

The Hong Kong Polytechnic University

Subject Description Form

Please read the notes at the end of the table carefully before completing the form.

Subject Code	APSS1A38											
Subject Title	Ethics of Artificial Intelligence											
Credit Value	3											
Level	1											
GUR Requirements Intended to Fulfill	<p>This subject intends to fulfill the following requirement(s) :</p> <ul style="list-style-type: none"> <input type="checkbox"/> Healthy Lifestyle <input type="checkbox"/> AI and Data Analytics (AIDA) <input type="checkbox"/> Innovation and Entrepreneurship (IE) <input type="checkbox"/> Languages and Communication Requirement (LCR) <input type="checkbox"/> Leadership Education and Development (LEAD) <input type="checkbox"/> Service-Learning <input checked="" type="checkbox"/> Cluster-Area Requirement (CAR) <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Human Nature, Relations and Development [CAR A] <input type="checkbox"/> Science, Technology and Environment [CAR D] <input type="checkbox"/> Chinese History and Culture [CAR M] <input type="checkbox"/> Cultures, Organizations, Societies and Globalization [CAR N] <input type="checkbox"/> China-Study Requirement <ul style="list-style-type: none"> <input type="checkbox"/> Yes or <input checked="" type="checkbox"/> No <input checked="" type="checkbox"/> Writing and Reading Requirements <ul style="list-style-type: none"> <input checked="" type="checkbox"/> English or <input type="checkbox"/> Chinese 											
Pre-requisite/ Co-requisite/ Exclusion	Nil											
Assessment Methods	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 50%;">100% Continuous Assessment</th> <th style="width: 25%;">Individual Assessment</th> <th style="width: 25%;">Group Assessment</th> </tr> </thead> <tbody> <tr> <td style="text-align: left;">1. Term Paper [EW Assessment: 40% to be assessed by subject teacher and 10% assessed by ELC]</td> <td>50%</td> <td style="text-align: right;">/</td> </tr> <tr> <td style="text-align: left;">2. Quizzes [ER Assessment]</td> <td>40% [4 x 10%]</td> <td style="text-align: right;">/</td> </tr> </tbody> </table>			100% Continuous Assessment	Individual Assessment	Group Assessment	1. Term Paper [EW Assessment: 40% to be assessed by subject teacher and 10% assessed by ELC]	50%	/	2. Quizzes [ER Assessment]	40% [4 x 10%]	/
100% Continuous Assessment	Individual Assessment	Group Assessment										
1. Term Paper [EW Assessment: 40% to be assessed by subject teacher and 10% assessed by ELC]	50%	/										
2. Quizzes [ER Assessment]	40% [4 x 10%]	/										

	3. Class Participation (included take-home assignments)	10%	/
Objectives	<p>This course aims to raise students’ awareness of the ethical issues posed by AI technologies and their integration into modern life and society. Students will learn basic concepts about artificial intelligence, ethical theories, as well as issues related to the ethical and social implications of AI. Students will be encouraged to critically reflect on their own use of and engagement with AI technologies as well as the likely and potential future applications of AI and their societal impacts. More specifically, the aims of this course are as follows:</p> <ol style="list-style-type: none"> 1) To improve students’ understanding of the ethical and social issues raised by AI and related technologies. 2) To equip students with ethical concepts and theories useful for analysing the impact of AI and related technologies on human life and society. 3) To instigate students to critically reflect on their own use of and exposure to AI technologies in various personal and social domains. 		
Intended Learning Outcomes <i>(Note 1)</i>	<p>On completing the subject, students will be able to:</p> <ol style="list-style-type: none"> a) Demonstrate an understanding of the ethical issues and social impacts raised by AI technologies and their deployments in society. b) Apply philosophical concepts in analysing the ethical, existential, and social implications of AI and related technologies. c) Reflect critically on their own use of and exposure to AI technologies in everyday life and society. d) Enhance their skills in English reading and writing. 		
Subject Synopsis/ Indicative Syllabus <i>(Note 2)</i>	<p>Breakthroughs in the development and deployment of artificial intelligence (AI) have already changed the world and will, without a doubt, precipitate more radical changes. This course addresses itself to analyse and assess the ethical and social implications of these technologies and their deployments. Topics will include the social issues surrounding the production and classification of data, transparency/ ‘black box’ issues, algorithmic bias, responsibility gaps as well as the social implications of AI including technological</p>		

unemployment and its connection to economic inequality and crises of meaning and purpose, asymmetries of information and power, and ‘surveillance capitalism’ (Zuboff). The course will also explore AI’s application to social robotics and autonomous machines. Topics to be considered, include self-driving vehicles, robots in eldercare, robot friends, questions of moral agency/patency, and how embodied AI may alter human relationships. The course will end by considering various images of how AI can be integrated well into our social world, including both utopia and dystopian scenarios.

Indicative Outline:

1. Introduction to AI Ethics

AI in the Popular Imaginary
Brief History of AI
Machine Learning; Deep Learning
ANI and AGI; Weak and Strong AI
Technological Ethics: Basic Theories

2. Ethics of Data and Algorithms

Social Origins of Data and Microwork
The Ethics of Data and Classification
Algorithmic Bias
AI and Racism/Sexism
Feedback Loops
Black Box/Transparency Issues
Privacy
Ethics of ChatGPT and LLMs

3. Ethics of Autonomous Machines

Self-Driving Vehicles
Responsibility and Retribution Gaps
Comparison with Trolley-Problems
Autonomous Weapons Systems

4. Ethics of AI & Social Robotics

AI Assistants
Social Robotics and Eldercare
Robot Friends and Lovers
Moral Status of Artificial Agents

5. Social Implications of AI

AI and Environmental Issues
AI in decision-making (hiring, policing, education, law)
Facial Recognition, Affect Recognition, & Surveillance
Information Asymmetries and Power
AI and Technological Unemployment
AI and Economic Inequality
Crises of Meaning and Purpose
Surveillance Capitalism (Zuboff)

6. A.I. & Utopian/Dystopian Scenarios

	AI Superintelligence Domsday Scenarios Value (Mis)alignment AI Utopias (Danaher)																																																
Teaching/Learning Methodology <i>(Note 3)</i>	<p>The class will be organized primarily around lectures, which will involve class discussions and opportunities for students to express their own views and actively participate. Feedback and guidance will be given by the lecturer throughout the class, including but not limited to consultations.</p> <p>In order to meet the requirement for the ‘EW’, students should use resources from the ELC to improve the quality of their writing. This involves, more specifically: (1) accessing additional online materials for improving writing that are provided by the ELC and (2) students will be required to submit a plan and a draft to the ELC for feedback and 10% of the total grade will be constituted by the ELC grade. The second draft should be a minimum of 1,500 words.</p>																																																
Assessment Methods in Alignment with Intended Learning Outcomes <i>(Note 4)</i>	<table border="1" data-bbox="533 913 1375 1727"> <thead> <tr> <th data-bbox="533 913 850 1122" rowspan="2">Specific assessment methods/tasks</th> <th data-bbox="850 913 1007 1122" rowspan="2">% weighting</th> <th colspan="5" data-bbox="1007 913 1375 1048">Intended subject learning outcomes to be assessed (Please tick as appropriate)</th> </tr> <tr> <th data-bbox="1007 1048 1075 1122">a</th> <th data-bbox="1075 1048 1144 1122">b</th> <th data-bbox="1144 1048 1212 1122">c</th> <th data-bbox="1212 1048 1281 1122">d</th> <th data-bbox="1281 1048 1350 1122"></th> <th data-bbox="1350 1048 1375 1122"></th> </tr> </thead> <tbody> <tr> <td data-bbox="533 1122 850 1391"> 1. Term Paper [EW Assessment: 40% to be assessed by subject teacher and 10% assessed by ELC] </td> <td data-bbox="850 1122 1007 1391">50%</td> <td data-bbox="1007 1122 1075 1391">✓</td> <td data-bbox="1075 1122 1144 1391">✓</td> <td data-bbox="1144 1122 1212 1391">✓</td> <td data-bbox="1212 1122 1281 1391">✓</td> <td data-bbox="1281 1122 1350 1391"></td> <td data-bbox="1350 1122 1375 1391"></td> </tr> <tr> <td data-bbox="533 1391 850 1514"> 2. Quizzes [ER Assessment] </td> <td data-bbox="850 1391 1007 1514">40%</td> <td data-bbox="1007 1391 1075 1514">✓</td> <td data-bbox="1075 1391 1144 1514">✓</td> <td data-bbox="1144 1391 1212 1514"></td> <td data-bbox="1212 1391 1281 1514">✓</td> <td data-bbox="1281 1391 1350 1514"></td> <td data-bbox="1350 1391 1375 1514"></td> </tr> <tr> <td data-bbox="533 1514 850 1659"> 3. Class Participation (included take-home assignments) </td> <td data-bbox="850 1514 1007 1659">10%</td> <td data-bbox="1007 1514 1075 1659">✓</td> <td data-bbox="1075 1514 1144 1659">✓</td> <td data-bbox="1144 1514 1212 1659">✓</td> <td data-bbox="1212 1514 1281 1659">✓</td> <td data-bbox="1281 1514 1350 1659"></td> <td data-bbox="1350 1514 1375 1659"></td> </tr> <tr> <td data-bbox="533 1659 850 1727"> Total </td> <td data-bbox="850 1659 1007 1727">100 %</td> <td colspan="5" data-bbox="1007 1659 1375 1727"></td> </tr> </tbody> </table> <p data-bbox="533 1749 1375 1816">Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:</p> <ol data-bbox="533 1839 1375 2114" style="list-style-type: none"> The <i>term paper</i> (1,500 – 2,500 words) will assess how students reflect on the ethical issues and social implications related to AI technologies. The term paper will give students an opportunity to demonstrate comprehension, critical engagement, and integration of key concepts from lecture and class reading. The term paper will assess both comprehension of key philosophical/ethical concepts relevant to AI technology as well as the ability to apply these concepts critically to one’s own life and social context. 					Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)					a	b	c	d			1. Term Paper [EW Assessment: 40% to be assessed by subject teacher and 10% assessed by ELC]	50%	✓	✓	✓	✓			2. Quizzes [ER Assessment]	40%	✓	✓		✓			3. Class Participation (included take-home assignments)	10%	✓	✓	✓	✓			Total	100 %					
Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)																																															
		a	b	c	d																																												
1. Term Paper [EW Assessment: 40% to be assessed by subject teacher and 10% assessed by ELC]	50%	✓	✓	✓	✓																																												
2. Quizzes [ER Assessment]	40%	✓	✓		✓																																												
3. Class Participation (included take-home assignments)	10%	✓	✓	✓	✓																																												
Total	100 %																																																

	<p>Students will also be given a supplementary bibliography and encouraged to do additional research to enhance their papers. In order to meet the ‘EW’ (English Writing) requirement, students will be required to submit a writing plan and a draft to the ELC for feedback and 10% of the total grade will be constituted by the ELC grade.</p> <p>2. <i>Quizzes</i> will assess whether students have understood basic philosophical concepts concerning the ethical and social implications of artificial intelligence and its social applications. This will be based on lecture material as well as required reading in accordance with the ‘ER’ (English Reading) requirement.</p> <p>3. <i>Class participation</i> will be assessed based on in-class and take-home activities, e.g., responses to reflective prompts, class discussion, take-home assignments, etc. These activities will ask students to connect philosophical concepts to their own use of AI technologies and involvement in a society where these technologies are used. They will be asked to identify ethical issues in their own immediate social contexts.</p>	
Student Study Effort Expected	Class contact:	
	<ul style="list-style-type: none"> ▪ Lecture 	39 Hrs.
	Other student study effort:	
	<ul style="list-style-type: none"> ▪ Course Reading 	30 Hrs.
	<ul style="list-style-type: none"> ▪ Writing Term Paper 	30 Hrs.
	<ul style="list-style-type: none"> ▪ Other Take-Home Assignments 	10 Hrs.
	<ul style="list-style-type: none"> ▪ Studying course materials 	26 Hrs.
	Total student study effort	135 Hrs.
Reading List and References	<p><i>Required Reading:</i></p> <ul style="list-style-type: none"> ▪ Crawford, K. (2021). <i>Atlas of AI</i>. Yale University Press [pp. 1-237] <p><i>Suggested Supplementary Readings:</i></p> <ul style="list-style-type: none"> ▪ Boden, M. (2018). <i>Artificial intelligence: A very short introduction</i>. Oxford University Press. ▪ Borgmann, A. (1984). <i>Technology and the character of contemporary life: A philosophical inquiry</i>. University of Chicago Press ▪ Bostrom, N. (2014). <i>Superintelligence: paths, dangers, strategies</i>. Oxford University Press. ▪ Broussard, M. (2018). <i>Artificial unintelligence: How computers misunderstand the world</i>. MIT Press. ▪ Chan, B. (2020). The rise of artificial intelligence and the crisis of moral passivity. <i>AI and Society</i>, 35: 991-993. 	

- Christian, B. (2020). *The alignment problem: Machine learning and human values*. Norton.
- Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
- Coeckelbergh, M. (2022). *The Political Philosophy of AI*. Polity Press.
- Danaher, J. (2019). *Automation and utopia*. Harvard University Press.
- Danaher, J. (2018). Toward an ethics of AI assistants: An initial framework. *Philosophy and Technology* 31: 629-653.
- Danaher, J. (2017). Will life be worth living in a world without work? Technological unemployment and the meaning of life. *Science and Engineering Ethics* 23, 41-64.
- Dreyfus, H. (1992). *What computers still can't do*. MIT Press.
- Dubner, M., Pasquale, F., Sunit, D. (2021). *The Oxford handbook of ethics of AI*. Oxford University Press.
- Dumouchel, P. & Damaiano, L. (2017). *Living with robots*. Harvard University Press.
- Floridi, L. (2019) Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology* 32: 185-193.
- Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press.
- Kearns, M. & Roth, A. (2020). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.
- Larson, E. (2021). *The myth of artificial intelligence: why computers can't think the way we do*. Belknap Press.
- Lee, K. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Mariner.
- Lee, K. and Chen, Q. (2021). *AI 2041: Ten visions for our future*. Currency.
- Liao, M. (ed.) (2020). *Ethics of artificial intelligence*. Oxford University Press.
- Lin, P. (2016) Why ethics matters for autonomous cars, In *Autonomous driving* (ed.) M. Mauer et. al. Springer.
- Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- O'Neil, C. (2016). *Weapons of math destruction*. Crown.
- Pasquale, F. (2020). *New laws of robotics: Defending human expertise in the age of AI*. Belknap Press.
- Pasquale, (2016). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

	<ul style="list-style-type: none"> ▪ Ramsey, W. and Frankish, F. (ed.) (2011) <i>Cambridge handbook of artificial intelligence</i>, Cambridge University Press. ▪ Schneider, S (2019). <i>Artificial you: AI and the future of your mind</i>, Princeton University Press. ▪ Schellman, H. (2024). <i>The Algorithm: How AI Decides Who Gets Hired, Monitored, Promoted and Fired and Why We Need to Fight Back Now</i>. Hachette Books. ▪ Sharkey, N. and Sharkey, A. (2012). Granny and the robots: ethical issues in robot care for the elderly. <i>Ethics of Information Technology</i> 14: 27-40. ▪ Sparrow, R. (2007). Killer robots. <i>Journal of Applied Philosophy</i> 24: 62-77. ▪ Susskind, D. (2020). <i>A world without work</i>. Allen Lane. ▪ Tegmark, M. (2017). <i>Life 3.0: Being human in the age of artificial intelligence</i>. Penguin. ▪ Turkle, S. (2011). <i>Alone together: Why we expect more from technology and less from each other</i>. Basic Books. ▪ Vallor, S. (2016). <i>Technology and the virtues: A philosophical guide to a future worth wanting</i>. Oxford University Press. ▪ Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. <i>Philosophy and Technology</i>. 24 (3): 251-268. ▪ Vallor, S. (2024). <i>The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking</i>. Oxford University Press. ▪ Woodridge, M. (2020). <i>The road to conscious machines: The Story of AI</i>. Penguin. ▪ Zuboff, S. (2019). <i>The age of surveillance capitalism</i>. Profile Books.
--	--

Note 1: Intended Learning Outcomes

Intended learning outcomes should state what students should be able to do or attain upon subject completion. Subject outcomes are expected to contribute to the attainment of the overall programme outcomes.

Note 2: Subject Synopsis/Indicative Syllabus

The syllabus should adequately address the intended learning outcomes. At the same time, overcrowding of the syllabus should be avoided.

Note 3: Teaching/Learning Methodology

This section should include a brief description of the teaching and learning methods to be employed to facilitate learning, and a justification of how the methods are aligned with the intended learning outcomes of the subject.

Note 4: Assessment Method

This section should include the assessment method(s) to be used and its relative weighting, and indicate which of the subject intended learning outcomes that each method is intended to assess. It should also provide a brief explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes.

(Form AR 140) 8.2020