



**Carleton**  
University

**Sprett**  
School of Business



Technology Innovation  
Management



**AACSB**  
ACCREDITED

## **Institute of Technology Entrepreneurship and Commercialization**

### **TIMG 5204, Responsible Artificial Intelligence - 21176 - A [0.5 credit]**

#### **Summer 2024**

##### **Time and Place**

May 06, 2024 - Jun 16, 2024, Mondays and Wednesdays, Nicol Building 4030

18:55pm to 20:55pm

IN-PERSON with flexible online/on campus.

##### **Instructor**

Somaieh Nikpoor

##### **Course materials**

Access to online course sessions, course materials and recorded videos will be provided through the new CU Brightspace system: <https://carleton.ca/brightspace/>. To access Brightspace you should use your CU credentials and select the TIMG5204A Responsible Artificial Intelligence (SEM) Summer 2024

##### **Office hours**

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

##### **Target audience**

The course addresses the need of graduate students registered in the M.A.B.A. option of the the Technology Innovation Management (TIM) program. It is however open for other TIM students from the thesis (M.A.Sc.) and project (M.Eng., M.Ent., M.A.B.A., M.D.T.E.) options. Students in other programs are welcome to attend this course depending on space availability. However, all students will need to meet the high academic standards of the TIM program. For non-TIM students a preliminary meeting with the professor will be required before admission in the course is granted.

##### **Calendar description**

Responsible Artificial Intelligence - TIMG 5204 - A

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**

The course offers an opportunity for students to engage in a critical reflection on the ethical dimensions of the global AI revolution and related phenomena in various aspects of technology-driven businesses, private, public, and political life. The focus however is on the ethical implications and frameworks for evaluating trustworthy AI targeted for managers of technology firms and entrepreneurs. Students will engage in a combination of lectures, open discussions and projects that will help them to get familiarized with the complex nature of the ethical challenges of emerging AI technologies and the continuous efforts required to address them. Classes will be used to drill down into case studies, structured around core concerns being raised by society, governments and industry, such as bias, fairness, human rights and fundamental rights, data re-use, data protection and data privacy, discrimination, transparency and accountability. Particular attention will be paid to responsible AI frameworks that could be used in examining the relationship between the adoption of specific AI capabilities and the corresponding ethical issues emerging from their application in specific business domains. Students will also engage in a research project examining recent literature on responsible AI and ethics, identifying open research questions, and generating insights that could inform the responsible use of AI technologies in ethical business practice.

### **Objectives / Learning outcomes**

Students will be expected to:

- Develop a sense of responsibility and ethical norms in the field of AI technological applications.
- Achieve a deeper familiarization with ethical issues emerging from exemplary real-life cases of AI applications in specific business domains.
- Acquire an understanding of the complexities of making ethical decisions in a landscape of competing business and stakeholder interests.
- Describe and critically discuss the social, political, legal and ethical issues surrounding data-driven innovation, including those posed by big data, AI systems, and machine learning systems.
- Get familiarized with the ethical implications for managers of technology firms interested in adopting or developing AI technologies, and turning potential ethical challenges into business opportunities.
- Examine the relationship between the practical adoption of specific AI capabilities and the corresponding ethical implications.
- Enhance their critical thinking, analytical ability and research skills focusing on searching for and finding most recent academic articles on responsible AI and ethics, examining the relevance of their findings, identifying open research questions, and generating technological and business insights that could inform the responsible use of AI technologies in ethical business practice.

### **Class sessions**

Class sessions will include a combination of online lectures, video sessions, interactive discussions, student workshops & 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 TIMG 5103 course will work individually and in a group of up to 6 students to perform tasks in class, participate in informal group meetings and contribute to delivering the assignments. Once formed, each group should establish a clear project management structure that will help collaboration between group members and maximize the value of the deliverables. The group member tasks should be defined and agreed on a weekly basis. The communication with any external clients should be concise, clear and meaningful.

#### **1. Assignment # 1 (individual assignment, 25%): Formulate actionable insights based on video recording of the webinar “The Challenges of Applying AI Ethics”**

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

This is the link to the recording of the webinar on YouTube: <https://youtu.be/8iULaxKDhuE>.

The assignment is an individual assignment and consists of the following tasks:

- Watch the recording of the webinar to familiarize yourself with the content of the talks.
- Formulate min 5 actionable statements on what you suggest should be done with what you learned from the webinar.
- Each statement should include the following ingredients: [who] + [should/could/will] + [action to take]. See examples of statements below.
- Submit your statements here:  
[https://docs.google.com/forms/d/e/1FAIpQLSdxrhNytNX6UyVTXCaTJHJLb3Vp7eP6gyW4rgK-0b7faQF7Ag/viewform?usp=pp\\_url](https://docs.google.com/forms/d/e/1FAIpQLSdxrhNytNX6UyVTXCaTJHJLb3Vp7eP6gyW4rgK-0b7faQF7Ag/viewform?usp=pp_url)
- 

### *Grading*

The corpus of all submitted actionable statements will be used by the instructor to perform a topic modeling analysis. The grading of the assignment will be based on the quality of the submitted statements including:

- Linguistic clarity of the statements

- Degree to which statements follow the format: [who] + [should/could/will] + [action to take]
- Actionable value and quality of the insights.

### *Slides provided from Invited Speakers*

The speakers have generously shared their slides and some other suggestions. Please see these links:

- Svetlana Kiritchenko – [slide deck](#)
- Norm Vinson – [slide deck](#)
- Margaret McKay – [slide deck](#)

### *Examples of statements*

- 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.

## **2. Assignment # 2 (group assignment, 25%): Analyzing bias**

### **2A. Select one of the following case studies:**

- Racial Bias in a health algorithm: [https://www.ftc.gov/system/files/documents/public\\_events/1548288/privacycon-2020-ziad\\_obermeyer.pdf](https://www.ftc.gov/system/files/documents/public_events/1548288/privacycon-2020-ziad_obermeyer.pdf)
- AI bias is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions): <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
- Facebook's Ad's algorithm: <https://arxiv.org/abs/1904.02095>
- ChatGPT fails at diagnosing pediatric case studies: <https://jamanetwork.com/journals/jamapediatrics/article-abstract/2813283>
- iTotur Group AI system rejects applicants due to age: <https://www.eeoc.gov/newsroom/itotorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit#:~:text=According%20to%20the%20EEOC's%20lawsuit,States%20because%20of%20their%20age.>
- 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>

**2B.** Familiarize yourself individually with the selected case study and set up group meetings to discuss and answer the following questions:

- What type of data is used?
- What types of biases exist in the data/system?
- What machine learning technique is deployed?
- What are the consequences of these biases on human life and human dignity?
- How were biases in the system addressed?
- Based on what you learned in this lecturer, are there any types of biases that should be considered?
- How would you mitigate these biases?

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

**2C.** Summarize your group discussions and prepare a 3-7 pages report. You can use the following format to organize your content:

- **Introduction:** Explain the background and the use case and formulate the problem
- **Data and Mechanism of Bias:** List and explain all the existing biases and any additional ones you discovered through your research and explain how they impact the outcome of the model (or decision making).
- **Mitigation:** What mechanism was used to address the biases? If there are none, how do you propose to mitigate these biases?

### **3. Assignment # 3 (group assignment, 30%): Trustworthy AI assessment**

Select one of the AI systems/solutions from assignment 2.

- Assess the AI system for “trustworthiness” based on the EU “Framework for Trustworthy AI”, using the Z-inspection® process.
- Summarize your analysis in no more than 3-7 pages.

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>

### **4. Take home exam (individual assignment, 20%):** Will be provided at the last class session.

## Student evaluation and assignment grading

Final grade will be assigned using the following mark allocation:

	Assignment	Deliverable	Date	%
1	Formulate actionable insights on the recording of the webinar “The Challenges of Applying AI Ethics”	Min 5 actionable statements on what you suggest should be done with what you learned from the webinar.	Monday, May 20	25
2	<i>Analyzing bias</i>	A 3-7 pages report including: <ul style="list-style-type: none"> <li>• <u>Introduction</u> explaining the background and the use case and providing a formulation of the problem</li> <li>• <u>Data and Mechanism of Bias</u> providing a list and explaining all the existing biases and any additional ones you have discovered through your research, and explaining how they impact the outcome of the model (or decision making).</li> <li>• <u>Mitigation</u> describing what mechanism was used to address the biases. If there were no such mechanisms used, explain how you propose to mitigate the existing biases.</li> </ul>	Monday, June 12	25
3	<i>Trustworthy AI Assessment</i>	A 3-7 pages report: <ul style="list-style-type: none"> <li>• Select one of the AI system/solution from assignment 2.</li> <li>• Assess the AI system for “trustworthiness” based on the EU “Framework for Trustworthy AI”, using the Z-inspection® process.</li> </ul>	Friday, June 14	30
4	<i>Take home exam</i>	Details will be provided at the last class session.	Monday, June 24	20
Total				100

## Plagiarism

Plagiarism, including copying and handing in for credit someone else's work, is a serious instructional offense that will not be tolerated. Please refer to the section on instructional offenses in the Graduate Calendar for additional information. A case of plagiarism will be referred to the Chair of the department and the Carleton University Ethics Committee. The instructor will not deal with the matter directly. The University has clear processes to deal with students who are suspected of plagiarism.

## Group work and free loaders

There is zero tolerance for free loaders. A free loader refers to an individual who takes advantage of team members’ efforts without contributing much in return. Group work is an important component of this course. Group conflicts are to be dealt with by the group in a way that is fair, respectful and fast.

## Class schedule

#	Date	Topic	Assigned readings & details
1	Monday May 6	<ul style="list-style-type: none"> <li>● Introduction to course objectives</li> <li>● Detailed description of Assignments</li> <li>● Introduction to AI in the context of responsibility and ethics</li> <li>● Student group formation</li> </ul>	<ul style="list-style-type: none"> <li>● Course outline</li> <li>● Dignum, Virginia (2019). Ch. 1: Introduction &amp; Ch. 2: What is Artificial Intelligence? In: <i>Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way</i>, pp. 1-34.</li> <li>● Wagner, D. (2020). The nature of the Artificially Intelligent Firm – An economic investigation into changes that AI brings to the firm. <i>Telecommunications Policy</i>, 44, 101954.</li> <li>● Morgan et al. (2019). Introduction. In: Morgan, Editor. (2019). <i>Responsible AI. A Global Policy Framework</i>, pp. 16-29.</li> </ul>
2	Wednesday May 8	<ul style="list-style-type: none"> <li>● Why Trustworthy AI?</li> <li>● Trustworthy AI frameworks</li> <li>● The EU "Framework for Trustworthy AI"</li> <li>● EU Ethics Guidelines for Trustworthy AI</li> <li>● EU AI Act – The first AI regulation</li> <li>● Q&amp;A about assignments</li> </ul>	<ul style="list-style-type: none"> <li>● Lecture slides provided by instructor <ul style="list-style-type: none"> <li>● Dignum (2019). Can AI Systems Be Ethical? Ch. 5 in: <i>Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way</i>, pp. 71-92.</li> </ul> </li> </ul> <p>Reading materials for the three sessions on human rights:</p> <ul style="list-style-type: none"> <li>● UN Special Rapporteur Report concerning AI and Human Rights: UN_AI and Human Rights</li> <li>● The Fundamental Rights and Algorithm Impact Assessment (FRAIA): <a href="https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms">https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms</a></li> <li>● [Optional] Human Rights Handbook for UN Staffs (Read only p.20-34 / Part 1: International Human Rights Standard and Their Development): <a href="https://www.ohchr.org/sites/default/files/Documents/Publications/HRhandbooken.pdf">https://www.ohchr.org/sites/default/files/Documents/Publications/HRhandbooken.pdf</a></li> <li>● [Optional] McGregor et al, International Human Rights Law as a Framework for Algorithmic Accountability, <a href="https://www.researchgate.net/publication/332457262_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</a></li> <li>● [Optional] Human Rights Instruments (You can find the texts of the core human rights treaties adopted under the UN here): <a href="https://www.ohchr.org/en/instruments-listings#tab-1">https://www.ohchr.org/en/instruments-listings#tab-1</a></li> <li>● Ethics Guidelines for Trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence. European commission, 8 April, 2019: <a href="https://ai.bsa.org/wp-content/uploads/2019/09/AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf">https://ai.bsa.org/wp-content/uploads/2019/09/AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf</a></li> </ul>
3	Monday May 13	<ul style="list-style-type: none"> <li>● EU Ethics Guidelines for Trustworthy AI – cont'ed</li> </ul>	<p>Lecture slides provided by instructor</p> <ul style="list-style-type: none"> <li>● High-Level Expert Group on Artificial Intelligence (2020) – Assessment List for Trustworthy Artificial Intelligence: <a href="https://ai.bsa.org/wp-">https://ai.bsa.org/wp-</a></li> </ul>

		<ul style="list-style-type: none"> <li>● Assessing Trustworthy AI in Practice</li> <li>● Intro Z-Inspection®: A Process to Assess Trustworthy AI</li> <li>● Use Case. On Assessing Trustworthy AI in Healthcare.</li> <li>● Introduction to assignment 3</li> </ul>	<ul style="list-style-type: none"> <li>● <a href="https://www.mdpi.com/2199-8531/5/1/18">content/uploads/2019/09/AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf</a></li> <li>● Ethical Framework for Designing Autonomous Intelligent Systems. J Leikas et al. J. of Open Innovation, 2019, 5, 1: <a href="https://www.mdpi.com/2199-8531/5/1/18">https://www.mdpi.com/2199-8531/5/1/18</a></li> <li>● Z-Inspection®: A Process to Assess Trustworthy AI: <a href="https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&amp;arnumber=9380498">https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&amp;arnumber=9380498</a></li> <li>● Whittlestone, J et al (2019) – Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. Whittlestone, J, Nyrupe, R, Alexandrova, A, Dihal, K, Cave, S. (2019), London. Nuffield Foundation.</li> </ul>
4	Wednesday May 15	<ul style="list-style-type: none"> <li>● Assessing Trustworthy AI – cont’ed</li> <li>● Z-Inspection® - Overview and Socio-Technical Scenarios</li> <li>● Ethical Tensions</li> <li>● Q&amp;A Assignment 3</li> </ul>	<ul style="list-style-type: none"> <li>● Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. Whittlestone et al. (2019): <a href="https://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf">https://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf</a></li> <li>● How to Assess Trustworthy AI in Practice: <a href="https://arxiv.org/ftp/arxiv/papers/2206/2206.09887.pdf">https://arxiv.org/ftp/arxiv/papers/2206/2206.09887.pdf</a></li> <li>● Ethical Framework for Designing Autonomous Intelligent Systems, by Jaana Leikas , Raija Koivisto and Nadezhda Gotcheva: <a href="https://www.mdpi.com/2199-8531/5/1/18">https://www.mdpi.com/2199-8531/5/1/18</a></li> <li>● Z-Inspection®: A Process to Assess Trustworthy AI, <a href="https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&amp;arnumber=9380498">https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&amp;arnumber=9380498</a></li> </ul>
	Monday May 20	Victoria Day Holiday	<ul style="list-style-type: none"> <li>● No class</li> <li>● <b>Delivery of Assignment 1</b></li> </ul>
5	Wednesday May 22	Guest speaker	TBD
6	Monday May 27	<ul style="list-style-type: none"> <li>● Bias and Fairness in AI</li> <li>● Biases in Large Language Models (LLMs)</li> <li>● Introduction to Assignment 2</li> </ul>	<ul style="list-style-type: none"> <li>● Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A. (2019). A Survey on Bias and Fairness in Machine Learning. <i>arXiv:1908.09635</i>: <a href="https://doi.org/10.48550/arXiv.1908.09635">https://doi.org/10.48550/arXiv.1908.09635</a></li> <li>● Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., Crawford, K. (2018). Datasheets for Datasets. <i>arXiv:1803.09010</i>: <a href="https://doi.org/10.48550/arXiv.1803.09010">https://doi.org/10.48550/arXiv.1803.09010</a></li> <li>● Navigli, R. and Conia, S. (2023), Biases in Large Language models: Origins, Inventory and discussion, <a href="https://dl.acm.org/doi/pdf/10.1145/3597307">https://dl.acm.org/doi/pdf/10.1145/3597307</a></li> <li>● Suresh, H., Guttag, J. V. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. <i>arXiv:1901.10002</i>: <a href="https://doi.org/10.48550/arXiv.1901.10002">https://doi.org/10.48550/arXiv.1901.10002</a></li> <li>● Passi, S., Barocas, S. (2019). Problem Formulation and Fairness. <i>arXiv:1901.02547</i>: <a href="https://doi.org/10.48550/arXiv.1901.02547">https://doi.org/10.48550/arXiv.1901.02547</a></li> </ul>
7	Wednesday, May 29	● Introduction to Algorithmic Impact Assessment	● Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittake (2018), ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL



		<ul style="list-style-type: none"> <li>● Q&amp;A about Assignment 2</li> </ul>	<p>FRAMEWORK FOR PUBLIC AGENCY, ACCOUNTABILITY.</p> <p><a href="https://ainowinstitute.org/aiareport2018.pdf">https://ainowinstitute.org/aiareport2018.pdf</a></p> <ul style="list-style-type: none"> <li>● Lara Groves, (2022), Algorithmic impact assessment: a case study in healthcare, Ada Lovelace Institute, <a href="https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/">https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/</a></li> <li>● Andrew D. Selbst, (2021), An Institutional View of Algorithmic Impact Assessment, Harvard Journal of Law &amp; Technology Volume 35, Number 1 Fall 2021, <a href="https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf">https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf</a></li> </ul>
8	Monday, June 3	<ul style="list-style-type: none"> <li>● Introduction to Generative AI</li> <li>● Ethical Implication of Large Language Models (ex:ChatGPT)</li> </ul>	<ul style="list-style-type: none"> <li>● Weidinger, et all (2021), Ethical and social risks of harm from Language Models, <a href="https://arxiv.org/abs/2112.04359">https://arxiv.org/abs/2112.04359</a></li> <li>● Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Shmargaret Shmitchel, On the Dangers of Stochastic Parrots? Can language models be too big?, <a href="https://dl.acm.org/doi/10.1145/3442188.3445922">https://dl.acm.org/doi/10.1145/3442188.3445922</a></li> <li>●</li> </ul>
9	Wednesday June 5	<ul style="list-style-type: none"> <li>● AI and responsible innovation</li> <li>● Leveraging responsible innovation to create business opportunities</li> <li>● Q&amp;A about Assignment # 3</li> </ul>	<ul style="list-style-type: none"> <li>● Jarmai, Tharani &amp; Nwafor (2020). Responsible Innovation in Business. Ch. 2 in: Jarmai, Editor. (2020). <i>Responsible Innovation. Business Opportunities and Strategies for Implementation</i>, Springer, pp. 7-17.</li> <li>● Schroeder (2020). RI – A Drain on Company Resources or a Competitive Advantage? Ch. 5 in: Jarmai, Editor. (2020). <i>Responsible Innovation. Business Opportunities and Strategies for Implementation</i>, Springer, pp. 51-69.</li> <li>● Schönherr, Martinuzzi &amp; Jarmai (2020). Towards a Business Case for Responsible Innovation. Ch. 7 in: Jarmai, Editor. (2020). <i>Responsible Innovation. Business Opportunities and Strategies for Implementation</i>, Springer, pp. 85-97.</li> <li>● Dignum (2019). Taking Responsibility. Ch. 4 in: <i>Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way</i>, pp. 47-69.</li> <li>● Brundage (2016). Artificial Intelligence and Responsible Innovation. Ch. 32 in: Müller, Editor. (2016). <i>Fundamental Issues of Artificial Intelligence</i>, pp. 543-554.</li> </ul>
10	Monday, June 10	Individual group workshops on assignment # 2	<ul style="list-style-type: none"> <li>● Group work presentation</li> <li>● <b>Delivery of Assignment # 2 -</b></li> </ul>
11	Wednesday June 12	<ul style="list-style-type: none"> <li>● Global AI Governance</li> <li>● The 8 principles of a Responsible AI Policy Framework</li> <li>● Q&amp;A about Assignment # 3</li> </ul>	<ul style="list-style-type: none"> <li>● OECD Recommendation of the Council on AI: <a href="https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449">https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449</a></li> <li>● Zhang, L. 2020, Initiatives in AI Governance: <a href="https://static1.squarespace.com/static/5ef0b24bc96ec4739e7275d3/t/5fb58df18fbd7f2b94b5b5cd/1605733874729/SRI+1+-+Initiatives+in+AI+Governance.pdf">https://static1.squarespace.com/static/5ef0b24bc96ec4739e7275d3/t/5fb58df18fbd7f2b94b5b5cd/1605733874729/SRI+1+-+Initiatives+in+AI+Governance.pdf</a></li> <li>● Schneider, J., Abraham, R., &amp; Meske, Ch. 2020: <a href="https://www.researchgate.net/publication/346045591_AI_Governance_for_Businesses">https://www.researchgate.net/publication/346045591_AI_Governance_for_Businesses</a></li> <li>● Summary of a Responsible AI Policy Framework. In: Morgan, Editor. (2019). <i>Responsible AI. A Global Policy Framework</i>, pp. 289-303.</li> </ul>

12	Friday June 14	Student presentations of Assignment # 3: Cases in responsible AI and ethics	Group work presentation • <b>Delivery of assignment # 3</b>

Take home exam should be delivered by Sunday midnight, June 23.

### 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.

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](https://doi-org.proxy.library.carleton.ca/10.1007/978-3-030-01800-9_14).

Ntoutsis, 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](http://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](http://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](mailto: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](http://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 its 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](http://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](http://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](http://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/>
-