



AI and Data Ethics Course Descriptor

| | | | |
|-------------------------|--------------------|-----------------|-----------------|
| Course Title | AI and Data Ethics | Faculty | Philosophy |
| Course Code | NCHAI749 | Course Leader | Alice Helliwell |
| Credit Points | 15 | Teaching Period | Either |
| FHEQ Level | Level 7 | Date Approved | June 2020 |
| Compulsory/ Optional | Compulsory | | |
| Pre-requisites | None | | |
| Co-requisites | None | | |

COURSE SUMMARY

This course will give students a strong foundation in key ethical issues that are emerging in data science and artificial intelligence (AI). They will gain a critical understanding of how these technologies present ethical questions which require a deep understanding both of the detailed context of application, and of pertinent philosophical questions. This encompasses not simply practical ethics as generally conceived but requires consideration of other areas of philosophy such as metaethics, philosophy of mind, and epistemology. Methodological questions concerning the application of ethical theory to practice will also be discussed. As well as developing an appreciation the background philosophical debates, students will also be introduced to the history of work on the ethics of AI, and the considerable current work in policy and regulation for this area.

Students will develop skills in the application of abstract philosophical and ethical concepts to practical uses, including the analysis of cases and the production of relevant policy for AI and data ethics. They will gain an appreciation of how engagement in ethical debates can contribute to the beneficial development of such technologies.

COURSE AIMS

- To provide a strong foundation in the major ethical theories, concepts, and values relevant to current and emerging ethical issues in data science and AI.
- To develop the critical skills needed to apply theoretical concepts to ethical questions in relevant practical and technical applications.
- To understand how developments in technology may impact upon ethical reasoning and philosophical debates on specific issues, such as moral agency, accountability, privacy, responsibility, and fairness.

- To enable students to form, elaborate, and defend their own views in this field, to appreciate the relevance and value of diverse viewpoints, and contribute constructively to collaborate work aimed at addressing concrete problems.
- To develop an understanding of how philosophical thinking can contribute to the beneficial development of artificial intelligence.

LEARNING OUTCOMES

On successful completion of the course, students will be able to:

KNOWLEDGE AND UNDERSTANDING

- K1d Demonstrate a critical awareness of important ideas and debates in ethics and their relation to wider philosophical questions.
- K2d Demonstrate originality in applying this knowledge to illuminate key contemporary ethical and policy debates in data science and AI, gaining an understanding of the application of theory to practice and the particular roles of philosophy.
- K3d Understand and critically evaluate how developing technologies present ethical questions including the analysis of the complex nature of such questions.

SUBJECT SPECIFIC SKILLS

- S1d Demonstrate a comprehensive understanding of how different aspects of artificial intelligence and techniques of data science can generate ethical questions.
- S2d Show an awareness of the global landscape of ethical debate, regulation and policy in artificial intelligence and data science, and the ability to contribute constructively.
- S3d Show an understanding of the historical and conceptual background of ethical issues in AI and data science.

TRANSFERABLE AND PROFESSIONAL SKILLS

- T1d Demonstrate a comprehensive understanding of the nature of professional ethics including the basics of GDPR.
- T2d Demonstrate originality in applying theoretical knowledge to case analysis and detailed policy.
- T3d Consistently apply an excellent level of technical proficiency in written English, using an advanced application of scholarly terminology, that demonstrates the ability to deal with complex issues both systematically and with sophistication.
- T4d Ability to present an argument clearly and succinctly in both written and oral form.

TEACHING AND LEARNING

Teaching and learning strategies for this course will include:

- 15 hours of full-cohort lectures

- One one-hour one-to-one tutorials per student

Course information and supplementary materials are available on the College’s Virtual Learning Environment (VLE).

Students are required to attend and participate in all the formal and timetabled sessions for this course. Students are also expected to manage their directed learning and independent study in support of the course.

EMPLOYABILITY SKILLS

- Ability to present succinct argument and analysis in both written and oral form
- Ability to work in collaborative groups
- Ability to find practical solutions and acceptable compromises to complex problems
- Awareness of how to track and understand contemporary issues in the ethics of technology
- Basic understanding of the main points of GDPR

ASSESSMENT

FORMATIVE

Students will be formatively assessed during the course by means of set assignments. These do not count towards the end of year results, but will provide students with developmental feedback, both written and oral.

SUMMATIVE

Assessment will be in two forms:

| AE: | Assessment Activity | Weighting (%) | Online submission | Duration | Length |
|-----|---------------------------------|---------------|-------------------|----------|-----------------|
| 1 | Written assignment (case study) | 30% | Yes | N/A | 1000 words |
| 2 | Written assignment (essay) | 70% | Yes | N/A | 2500-3000 words |

FEEDBACK

Students will receive formal feedback in a variety of ways: written (including via email correspondence); oral (within one-to-one tutorials or on an *ad hoc* basis) and indirectly through discussion during group tutorials. Student’s will also attend the formal meeting, Collections, in which they will receive constructive and developmental feedback on their performance.

Feedback is provided on written assignments (including essays, briefings and reports) and through generic internal examiners’ reports, both of which are posted on the College’s VLE.

INDICATIVE READING

Note: Comprehensive and current reading lists for courses are produced annually in the Course Syllabus or other documentation provided to students; the indicative reading list provided below is used as part of the approval/modification process only.

The literature in this field is developing very rapidly and reading for the course will reflect this.

BOOKS

Anderson, M. and Anderson, S.L. eds., 2011. *Machine ethics*. Cambridge University Press.

Boddington, P., 2017. *Towards a code of ethics for artificial intelligence*. Springer International Publishing.

Dubber, M., Pasquale, F., and Dased, S. 2020. *The Oxford Handbook of the Ethics of AI*, Oxford, Oxford University Press.

Wallach, W. and C. Allen. 2008. *Moral Machines: Teaching Robots Right from Wrong*, Oxford: Oxford University Press.

JOURNALS

Chouldechova, A. and Roth, A., 2018. The frontiers of fairness in machine learning. arXiv preprint arXiv:1810.08810.

Just, N., & Latzer, M. (2017). "Governance by algorithms: reality construction by algorithmic selection on the Internet." *Media, Culture & Society*, 39(2), 238-258. doi: 10.1177/0163443716643157

Lawrence, N.D., 2017. Living together: Mind and machine intelligence. arXiv preprint arXiv:1705.07996.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L., 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), p.2053951716679679.

Nissenbaum H. 2004. Privacy as contextual integrity. *Wash Law Rev* 79(1):119–158

Rawahn, I. et al. 2019. Machine behaviour. *Nature*, 568, 477-486.
<https://doi.org/10.1038/s41586-019-1138-y>

Russell, S., Hauert, S., Altman, R. and Veloso, M., 2015. Ethics of artificial intelligence. *Nature*, 521(7553), pp.415-416.

ELECTRONIC RESOURCES

HLEG. 2019. *Ethics guidelines for trustworthy AI*. Brussels: Independent High-Level Expert Group on Artificial Intelligence. <https://ec.europa.eu/futurium/en/ai-alliance-consultation>

IEEE *Ethically Aligned Design: A vision for prioritising human wellbeing for autonomous and intelligent systems*, v 2 https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

Science and technology committee. 2016. *Robotics and Artificial Intelligence. Fifth Report of Session, 17*. Science and technology committee. House of Commons, London.
<https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/896/896.pdf>

Talking Machines podcast <http://www.thetalkingmachines.com>

INDICATIVE TOPICS

Students will study the following topics:

- Ethical frameworks for AI: ethical theories and concepts, policy and regulation landscapes for data and AI ethics
- Human and machine agency, human enhancement
- Autonomy, control, and trust in AI
- 'Moral' machines, accountability and responsibility
- Transparency and explainability in Machine Learning, data, and AI
- Data privacy, data ownership, the digital self
- Bias of algorithms and fairness in the use of AI
- Access to and control over information, governance of social media
- AI, data science, the future of employment, and the value of work

| Title: NCHAI749 AI and Data Ethics Course Descriptor | | | | | |
|---|---------------|----------------|------------|---------------------------|--|
| Approved by: Academic Board | | | | | |
| Location: Academic Handbook/Programme specifications and Handbooks/ Postgraduate Programme Specifications/MA Philosophy & Artificial Intelligence Specification/Course Descriptors | | | | | |
| Version number | Date approved | Date published | Owner | Proposed next review date | Modification (As per AQF4) & category number |
| 2.0 | January 2022 | May 2022 | Brian Ball | April 2025 | Category 3: Changes to Learning Outcomes |
| 1.2 | November 2021 | November 2021 | Brian Ball | April 2025 | Category 1: Change to (Teaching and Learning). Category 1: - Course Guide to Course Syllabus. |
| 1.1 | October 2021 | October 2021 | Brian Ball | April 2025 | Category 1: Typographical error |
| 1.0 | June 2020 | June 2020 | Brian Ball | April 2025 | |