AI Governance and Regulation (SOSC3000J)

Instructor: Gleb Papyshev (gleb@ust.hk)
Time: Monday, Wednesday, Friday, 09:00-12:50
Location: Rm 2302, Lift 17-18
Teaching Assistant: Pat Shu Roy Ho (psrho@ust.hk)
Credits: 3 credits
Enrollment requirements: None

Office: Room 3351
Office Hours: Mon, 14:00-17:00
TA Hours: Tuesday/Thursday 3- 4:30

COURSE DESCRIPTION

How can we steer the development and application of AI technologies for the benefit of society? How do we balance the need for AI innovation with the protection from its potential risks? This course serves as a compass, guiding you through the intricate maze of the governance structures surrounding this technology. By delving into a variety of governance and regulatory instruments proposed by the private sector, national governments, and international organizations, you'll gain a comparative understanding of the global AI landscape.

Every week, we'll unpack the complexities of AI governance through engaging, real-world scenarios. You'll explore AI-related issues such as the ethical implications, geopolitics, corporate governance, and global governance. We'll examine the tradeoff between promoting AI innovation and mitigating its risks, offering you a nuanced, balanced perspective. You'll gain insights into different governance principles and frameworks, equipping you with the skills to navigate the complex world of AI policy.

Throughout the course, we'll emphasize diversity and inclusivity, bringing in perspectives from around the globe. We'll contrast Eastern and Western approaches to AI governance, fostering a rich, multifaceted understanding of the topic.

The course culminates with a final project, providing a platform for you to apply your newfound understanding of AI governance to a real-world problem or scenario. Here, you'll have the chance to critically analyze, debate, and propose solutions, integrating the knowledge and skills you've developed throughout the course.

Whether you come from a technical or social sciences background or are simply interested in the intersection of AI and society, this course is designed to cater to your interests. An open, curious mind is the only prerequisite.

INTENDED LEARNING OUTCOMES

By the end of the course, students will have achieved the following learning outcomes:

1) Acquire a comprehensive understanding of various governance structures and regulatory instruments surrounding AI technologies proposed by different entities such as the private sector, national governments, and international organizations.

2) Develop a nuanced perspective on the balance between promoting AI innovation and mitigating its potential risks, and understand how this trade-off is navigated within different governance frameworks.

3) Analyze the implications of AI from a global perspective, contrasting Eastern and Western approaches to AI governance.

4) Apply their knowledge to real-world AI-related issues, using their understanding of governance principles and frameworks to navigate complex scenarios and propose well-informed solutions.

5) Enhance their communication skills, effectively articulating complex ideas related to AI governance and policy both in written and verbal form.

6) Conduct independent research and utilize teamwork skills to collaboratively investigate a real-world problem or scenario related to AI governance, demonstrating their ability to integrate and apply their knowledge and skills.

ASSESSMENT AND GRADING

Participation	20%	Before 11 July
Reading Response (Individual)	15%	Before 11 July
Role Playing Games	15%	20 June and 30 June
News Presentation (Individual)	5%	Before 11 July
Policy Memo (Individual)	20%	Before 7 July
Final Group Project	25%	Before 18 July

MAPPING OF COURSE ILOS TO ASSESSMENT TASKS

Assessed Task	Mapped ILOs	Explanation
Reading Response	ILO1, ILO5, ILO6	Every week, a curated selection of readings will be provided as part of the course curriculum. Each student is expected to select one piece from the presented options and compose a critical response essay (500 words). In your essay, you should offer a detailed critique of the chosen reading. This includes articulating a nuanced analysis of the author's arguments, identifying and discussing the elements you concur with, as well as those you find disputable or unconvincing.
Role Playing Games	ILO 2-6	Throughout the course, there will be two interactive group sessions dedicated to the role-playing games. Students will assume the roles of different stakeholders in AI governance, collaboratively dissect the complexities of the presented issues and propose viable resolutions to the challenges highlighted. Specific instructions and further details regarding these activities will be conveyed and discussed during class meetings.
News Presentation	ILO 5-6	In light of the rapid advancements in AI technologies, a wealth of AI-related news emerges on a daily basis. Each student is tasked with preparing a succinct presentation (maximum of 5 minutes) focused on a significant AI news event from the previous week. Your presentation should address two fundamental questions: *What happened?* *Why is it significant?* Students have the autonomy to schedule the timing of their presentations at their convenience.

Policy Memo	ILO 1-2; ILO 4-6	Students are required to compose a policy memo concerning the application of generative AI technologies at HKUST. The memo should be around 1,000 words in length. To assist students in adhering to the appropriate structure and style of a policy memo, an instructional tutorial will be conducted during a class session.
Final Group Project	ILO 2-6	The culmination of this course involves a two-part group project centered on AI governance and regulation. *Part One: Case Study Preparation* Groups will develop a comprehensive case study within the specified field. Integral to this case study is the formulation of five open-ended questions that probe deeply into its content, prompting critical thinking and further exploration. *Part Two: Case Analysis* The second segment of the project entails a thorough analysis of case studies prepared by other student groups. This analysis will primarily take the form of responses to the five open-ended questions posed in each case study, demonstrating an understanding and thoughtful examination of the issues at hand.

GRADING RUBRICS

Detailed rubrics for each assignment will be provided. These rubrics clearly outline the criteria used for evaluation. Students can refer to these rubrics to understand how their work will be assessed.

FINAL GRADE DESCRIPTION

A Excellent Performance Demonstrates a comprehensive grasp of subject matter, expertise in problem-solving, and significant creativity in thinking. Exhibits a high capacity for scholarship and collaboration, going beyond core requirements to achieve learning goals.

B Good Performance Shows good knowledge and understanding of the main subject matter, competence in problem-solving, and the ability to analyze and evaluate issues. Displays high motivation to learn and the ability to work effectively with others.

C Satisfactory Performance Possesses adequate knowledge of core subject matter, competence in dealing with familiar problems, and some capacity for analysis and critical thinking. Shows persistence and effort to achieve broadly defined learning goals.

D Marginal Pass Has threshold knowledge of core subject matter, potential to achieve key professional skills, and the ability to make basic judgments. Benefits from the course and has the potential to develop in the discipline.

F Fail Demonstrates insufficient understanding of the subject matter and lacks the necessary problem-solving skills. Shows limited ability to think critically or analytically and exhibits minimal effort towards achieving learning goals. Does not meet the threshold requirements for professional practice or development in the discipline.

COURSE AI POLICY

In this course, students are permitted to engage with generative AI technologies as a resource in the preparation of their assignments. However, the direct submission of output from such AI tools as the final work for any assignment is strictly prohibited.

Students must also provide a thorough account of how they utilized generative AI in the creation of their work. This documentation should include reflective insights and a critical evaluation of the role generative AI played in their assignment preparation process.

ACADEMIC INTEGRITY

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Academic Integrity | HKUST – Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

STRUCTURE OF THE COURSE

Class 1 (16 June) - Defining AI, Governance, Regulation, and Ethics

Readings:

a) Ansell, C., & Torfing, J. (2022). Introduction to the Handbook on Theories of Governance. In Handbook on Theories of Governance (pp. 1-16). Edward Elgar Publishing.
b) Yeung, K. (2018). Algorithmic regulation: A critical interrogation. Regulation & governance, 12(4), 505-523.
c) Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature machine intelligence, 1(9), 389-399.

Class 2 (18 June) – Why We Need to Govern AI: Opportunities and Challenges

Readings:

a) Dafoe, A. (2018). AI governance: a research agenda (pp. 1-14). Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK, 1442, 1443.
b) Tan, S., Taeihagh, A., & Baxter, K. (2022). The risks of machine learning systems. arXiv preprint arXiv:2204.09852.
c) Automated Healthcare App (Canvas)

Class 3 (20 June) – Mapping Global AI Governance & Role Playing Game 1

Readings:

a) The Bumpy Road Toward Global AI Governance by Miranda Gabbot: https://www.noemamag.com/the-bumpy-road-toward-global-ai-governance/
b) Doomsday to Utopia: Meet AI's Rival Factions by Nitasha Tiku: https://www.washingtonpost.com/technology/2023/04/09/ai-safety-openai/
c) The A.I. Wars Have Three Factions, and They All Crave Power: https://www.nytimes.com/2023/09/28/opinion/ai-safety-ethics-effective.html

Recommended reading: How Elite Schools like Stanford Became Fixated on the AI Apocalypse by Nitasha Tiku: https://www.washingtonpost.com/technology/2023/07/05/ai-apocalypse-college-students/

Role Playing Game 1

Class 4 (23 June) – International AI Governance

Readings:

a) Schmitt, L. (2022). Mapping global AI governance: a nascent regime in a fragmented landscape. AI and Ethics, 2(2), 303-314.
b) Roberts, H., Hine, E., Taddeo, M., & Floridi, L. (2024). Global AI governance: barriers and pathways forward. International Affairs, 100(3), 1275-1286.
c) Zaidan, E., & Ibrahim, I. A. (2024). AI governance in a complex and rapidly changing regulatory landscape: A global perspective. Humanities and Social Sciences Communications, 11(1), 1-18.

Class 5 (25 June) – National AI Governance in the EU and the US

Readings:

a) Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, 22(4), 97-112.
b) Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. Regulation & Governance, 18(1), 3-32.
c) Technology Federalism: U.S. States at the Vanguard of AI Governance: https://carnegieendowment.org/research/2025/02/technology-federalism-us-states-at-the-vanguard-of-ai-governance?lang=en
d) Regulating general-purpose AI: Areas of convergence and divergence across the EU and the US: https://www.brookings.edu/articles/regulating-general-purpose-ai-areas-of-convergence-and-divergence-across-the-eu-and-the-us/

Class 6 (27 June) – AI Governance in China and the Global South

Readings:

a) State of AI Safety in China by Concordia AI (pp. 1-21): https://concordia-ai.com/wp-content/uploads/2023/10/State-of-AI-Safety-in-China.pdf
b) Cheng, J., & Zeng, J. (2023). Shaping AI's future? China in global AI governance. Journal of Contemporary China, 32(143), 794-810.
c) Png, M. T. (2022, June). At the tensions of south and north: Critical roles of global south stakeholders in AI governance. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 1434-1445).

Class 6 (30 June) – Policy Memo Writing Workshop & Role-Playing Game 2

Class 8 (7 July) – Corporate AI Governance

Readings:

a) Cihon, P., Schuett, J., & Baum, S. D. (2021). Corporate governance of artificial intelligence in the public interest. Information, 12(7), 275.

b) Schuett, J., Dreksler, N., Anderljung, M., McCaffary, D., Heim, L., Bluemke, E., & Garfinkel, B. (2023). Towards best practices in AGI safety and governance: A survey of expert opinion. arXiv preprint arXiv:2305.07153.

Role-Playing Game 2

Class 9 (9 July) – AI Alignment and Interpretability Issues

a) Ji, J., Qiu, T., Chen, B., Zhang, B., Lou, H., Wang, K., ... & Gao, W. (2023). AI alignment: A comprehensive survey. arXiv preprint arXiv:2310.19852. (pp. 4-15)
b) Erasmus, A., Brunet, T. D., & Fisher, E. (2021). What is interpretability?. Philosophy & Technology, 34(4), 833-862.

Class 10 (11 July) – Final Presentations