



| | | | |
|--|---|-----------------------------|---|
| Course Code & Title | LISS2119 Introduction to the Ethical & Societal Impacts of AI | | |
| Convenor(s) | Boxi Wu | | |
| Institution | University of Oxford | Department | Oxford Internet Institute |
| Academic Year | 2025-26 | Term | Summer 2026 (June) |
| Number of Sessions | 4 | Length of Session(s) | 2 hours |
| Day, Date | Start and End time | | Room Location |
| 09/06/2026 16/06/2026 23/06/2026 30/06/2026 | 10am - 12pm | | G80, Franklin-Wilkins Building, KCL Franklin-Wilkins Building King's College London |
| Enrolment Link: | Available to book on SkillsForge from 2 March 2026 . Click the course link to log in and register: Questions? Visit our Training FAQ here: Frequently Asked Questions - LISS DTP (liss-dtp.ac.uk) | | |

Course Description

Technological advancements in advanced Artificial Intelligence (AI) systems are increasingly mediating social, political and environmental worlds. From the datafication of our online lives to the hype around large language models (LLMs) and generative AI, the question of what it means for technology to be ethical and socially beneficial is increasingly important.

This course aims to introduce students to the ethical and societal impacts of AI technologies and their underlying data, models and infrastructure, drawing from the interdisciplinary fields of AI Ethics, Fairness Accountability and Transparency (FAcCT), and Science and Technology Studies (STS). Each week students will be introduced to key papers often cited in sociotechnical studies of AI alongside real-world examples of digital technology’s ethical and social impacts. Through weekly lectures and interactive discussions, we will explore critical approaches to evaluating technology through the lens of algorithmic bias and fairness, the materiality of digital technologies and issues of corporate and state power. In doing so, students will interrogate the field of AI ethics and emerging paradigms of ‘Responsible AI’ embedded within technological development. Throughout the course, students will be encouraged to reflect on how their respective disciplinary backgrounds inform how they think about contemporary issues related to AI, as well as how digital technologies shape their research.



Meet your course convenor(s)

Boxi A. Wu is an AI researcher focused on the ethical and societal impacts of AI. They are currently a Research Fellow at the AI Now Institute where they focus on public interest AI policy research. Boxi has over 10 years of experience working in technology, policy and academia which they draw on in their research and teaching practices. Most recently they worked in AI ethics at Google DeepMind where they specialised in the ethics of large language models (LLMs).

Boxi has published extensively on AI ethics and the inequalities of digital infrastructure in academic journals including *Big Data & Society*, *Philosophy & Technology*, *ACM Fairness, Accountability & Transparency* and *AAAI/ACM AI, Ethics & Society* and *Review of International Political Economy*. They have contributed to AI Policy research with the OECD's AI Policy Observatory (OECD.AI), Centre for the Governance of AI, the Oxford Martin School and the Think20 engagement group. They are also reading a DPhil in Information, Communication and Social Science at the Oxford Internet Institute (OII) where they are a member of the Oxford-Aalto University Digital Economic Security Lab (DIESL).

Learning Outcomes:

This course will introduce key texts within AI ethics, Fairness Accountability and Transparency (FAcCT), and Science and Technology Studies (STS) that analyse the ethical and societal impacts of AI within a real-world context. Learning outcomes include:


- 1) Familiarity with the academic and real-world landscape of AI ethics and Responsible AI and the key issues discussed within this field.
- 2) Engagement with critical debates about algorithmic bias and fairness, the environmental impacts of AI, and issues of corporate and state power in technological development.
- 3) Engagement with the core sociotechnical concepts relevant to a critical analysis of big data, AI and machine learning.

Course Outline and Reading:

Week 1: AI as a sociotechnical infrastructure

The first session will orient the course with a brief history and introduction to AI as both an ideology and technology, tracing key actors in the development of contemporary big data, AI and machine learning systems. Through explorations of AI as a sociotechnical infrastructure students will discuss the human and ecological inputs to algorithmic technologies as well as the data, models and infrastructure required to train, develop and deploy AI systems. Reflecting on Emily Bender and colleague's seminal *Stochastic Parrots* paper we will critically appraise the current paradigm of large-scale machine learning, questioning the values that are embedded within AI systems through organisational and design decisions.

Readings:

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? . In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610-623. <https://doi.org/10.1145/3442188.3445922>

Odura, S., & Kneese, T. (2024, May 15). AI Governance Needs Sociotechnical Expertise: Why the Humanities and Social Sciences Are Critical to Government Efforts, *Data & Society*. <https://datasociety.net/library/ai-governance-needs-sociotechnical-expertise/>



Week 2: What does it mean for AI to be ethical?

In week 2, we will explore the burgeoning field of AI ethics, reflecting on academic literature, journalistic reporting on algorithmic harms and industry frameworks for “Responsible AI”. We will begin by mapping the key issues and actors within the field, focusing on real-world issues of algorithmic bias and fairness. Touching on issues of policing, border control and surveillance, we will discuss how algorithmic tools and practices can magnify existing social biases. Students will critically appraise who is disproportionately impacted by algorithmic harms and what this means for the solutions posed by academic, policy and industry circles. We will then consider how AI ethics is implemented in practice, focusing on the frameworks used by Big Tech and government actors to identify the benefits and harms of a technology. In doing so, we will critically analyse the potential limitations of these frameworks in the real-world design, development and deployment of AI technologies, and consider what it means for technology to be socially beneficial.

Readings:

Phan, T., Goldenfein, J., Mann, M., & Kuch, D. (2022). Economies of virtue: The circulation of ‘ethics’ in big tech. *Science as culture*, 31(1), 121-135.

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for a New Jim Code*. Polity. (Read Introduction and Chapter 1)

Dave, P., & Haskins, C. (2025, February 4). Google Lifts a Ban on Using Its AI for Weapons and Surveillance. *Wired*. <https://www.wired.com/story/google-responsible-ai-principles/>

Week 3: AI for sustainability and the sustainability of AI

Week 3 will introduce key arguments related to the environmental impacts of AI. Reading van Wynsberghe’s seminal paper on sustainable AI we will discuss both the application of AI to issues of sustainability and the sustainability of current AI systems within the current climate crisis. Drawing from STS traditions of infrastructure studies we will consider the infrastructural qualities of AI. We will map the material dimensions of the AI supply chain from the rare minerals and natural resources required to build and run AI’s physical infrastructure (AI chips and data centres) through to the energy intensive nature of large-scale algorithms. Drawing from the prior weeks discussions on AI histories and ethics we will consider the visibility of AI’s materiality and how this shapes normative understandings of the environmental impacts of AI.

Readings:

Van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1(3), 213-218.

Valdivia, A. (2022). Silicon valley and the environmental costs of AI. *Political Economy Research Centre*. https://www.perc.org.uk/project_posts/silicon-valley-and-the-environmental-costs-of-ai/

Pitt, M., Lucivero, F., & Samuel, G. (2024, December 5). More data on AI’s climate impact won’t save us from inaction. *Ada Lovelace*. <https://www.adalovelaceinstitute.org/blog/data-ai-climate-impact/>

Week 4: AI narratives and corporate power

The final week will consider how corporate power has shaped the development of contemporary big data, AI and machine learning technologies; reflecting on notions of ‘AI safety’, ‘AI governance’ and ‘artificial general intelligence’ (AGI). Returning to a historical analysis of AI as an ideology and technology, we will discuss how narratives about technology have informed corporate and state investment into AI and its underlying infrastructure. Students will explore the landscape of AI safety and governance, tracing narratives around the governance of data, models and infrastructure. In doing so, we will contrast different governance



approaches embedded in Big Tech self-regulation, emerging regulatory frameworks and public interest oriented advocacy from civil society. Students will critically appraise the dynamics of market concentration, regulatory capture and AI narratives; such as 'frontier AI' and 'AGI'. Through an analysis of the corporate narratives that drive AI development and investment, we will question the objectives and incentives of corporations.

Readings:

Whittaker, M. (2021). The Steep Cost of Capture, *Interactions* 28, 6 50–55. <https://doi.org/10.1145/3488666>

Merchant, B. (2024). AI Generated Business: The Rise of AGI and the Rush to Find a Working Revenue Model, *AI Now*.

Pre course preparation:

The weekly readings will be essential for students' participation in the weekly sessions as lectures and workshops will build on these materials. Ahead of Week 1, students are encouraged to a) Consider their existing understanding of AI and its ethical and societal impacts (to discuss during Week 1) and b) ensure reading materials have been read ahead of class.

Eligibility:

All PhD students are welcome to join this course which is designed to introduce humanities and social sciences scholars to core concepts in the critical study of AI and digital technologies. Students should have a willingness to engage with sociotechnical approaches to AI, keeping in mind differing levels of comfort with technical knowledge amongst classmates. Note that an existing technical knowledge of big data, AI and machine learning is *not* a prerequisite for this course.

Number of students:

Min: 8

Max: 25