Course Title	: Artificial Intelligence and Society
Course Code	: CLD9020
Recommended Study Year	: Any
No. of Credits/Term	: 3
Mode of Tuition	: Sectional approach
Class Contact Hours	: 3 hours per week
Category in Major Programme	: CLD – Cluster Science, Technology, and Society
Prerequisite(s)	: N/A
Co-requisite(s)	: N/A
Exclusion(s)	: N/A
Exemption Requirement(s)	: N/A

Brief Course Description

This course provides an accessible introduction to the main methods of applied artificial intelligence as they are used in commonly available household products, care robots, self-driving cars, and automated weapons. After establishing some theoretical basis (theories of intelligence, methods of AI), it goes on to describe in detail the architecture of a number of easy, introductory AI systems (expert systems, computer game players, chat bots, and robotic pets), while also presenting these systems in the classroom and giving the students the opportunity to interact with them and (give sufficient interest and ability of the students) to make some easy modifications and adaptations. This hands-on approach is a unique feature of this course, and will give the students some actual experience in handling, analysing, and understanding the workings of basic AI systems. No previous programming knowledge is needed, but students should be open to learning basic concepts from computer science. All actual interaction with computers will be in form of guided exercises that do not require special skills. Finally, the course will address the social and ethical implications of autonomous machines, using examples such as household assistants, care robots, self-driving cars, and autonomous weapons.

No prior programming experience is needed, but participants should have a basic familiarity with computers and an interest in information technology.

Aims

This course has been explicitly designed to not only provide theoretical insight into ethical and social problems, but to base these insights on the concrete and practical experience of the students, by presenting to them and analysing in class a number of small but working AI systems that demonstrate some of the main techniques of Artificial Intelligence: expert systems, search heuristics, natural language processing, and household robotics. Based on the insights gained from these systems, students will be able to get a deeper, more informed and more solid understanding of the problems that arise from such technologies, and how these problems can be addressed. The second part of the course will therefore discuss some of the main social and ethical issues posed by AI technologies today, using standard literature from the areas of technoregulation, philosophy of technology, including technoethics, affective computing, and the philosophy of AI.

Learning Outcomes

On completion of this course, students who successfully engage with the course material will be able to:

1. gain a basic understanding of the concepts used in AI programming and the basic architecture of AI systems, and experience a number of demo systems in action;
2. explain and evaluate critically the basic principles upon which AI systems operate, insofar

as these principles are relevant to societal or ethical issues;
3. apply the knowledge gained to the analysis of AI-based systems encountered in everyday life, and give a basic assessment of the social and ethical issues that might arise from the use of these systems;
4. understand and assess critically and in an informed way the role of AI in society and the possible dangers arising from it.

Indicative Content

I. Basic theory background: What is intelligence?
- Definitions
- Turing Test and the Chinese Room argument
- What is consciousness?
- Introduction to biological and artificial neural networks

II. Methods of Artificial Intelligence (overview with demonstrations):
- Symbolic AI and GOFAI, expert systems
- Artificial neural networks
- Reinforcement learning
- Deep learning

III. Practical demonstrations with student exercises, such as:
- Expert systems, e.g. a Starbucks coffee recommendation assistant
- Computer game players, alpha/beta search, e.g. playing tic-tac-toe
- Chat bots and natural language parsing principles, e.g. a customised exam revision assistant
- Anatomy and basic software architecture of a robotic pet (with in-class demonstrations)

IV. Applications of AI. Discussion of use cases of AI, including:
- Household assistants
- Care robots
- Self-driving cars
- War robots

V. Social and Ethical issues (a selection among topics like:)
- Responsibility ascription to machines; personhood and rights of machines
- Affective computing and the emotional bonding with machines (Turkle, Picard)
- The bureaucracy of mind (Winograd, Dreyfus), the regulatory issues facing AI (Lessig), and the relation of AI technologies to society (Latour, Sclove, Joy)
- Dangers of robot misuse (Denning et al., Yampolskiy)
- Moral machines and machine ethics (Arkin, Anderson&Anderson)
- The debate around autonomous weapons (Arkin, Singer)
- The Singularity and its critics (Kurzweil, Chalmers, Bostrom)

Teaching Method

Sectional approach.

Measurement of Learning Outcomes

1. Students are required to regularly attend classes and to participate in the practical exercises (#1). Their understanding is documented in the student presentation and the midterm and final examinations (#1).
2. In-class discussions and the presentation or term paper give the students an opportunity to personally and critically engage with the theories, and to test their understanding of the basic

principles upon which AI systems operate (#2).

3. Term paper and student presentations will demonstrate and deepen the students' abilities to analyse AI-based systems encountered in everyday life, and to give a basic assessment of the social and ethical issues that might arise from the use of these systems (#3).
4. Active attendance and participation in class discussions will confront students with the diversity of opinions in class (#4) and also allow them to test their understanding of the theories by applying them to particular cases during discussions (#3, #4).
5. The mid-term and final examinations measure to what extent students have achieved a general understanding of the general principles of AI discussed (#1) and the theories pertaining to the significance of AI for society, as well as the ethical problems arising from the deployment of particular AI systems (#2, #3, #4).

Assessment

- In-class participation: 10%
- Mid-term examination: 10%
- Student presentation: 20%
- Term paper: 30%
- Final examination: 30%

Required Readings (extracts; in order of importance)

- George B. and Carmichael G. (2016). *Artificial Intelligence Simplified. Understanding Basic Concepts.* CSTrends LLP.
- Whitby, Blay (2008). *Artificial Intelligence: A Beginner's Guide.* Oneworld Beginners' Guides. OneWorld Publications.
- Finlay J. and Dix A. (1996). *An Introduction To Artificial Intelligence.* CRC Press.
- Floridi, L. (ed.) (2004). *The Blackwell Guide to the Philosophy of Computing and Information.* Oxford: Blackwell Publishing.

Supplementary Readings (selected articles only)

- Arkin, R. (2009). Governing Lethal Behaviour in Autonomous Robots. CRC Press.
- Arkin, R. et al. (2009). *An Ethical Governor for Constraining Lethal Action in Autonomous Systems.* Tech Report No GIT-GVU-09-02, GVU Center, Georgia Institute of Technology.
- Boden, M.A. (ed.) (2005). *The Philosophy of Artificial Intelligence.* (Reprint). Oxford: Oxford University Press.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. OUP Oxford.
- Fellous, J.M. and Arbib M. (eds.) (2005). *Who Needs Emotions? The Brain Meets the Robot.* Oxford: Oxford University Press.
- Kaplan, D. (ed.) (2009). *Readings in the Philosophy of Technology.* 2nd ed. Lanham: Rowman & Littlefield.
- Lessig, L. (2006). *Code Version 2.0* Basic Books.
- Lessig L. (1999). *The Law of the Horse: What Cyberlaw Might Teach.* Harvard Law Review Vol. 113:501.
- Russell S.J. and Norvig, P. (1995). *Artificial Intelligence. A Modern Approach.* Englewood Cliffs, New Jersey : Prentice Hall.
- Sharkey, N. E. (2008). *Grounds for Discrimination: Autonomous Robot Weapons.* RUSI Defence Systems 86, 11(2).
- Singer, P.W. (2009). *Robots at War: The New Battlefield.* The Wilson Quarterly 30–48
- Whitby, B. (2008). *Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents.* Interacting with Computers, 20 (3). pp. 326-333.
- Winograd, T. (1991). *Thinking Machines: Can There Be? Are We?* in James J Sheehan and Morton Sosna (eds), The Boundaries of Humanity: Humans, Animals, Machines (University of California Press).