The Hong Kong Polytechnic University

Subject Description Form

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

<table>
<tr><td>Subject Code</td><td>APSS1A38</td></tr>
<tr><td>Subject Title</td><td>Ethics of Artificial Intelligence</td></tr>
<tr><td>Credit Value</td><td>3</td></tr>
<tr><td>Level</td><td>1</td></tr>
<tr><td>GUR Requirements Intended to Fulfill</td><td>This subject intends to fulfill the following requirement(s) :

☐ Healthy Lifestyle
☐ AI and Data Analytics (AIDA)
☐ Innovation and Entrepreneurship (IE)
☐ Languages and Communication Requirement (LCR)
☐ Leadership Education and Development (LEAD)
☐ Service-Learning
☒ Cluster-Area Requirement (CAR)
☒ Human Nature, Relations and Development [CAR A]
☐ Science, Technology and Environment [CAR D]
☐ Chinese History and Culture [CAR M]
☐ Cultures, Organizations, Societies and Globalization [CAR N]
☐ China-Study Requirement
☐ Yes or ☒ No
☒ Writing and Reading Requirements
☒ English or ☐ Chinese</td></tr>
<tr><td>Pre-requisite/ Co-requisite/ Exclusion</td><td>Nil</td></tr>
<tr><td>Assessment Methods</td><td>
<table>
<tr><th>100% Continuous Assessment</th><th>Individual Assessment</th><th>Group Assessment</th></tr>
<tr><td>1. Term Paper
[EW Assessment: 40% to be assessed by subject teacher and 10% assessed by ELC]</td><td>50%</td><td></td></tr>
<tr><td>2. Quizzes
[ER Assessment]</td><td>40%
[4 x 10%]</td><td></td></tr>
</table>
</td></tr>
</table>

	<table><tr><td>3. Class Participation (included take-home assignments)</td><td>10%</td><td></td></tr></table> • The grade is calculated according to the percentage assigned; • The completion and submission of all component assignments are required for passing the subject; • Student must pass all component(s) if he/she is to pass the subject; and • Student must obtain a D grade or above on the Writing Requirement assignment to pass the subject.
Objectives	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: 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 *(Note 1)*	On completing the subject, students will be able to: 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 *(Note 2)*	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

	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/patiency, 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**

<table>
<tr><td></td><td>AI Superintelligence
Doomsday Scenarios
Value (Mis)alignment
AI Utopias (Danaher)</td></tr>
<tr><td>Teaching/Learning Methodology
(Note 3)</td><td>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.

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.</td></tr>
<tr><td>Assessment Methods in Alignment with Intended Learning Outcomes
(Note 4)</td><td>
<table>
<tr><td rowspan="2">Specific assessment methods/tasks</td><td rowspan="2">% weighting</td><td colspan="6">Intended subject learning outcomes to be assessed (Please tick as appropriate)</td></tr>
<tr><td>a</td><td>b</td><td>c</td><td>d</td><td></td><td></td></tr>
<tr><td>1. Term Paper
[EW Assessment: 40% to be assessed by subject teacher and 10% assessed by ELC]</td><td>50%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td></td><td></td></tr>
<tr><td>2. Quizzes
[ER Assessment]</td><td>40%</td><td>✓</td><td>✓</td><td></td><td>✓</td><td></td><td></td></tr>
<tr><td>3. Class Participation (included take-home assignments)</td><td>10%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td></td><td></td></tr>
<tr><td>Total</td><td>100 %</td><td colspan="6"></td></tr>
</table>
Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:

1. The term paper (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.</td></tr>
</table>

<table>
<tr><td></td><td colspan="2">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.

2. Q*uizzes* 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.

3. *Class participation* 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.</td></tr>
<tr><td rowspan="8">Student Study Effort Expected</td><td>Class contact:</td><td></td></tr>
<tr><td>▪ Lecture</td><td>39 Hrs.</td></tr>
<tr><td>Other student study effort:</td><td></td></tr>
<tr><td>▪ Course Reading</td><td>30 Hrs.</td></tr>
<tr><td>▪ Writing Term Paper</td><td>30 Hrs.</td></tr>
<tr><td>▪ Other Take-Home Assignments</td><td>10 Hrs.</td></tr>
<tr><td>▪ Studying course materials</td><td>26 Hrs.</td></tr>
<tr><td>Total student study effort</td><td>135 Hrs.</td></tr>
<tr><td>Reading List and References</td><td colspan="2">*Required Reading:*
▪ Vallor, S. (2024). *The AI mirror: How to reclaim our humanity in an age of machine thinking*. Oxford University Press [pp. 1–225]
Suggested Supplementary Readings:
▪ Boden, M. (2018). *Artificial intelligence: A very short introduction.* Oxford University Press.
▪ Borg, JS, Sinnott-Armstrong, W & Conitzer, V. (2024). *Moral AI: And how we get there*. Pelican.
▪ Borgmann, A. (1984). *Technology and the character of contemporary life*: *A philosophical inquiry*. University of Chicago Press
▪ Bostrom, N. (2014). *Superintelligence*: p*aths, dangers, strategies.* Oxford University Press.
▪ Bostrom, N, (2024). *Deep utopia: Life and meaning in a solved world*. Ideapress Publishing.</td></tr>
</table>

	▪ Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world.* MIT Press. ▪ Chan, B. (2020). The rise of artificial intelligence and the crisis of moral passivity. *AI and Society*, 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. ▪ Crawford, K. (2021). *Atlas of AI.* Yale University 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. ▪ 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. ▪ Ramsey, W. and Frankish, F. (ed.) (2011) *Cambridge handbook of artificial intelligence*, Cambridge University Press. ▪ Schneider, S (2019). *Artificial you: AI and the future of your mind*, Princeton University Press. ▪ Schellman, H. (2024). *The Algorithm: How AI Decides Who Gets Hired, Monitored, Promoted and Fired and Why We Need to Fight Back Now*. Hachette Books. ▪ Sharkey, N. and Sharkey, A. (2012). Granny and the robots: ethical issues in robot care for the elderly. *Ethics of Information Technology* 14: 27-40. ▪ Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy* 24: 62-77. ▪ Susskind, D. (2020). *A world without work.* Allen Lane. ▪ Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Penguin. ▪ Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting.* Oxford University Press. ▪ Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. *Philosophy and Technology*. 24 (3): 251-268. ▪ Woodridge, M. (2020). *The road to conscious machines: The Story of AI.* Penguin. ▪ Zuboff, S. (2019). *The age of surveillance capitalism*. 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