Institute of Technology Entrepreneurship and Commercialization

TIMG 5204, Responsible Artificial Intelligence - 21429 - A [0.5 credit]

Summer 2025

Time and Place
Jul 02 to Aug 14, Mondays and Wednesdays, Nicol Building 4030
18:05 - 20:55
IN-PERSON with flexible online/on campus (see schedule).
Zoom link: https://carleton-ca.zoom.us/j/96859969788

Instructor
Gloria Sánchez-Cuevas, *MA (Public Ethics; Philosophy), MBA (Entrepreneurship and Strategic Innovation), PhD in Ethics & Public Affairs (Candidate)*

Course Materials
Access to online course sessions, course materials and recorded videos will be provided through the CU Brightspace system: https://carleton.ca/brightspace/. To access Brightspace, use your CU credentials and select the Responsible Artificial Intelligence (SEM) Summer 2025 course. The course will be available on Brightspace starting June 27th.

Office Hours
The instructor could be contacted via email. Email is the preferred mode of communication because it provides a record of the content exchanged. The instructor will be available during class and for online meetings by appointment, depending on availability.

Target Audience
This course is designed for graduate students enrolled in the Technology Innovation Management (TIM) program. Students from other programs may also attend, space permitting. However, all participants must meet the high academic standards set by the TIM program. Non-TIM students are required to have a preliminary meeting with the professor before they can be granted admission to the course.

Calendar Description
Responsible Artificial Intelligence - TIMG 5204A

Ethical aspects of development/adoption of Artificial Intelligence (AI) and digital technologies in business practice. Responsible AI business opportunities in cross-border businesses. Responsible AI governance frameworks. AI inclusiveness, bias, fairness, transparency, explainability, accountability, data re-use, protection, and privacy. Assessment of trustworthy AI systems. Prerequisite(s): TIMG 5002 or TIMG 5008.

Responsible Artificial Intelligence Course Description

This course offers students the opportunity to critically examine the ethical dimensions of the global AI revolution and its growing influence across technology-driven business, public policy, and social life. Designed with future technology managers and entrepreneurs in mind, the course bridges ethical theory and practical application, equipping students with the tools to navigate and evaluate the societal implications of emerging AI technologies.

Through a combination of lectures, case-based discussions, and collaborative projects, students will engage with real-world ethical challenges, such as algorithmic bias, discrimination, fairness, data reuse and privacy, surveillance, transparency, and accountability. Each session will examine these issues from the perspective of multiple stakeholders, including industry, government, and affected communities, with a focus on understanding the trade-offs and responsibilities involved in AI design and deployment.

A central emphasis will be placed on current frameworks and standards for responsible and trustworthy AI used to assess ethical risk and guide governance in practice. Students will learn to apply these frameworks to analyze specific AI use cases in various domains evaluating not only their technical performance but also their ethical and social impact.

Throughout the course, students will also conduct a research-based project, identifying emerging issues in the field of AI ethics and proposing actionable strategies for responsible adoption. This process will strengthen their ability to formulate ethical recommendations grounded in both normative reasoning and real-world constraints.

The course will be delivered in a flexible format with a mix of in-person and live online (Zoom) sessions, complemented by asynchronous materials. It will also feature guest speakers from academia, policy, and industry, offering students a well-rounded view of how ethical considerations are addressed across different sectors. By the end of the course, students will be prepared to play an active role in shaping the ethical development and use of AI in their organizations and fields.

Learning Objectives

By the end of this course, students will be able to:

- Understand major ethical theories and social justice frameworks and apply these to AI scenarios and decision-making in technology contexts.
- Achieve a deep familiarity with ethical issues emerging from real-life AI cases in various domains, identifying biases, harms, and stakeholder impacts in each.
- Comprehend the complexities of ethical decision-making in AI when balancing competing interests of businesses, users, and society, and propose strategies that align innovation with ethical norms.
- Describe and critically discuss the social, political, legal, and ethical issues surrounding data-driven innovation with attention to fairness, transparency, accountability, and privacy.
- Evaluate the ethical implications for tech managers and entrepreneurs when adopting or developing AI and learn to turn ethical challenges into opportunities.
- Demonstrate knowledge of leading AI governance frameworks and regulations and how to implement these in organizational settings.
- Utilize frameworks for assessing AI system trustworthiness and carry out structured evaluations or audits of AI solutions against these criteria in a given context.
- Strengthen research and analytical skills by engaging with the latest research on responsible AI and ethics, identifying emerging issues and open questions, and generating business and policy insights that could shape the responsible use of AI in the future.

Class sessions

Class sessions will include a combination of online/in-person sessions, interactive discussions, student workshops and presentations.
Video recordings of class sessions will be available on the course site within Brightspace.

Paul Menton Centre

Students with disabilities requiring academic accommodations in this course are encouraged to contact a coordinator at the Paul Menton Centre (PMC) for Students with Disabilities to complete the necessary letters of accommodation. After registering with PMC, make an appointment to meet and discuss your needs with your

instructor at least two weeks prior to requiring accommodation for assignments or presentations. This is necessary in order to ensure sufficient time to make the necessary arrangements.

Course assignments
Each TIM student registered in the course will work individually and in groups of up to 6 students to perform tasks in class, participate in informal group meetings, and contribute to delivering assignments. Once formed, each group should establish a clear project management structure that facilitates collaboration between group members and maximizes the value of the deliverables. The group members' tasks should be defined and agreed on a weekly basis. Communication with external clients/stakeholders should be concise, clear and meaningful.

Assignment 1 (individual assignment, 25%): AI Ethics Insights
Due: Monday, July 14th, 2025, via Brightspace

Background: On March 11, 2022, the TIM program organized an international webinar titled "The Challenges of Applying AI Ethics." The Invited Speakers were from the Digital Technologies Research Center of the National Research Council of Canada:

- Dr. Joel Martin, Chief Digital Research Officer
- Dr. Svetlana Kiritchenko, Senior Research Scientist
- Dr. Norm Vinson, Research Officer
- Margaret McKay, Program Manager, AI for Logistics

The link to the recording of the webinar is **https://youtu.be/8iULaxKDhuE**. The speakers also generously shared their slides and some other suggestions:

- Svetlana Kiritchenko – slide deck
- Norm Vinson – slide deck
- Margaret McKay – slide deck

Instructions: Each student will watch the webinar "The Challenges of Applying AI Ethics" and formulate **a minimum of five actionable insights/recommendations** based on the discussion. Statements should follow the format "[**Who**] should/could + [action to take] and reflect lessons from the panel. Some examples of statements are:

- Anyone dealing with AI - researchers, teachers, developers, business experts etc., **should** think about ethics during the AI system design and not during the AI system implementation phases. Thinking about ethics during the implementation phase of AI systems is too late.
- Companies and organizations interested in adopting commercially available AI systems **must** develop the skills and the mindset to negotiate the subtleties of the specific terms of use with the providers of AI systems.
- Companies interested in adopting AI systems **need to** learn current data protection legislation to avoid future legal issues.

Assessment criteria: clarity of writing, the degree to which statements follow the format (i.e., [who] + [should/could/will] + [action to take]), correct interpretation of ethical issues raised, and the actionability and originality of the suggested insights.

Assignment 2 (group assignment, 25%): Analyzing bias
Due: Wednesday, July 30th, 2025, via Brightspace

Instructions:
a. In teams, select one of the following case studies:

1. Racial Bias in a health algorithm: https://www.ftc.gov/system/files/documents/public_events/1548288/privacycon-2020-ziad_obermeyer.pdf
2. AI bias is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions): https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
3. Facebook's Ad's algorithm: https://arxiv.org/abs/1904.02095
4. iTotur Group AI system rejects applicants due to age: https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit#:~:text=According%20to%20the%20EEOC's%20lawsuit,States%20because%20of%20their%20age.
5. Lawyer Used ChatGPT in Court and cited fake case: https://www.nbcnews.com/tech/tech-news/chatgpt-cited-bogus-cases-new-york-federal-court-filing-rcna86843
6. SafeRent Tenant Screening: https://www.theguardian.com/technology/2024/dec/14/saferent-ai-tenant-screening-lawsuit
7. TikTok face filter bubbles accused of reinforcing bias, dividing users: https://www.wired.com/story/tiktok-filter-bubbles/
8. Programs to detect AI discriminate against non-native English speakers, shows study: https://www.theguardian.com/technology/2023/jul/10/programs-to-detect-ai-discriminate-against-non-native-english-speakers-shows-study

b. Familiarize yourself individually with the selected case study and set up group meetings to discuss and answer the following questions to perform an in-depth case analysis:
   - *System context*: What type of data is used? What machine learning technique is deployed?
   - *Types of biases or discrimination observed*: What types of biases exist in the data/system? What are the consequences of these biases on human life and human dignity? Based on what you learned in class, are there any types of biases that should be considered?
   - *Mitigation measures attempted*: How were biases in the system addressed? How would you mitigate these biases?

   Important: You should deliberate on the ethical dimensions of your topic, rather than getting hung up on the technical issues of a topic.

c. Summarize your group discussions and prepare a report of 3–7 pages. You can use the following format to organize your content:
   - *Introduction:* Explain the background and the use case, and elaborate on the main ethical issue (i.e., what is fundamentally wrong or ethically problematic about the use case?).
   - *Data and Bias Analysis*: List and explain all existing biases and any additional ones you discovered through your research, detailing how they impact the outcome of the model or decision-making.
   - *Mitigation Strategies*: How biases were addressed by the developers or how *you* propose they be mitigated, and any remaining challenges (i.e., what mechanism was used to address the biases? If there are none, how do you propose to mitigate these biases?)

Assessment criteria: Depth of analysis, use of evidence (references to studies or news), understanding of bias/fairness concepts, and quality of recommendations. Each group will also give a short, informal presentation or summary of their findings in class for peer feedback (ungraded).

Assignment 3 (group assignment, 30%): **Trustworthy AI assessment**
Due: Friday, August 8th, 2025, via Brightspace

Instructions:

a. Continuing from your case study, each group will conduct a formal trustworthiness audit of one selected AI system (for consistency, you may choose the same system analyzed in Assignment 2, another from the list or another approved by the instructor). Apply a recognized *Responsible AI framework* to assess the system across key dimensions, including transparency, fairness, accountability, robustness, and privacy. For example, groups can use the EU's Ethics Guidelines for Trustworthy AI (and its Assessment List) or an equivalent (IEEE, NIST, etc.), and optionally the Z-Inspection® process (a socio-technical assessment method) introduced in class.
b. Summarize your group discussions and prepare a report of 3–7 pages. You can use the following format to organize your content:
    - Summarize the system's purpose and context
    - Evaluate the system's adherence to *trustworthy AI* criteria (noting strengths and weaknesses for each criterion)
    - Propose actionable recommendations for improving its ethical alignment or compliance.

The following two papers are useful resources to assist you in completing Assignment 3.

- On Assessing Trustworthy AI in Healthcare. Machine Learning as a Supportive Tool to Recognize Cardiac Arrest in Emergency Calls. Front. Hum. Dyn., Human and Artificial Collaboration for Medical Best Practices, 08 July 2021, https://www.frontiersin.org/articles/10.3389/fhumd.2021.673104/full
- Co-Design of a Trustworthy AI System in Healthcare: Deep Learning Based Skin Lesion Classifier. Front. Hum. Dyn. |Human and Artificial Collaboration for Medical Best Practices, July 13, 2021, https://www.frontiersin.org/articles/10.3389/fhumd.2021.688152/full

Final Take-Home Exam (individual, 20%): Will be provided at the last class session.
Due: Saturday, August 23rd, via Brightspace.

The final assessment will be an open-book, take-home examination consisting of one or more integrative essay questions or case scenarios. Students will have several days to prepare a written response (approx. 2000 words total). Questions will require students to synthesize course concepts, apply ethical theories to novel AI dilemmas, evaluate trade-offs, and propose well-argued solutions or policy recommendations. *Example:* analyze the ethical and governance issues in a hypothetical rollout of a new AI-driven public service (drawing on course frameworks like the AI Act risk categories, etc.), and articulate a responsible implementation plan.

Assessment criteria: Demonstrated understanding of course material (breadth and depth), quality of ethical reasoning, use of evidence or references to support arguments, and clarity of writing.

Student Evaluation and Assignment Grading
The final grade will be assigned using the following mark allocation:

Item #	Assignment	Type	Date	%
1	AI Ethics Insights	Individual	Monday, July 14	25
2	Analyzing Bias and Fairness	Group	Wednesday, July 30th	25
3	Trustworthy AI Assessment Project	Group	Friday, August 8th	30
4	Final Take-Home Exam	Individual	Saturday, August 23rd	20
				100

Class schedule

#	Date	Delivery	Topic	Agenda	Readings
1	Wed, July 2	Fully online	Course Introduction	• Course overview • Key ethical issues in AI • Moral philosophy and ethics of technology	• Dignum, Virginia (2019). Ch. 1: Introduction & Ch. 2: What is Artificial Intelligence? In: *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*, pp. 1-34. • Morgan et al. (2019). Introduction. In: Morgan, Editor. (2019). *Responsible AI. A Global Policy Framework*, pp. 16-29. • Dignum (2019). Can AI Systems Be Ethical? Ch. 5 in: Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way, pp. 71-92.
2	Mon, July 7	Fully online	Ethical Theory Applied to AI	• Consequentialism • Deontological Ethics • Virtue Ethics • Social Justice	• Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). Ch. 3 What is Ethics? In *An Introduction to Ethics in Robotics and AI* (1st ed. 2021.). Springer Nature • Gabriel, I. (2022). Toward a Theory of Justice for Artificial Intelligence. *Daedalus*, *151*(2), 218–231. • Buijsman, S., Klenk, M. & Jvan den Hoven, J. (2025). Ch.4 Ethics of AI In *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*. Cambridge University Press Optional: • Ethical Framework for Designing Autonomous Intelligent Systems. J Leikas et al. J. of Open Innovation, 2019, 5, 1: https://www.mdpi.com/2199-8531/5/1/18
3	Wed, July 9	Fully online	Fairness and Bias	• Types of bias in AI systems and potential discriminatory outcomes • Ethical obligations to mitigate bias • Fairness in practice	• Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A. (2019). A Survey on Bias and Fairness in Machine Learning. *arXiv:1908.09635*: https://doi.org/10.48550/arXiv.1908.09635 • Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., Crawford, K. (2018). Datasheets for Datasets. *arXiv:1803.09010*: https://doi.org/10.48550/arXiv.1803.09010 • Navigli, R.and Conia, S. (2023), Biases in Large Language models: Origins, Inventory and discussion, https://dl.acm.org/doi/pdf/10.1145/3597307 • Suresh, H., Guttag, J. V. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *arXiv:1901.10002*: https://doi.org/10.48550/arXiv.1901.10002 • Passi, S., Barocas, S. (2019). Problem Formulation and Fairness. arXiv:1901.02547: https://doi.org/10.48550/arXiv.1901.02547 Optional: • Herzog, Lisa, 'Algorithmic Bias and Access to Opportunities', in Carissa Véliz (ed.), *Oxford Handbook of Digital Ethics*, Oxford Handbooks (2023; online edn, Oxford Academic, 10 Nov. 2021)

4	Mon, July 14	Fully online	Privacy, Data Protection & Surveillance	• Consent and data ownership • Anonymization limits • Mass surveillance vs individual rights • Regulatory approaches to privacy	• **Guest Speaker:** Sahar Rahmani (PhD), Director of AI at Fullscript. (Additional readings TBD) • Montreal Declaration for Responsible AI (2018), Principles on Privacy and Autonomy (short excerpt). • Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). Ch. 8 Privacy Issues of AI In *An Introduction to Ethics in Robotics and AI* (1st ed. 2021.). Springer Nature. • Véliz, Carissa, 'The Surveillance Delusion', in Carissa Véliz (ed.), *Oxford Handbook of Digital Ethics*, Oxford Handbooks (2023; online edn, Oxford Academic, 10 Nov. 2021) Optional: • Zuboff, S. (2019). *The Age of Surveillance Capitalism*, Ch. 1 (excerpt) • Selinger, Evan, and Brenda Leong, 'The Ethics of Facial Recognition Technology', in Carissa Véliz (ed.), *Oxford Handbook of Digital Ethics*, Oxford Handbooks (2023; online edn, Oxford Academic, 10 Nov. 2021
5	Wed, July 16	Flexible (in-person/online)	AI Governance: Regulations, Standards and Policies	• Global and regional regulatory landscape • Canadian regulatory frameworks	• The EU Artificial Intelligence Act: https://artificialintelligenceact.eu • IEEE "Ethically Aligned Design" https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf • NIST AI Risk Management Framework 1.0: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf • The Fundamental Rights and Algorithm Impact Assessment (FRAIA): https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms • OECD Recommendation of the Council on AI: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 • UNESCO Recommendation on the Ethics of Artificial Intelligence: https://unesdoc.unesco.org/ark:/48223/pf0000381137 • Responsible use of artificial intelligence in government: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html Optional: • Ethics Guidelines for Trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence. European commission, 8 April, 2019: https://ai.bsa.org/wp-content/uploads/2019/09/AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf • Zhang, L. 2020, Initiatives in AI Governance: https://static1.squarespace.com/static/5ef0b24bc96ec4739e7275d3/t/5fb58df18fbd7f2b94b5b5cd/1605733874729/SRI+1+-+Initiatives+in+AI+Governance.pdf • Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.

6	Mon, July 21	Flexible (in-person/online)	AI Privacy, Governance and Ethics:	• Cases and Applications • Human rights as an ethical AI framework	• **Guest Speaker:** Diane Gutiw (PhD), VP AI Research Center Lead, Global AI Enablement Center of Expertise, CGI; Member of the Advisory Council on Artificial Intelligence, Innovation, Science and Economic Development Canada. (Additional readings TBD) • UN Special Rapporteur Report concerning AI and Human Rights: UN_AI and Human Rights • Human Rights Handbook for UN Staffs (Read only p.20-34 / Part 1: International Human Rights Standard and Their Development): https://www.ohchr.org/sites/default/files/Documents/Publications/HRhandbooken.pdf • McGregor et al, International Human Rights Law as a Framework for Algorithmic Accountability, https://www.researchgate.net/publication/332457262_International_human_rights_law_as_a_framework_for_algorithmic_accountability Optional: • Human Rights Instruments (You can find the texts of the core human rights treaties adopted under the UN here): https://www.ohchr.org/en/instruments-listings#tab-1
7	Wed, July 23	Fully online	Accountability, Transparency and Explainable AI	• Accountability in AI and "black box" issues • Right to explanation • Algorithmic Impact Assessments	• Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittake (2018), Algorithmic Impact Assessments: A Practical Framework for Public Agency, Accountability. https://ainowinstitute.org/aiareport2018.pdf • Lara Groves, (2022), Algorithmic impact assessment: a case study in healthcare, Ada Lovelace Institute, https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/ • Andrew D. Selbst, (2021), An Institutional View of Algorithmic Impact Assessment, Harvard Journal of Law & Technology Volume 35, Number 1 Fall 2021, https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf Optional: • Algorithmic impact assessments at scale: Practitioners' challenges and needs. (2024). *Journal of Online Trust and Safety, 2*(4) • Mittelstadt, Brent, 'Interpretability and Transparency in Artificial Intelligence', in Carissa Véliz (ed.), *Oxford Handbook of Digital Ethics*, Oxford Handbooks (2023; online edn, Oxford Academic, 10 Nov. 2021),
8	Mon, July 28	Flexible (in-person/online)	Operationalizing Ethics – Tools and Techniques for Responsible AI	• Bias mitigation techniques • Algorithmic audits • Model documentation • Governance processes • Z-Inspection® approach	• Z-Inspection®: A Process to Assess Trustworthy AI: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9380498 • Wade, M & Yokoi, T. (2024). How to Implement AI—Responsibly/ Harvard Business Review https://hbr.org/2024/05/how-to-implement-ai-responsibly

					• Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations. *Science and Engineering Ethics*, *27*(4), Article 44. Optional: • Eticas (2021). Guide to Algorithmic Auditing. https://eticas.ai/wp-content/uploads/2024/04/Guide-to-Algorithmic-Auditing-EN.pdf
9	Wed, July 30	Fully online	Ethical AI in Corporate Management and Resposible Innovation	• AI in business • Responsible innovation • Quick review of "Analyzing bias" assignments	• Schönherr, Martinuzzi & Jarmai (2020). Towards a Business Case for Responsible Innovation. Ch. 7 in: Jarmai, Editor. (2020). Responsible Innovation. Business Opportunities and Strategies for Implementation, Springer, pp. 85-97. • Jarmai, Tharani & Nwafor (2020). Responsible Innovation in Business. Ch. 2 in: Jarmai, Editor. (2020). *Responsible Innovation. Business Opportunities and Strategies for Implementation*, Springer, pp. 7-17. • Schroeder (2020). RI – A Drain on Company Resources or a Competitive Advantage? Ch. 5 in: Jarmai, Editor. (2020). Responsible Innovation. Business Opportunities and Strategies for Implementation, Springer, pp. 51-69. • Brundage (2016). Artificial Intelligence and Responsible Innovation. Ch. 32 in: Müller, Editor. (2016). *Fundamental Issues of Artificial Intelligence*, pp. 543-554. Optional: • Dignum (2019). Taking Responsibility. Ch. 4 in: Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way, pp. 47-69. • Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2020).Ch.6 Risk in the Business of AI. In *An introduction to ethics in robotics and AI*. Springer International Publishing AG.
10	Wed, Aug 6	Fully online	The Rise of Generative AI & Ethics of ChatGPT	• Misinformation and truth • Bias and hate content • Content moderation • Intellectual property • Employment impact	• Weidinger, et all (2021), Ethical and social risks of harm from Language Models, https://arxiv.org/abs/2112.04359 • Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Shmargaret Shmitchel, On the Dangers of Stochastic Parrots? Can language models be too big?, https://dl.acm.org/doi/10.1145/3442188.3445922 • Levy, Neil, 'Fake News: Rebuilding the Epistemic Landscape', in Carissa Véliz (ed.), *Oxford Handbook of Digital Ethics*, Oxford Handbooks (2023; online edn, Oxford Academic, 10 Nov. 2021). Optional: • OpenAI (2024). *"GPT-4o System Card."* https://cdn.openai.com/gpt-4o-system-card.pdf
11	Mon, Aug 11	Fully online	Emerging Ethical Issues:	• AI consciousness and moral status • Emerging AI applications	• **Guest Speaker:** Myrto Mylopoulos (PhD), Associate Professor of Philosophy and Cognitive Science at Carleton University (Additional readings TBD)

					• Robillard, Michael, 'The Ethics of Weaponized AI', in Carissa Véliz (ed.), *Oxford Handbook of Digital Ethics*, Oxford Handbooks (2023; online edn, Oxford Academic, 10 Nov. 2021) • Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2020).Ch.9 Application Areas of AI. In *An introduction to ethics in robotics and AI*. Springer International Publishing AG.
12	Wed, Aug 13	Fully online	AI Ethical AI for Social Good & the Future	• AI's impact on society and the global stage • International cooperation on AI ethics • AI for social good • Course recap	• Tomašev, N., Cornebise, J., Hutter, F. *et al.* (2020). AI for social good: unlocking the opportunity for positive impact. *Nat Commun* 11, 2468 • Floridi, L., Cowls, J., Beltrametti, M. *et al.* (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* **28**, 689–707. • For class discussion: Tegmark, M. (2023). "*The Case for a Pause on Giant AI Experiments*." https://futureoflife.org/wp-content/uploads/2023/05/FLI_Pause-Giant-AI-Experiments_An-Open-Letter.pdf

Plagiarism
Plagiarism, which includes copying someone else's work and submitting it for credit, is a serious academic offence that will not be tolerated. For more information, please refer to the section on academic offences in the Graduate Calendar. Any case of plagiarism will be reported to the Chair of the department and the Carleton University Ethics Committee; the instructor will not address the matter directly. The University has established procedures to handle cases of suspected plagiarism.

Late submissions
To ensure fairness for all students, penalties will be applied to late assignments without medical certificates: Failure to submit an assignment on time will result in an initial penalty of five (5) percentage points, followed by an additional (5) percentage points per day thereafter. For example, an assignment that would normally merit a grade of 20% would receive a grade of 15% if submitted after the deadline on the due date, 10% if submitted the following day, and so on.

Preparation
- **In-Class Participation**: You will be expected to participate actively in class. In addition to voluntary class participation, I might randomly call on students to answer questions or share their insights. High-quality contributions include insightful questions, thoughtful critiques, relevant examples, and connections to course theories. To get the most value from this course, I expect everyone to complete the readings, be prepared to engage in class discussions, and treat classes as a networking opportunity.
- **Reading:** Students are expected to devote time to reading the material and staying current with business and tech news for their class participation. The course readings provide an excellent opportunity to learn to read efficiently and extract the most important information in less time. Detailed instructions regarding the required and optional reading material will be provided before each session.

Group work and free loaders
There is a strict zero-tolerance policy for freeloaders. A freeloader is an individual who exploits the efforts of their team members without making a meaningful contribution in return. Group work is a crucial part of this course. Conflicts that arise within the group should be resolved by the members in a manner that is fair, respectful, and timely.

Recommended books

Bartneck, Ch., Wagner, A., Lütge, Ch., & Welsh, S. (2021). An Introduction to Ethics in Robotics and AI. Springer.

Boddington, P. (2017). *Towards a Code of Ethics for Artificial Intelligence*. Springer.

Chishti, S. (2021). The AI book: the artificial intelligence handbook for investors, entrepreneurs and fintech visionaries. Wiley.

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer.

Dubber, M., Pasquale, F., & Das, S. (2020). *The Oxford Handbook of Ethics of AI*. Oxford University Press.

Jarmai, Katharina, Editor. (2020). *Responsible Innovation. Business Opportunities and Strategies for Implementation*. SpringerOpen.

Morgan, Ch., Editor (2019). *Responsible AI. A Global Policy Framework*. International Technology Law Association.

Recommended articles and book chapters

Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, A., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts,

taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58: 82-115. https://doi.org/10.1016/j.inffus.2019.12.012.

Bender, E., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? FAccT '21, March 3–10, 2021, Virtual Event, Canada. https://doi.org/10.1145/3442188.3445922.

Ghallab, M. (2019). Responsible AI: requirements and challenges. *AI Perspect*, 1(3). https://doi.org/10.1186/s42467-019-0003-z.

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law & Security Review*, 35(4): 410-422. https://doi.org/10.1016/j.clsr.2019.04.007.

Etzioni, A. & Etzioni, O. (2017). Incorporating Ethics into Artificial Intelligence. *Journal of Ethics*, 21:403–418. DOI 10.1007/s10892-017-9252-2.

King O. (2019). Machine Learning and Irresponsible Inference: Morally Assessing the Training Data for Image Recognition Systems. In: Berkich D., d'Alfonso M. (eds) *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence*. Philosophical Studies Series, vol 134. Springer, Cham. https://doi-org.proxy.library.carleton.ca/10.1007/978-3-030-01800-9_14.

Ntoutsi, E., et al. (2020). Bias in data-driven artificial intelligence systems— An introductory survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1356: https://doi-org.proxy.library.carleton.ca/10.1002/widm.1356.

Peters, D., Vold, K., Robinson, D., & Calvo, R. (2020). Responsible AI – Two Frameworks for Ethical Design Practice. *IEEE Transactions on Technology and Society*, 1(1): 34-47. DOI 10.1109/TTS.2020.2974991.

Robbins, S. (2020). AI and the path to envelopment: knowledge as a first step towards the responsible regulation and use of AI-powered machines. *AI & Society*, 35: 391–400. https://doi-org.proxy.library.carleton.ca/10.1007/s00146-019-00891-1.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42: 1568– 1580. http://dx.doi.org/10.1016/j.respol.2013.05.008.

Vetrò, A., Santangelo, A., Beretta, E. and De Martin, J.C. (2019). AI: from rational agents to socially responsible agents. *Digital Policy, Regulation and Governance*, 21(3): 291-304. https://doi.org/10.1108/DPRG-08-2018-0049.

Wagner, D. (2020). The nature of the Artificially Intelligent Firm – An economic investigation into changes that AI brings to the firm. *Telecommunications Policy*, 44: https://doi.org/10.1016/j.telpol.2020.101954.

Additional resources

Ethics of AI in Context website: https://c4ejournal.net/category/ethics-of-ai/

Online companion to the Oxford Handbook of Ethics of AI: https://c4ejournal.net/the-oxford-handbook-of-ethics-of-ai-online-companion/

- Video: Jason Millar, Social Failure Modes in Technology – Implications for AI: https://c4ejournal.net/2019/03/08/jason-millar-social-failure-modes-in-technology-implications-for-ai-2019-c4ej-14/
- Video: Nagla Rizk, *Artificial Intelligence and Inequality in the Middle East*: https://www.youtube.com/watch?v=qeFRu-_4Uec&feature=youtu.be

- Video: Tom Slee, *Private Sector AI: Ethics and Incentives*: https://c4ejournal.net/2019/03/08/tom-slee-private-sector-ai-ethics-and-incentives-2019-c4ej-12/
- Video: Anton Korinek, Economic and Ethical Perspectives on the Rise of Artificial Intelligence: https://c4ejournal.net/2019/03/08/anton-korinek-economic-and-ethical-perspectives-on-the-rise-of-artificial-intelligence-2019-c4ej-16/

Appendix: ADDITIONAL INFORMATION

Group work

The Sprott School of Business encourages group assignments in the school for several reasons. They provide you with opportunities to develop and enhance interpersonal, communication, leadership, follower-ship and other group skills. Group assignments are also good for learning integrative skills for putting together a complex task. Your professor may assign one or more group tasks/assignments/projects in this course. Before embarking on a specific problem as a group, it is your responsibility to ensure that the problem is meant to be a group assignment and not an individual one.

In accordance with the Carleton University Undergraduate Calendar (p. 34), the letter grades assigned in this course will have the following percentage equivalents:

A+ = 90-100	**B+ = 77-79**	C+ = 67-69	D+ = 57-59
A = 85-89	**B = 73-76**	C = 63-66	D = 53-56
A - = 80-84	**B - = 70-72**	C - = 60-62	D - = 50-52

F = Below 50

The minimum passing grade for the course is B-.

Grades entered by Registrar:

WDN = Withdrawn from the course

DEF = Deferred

Academic Regulations

University rules regarding registration, withdrawal, appealing marks, and most anything else you might need to know can be found on the university's website, here:

http://calendar.carleton.ca/undergrad/regulations/academicregulationsoftheuniversity/

Requests for Academic Accommodation

You may need special arrangements to meet your academic obligations during the term. For an accommodation request, the processes are as follows:

Pregnancy obligation

Please contact your instructor with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. For more details, visit the Equity Services website: carleton.ca/equity/wp-content/uploads/Student-Guide-to-Academic-Accommodation.pdf

Religious obligation

Please contact your instructor with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. For more details, visit the Equity Services website: carleton.ca/equity/wp-content/uploads/Student-Guide-to-Academic-Accommodation.pdf

Academic Accommodations for Students with Disabilities

If you have a documented disability requiring academic accommodations in this course, please contact the Paul Menton Centre for Students with Disabilities (PMC) at 613-520-6608 or pmc@carleton.ca for a formal evaluation or contact your PMC coordinator to send your instructor your Letter of Accommodation at the beginning of the term. You must also contact the PMC no later than two weeks before the first in-class scheduled test or exam requiring accommodation (if applicable). After requesting accommodation from PMC, meet with your instructor as soon as possible to ensure accommodation arrangements are made. carleton.ca/pmc

Survivors of Sexual Violence

As a community, Carleton University is committed to maintaining a positive learning, working and living environment where sexual violence will not be tolerated, and is survivors are supported through academic accommodations as per Carleton's Sexual Violence Policy. For more information about the services available at the university and to obtain information about sexual violence and/or support, visit: carleton.ca/sexual-violence-support

Accommodation for Student Activities

Carleton University recognizes the substantial benefits, both to the individual student and for the university, that result from a student participating in activities beyond the classroom experience. Reasonable accommodation must be provided to students who compete or perform at the national or international level. Please contact your instructor with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. https://carleton.ca/senate/wp-content/uploads/Accommodation-for-Student-Activities-1.pdf

For more information on academic accommodation, please contact the departmental administrator or visit: students.carleton.ca/course-outline

Academic Integrity

Violations of academic integrity are a serious academic offence. Violations of academic integrity – presenting another's ideas, arguments, words or images as your own, using unauthorized material, misrepresentation, fabricating or misrepresenting research data, unauthorized co-operation or collaboration or completing work for another student – weaken the quality of the degree and will not be tolerated. Penalties may include; a grade of Failure on the submitted work and/or course; academic probation; a refusal of permission to continue or to register in a specific degree program; suspension from full-time studies; suspension from all studies at Carleton; expulsion from Carleton, amongst others. Students are expected to familiarize themselves with and follow the Carleton University Student Academic Integrity Policy which is available, along with resources for compliance at: https://carleton.ca/registrar/academic-integrity/.

Sprott Student Services

The Sprott student services office, located in 710 Dunton Tower, offers academic advising, study skills advising, and overall academic success support. If you are having a difficult time with this course or others, or just need some guidance on how to successfully complete your Sprott degree, please drop in any weekday between 8:30am and 4:30pm. Our advisors are happy to discuss grades, course selection,

tutoring, concentrations, and will ensure that you get connected with the resources you need to succeed! http://sprott.carleton.ca/students/undergraduate/learning-support/

Centre for Student Academic Support

The Centre for Student Academic Support (CSAS) is a centralized collection of learning support services designed to help students achieve their goals and improve their learning both inside and outside the classroom. CSAS offers academic assistance with course content, academic writing and skills development. Visit CSAS on the 4th floor of MacOdrum Library or online at: carleton.ca/csas.

Important Information:

- Students must always retain a hard copy of all work that is submitted.

- All final grades are subject to the Dean's approval.

- For us to respond to your emails, we need to see your full name, CU ID, and the email must be written from your valid CARLETON address. Therefore, in order to respond to your inquiries, please send all email from your Carleton CMail account. If you do not have or have yet to activate this account, you may wish to do so by visiting http://carleton.ca/ccs/students/