The Ethics and Philosophy of Digital Technologies

PHIL 2653

University of Hong Kong

Spring 2023

Wednesdays, 08:30 a.m. – 10:20 a.m.

Centennial Campus, Central Podium, CPD-2.37

Instructor:	David Villena
Email:	dvillena@hku.hk
Office:	10.13, 10/F, Run Run Shaw Tower, Centennial Campus
Office Hours:	Tuesdays 09:15 a.m. – 10:15 a.m., and by appointment

Course Description

New digital technologies, from driverless cars to twitter to facial recognition software, pose a variety of new philosophical and ethical challenges. What are these challenges, and how can we respond to them to ensure informed and responsible innovation? This course will consider case studies covering three broad philosophical/ethical themes: (1) fairness in AI systems, (2) fake news and epistemic infrastructure, and (3) bias and discrimination in the digital age.

Course Goals

- Develop knowledge and understanding of philosophical and ethical challenges posed by new and emerging digital technologies
- Develop skills for critically engaging with the digital ethics debate.
- Develop skills for reconstructing others' arguments as well as forming and defending one's own.

Learning Outcomes

On completion of the course, students will be able to

- Demonstrate understanding of and identify a variety of new digital technologies, their applications, and their potential positive and negative outcomes.
- Support claims about the benefits and drawbacks of new digital technologies using good reasoning and argumentation.
- Interpret and analyse the theoretical and ethical significance of different applications of new digital technologies.

- Apply critical thinking in assessing contemporary debates over the value and place of new forms of digital technology in society at large.

Measurement of Learning Outcomes

- Students are required to write two short papers (1100 words not including references each), and one term paper (2200 words not including references).
- Students' term papers will be assessed on their ability to write their paper based on an effective introduction, a central thesis, the way that thesis is supported, the quality of research, and their overall writing skills.

Disabilities

Let me know if you have a visible or invisible disability (e.g., mental health condition) that may have some impact on your work in this class so reasonable accommodations and assistance can be provided to you.

Academic Honesty

Plagiarism is a serious academic offense. Coursework submitted for assessment purposes must be the student's own work and properly acknowledge all sources. Students who plagiarize might be referred to the University's disciplinary committee and might be expelled from the University as a result. You can find more information at http://www.hku.hk/plagiarism. (Also, visit this: http://arts.hku.hk/current-students/undergraduate/assessment/plagiarism) If you have any doubts as to whether the work you're handing in constitutes plagiarism, please don't hesitate to consult with me beforehand.

Assessment

All assignments must be submitted to pass the course

- First Short Paper (30%): Due NOON, 12:00 p.m., March 06
- Second Short Paper (30%): Due NOON, 12:00 p.m., April 03
- Term Paper (40%): Due NOON, 12:00 p.m., May 08

Grading scale

≥ 90%	A+	≤ 100%
≥ 85%	A	< 90%
≥ 80%	A-	< 85%
≥ 75%	B+	< 80%
≥ 70%	B	< 75%

≥ 65%	B-	< 70%
≥ 60%	C+	< 65%
≥ 55%	C	< 60%
≥ 50%	C-	< 55%
≥ 45%	D+	< 50%
≥ 40%	D	< 45%
≥ 0%	F	< 40%

Here are the Faculty of Arts grade descriptions:
http://www.arts.hku.hk/grade_expectations.pdf

Tentative schedule

Fake news and epistemic infrastructure

Weeks 1 & 2: Deepfakes and fake news

Suggested readings:

- D. Villena, Deepfakes, deception, and distrust: Epistemic and social concerns
- R. Jaster & D. Lanius, Speaking of fake news: Definitions and dimensions
- J. Pepp, E. Michaelson & R. K. Sterken, What's new about fake news?
- J. Habgood-Coote, Stop talking about fake news!
- J. Pepp, E. Michaelson & R. K. Sterken, Why we should keep talking about fake news

Weeks 3 & 4: Privacy and surveillance

Suggested readings:

- C. Véliz, The surveillance delusion
- C. Véliz, *Privacy is power* (Excerpts)
- D. Moore, Privacy, security, and surveillance
- E. Selinger & B. Leong, The ethics of facial recognition technology

Week 5: Online manipulation

Suggested readings:

- A. Barnhill, How philosophy might contribute to the practical ethics of online manipulation
- M. Ienca & E. Vayena, Digital nudging: Exploring the ethical boundaries
- J. Williams, Ethical dimensions of persuasive technology

- J. Williams, *Stand out of our light: Freedom and resistance in the attention economy* (Excerpts)
- T. Harris, How technology is hijacking your mind—from a magician and Google design ethicist
- D. Susser, B. Roessler & H. F. Nissenbaum, Online manipulation: Hidden influences in a digital world

Week 6: Social media issues

Suggested readings

- V. R. Bhargava & M. Velasquez, Ethics of the attention economy: The problem of social media addiction
- A. Pham, A. Rubel & C. Castro, Social media, emergent manipulation, and political legitimacy
- R. Roache, What's wrong with trolling?
- D. M. Douglas, Doxing: A conceptual analysis
- K. K. Thomason, The moral risks of online shaming
- A. Marmor, Privacy in social media

Week 7: No Class (Reading week)

Fairness in AI Systems

Week 8: AI fairness

Suggested readings:

- R. Binns, Fairness in machine learning: Lessons from political philosophy
- K. Vredenburgh, Fairness
- Gabriel & V. Ghazavi, The challenge of value alignment: From fairer algorithms to AI safety
- J.-M. John-Mathews, D. Cardon & C. Balagué, From reality to world. A critical perspective on AI fairness
- M. Le Bui & S. Umoja Noble, We're missing a moral framework of justice in Artificial Intelligence: On the limits, failings, and ethics of fairness
- N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman & A Galstyan, A survey on bias and fairness in machine learning

Weeks 9 & 10: Algorithmic decision-making: Opacity, accountability, and transparency

Suggested readings:

- H. Kissinger, How the Enlightenment ends: Philosophically, intellectually—in every way—human society is unprepared for the rise of Artificial Intelligence
- B. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, & L. Floridi. The ethics of algorithms: Mapping the debate
- J. Zerilli, J. M. Alistair Knott, & C. Gavaghan, Transparency in algorithmic and human decision-making: Is there a double standard?
- J. A. Kroll, Accountability in computer systems
- N. Diakopoulos, Transparency
- J. Burrell, How the machine 'thinks': Understanding opacity in machine learning algorithms

Week 11: No Class (Ching Ming Festival)

Bias and discrimination in the digital age

Week 12: Algorithmic bias and injustice

- L. Herzog, Algorithmic bias and access to opportunities
- V. Dignum, Responsibility and artificial intelligence
- J. Himmelreich & D. Lim, AI and structural injustice: Foundations for equity, values, and responsibility
- T. Gebru, Race and gender
- K. Lippert-Rasmussen & L. Aastrup Munch, Price discrimination in the digital age
- C. Barabas, Beyond bias: "Ethical AI" in criminal law

Weeks 13 & 14: Digital health: Discrimination and the ethics of enhancement

Suggested readings:

- A. Mishra, J. Savulescu, & A. Giubilini The ethics of medical AI
- C. Johnston, *Digital health technologies: Law, ethics and the doctor-patient relationship*, Chapter 5
- L. Specker Sullivan, Health and digital technology partnerships: Too close for comfort?

- D. Moseley & C. Murray, Biomedical technology and the ethics of enhancement
- J. Brennan, Genetic enhancement: Just say yes
- Blasimme & E. Vayena, The ethics of AI in biomedical research, patient care, and public health

Additional resources

Papers and chapters in edited books

- Floridi, L., Cath, C., & Taddeo, M. (2019). Digital ethics: Its nature and scope. In: C. Öhman & D. Watson (Eds), *The 2018 yearbook of the Digital Ethics Lab*. Springer. https://doi.org/10.1007/978-3-030-17152-0_2
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P. , Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds & Machines*, *28*, 689-707. https://doi.org/10.1007/s11023-018-9482-5
- Gunkel, D. J. (2020). Perspectives on ethics of AI: Philosophy. In: M. D. Dubber, F. Pasquale & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 539-553). Oxford University Press.
- Hanna, R., & Kazim, E. (2021). Philosophical foundations for digital ethics and AI ethics: A dignitarian approach. *AI Ethics*, *1*, 405-423. https://doi.org/10.1007/s43681-021-00040-9
- Powers, T. M., & Ganascia, J.-G. (2020). The ethics of the ethics of AI. In: M. D. Dubber, F. Pasquale & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 26-51). Oxford University Press.

Books, reports, and edited books

- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Bullock, J. B., Chen, Y.-C., Himmelreich, J., Hudson, V. M., Korinek, A., Young, M. M., & Zhang, B. (Eds.). (2022). *The Oxford handbook of AI governance*. Oxford University Press.
- DiMatteo, L. A., Poncibò, C., & Cannarsa, M. (Eds.). (2022). *The Cambridge handbook of Artificial Intelligence: Global perspectives on law and ethics*. Cambridge University Press.

- Dubber, M. D., Pasquale, F., & Das, S. (Eds.). (2020). *The Oxford handbook of ethics of AI*. Oxford University Press.
- European Group on Ethics in Science and New Technologies. (2018). *Statement on Artificial Intelligence, robotics and 'autonomous' systems*. European Commission. https://data.europa.eu/doi/10.2777/531856
- Frankish, K., & Ramsey, W. M. (Eds.). (2014) *The Cambridge handbook of Artificial Intelligence*. Cambridge University Press.
- High-Level Expert Group on Artificial Intelligence (2019) *Ethics guidelines for trustworthy AI.* European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419
- Jongepier, F., & Klenk, M. (Eds.). (2022) *The philosophy of online manipulation*. Routledge.
- Kissinger, H., Schmidt, E., & Huttenlocher, D. (2021). *The age of AI: And our human future*. Little, Brown and Company.
- Vallor, S. (Ed.). (2022) *The Oxford handbook of philosophy of technology*. Oxford University Press.
- Véliz, C. (Ed.). (2021) *The Oxford handbook of digital ethics*. Oxford University Press.