AI Governance and Regulation (SOSC3000J)

Instructor: Gleb Papyshev (gleb@ust.hk)　　　　　　　Office: Room 3351
Time: Wednesday, Friday, 15:00-16:20　　　　　　　　Office Hours: Wed, Fri, 10:00-12:00
Location: Rm 4502, Lift 25-26　　　　　　　　　　　　TA Hours: Tue, Thu, 15:00-16:00
Teaching Assistant: Pat Shu Roy Ho (psrho@ust.hk)
Credits: 3 credits
Enrollment requirements: None

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.

GENERATIVE AI USAGE 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.

EVALUATION

Participation	10%
Reading Response (Individual)	10%
Group Case Analysis	20%
News Presentation (Individual)	5%
Policy Memo (Individual)	15%
Final Group Project	40%

Reading Response (Individual)

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.

Group Case Analysis

Throughout the course, there will be two interactive group sessions dedicated to the examination of case studies. Students will collaboratively dissect the complexities presented in these cases 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 (Individual)

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 (Individual)

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

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.

STRUCTURE OF THE COURSE

Introduction and Background

Week 1 (Jan. 31 and Feb. 2) – Defining AI, Governance, and Regulation

Required reading: 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.

Recommended reading: Yeung, K. (2018). Algorithmic regulation: A critical interrogation. Regulation & governance, 12(4), 505-523.

Week 2 (Feb. 7 and Feb. 9) – Why We Need to Govern AI: Opportunities and Challenges

Required reading: 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.

Case Analysis Tutorial on Feb 9

The case study will be uploaded to Canvas.

Week 3 (Feb. 14 and Feb. 16) – Mapping Global AI Governance

Required readings: 1) The Bumpy Road Toward Global AI Governance by Miranda Gabbot: https://www.noemamag.com/the-bumpy-road-toward-global-ai-governance/

2) Doomsday to Utopia: Meet AI's Rival Factions by Nitasha Tiku: https://www.noemamag.com/the-bumpy-road-toward-global-ai-governance/

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/

First In-class Group Work on Feb. 16

International AI Governance

Week 4 (Feb. 21 and Feb. 23) – International AI Governance

Required reading: Schmitt, L. (2022). Mapping global AI governance: a nascent regime in a fragmented landscape. AI and Ethics, 2(2), 303-314.

Policy Memo Writing Tutorial on Feb. 23

National AI Governance

Week 5 (Feb. 28 and Mar. 1) – AI Governance in the EU

Required reading: 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.

Recommended reading: Policy documents uploaded to Canvas.

Week 6 (Mar. 6 and Mar. 8) – AI Governance in the US

Required reading: Zhang, B., & Dafoe, A. (2020, February). US public opinion on the governance of artificial intelligence. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (pp. 187-193).

Recommended reading: Policy documents uploaded to Canvas.

Week 7 (Mar. 13 and Mar. 15) – AI Governance in China

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

Recommended reading: Policy documents uploaded to Canvas.

Week 8 (Mar. 20 and Mar. 22) – AI Governance in Other Countries

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).

Recommended reading: Policy documents uploaded to Canvas.

Second In-class Group Work on Mar. 22

Corporate AI Governance

Week 9 (Mar. 27 and Apr. 10) – Big Tech and AI Governance

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

Recommended reading: Policy documents uploaded to Canvas.

Week 10 (Apr. 12 and Apr. 17) – AGI Governance

Required reading: 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.

Recommended reading: Policy documents uploaded to Canvas.

Week 11 (Apr. 19 and Apr. 24) – Case Presentation & Case In-class Discussion

Final Presentations and Conclusion

Week 12 (Apr. 26 and May 3) – Conclusion

Week 13 (May 8 and May 10) – Final Presentations