THE ETHICS and GOVERNANCE of ARTIFICIAL INTELLIEGENCE

Semester No 7	Course Code- LA- 411	Credit Hours 3-1

COURSE OBJECTIVES:

1. Identify and remember core principles, challenges, and developments in AI governance, including essential policy milestones and landmark case studies.

2. Understand the complex interplay of AI development and regulatory practices, comprehending how emerging technologies influence and are influenced by public norms and values.

3. Apply interdisciplinary approaches integrating knowledge from the humanities, social sciences, and public policy to evaluate the social impact of emerging AI applications in real world scenarios.

COURSE LEARNING OUTCOMES:

By the end of this course, students will be able to:

1. Analyze emerging AI governance strategies, critically assessing their origins, applications, and effectiveness for responsible AI management and regulation.

2. Exercise leadership and communication skills necessary for advocating ethical decision making and public interest in AI governance, considering both global impact and specific community needs.

3. Create an original project around a specific challenge that demonstrates critical thinking and innovation in promoting the strategic use of AI for the public good

PRESCRIBED TEXT:

1. Coeckelbergh, M. (2020) AI Ethics. The MIT Press. Cambridge, Massachusetts. ISBN - 9780262538190

2. Additional required readings. Posted within each module on the LMS course website. Readings are also listed in the topical outline/course schedule table.

COURSE DESCRIPTION:

This course observes the emergence of AI ethics in law and public policy, examining the norms, values, and political strategies involved in the consensus-building processes that shape the development and governance of AI systems. Students will engage in critical analysis of AI policy documents and delve into core principles such as fairness, accountability, and transparency, exploring their origins and practical applications. The curriculum centers on a series of design thinking workshops that will challenge students to debate over the responsible use of AI in real world case studies, ranging from issues of human rights to sustainable development and geopolitics. By exercising analytical and rhetorical excellence while engaging with the regulatory complexities inherent in the development of advanced technologies, this course is considered a bootcamp for transformative leaders capable of governing AI for the public good.

ASSESSMENT SYSTEM:

Quizzes	15%
Assignments	10%
Presentations	10%
Mid Term/	20%
ESE / Final Term Project	40%

Weekly breakdown of course contents is as follows:

WEEK	TOPICS	ACTIVITIES	READINGS/ ASSIGNMENTS
1	Syllabus Discussion and Development of Course Ground Rules: -Syllabus, course activities, and course content discussion Students and professor will work collaboratively to define expectations for course, conduct, timeliness/late policies, and interactions INTEGRITY	Class Discussion – course ground rules and introduction to course Course Introductions Pre-Course Survey What are the challenges of evaluating AI augmented knowledge; the value of intellectual property in the age of knowledge extractivism; and the relevance of academic traditions in emerging knowledge landscapes	Syllabus Joler, Vladan, and Matteo Pasquinelli. "The Nooscope Manifested: AI as Instrument of Knowledge Extractivism." AI & Society 36, no. 4 (2021): 1263–1280.
2	The Human/Computer Relationship: ✦ What are the fundamental differences between humans and machines? ✦ Is General AI possible?	Class Discussion – What are the ramifications to using AI in healthcare or public health when adopting specific philosophical assumptions regarding AI? (symbolic intelligence, phenomenology, transhumanism, post humanism) Small Group Activity: What types of	Coeckelbergh: Chapters 3 & 4 **Weekly Quiz 1**

		safeguards would be necessary to implement general AI in health settings?	
3	CONTROL The balance between using technology for collective governance and the potential risks to personal liberty; the impact of algorithmic bias on such systems; and the meaning of user-centric data agency.	Class Discussion –	Joque. Justin. Revolutionary Mathematics: Artificial Intelligence, Statistics, and the Logic of Capitalism. New York: Verso Books, 2022. Nissenbaum. Helen Fay. Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford, CA: Stanford Law Books, 2010. Pasquale, Frank. The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge, MA; London, England: Harvard University Press, 2015.

4	Health,Superinte llig- ence and the AI Apocalypse What are the drivers of AI integration within healthcare and the public health?	Class Discussion – What are the building blocks of responsible AI in healthcare and public health? Small Group Activity:	Coeckelbergh: Chapters 1 & 2
		Ethical considerations and Drug Discovery using cloud based prediction technology (polymorph)	
5	Who Doesn't Want to Live Longer? --- Artificial Intelligence Applications for Healthcare and Public Health: How does AI support the delivery of healthcare and the practice public health? What is precision medicine? What is precision public health?	Class Discussion – What types of support does AI provide in healthcare and public health? (surveillance, research, decision making, prediction) What are the potential risks for each? Small Group Activity: What are the ethical considerations when Integrating AI to reduce human error?	Coeckelbergh: Chapter 5 Khoury MJ, Engelgau M, Chambers DA, Mensah GA. Beyond Public Health Genomics: Can Big Data and Predictive Analytics Deliver Precision Public Health? Public Health Genomics. 2018;21(5 6):244-250.

6	Garbage In --- Garbage Out: Data and the fit Data Sources Training Algorithms Morally	Class Discussion – What are the limitations of machine learning and how do these translate to the ethical use of AI in healthcare and public health? Class Debate: The pros and cons of requiring machine learning instances to meet predetermined ethical	Coeckelbergh: Chapter 6 **Weekly Quiz 2**
		standards before integration.	
7	The Usual Suspects and Privacy Protection for the Vulnerable: Who determines which data to use? How can exploitation occur without persons ever knowing?	Class Discussion – At what point can AI become a manipulator of human behavior? Surveillance, Spying, or Safeguarding? Small Group Activity: Defining criteria for allowable Surveillance	Coeckelbergh: Chapter 7

8	. TRANSPARENCY	The use of technology for professional judgment in critical decision-making processes; the potential for deceptive outputs of machine learning algorithms; and the essential role of human oversight in mitigating the limitations of AI.	Christian, Brian, and Tom Griffiths. Algorithms to Live By: The Computer Science of Human Decisions. New York: Henry Holt and Company, 2016. ▪ Eubanks, Virginia. Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor. First Edition. New York, NY: St. Martin's Press, 2018. ▪ Natale, Simone. Deceitful Media: Artificial Intelligence and Social Life After the Turing Test. New York, NY: Oxford University Press, 2021
9	**MID TERM**		

10	Safety Doesn't Happen By Accident: The physical nature of virtual safety The damage that hackers can do through devices Security Failure, Disruptions, and Health Consequences	Class Discussion – Planning for safety and security: What are the steps for each level of AI implementation? Class Debate: The pros and cons of mining and managing medical data with AI	Ellahham S, Ellahham N, Simsekler MCE. Application of Artificial Intelligence in the Health Care Safety Context: Opportunities and Challenges. American Journal of Medical Quality. 2020;35(4):341-348. **Weekly Quiz 3**
11	Agency and the Moral Dilemma: Who is responsible for decisions made by machines? The ethical challenges of programming moral decisions into machines; the principles that should guide such decisions; and the legal and ethical frameworks for accountability when those decisions lead to harm.	Class Discussion – Can machines be taught to be moral agents? What are the ramifications of this? Small Group Activity: Designing an oversight committee to address ethical integration of AI in health contexts. What roles and expertise are needed?	Coeckelbergh: Chapter 8 Elish, Madeleine Clare. "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction." Engaging Science, Technology, and Society 5 (2019): 40-60. Mosqueira-Rey, E., E. Hernández-Pereira, D. Alonso-Ríos, et al. "Human-in-the-Loop Machine Learning: A State of the Art." Artificial Intelligence Review 56 (2023): 3005–3054. O'Neil, Cathy. Weapons of Math Destruction: How

			Big Data Increases Inequality and Threatens Democracy. New York, NY: Crown, 2016. Scharre, Paul. Army of None: Autonomous Weapons and the Future of War. New York, NY: W.W. Norton, 2019.
12	✚ Don't Go Behind the Curtain – Explainable AI in Public Health and Healthcare: Decision Trees, Transparency, and Explainability The Problem of the Black Box	Class Discussion – How can machine learning or deep learning using neural networks be made transparent? What safeguards can help? Which is more important: performance or explainability? Small Group Activity: What are the ethical considerations that should be addressed when using AI to predict outbreaks?	U. Pawar, D. O'Shea, S. Rea and R. O'Reilly, "Explainable AI in Healthcare," 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), Dublin, Ireland, 2020, pp. 1-2.
13	✚ Looking in the Mirror --- Bias in the Machine: The sources of bias: training data, data sampling, defined variables, human error What role should AI fulfil in promoting a just, fair, and healthy society?	Class Discussion – Should algorithms be designed to mirror the real world or should they look to advantage the already disadvantaged? Small Group Activity: Setting the priorities of an AI project.	Coeckelbergh: Chapter 9 **Weekly Quiz 4**

14	Tyranny over the Ingenuous --- Policy, Process, and Responsible AI: Who are the affected when it comes to AI and health? Who is responsible for the affected? What role does policy play in equitable AI?	Class Discussion - How does policy impact the downstream effects of integrating AI into public health decisions or clinical workflows? Small Group Activity: Compose a policy that should be used by a healthcare provider or community health unit that would guide the ethical implementation of AI.	Coeckelbergh: Chapter 10
15	The Utopian Influence over the Dystopia: The pace and the progress of AI are begging for responsible oversight. What are the larger issues and possibilities that are yet to be discovered?	Class Discussion – What would an artificial intelligence bill of rights look like? Who should have their finger on the “off” button? Class Debate: The Pros and Cons of general AI for healthcare and public Health	Coeckelbergh: Chapter 11 & 12 Weekly Quiz 5

16	INTEROPERABILITY The global context of AI governance; the feasibility of harmonizing policies between nations with fundamentally different norms and values;	What is the potential for creating standards that respect both technological progress and cultural diplomacy?	Birhane, Abeba. "Algorithmic Colonization of Africa." SCRIPTed 17, no. 2 (2020): 389. Bratton, Benjamin H. The Stack: On Software and Sovereignty. Cambridge, MA: MIT Press, 2015. Gal, Danit. "Perspectives and Approaches in AI Ethics: East Asia." In Markus D. Dubber, Frank Pasquale, and Sunit Das (eds). The Oxford Handbook of Ethics of AI. Oxford Academic, 2020. Lee, Kai-Fu. AI Superpowers: China, Silicon Valley, and the New World Order. Boston: Houghton Mifflin Harcourt, 2018.
17	AI Application Pitch Presentations: The class session will be used for presentations	In-Class Presentation	No Readings or Quiz this Week
18	**END SEMESTER EXAMINATION**		