from the Office of Service-Learning. Workshops will be developed in consultation with partner agencies to ensure that they address the specific needs and concerns of participants. Students will also gather feedback from participants, which will inform both the delivery of the workshops and students' subsequent analysis. Students must spend at least 30 hours total over the term on Service-Learning components to meet the SL requirements for the course.

Topics may include: the use of AI in education (e.g. positive and negative uses, issues relating to deskilling and dependency, and how education systems should adapt); the creation of AI art and its implications (can AI create art? Do these tools enhance or endanger human creativity? How should we approach copyright and intellectual property law in light of AI?); AI relationships, including so-called "grief bots," and their possible impact on human relationships; AI-enabled misinformation and how it should be managed; and automation and the replacement of human workers by AI (how will the economic effects be managed? Will people be able to live meaningful lives without work?).

The workshops may include hands-on components, in which students introduce participants to different uses of AI tools; presentations on the ethical and social issues these tools raise; and student-led discussion sessions, in which students and participants engage with these issues together. Insights gained from these workshops will form a core component of the students' assessed work, particularly the research paper. Students will be required to analyse the concerns, values, and experiences expressed by participants and to relate these to the ethical frameworks and policy proposals studied in the course. This will enable them to evaluate whether existing approaches to AI governance adequately address real-world concerns and, where appropriate, to propose alternative or revised policy responses grounded in stakeholder perspectives.

Measurement of Learning Outcomes

Learning Outcomes	Assessment Methods				
	Attendance and participation	Quizzes	In-class presentation/ S-L Presentation	Research paper/ Research paper with reflection	Midterm and final examinations
LO1: Characterize some main uses of AI tools as well as the potential risks deriving from them.	✓	✓	✓	✓	✓
LO2: Evaluate and explain the impact of various AI tools from a variety of ethical, political, legal, and policy perspectives.	✓	✓	✓	✓	✓
LO3: Anticipate future applications of AI and plan for their potential impacts			✓	✓	
LO4: Interpret and assess existing regulatory proposals put forth by governments, corporations, and other actors for addressing the social risks posed by AI systems.	✓	✓	✓	✓	✓

LO5: Formulate and articulate one's own perspective on the correct social policies to implement regarding the ethical and social challenges posed by AI	✓		✓	✓	
LO6: Apply philosophical analysis of AI ethics to real-world contexts through engagement with community stakeholders.			✓	✓	

Assessment		Without S-L	With S-L
Participation and attendance	:	10%	10%
Quizzes	:	10%	10%
In-Class presentation	:	10%	-
S-L Presentation	:	-	10%
Research paper	:	20%	-
Research paper with reflection	:	-	20%
Midterm examination	:	20%	20%
Final examination	:	30%	30%

Participation and attendance – Students are expected to actively participate in classroom discussions by asking and answering questions and by raising objections to arguments in the materials studied.

Quizzes – Three short in-class quizzes will be conducted during the course. These will include multiple choice and short answer questions relating to the readings covered in the course.

In-Class Presentation/ S-L Presentation – Students will present the contents of their final research paper project in the form of a short (10 minute) class presentation. Students will be expected to present their ideas clearly. For students opting for Service-Learning, the presentation will be on the Service-Learning project, including details of the workshops and the lessons students learned

from them.

Research paper/ Research paper with reflection– Each student will write a research paper exploring an ethical, social, or policy issue raised by the development of AI technology. For students opting for Service-Learning, the research paper must be based on and informed by lessons learned during the educational workshops in partner agencies. This reflection will require students to critically assess how real-world concerns align with or challenge theoretical frameworks studied in the course.

Midterm and Final Examination – Midterm and final examinations will consist of a combination of multiple choice, short answer, and extended response questions. These will include case-based questions requiring students to apply ethical and policy frameworks to real-world and hypothetical scenarios involving AI systems.

Required/Essential Readings

Anderljung, M., Barnhart, J., Korinek, A., Leung, J., O’Keefe, C., & Whittlestone, J., et al. (2023). *Frontier AI regulation: Managing emerging risks to public safety* (White paper).
<https://arxiv.org/pdf/2307.03718>

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias*. ProPublica.
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Bullock, J. B., Chen, Y.-C., Himmelreich, J., Hudson, V. M., Korinek, A., Young, M. M., & Zhang, B. (Eds.). (2023). *The Oxford handbook of AI governance*. Oxford University Press.

Dubber, M. D., Pasquale, F., & Das, S. (Eds.). (2020). *The Oxford handbook of ethics of AI*. Oxford University Press.

Fazelpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, 16(8), e12760.

Habgood-Coote, J. (2023). Deepfakes and the epistemic apocalypse. *Synthese*, 201(3), 1–23.

Marchegiani, B. (2025). Anthropomorphism, false beliefs, and Conversational AIs: How Chatbots undermine users' autonomy. *Journal of Applied Philosophy*, 42(5):, 1399-1419.

Müller, V. C. (2021). Ethics of artificial intelligence and robotics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 ed.).
<https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/>

O’Brien, G. D. (2025). All play and no work? AI and existential unemployment. *Journal of Ethics*, 29, 747–771.

United Nations. (2024). *United Nations system white paper on AI governance*.
<https://unsceb.org/sites/default/files/2024-05/United%20Nations%20System%20White%20Paper%20on%20AI%20Governance.pdf>

Supplementary/Optional Readings

Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). *Hello, world: Artificial intelligence and its use in the public sector* (OECD Working Papers on Public Governance No. 36).
<https://doi.org/10.1787/726fd39d-en>

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, 108, 43–47.

Buckner, C. (2023). *From deep learning to rational machines: What the history of philosophy can teach us about the future of artificial intelligence*. Oxford University Press.

Cholbi, M., & Weber, M. (2019). *The future of work, technology, and basic income*. Routledge.

Coeckelbergh, M. (2020). *AI ethics*. MIT Press.

Floridi, L. (Ed.). (2010). *The Cambridge handbook of information and computer ethics*. Cambridge University Press.

Kleinberg, J., Ludwig, J., & Mullainathan, S. (2016, December 8). A guide to solving social problems with machine learning. *Harvard Business Review*. <https://hbr.org/2016/12/a-guide-to-solving-social-problems-with-machine-learning>

Long, R. (2021). Fairness in machine learning: Against false positive rate equality as a measure of fairness. *Journal of Moral Philosophy*, 19(1), 49–78. <https://doi.org/10.1163/17455243-20213439>

Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Farrar, Straus and Giroux.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach*. Pearson.

United Nations AI Advisory Body. (2023). *Interim report: Governing AI for humanity*.
https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf

Weidinger, L., et al. (2021). *Ethical and social risks of harm from language models* (White paper). DeepMind.

The reading list will also be supplemented by recent press releases, technical reports, and research outputs from leading AI organisations (e.g. Anthropic, OpenAI, DeepMind), ensuring that students engage with the most current developments in AI ethics and governance.

Important Notes

1. Students are expected to spend a total of 9 hours (i.e. 3 hours of class contact and 6 hours of personal study) per week to achieve the course learning outcomes.

2. Students shall be aware of the University regulations about dishonest practice in coursework, tests and examinations, and the possible consequences as stipulated in the Regulations Governing University Examinations and Course Work. In particular, plagiarism, being a kind of dishonest practice, is “the presentation of another person’s work without proper acknowledgement of the source, including exact phrases, or summarised ideas, or even footnotes/citations, whether protected by copyright or not, as the student’s own work”. Students are required to strictly follow university regulations governing academic integrity and honesty.
3. Students are required to submit writing assignment(s) using Turnitin.
4. To enhance students’ understanding of plagiarism, a mini-course “Online Tutorial on Plagiarism Awareness” is available on <https://pla.ln.edu.hk/>.
5. Students must adhere to the University’s guidelines and practices when using Generative Artificial Intelligence (AI) tools. The official documents “Guidelines for Using GAI Tools at Lingnan University” and “Best Practices for Ethical and Responsible Use of GAI Tools in Course Assessments” can be found here: <https://www.ln.edu.hk/cital/generative-artificial-intelligence/gai-guidelines-best-practices>. Use of AI tools in the preparation of assessed material (presentations, and research papers) is strictly forbidden.

AI Ethics and Social Policy
Participation and Attendance Rubric (10%)

Excellent – A (85-100%)	Good - B (70-84%)	Fair - C (55-69%)	Pass - D (40-54%)	Failure - F (0-39%)
Answers or raises questions or comments in every class; comments are mostly cogent and relevant and advance the discussion; full attendance (except absence with prior notice and valid reasons.)	Attempts frequently to answer or raise questions; comments are often cogent and relevant and advance discussion; almost full attendance.	Attempts sometimes to answer or raise questions; comments are sometimes cogent and relevant; skips quite a few classes without notice.	Seldom attempts to answer or raise questions, but can sometimes offer relevant answers when asked; skips many classes.	Almost never makes cogent or relevant comment, even when asked; skips most classes.

AI Ethics and Social Policy
In-Class Presentation (Non-SL track) Rubric (10%)

Assessment Categories	Excellent (A)	Good (B)	Fair (C)	Pass (D)	Failure (F)
Communication (Weight: 25%)	Excellent presentation; closely engages audience; persuasive delivery.	Generally holds attention; generally good delivery.	Delivery is fair, somewhat engaging; makes an attempt to command attention of class	Delivery is apparently detached; fails to command attention of class	Little or no attempt to communicate with the class; non-existent presentation skills.
Organization (Weight: 25%)	Excellent structure; transitioning from topic to topic is clear and logical.	Generally good structure; transitioning from topic to topic is mostly clear.	Generally acceptable; some organization skills demonstrated.	Lack of organization; jumps from topic to topic without proper signals.	Ideas are apparently randomly expressed.
Argumentation (Weight: 25%)	Reasoning is almost impeccable; excellent usage and accreditation of supporting evidence.	Reasoning is strong; supporting evidence is demonstrated clearly.	Reasoning is fair; some usage of supporting evidence.	Contains major fallacies in reasoning; scant usage of supporting evidence	Little or no demonstration of reasoning skills.

		Good (B)	Fair (C)	Pass (D)	Failure (F)
		Paper demonstrates knowledge of most important points; shows understanding of the basic philosophical positions and disagreements.	Paper demonstrates knowledge of most important point, and the most important points of disagreement, but with some unclarity and inaccuracy.	Some of the basic points are understood, but the student shows substantial misunderstandings.	Student's understanding deviates significantly from the positions discussed in class/ possibly fails to write on the assigned topic.
	Essay presents compelling arguments, shows	Essay presents compelling arguments and shows	Essay presents arguments that successfully	Arguments are often vague and impressionistic, or	No discernible arguments/ no engagement

In-Class Presentation (SL track) Rubric (10%)

Assessment Categories	Excellent (A)	Good (B)	Fair (C)	Pass (D)	Failure (F)
Communication (Weight: 25%)	Excellent presentation; closely engages audience; persuasive delivery.	Generally holds attention; generally good delivery.	Delivery is fair, somewhat engaging; makes an attempt to command attention of class	Delivery is apparently detached; fails to command attention of class	Little or no attempt to communicate with the class; non-existent presentation skills.

Organization (Weight: 25%)	Excellent structure; transitioning from topic to topic is clear and logical.	Generally good structure; transitioning from topic to topic is mostly clear.	Generally acceptable; some organization skills demonstrated.	Lack of organization; jumps from topic to topic without proper signals.	Ideas are apparently randomly expressed.
Workshop activities (Weight: 50%)	Excellent engagement with the community and critical reflection on their concerns and values around AI policy.	Very good engagement with the community and critical reflection on their concerns and values around AI policy.	Moderate engagement with the community and some critical reflection on their concerns and values around AI policy.	Adequate engagement with the community and some reflection on their concerns and values around AI policy.	Little or no demonstration of engagement with the community or reflection on their concerns and values around AI policy.

I Ethics and Social Policy Research Paper (Non-SL track) (20%)

Assessment Categories	Excellent (A)	Good (B)	Fair (C)	Pass (D)	Failure (F)
Knowledge of Course Material (Weight: 40%)	Knowledge goes substantially beyond what is expected of students in this course.	Paper demonstrates knowledge of most important points; shows understanding of the basic philosophical positions and disagreements.	Paper demonstrates knowledge of most important point, and the most important points of disagreement, but with some unclarity and inaccuracy.	Some of the basic points are understood, but the student shows substantial misunderstandings.	Student's understanding deviates significantly from the positions discussed in class/ possibly fails to write on the assigned topic.

Argumentation and Originality (Weight: 40%)	Essay presents compelling arguments, shows understanding of important objections, and offers novel arguments or ideas that show a high degree of understanding.	Essay presents compelling arguments and shows understanding of important objections. Novel portions are less compelling or absent.	Essay presents arguments that successfully motivate the main claims, but details are overlooked, and objections are not handled as well.	Arguments are often vague and impressionistic, or supported with dubious claims or fallacious modes of reasoning.	No discernible arguments/ no engagement with the ideas of the class.
Organization (Weight: 10%)	Ideas are clearly arranged to allow easy understanding of their relations.	As before, but with slightly more mistakes in arrangement of ideas.	Basic structure is good, but sometimes the arrangement of ideas is confusing.	Essay seems to switch topics at points.	Total lack of organizational structure/ incomplete.
Writing (Weight: 10%)	Flawless, engaging writing.	Occasional mistakes which do not impede comprehension.	More substantial mistakes which make comprehension difficult.	Portions of the paper cannot be assessed because of problems with the writing.	The content of the paper cannot be evaluated.

AI Ethics and Social Policy Research Paper (SL track) Rubric (20%)

Assessment Categories	Excellent (A)	Good (B)	Fair (C)	Pass (D)	Failure (F)
Knowledge of Course Material (Weight: 30%)	Knowledge goes substantially beyond what is expected of students in this course.	Paper demonstrates knowledge of most important points; shows understanding of the basic philosophical positions and disagreements.	Paper demonstrates knowledge of most important point, and the most important points of disagreement, but with some unclarity and inaccuracy.	Some of the basic points are understood, but the student shows substantial misunderstandings.	Student's understanding deviates significantly from the positions discussed in class/ possibly fails to write on the assigned topic.

Community Engagement and Reflection (Weight: 30%)	Paper draws extensively and specifically on community perspectives encountered during workshops. Reflection on how engagement shaped or challenged the student's argument is detailed, honest, and intellectually substantive.	Paper references community engagement meaningfully and reflection is present, though connections to the academic argument could be developed further.	Paper makes some reference to community engagement, but reflection is superficial or the connection to the academic argument is underdeveloped	Community engagement is mentioned only in passing with little or no genuine reflection on how it informed the paper's argument.	No reference to community engagement or reflection on the S-L experience.
Argumentation and Originality (Weight: 20%)	Essay presents compelling arguments, shows understanding of important objections, and offers novel arguments or ideas that show a high degree of understanding.	Essay presents compelling arguments and shows understanding of important objections. Novel portions are less compelling or absent	Essay presents arguments that successfully motivate the main claims, but details are overlooked, and objections are not handled as well.	Arguments are often vague and impressionistic, or supported with dubious claims or fallacious modes of reasoning.	No discernible arguments/ no engagement with the ideas of the class.
Organization (Weight: 10%)	Ideas are clearly arranged to allow easy understanding of their relations.	As before, but with slightly more mistakes in arrangement of ideas.	Basic structure is good, but sometimes the arrangement of ideas is confusing.	Essay seems to switch topics at points.	Total lack of organizational structure/ incomplete.
Writing (Weight: 10%)	Flawless, engaging writing.	Occasional mistakes which do not impede comprehension.	More substantial mistakes which make comprehension difficult.	Portions of the paper cannot be assessed because of problems with the writing.	The content of the paper cannot be evaluated.

AI Ethics and Social Policy

Quizzes (10%), Midterm Examination (20%), Final Examination (30%) Rubrics

Rubric for Grading Examinations and Quizzes

For questions with clear-cut answers (e.g. definitions of key terms or explaining an argument or position from the literature):

Perfect (100%)	Partial Credit (20-80%)	Zero (0%)
The concluding answer is correct, with a clear, sound, and sufficient argument or explanation, if needed.	The concluding answer is correct, but the argument or explanation is unclear or insufficient, or the concluding answer is incorrect, but some intermediate steps are correct.	Neither the concluding answer nor anything in the argument or explanation is correct.

For essay questions or open-ended questions:

	Excellent (85-100%)	Good (70-84%)	Fair (55-69%)	Pass (40-54%)	Failure (0-39%)
Knowledge (Weight: 40%)	Demonstrates knowledge of all the relevant points, with a clear and accurate exposition.	Demonstrates knowledge of most relevant points, with a largely clear and accurate exposition.	Demonstrates knowledge of most relevant points, but the exposition is unclear or inaccurate.	Demonstrates only a limited understanding of the basic points.	Demonstrates very little understanding of even the basic points.
Argumentation and originality (Weight: 40%)	Presents novel, compelling arguments with enough details.	Presents compelling arguments, with less novelty or some gaps.	Presents reasonable arguments that are not novel or very compelling.	Attempts to express an argument, but the result is unclear or messy.	Makes little or no attempt to present arguments.
Clarity and organization (Weight: 20%)	Ideas are clearly arranged to allow easy understanding of their relations.	As excellent, but with slightly more mistakes in arrangement of ideas.	Basic structure is good, but sometimes arrangement of ideas is confusing.	Switches topics at points or is difficult to follow due to unclear structure.	Lack of organizational structure/ Largely incomplete.