

Module Details	
Module Title	Responsible AI: Ethics, Law and Governance
Module Code	OIM7510-B
Academic Year	2022/3
Credits	20
School	School of Management
FHEQ Level	FHEQ Level 7

Contact Hours	
Type	Hours
Independent Study	164
Lectures	12
Tutorials	24

Availability	
Occurrence	Location / Period
BDA	University of Bradford / Semester 1

Module Aims
The module aims to explore and judge the nature and scope of AI interventions, the legal, social, moral and ethical implications of these technologies and their impact on societies and government institutions. With the ever increasing adoption of AI systems for decision making, customer service, modelling and many other applications, the impact these systems have on our everyday lives must be examined and critically assessed in terms of their fairness, their compliance with legal frameworks and how they affect the diverse populations around the world. The module will provide a grounding in the fundamentals of cybersecurity and artificial intelligence, as well as in the strategic implementation of AI-enabled security methods.

Outline Syllabus

Topics covered in a typical semester include:

- Introduction to the ethics of AI: What is AI and what are the ethical AI-related challenges currently and in the next decade.
- Fundamental concepts in AI (cyber)security: identity, confidentiality, privacy, anonymity, integrity, etc.
- Ethical frameworks: virtue ethics, consequentialism, utilitarianism, deontology.
- Ethics and the alignment problem.
- Algorithmic bias; the accuracy-fairness trade-off.
- Accountability, moral responsibility, and law.
- Ownership and governance.
- Legal frameworks:
 - EU General Data Protection Regulation (GDPR);
 - The European Commission's policies and strategies (the proposal for one of the first legal frameworks in the world dealing with the regulation of AI; High-Level Expert Group on AI);
 - 2019 OECD Principles on Artificial Intelligence;
 - US (lack of) data policy contrasted with PRC's social credit system and EU/UK GDPR.
- The future of AI.

Learning Outcomes

Outcome Number	Description
1	Critically evaluate between the opportunities and potential harms that AI can create.
2	Hypothesise the impact of digitisation in the context of the economics of information.
3	Synthesise the ethical, legal and governance implications of AI applications in scenarios.

Learning, Teaching and Assessment Strategy

Learning will be directed, supported, and reinforced through a combination of lectures and tutorials, films, documentaries and discussion groups, as well as through directed and self-directed study. The classes will engage students in active learning sessions which draw on the knowledge acquired throughout the sessions as well as based on individual study to further deepen understanding of the topics.

These activities will be based on films/documentaries (Blade Runner, The Great Hack, The Social Dilemma, Coded Bias), case studies (Enron, Cambridge Analytica, AlphaGo), or group problem solving exercises. Formative feedback will be provided on ethical scenarios (The Milgram Experiment, The Stanford Prison Experiment, Five Steps to Tyranny, the Moral Machine simulation for automated vehicles) that aim to expose students to real-world situations illustrating legal, social, moral and ethical dilemmas and allow them to explore how what they have learned applies to such situations (e.g. A/B testing, differential privacy). This is a good opportunity to allow the application of theoretical concepts to be demonstrated, bridging the gap between theory and practice and their implications that inform ethical leadership.

Students will be assessed through an assignment (which contributes with 100% towards the module grade) that builds upon the material on which students would have received formative feedback. This report will be used as a means to evidence deeper understanding of the specific area of study and the implications of regulated and unregulated use of AI in the society.

Mode of Assessment

Type	Method	Description	Weighting
Summative	Coursework - Written	Critically explore AI opportunities and potential harms in a written response (d>	100%

Reading List

To access the reading list for this module, please visit <https://bradford.rl.talis.com/index.html>

Please note:

This module descriptor has been published in advance of the academic year to which it applies. Every effort has been made to ensure that the information is accurate at the time of publication, but minor changes may occur given the interval between publishing and commencement of teaching. Upon commencement of the module, students will receive a handbook with further detail about the module and any changes will be discussed and/or communicated at this point.

© University of Bradford 2022

<https://bradford.ac.uk>