Course Title	Philosophy and Ethics of Artificial Intelligence				
Course Code	AI655				
Course Type	Elective				
Level	Master (2nd Cycle)				
Year / Semester	1st Year/2nd Semester or 2nd Year/1st Semester				
Teacher's Name	TBA				
ECTS	7	Lectures / week	Up to 6 Teleconferences	Laboratories / week	None
Course Purpose and Objectives	This course examines the philosophical and ethical implications of the development of Artificial Intelligence systems and machines and their interaction with human life. It begins with introducing students to the broad issues of Philosophy of Mind and examining the possibility of human-like machine intelligence. Its second part deals with the ethical and social impact of the proliferation of autonomous systems and AI programs, examining issues such singularity (machine high-intelligence, autonomous weapons, big data processing, robots and human work and regulation of AI based on ethical principles				
Learning Outcomes	Upon succesful completion of this course students should be able to: • Discuss the broad issues of Philosophy of Mind • Recall the main topics of the philosophy of Artificial Intelligence and differences between machines and human minds • Summarize relevant examples of AI systems and became aware of the potential and also the limits of computation. • Illustrate dangers embedded in the development of AI implementations in social and economic life • Examine the main moral issues raised by AI systems such as the control of autonomous systems in warfare. • Reflect on the main issues in regulating AI systems and products along ethical lines • Develop tools for independent critical thinking, through philosophical and moral argumentation, reflection, discussions, and essay writing				
Prerequisites	None	Co-requisites	None		
Course Content	PART A Philosophical Issues including: <u>Introduction to the main themes of Philosophy of Mind:</u> What is a mind, Substance Dualism, Descartes' arguments in favour of Dualism, mind body interaction, the problem of other minds. <u>The main physicalist theories of Philosophy of Mind:</u> Behaviourism, Identity Theory, Eliminative Materialism, Functionalism				

	Can machines think? The Turing test. The Chinese Room (Searle attack on the possibility of AI). Differences between computing and human thought. Issues of Singularity (High AI). Mind uploading into a machine. Simulations: Is it possible to simulate the human mind? Do we live in a simulation? PART B: Ethical Issues including: Dangers from autonomous systems. Rational systems drive towards self-preservation, replication, and resource acquisition Dangers for human beings and ways to restrain them. Military and economic pressures are driving the rapid development of autonomous systems. Is it possible design these systems to approximating rational economic agents. The danger of rational systems leading to anti-social and dangerous behavior and the need to counter such tendencies. Military applications of AI "Drones," "unmanned systems," and "robots" capabilities are not merely coming from science-fiction. New technologies raise new human dilemmas concerning the use of autonomous weapons and robots. Robots at work. Effects on employment The effect of AI and extensive automation on the work environment, the nature of human work, the distribution of wealth and the extensive possibility that humans will have to work much less. Does AI destroy jobs or does it create more jobs than the ones it destroys? Will robots and machines have rights and responsibilities? How can we ensure that these systems respect our ethical principles when they make decisions? What, if any, legal rights and responsibilities should we grant them? And should we regard them merely as sophisticated tools or as a newly emerging form of life? Policy recommendations, regulation, codes of conduct, rules of behavior and decision making Policy recommendations and rules of decision making and behavior concerning choices in designing robots so that we avoid unnecessary policy friction over promising new technologies and help maximize human benefit. Guidelines that allow social and collaborative robotics to flourish.
Teaching Methodology	E-Learning
Bibliography	M. Carter, Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence, Edinburgh University Press, Latest Edition J. Kaplan, Artificial Intelligence, Oxford University Press, Latest Edition

	K. Warwick, Artificial Intelligence: The Basics, Routledge, Latest Edition A. Wallach, Moral Machines, Oxford University Press, Latest Edition N. Bostrom and E. Yudkowsky, “The ethics of artificial intelligence”. In W. M. Ramsey and K. Frankish, editors, The Cambridge Handbook of Artificial Intelligence, pages 316-334, Cambridge University Press, Cambridge, Latest Edition J. R. Searle, Review of Bostrom’s Superintelligence: Paths, Dangers, Strategies and Floridi’s The Fourth Revolution, NYRB, Latest Edition
Assessment	Final Examination 50% Assignments/On-going evaluation 50% 100%
Language	English