
UNIVERSITY of GUELPH

PHIL 6760 Science & Ethics Winter 2021

Schedule: Tuesdays 12:30-2:20 on Teams

Professors: Gus Skorburg & Stefan Linquist

E-mails: skorburg@uoguelph.ca & linquist@uoguelph.ca

Office Hours: By appointment

Course Description

This course will explore the broad philosophical implications (ethical, legal, social, political, epistemological, etc.) of recent developments in data science, artificial intelligence, and machine learning, especially as these pertain to the life sciences. We will consider important themes from the philosophy of science, the philosophy of technology, ethics, and social/political philosophy as these relate to recent developments data science, artificial intelligence, and machine learning.

Required Resources

For weeks 11 and 12, students should obtain a copy of *The Promise of Artificial Intelligence: Reckoning and Judgment* (MIT Press) by Brian Cantwell-Smith. All other readings will be posted as .pdfs on CourseLink

Learning Outcomes

By the end of this course, you should be able to:

1. Identify and critically evaluate a wide range of philosophical issues related to data science, artificial intelligence, and machine learning.
2. Produce professional, academic writing and commentary related to philosophical implications of data science, artificial intelligence, and machine learning.
3. Contribute to ongoing academic and popular discussions about the ethics of emerging technology.

Assessments

You need to develop the skills that are most important to launching and sustaining your career: scholarly research and the ability to present it. To that end, you will be assessed on course engagement, two short papers (2,000 words each), and one presentation, as follows:

1. Presentation (20% of final grade)

Over the course of the term, each of you will record a video presentation of around 15 minutes. These will be “asynchronous” (that is, you won’t present during the “live” class time, instead you will record and upload your presentation by **5PM on the Sunday before our “live” meeting on Tuesday**; more instructions about this below). These presentations are low-key and *simply meant to help stimulate discussion on the topics at hand*. You do not need to incorporate any fancy editing, graphics, animations, etc. Simply recording yourself discussing the issues will be totally fine.

The content is up to you, but in general, the presentations should identify and elaborate upon a central theme, argument, concept, etc. from the week’s readings. It may be helpful to think of your presentation as a cross between a teaching demonstration and a conference presentation. Taking the first week’s reading as an example, a good strategy for a presentation could be any of the following:

- Briefly summarizing the two main ways that Winner thinks values, principles, and power-relations are embedded within socio-technical systems and then developing an argument for where deepfake technology ought to fit in this framework, or developing an argument for why deepfakes don’t easily fit within Winner’s framework.
- Reviewing newer literature on deepfakes and developing an argument about how more recent findings might support or undermine Rini’s claims about the epistemic threats of deepfakes, perhaps providing specific examples of how deepfakes are currently being used.
- Summarize relevant technical literature on Generative Adversarial Networks to develop an argument that ethicists have misunderstood/ oversimplified/ overstated/ overlooked/ underappreciated, etc., some important aspect of the technology and its applications.
- And in general, you can always summarize the readings from the “additional literature” and connect them with themes from the required readings. In any case, you will ideally use your presentations as an opportunity to bring your unique research interests and background to bear on the assigned readings.

Logistics

First of all, a few days before your scheduled presentation date, please send a brief e-mail to Stefan and Gus with a few sentences about your plan for the presentation so that we don’t have too much overlapping content. Regarding the recording: There are many ways to record your presentation and we want to stress again that this is a low-key assignment. Basic methods will work just fine. You can simply write your presentation notes and then record yourself using the native “Camera” app on your laptop or smartphone. You can then easily upload powerpoint slides or “handouts” to accompany your video. Alternatively, you can feel free to use any of the available screencasting apps which allow you to record from your webcam while presenting powerpoint slides. UofG recommends Screencast-o-matic (<https://screencast-o-matic.com/>) but you can also use Teams for this (https://answers.microsoft.com/en-us/education_ms/forum/all/how-would-you-record-a-presentation-along-with/1f72e5d3-073e-470a-86ef-50004aeb6a88)

In order to minimize issues with compatibility, we ask that you upload your video file to Microsoft Stream (<https://docs.microsoft.com/en-us/stream/portal-upload-video>) and then submit a word .doc containing the Microsoft Stream link to the Courselink dropbox.

You will need to upload your presentation by **5:00 PM Sunday evening**, so that your classmates will have time to view it before our Tuesday meetings.

2. Research Papers (2 papers, each worth 25% of final grade)

The central assessments for this course are two short research papers of ~2,000 words each. The first research paper should deal with themes from the first unit, and the second research paper should deal with themes from the second unit. Ideally, your research papers could serve as the basis for a conference submission, thesis/dissertation chapter, etc. You should also plan to set up a short meeting with either Stefan or Gus to discuss your topic. More details about the papers will be discussed in class.

3. Engagement (30% of final grade)

This includes showing up on time, speaking up in class, paying attention to what the other students have to say, taking notes, asking questions, etc. Engagement also includes weekly discussion posts (submitted under the “discussions” tab on the CourseLink navigation bar), **due Sunday evening** about the readings for the upcoming week. These responses should be around 200 words. The content of the response is up to you, but all responses should deal directly with the reading assigned for that week. You may want to choose a sentence or paragraph you found especially provocative, difficult, or remarkable, then explain why you found it provocative, difficult, or remarkable. Alternatively, you may want to argue that one of the authors is right or wrong in making some particular claim. Or you may wish to connect two passages in the reading that illuminate each other. You may even just pick a passage and ask questions about it. Your responses will guide what we address in class, and we will sometimes quote from them, so please take them seriously.

Schedule (subject to change)

*always check CourseLink for most up-to-date information on scheduling, readings, assignments, etc.

**Please note that the “additional literature” for each week is meant to serve two purposes. First, it provides materials for the presenters to draw from in the presentations. Second, it provides jumping off points for students who want to write about the week’s topics in their research papers. Otherwise, they are totally optional and no component of your grade directly depends on having read them.

UNIT I: Philosophy of Technology

Week 1: January 12

Technological Neutrality

Assignments: None

Presentations: None

Readings

- Winner, L. (1980). Do artifacts have politics?. *Daedalus*, 121-136.
- Rini, R. (2020). Deepfakes and the epistemic backstop. *Philosopher's Imprint*, 1-16.
- Recommended: *RadioLab* “Breaking Bongo” podcast:
<https://www.wnycstudios.org/podcasts/radiolab/articles/breaking-bongo>

Additional Literature

- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2)
 - Floridi, L., & Taddeo, M. (2016). What is data ethics?. *Philosophical Transactions of the Royal Society of London Series A*, 374: 1-5.
 - Moor, J. H. (2005). Why we need better ethics for emerging technologies. *Ethics and Information Technology*, 7(3), 111-119.
 - Leonelli (2016) Locating ethics in data science: Responsibility and accountability. *Philosophical Transactions of the Royal Society of London Series A*, 374: 1-12.
-

Week 2: January 19

Addiction by design

Assignment: Discussion post 1 by 11:59 PM on Sunday January 17

Presentations: Clair

Readings

- Introduction, Chs. 1-3 from Natasha Dow Schüll's (2012) *Addiction by design*
- Alfano, M., Carter, J. A., & Cheong, M. (2018). Technological seduction and self-radicalization. *Journal of the American Philosophical Association*, 4(3), 298-322.

Additional Literature

- Castro, C., & Pham, A. (2020). Is the attention economy noxious? *Philosopher's Imprint*, 1-13.
 - Selections from Vallor (2016): New social media and the virtue of self-control
 - "How casinos enable gambling addicts" (John Rosengren, December 2016, *The Atlantic*)
 - "Addicted to your iPhone? You're not alone" (Bianca Bosker, November 2016, *The Atlantic*)
 - Screen Time Network's Open Letter to APA (<https://screentimenetwork.org/apa>)
-

Week 3: January 26

Privacy & Human Rights

Assignment: Discussion post 2 by 11:59 PM on Sunday, January 24

Presentations: Alexis

Reading

- Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus*, 140(4), 32-48.
- Amnesty International (2019) *Surveillance Giants: How the business model of Google and Facebook threatens human rights* (51 pages)

Additional Literature

- Rachels, J. (1975). Why privacy is important. *Philosophy & Public Affairs* 4(4): 323-333

- Altman et al. (2018). Practical approaches to big data privacy over time. *International Data Privacy Law* 8(1): 29-51
 - Bogen, M., & Rieke, A. (2018). Help wanted: an examination of hiring algorithms, equity, and bias. *Upturn*. (47 pages).
 - Metcalf, J., & Moss, E. (2019). Owing Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics. *Social Research: An International Quarterly*, 86(2), 449-476.
-

Week 4: February 2

Corporate Research

Assignment: Discussion post 3 by 11:59 PM on Sunday, January 31

Presentations: Vanita

Reading

- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790
- Meyer, M. N. (2015). Two cheers for corporate experimentation: The A/B illusion and the virtues of data-driven innovation. *Colo. Tech. Law Journal*, 13, 274-327

Additional Literature

- Bird, S., Barocas, S., Crawford, K., Diaz, F., & Wallach, H. (2016). Exploring or exploiting? Social and ethical implications of autonomous experimentation in AI. In *Workshop on Fairness, Accountability, and Transparency in Machine Learning* (4 pgs).
 - Grimmelmann, J. (2015). The law and ethics of experiments on social media users. *Colo. Tech. Law Journal*, 13, 219.
 - “OK Cupid plays with love in user experiments” (Molly Wood, 2014, *New York Times*)
 - Boyd, D. (2016). Untangling research and practice: What Facebook’s “emotional contagion” study teaches us. *Research Ethics*, 12(1), 4-13.
 - Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society*, 3(1), 2053951716650211.
-

Week 5: February 9

Responsibility and Accountability

Assignment: Discussion post 4 by 11:59 PM on Sunday, February 7; **PAPER ONE DUE Friday February 12 by 11:59 PM**

Presentations: Ashley

Reading

- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics & Information Technology*, 6(3): 175-183.
- Rubel, A., Castro, C., & Pham, A. (2019). Agency Laundering and Information Technologies. *Ethical Theory and Moral Practice*, 22(4), 1017-1041.

Additional Literature

- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020, January). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33-44).
- Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham L. Rev.*, 87, 1085.
- Gunkel, D. J. (2017). Mind the gap: responsible robotics and the problem of responsibility. *Ethics and Information Technology*, 1-14.
- Danaher, J. (2016). Robots, law and the retribution gap. *Ethics and Information Technology*, 18(4), 299-309.
- Sparrow, R. (2007). Killer robots. *Journal of applied philosophy*, 24(1), 62-77.
- Noorman (2018) - Computing and moral responsibility. *Stanford Encyclopedia of Philosophy*

Week 6: February 16
WINTER BREAK NO CLASS

UNIT II: Philosophy of Science

Week 7: February 23

Trust in Science and the “Black Box” problem in AI.

Assignment: Discussion post 5 by 11:59 PM on Sunday, February 21

Presentations: Goliath

Reading

- Schubbach, A. “Judging machines: philosophical aspects of deep learning.” *Synthese* (2019). <https://doi.org/10.1007/s11229-019-02167-z>
- Hardwig, John (1991), “The role of Trust in Knowledge.” *The Journal of Philosophy*, 88: 693-708 <https://doi.org/10.2307/2027007>

Additional Literature

- Castelvechi, Davide (2016), “Can we open the black box of AI?” *Nature* 538, p.20–23, <https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731>
- Wilholt, Torsten (2013), “Epistemic Trust in Science,” *The British Journal for the Philosophy of Science*, Volume 64: Pages 233–253, <https://doi.org/10.1093/bjps/axs007>

- Zednik, C. (2019). Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy & Technology*, 1-24.
-

Week 8: March 2

The Replication Crisis

Assignment: Discussion post 6 by 11:59 PM on Sunday, February 28

Presentations: Jahangir

Reading

- Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, 14(11).
- Hutson, M. (2018). Artificial intelligence faces reproducibility crisis. *Science* 359 (6377): 725-726.
- Warden, P. (2018) “The machine learning reproducibility crisis” (3 pgs)

Additional Literature

- Machery, E. (2020). What is a Replication?. *Philosophy of Science*.
 - McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafiyan, H., ... & Etemadi, M. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89-94.
 - Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Board, M. A. Q. C., Waldron, L., ... & Hoffman, M. M. (2020). The importance of transparency and reproducibility in artificial intelligence research. *arXiv preprint arXiv:2003.00898*.
 - Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8): 1-6.
 - Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.
-

Week 9: March 9

Big Data and the Epistemology of Science

Assignment: Discussion post 7 by 11:59 PM on Sunday, March 7

Presentations: Jacob

Reading:

- Anderson, Chris (2008), “The end of theory: The Data Deluge Makes the Scientific Method Obsolete.” *Wired Magazine* (06.23.2008). <https://www.wired.com/2008/06/pb-theory/>
- Lionelli, Sabina (2014), “What difference does quantity make? On the epistemology of Big Data in biology.” *Big Data & Society* 1: <https://doi.org/10.1177/2053951714534395>

Additional Literature:

- Kelling, S, Hochachka, WM, Fink, D (2009) Data-intensive science: A new paradigm for biodiversity studies. *BioScience* 59(7): 613–620.
 - Symons, John & Alvarado, Ramon (2016), “Can we trust Big Data? Applying philosophy of science to software.” *Big Data & Society*. <https://doi.org/10.1177/2053951716664747>
-

Week 10: March 16

Implications of Deep Learning for Cognitive Science

Assignment: Discussion post 8 by 11:59 PM on Sunday, March 14

Presentations: John

Reading:

- Buckner, C. Deep learning: A philosophical introduction. *Philosophy Compass*. 2019; 14:e12625. <https://doi.org/10.1111/phc3.12625>

Additional Literature:

- Buckner, Cameron (2018), “Empiricism without magic: transformational abstraction in deep convolutional neural networks.” *Synthese* (2018) 195:5339–5372
 - Buckner, C. (2020). Adversarial Examples and the Deeper Riddle of Induction: The Need for a Theory of Artifacts in Deep Learning. *arXiv preprint arXiv:2003.11917*.
 - Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512.
 - <https://web.media.mit.edu/~minsky/papers/ComputersCantThink.txt>
-

Week 11: March 23

TOPIC

Why neural networks (supposedly) can’t think: traditional philosophical objections to AI.

Assignment: Discussion post 9 by 11:59 PM on Sunday, March 21

Presentations: Rod and Nour

Reading:

- Cantwell-Smith, Brian (2019). “The Promise of Artificial Intelligence: Reckoning and Judgment.” Chapters 1-6 (p. 75 - p.144).

Additional Literature

- Haugeland, John (1990), “The Intentionality All-Stars.” *Philosophical Perspectives* (Vol.4) Action Theory and Philosophy of Mind pp. 383-427 (45 pages). <https://doi.org/10.2307/2214199>

- Jerry Fodor, Brian P. McLaughlin (1990), “Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work.” *Cognition*, 183-204. [https://doi.org/10.1016/0010-0277\(90\)90014-B](https://doi.org/10.1016/0010-0277(90)90014-B)
-

Week 12: March 30

Intentionality and object re-identification

Assignment: Discussion post 10 by 11:59 PM on Sunday March 28

Presentations: Lukas

Reading:

- Cantwell-Smith (2019), *The Promise of Artificial Intelligence Reckoning and Judgment*. Chapters 7– 13 (p. 75- 144).

Additional Literature

- Fields, Chris (2013), “How humans solve the frame problem.” *Journal of Experimental & Theoretical Artificial Intelligence*, 25: 441-456, DOI: [10.1080/0952813X.2012.741624](https://doi.org/10.1080/0952813X.2012.741624)
 - Salis, Pietro (2017) Conceptions of original intentionality (and social ontology), In *Mind, collective agency, norms. Essays on social ontology* (Ed: Pietro Salis, Guido Seddone) Shaker Verlag.
-

Week 13: April 6

Open class: Topic TBA

Assignment: Discussion post 11 by 11:59 PM on Sunday April 4

Presentations:

Reading:

Additional literature:

PAPER TWO: DUE DATE TBD

University Statements

Email Communication

As per university regulations, all students are required to check their e-mail account regularly: e-mail is the official route of communication between the University and its students.

When You Cannot Meet a Course Requirement

When you find yourself unable to meet an in-course requirement because of illness or compassionate reasons please advise the course instructor (or designated person, such as a teaching assistant) in writing, with your name, id#, and e-mail contact. The grounds for Academic Consideration are detailed in the Undergraduate

and Graduate Calendars. Undergraduate Calendar - Academic Consideration and Appeals

<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-ac.shtml>

Graduate Calendar - Grounds for Academic Consideration

<https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/index.shtml>

Associate Diploma Calendar - Academic Consideration, Appeals and Petitions

<https://www.uoguelph.ca/registrar/calendars/diploma/current/index.shtml>

Drop Date

Students will have until the last day of classes to drop courses without academic penalty. The deadline to drop two-semester courses will be the last day of classes in the second semester. This applies to all students (undergraduate, graduate and diploma) except for Doctor of Veterinary Medicine and Associate Diploma in Veterinary Technology (conventional and alternative delivery) students. The regulations and procedures for course registration are available in their respective Academic Calendars.

Undergraduate Calendar - Dropping Courses

<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-drop.shtml>

Graduate Calendar - Registration Changes

<https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/genreg-reg-regchg.shtml>

Associate Diploma Calendar - Dropping Courses

<https://www.uoguelph.ca/registrar/calendars/diploma/current/c08/c08-drop.shtml>

Copies of Out-of-class Assignments

Keep paper and/or other reliable back-up copies of all out-of-class assignments: you may be asked to resubmit work at any time.

Accessibility

The University promotes the full participation of students who experience disabilities in their academic programs. To that end, the provision of academic accommodation is a shared responsibility between the University and the student. When accommodations are needed, the student is required to first register with Student Accessibility Services (SAS). Documentation to substantiate the existence of a disability is required; however, interim accommodations may be possible while that process is underway.

Accommodations are available for both permanent and temporary disabilities. It should be noted that common illnesses such as a cold or the flu do not constitute a disability.

Use of the SAS Exam Centre requires students to book their exams at least 7 days in advance and not later than the 40th Class Day. For Guelph students, information can be found on the SAS website

<https://www.uoguelph.ca/sas> For Ridgetown students, information can be found on the Ridgetown SAS website <https://www.ridgetownc.com/services/accessibilityservices.cfm>

Academic Integrity

The University of Guelph is committed to upholding the highest standards of academic integrity, and it is the responsibility of all members of the University community-faculty, staff, and students-to be aware of what constitutes academic misconduct and to do as much as possible to prevent academic offences from occurring.

University of Guelph students have the responsibility of abiding by the University's policy on academic misconduct regardless of their location of study; faculty, staff, and students have the responsibility of supporting an environment that encourages academic integrity. Students need to remain aware that instructors have access to and the right to use electronic and other means of detection.

Please note: Whether or not a student intended to commit academic misconduct is not relevant for a finding of guilt. Hurried or careless submission of assignments does not excuse students from responsibility for verifying the academic integrity of their work before submitting it. Students who are in any doubt as to whether an action on their part could be construed as an academic offence should consult with a faculty member or faculty advisor.

Undergraduate Calendar - Academic Misconduct

<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-amisconduct.shtml>

Graduate Calendar - Academic Misconduct

<https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/index.shtml>

Recording of Materials

Presentations that are made in relation to course work - including lectures - cannot be recorded or copied without the permission of the presenter, whether the instructor, a student, or guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

Disclaimer

Please note that the ongoing COVID-19 pandemic may necessitate a revision of the format of course offerings and academic schedules. Any such changes will be announced via CourseLink and/or class email. All University-wide decisions will be posted on the COVID-19 website (<https://news.uoguelph.ca/2019-novel-coronavirus-information/>) and circulated by email.

Illness

The University will not normally require verification of illness (doctor's notes) for fall 2020 or winter 2021 semester courses. However, requests for Academic Consideration may still require medical documentation as appropriate

Resources

The Academic Calendars are the source of information about the University of Guelph's procedures, policies, and regulations that apply to undergraduate, graduate, and diploma programs. Academic Calendars <https://www.uoguelph.ca/academics/calendars>

Online Behaviour

Inappropriate online behaviour will not be tolerated. Examples of inappropriate online behaviour include:

- Posting inflammatory messages about your instructor or fellow students
- Using obscene or offensive language online
- Copying or presenting someone else's work as your own
- Adapting information from the Internet without using proper citations or references
- Buying or selling term papers or assignments
- Posting or selling course materials to course notes websites
- Having someone else complete your quiz or completing a quiz for/with another student
- Stating false claims about lost quiz answers or other assignment submissions
- Threatening or harassing a student or instructor online
- Discriminating against fellow students, instructors and/or TAs
- Using the course website to promote profit-driven products or services
- Attempting to compromise the security or functionality of the learning management system
- Sharing your user name and password
- Recording lectures without the permission of the instructor

Privacy

By enrolling in a course, unless explicitly stated and brought forward to their instructor, it is assumed that students agree to the possibility of being recorded during lecture, seminar or other "live" course activities, whether delivery is in-class or online/remote.

If a student prefers not to be distinguishable during a recording, they may:

1. turn off their camera
2. mute their microphone
3. edit their name (e.g., initials only) upon entry to each session
4. use the chat function to pose questions.

Students who express to their instructor that they, or a reference to their name or person, do not wish to be recorded may discuss possible alternatives or accommodations with their instructor.