Ethics and Data Science

Overview[1]

Details

Instructor: Rohan Alexander
Email: rohan.alexander@utoronto.ca
Term: Winter 2021
Mode: Online via Zoom by default at this stage, with an option for in-person discussions depending on the circumstances.
Synchronous discussion times: 9-11am on Fridays (timetable contingent).
Office hour: By appointment - please email the instructor.

Purpose

The purpose of this reading course is to develop students who can:
1. engage in thoughtful, ethical, critique of data science, its antecedents, current state, and likely evolution; and
2. work productively to implement existing data science methods, as well as contribute to the creation of novel methods or applications.

Each week students will read relevant papers and books, engage with them through discussion with each other and the instructor, learn related technical skills, and bring this together through on-going assessment. All students are expected to be prepared for each week's discussion through completing the readings and technical requirements. A specific student will act as the lead for each week.

Related courses

This course builds on the quantitative skills developed in INF2178 'Experimental design for data science' such as: gathering, cleaning, and preparing datasets in a reproducible way; analysing them using Bayesian hierarchical models; and finally communicating this analysis in an efficient manner.

Course learning objectives and their relationship to assessment and program learning outcomes[2]

1. To conduct original research applying sophisticated, quantitative, statistical analysis in situations involving real-world data that the student gathered.
   a. This objective is demonstrated through all assessment items, but in particular the two papers which focus explicitly on this objective.
   b. This objective is related to the Program Learning Outcomes #2, #3, #4 and #5.
2. To understand and identify ethical considerations in relation to quantitative statistical analysis and appropriately justify the choices made given competing ethical concerns.
   a. This objective is demonstrated through all assessment items, but in particular the two papers as they force this issue.
   b. This objective is related to the Program Learning Outcomes #2, #5, and #6.
3. To understand the way in which ethical considerations are an inherent aspect of every quantitative statistical analysis, and consider them appropriately.
   a. This objective is demonstrated through all assessment items.
   b. This objective is related to the Program Learning Outcomes #2, #5, and #6
4. To communicate in a way that engages a broad audience and establish a portfolio of independent research that could be further developed.
   a. This objective is demonstrated through the blog post assessment items.
   b. This objective is related to the Program Learning Outcomes #1, #2, #5, and #6.
5. To be able to implement state-of-the-art data science tools toward novel applications.
   a. This objective is demonstrated through the blog post assessment items.
   b. This objective is related to the Program Learning Outcomes #1, #3, #5, and #6.

Topics

Week 1 - General

Brian Cantwell Smith, Reid Hoffman Professor of Artificial Intelligence and the Human, University of Toronto, will be invited to join the discussion briefly this week.

---

[2] The program learning outcomes are available here:
https://ischool.utoronto.ca/areas-of-study/master-of-information/

Ethical

Core:
- Cantwell Smith, Brian, 2019, *The Promise of AI*, MIT Press.
- Healy, Kieran, 2020, 'The Kitchen Counter Observatory', 21 May, https://kieranhealy.org/blog/archives/2020/05/21/the-kitchen-counter-observatory/.
- O'Neil, Cathy, 2016, *Weapons of Math Destruction*, Crown Books.

Additional:
- Green, B, 2018, 'Data science as political action: Grounding data science in a politics of justice', *arXiv preprint arXiv:1811.03435*. https://arxiv.org/abs/1811.03435.
- Irving, Geoffrey and Amanda Askell, 2019, 'AI Safety Needs Social Scientists', *Distill*, 19 February, https://distill.pub/2019/safety-needs-social-scientists/.
- Keyes, Os, 2019, 'Counting the Countless', *Real Life*, 8 April, https://reallifemag.com/counting-the-countless/.
- Leslie, David, 2020, 'Tackling COVID-19 through Responsible AI Innovation: Five Steps in the Right Direction', *Harvard Data Science Review*, 5 June, https://hdsr.mitpress.mit.edu/pub/as1p81um.
- Suresh, Harini and John V. Guttag, 2019, 'A Framework for Understanding Unintended Consequences of Machine Learning', *arXiv preprint*, https://arxiv.org/abs/1901.10002. https://arxiv.org/pdf/1901.10002.pdf.
- Raji, Inioluwa Deborah, 2020, 'The Discomfort of Death Counts: Mourning through the Distorted Lens of Reported COVID-19 Death Data', *Patterns*, https://doi.org/10.1016/j.patter.2020.100066

Technical

- Review Part 1 of 'Telling Stories with Data' and address any short-comings.
- Set-up RStudio Cloud, if necessary.
- Set-up GitHub, if necessary (see Part 1 of 'Telling Stories with Data')
- Set-up a website, if necessary (see Part 3 of 'Telling Stories with Data')

Week 2 - Data and consent

Nina Yiannopoulos, Director, Data Protection and Privacy, Deloitte Australia, will be invited to join the discussion briefly this week.

Ethical

Core:
- Boykis, Vicki, 2019, 'Neural nets are just people all the way down', 16 October, https://vicki.substack.com/p/neural-nets-are-just-people-all-the.

- Crawford, Kate, and Vladan Joler, 2018, 'Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources', *AI Now Institute and Share Lab*, 7 September, https://anatomyof.ai.
- Kitchin, Rob, 2014, *The data revolution: Big data, open data, data infrastructures and their consequences*, Sage.

Additional:

- Boyd, Danah, and Kate Crawford, 2012, 'Critical Questions for Big Data', *Information, Communication & Society*, 15:55, 662-679.
- Denton, Emily, Alex Hanna, Razvan Amironesei, Andrew Smart, Hilary Nicole, Morgan Klaus Scheuerman, 2020, 'Bringing the People Back In: Contesting Benchmark Machine Learning Datasets', *arXiv*, 14 July, https://arxiv.org/abs/2007.07399.
- Eubanks, Virginia, 2019, 'Automating Inequality: How high-tech tools profile, police and punish the poor', Lecture, University of Toronto, 12 March, https://www.youtube.com/watch?v=g1ZZZ1QLXOI.
- Lemov, Rebecca, 2016, 'Big data is people!', *Aeon*, 16 June, https://aeon.co/essays/why-big-data-is-actually-small-personal-and-very-human.
- Office of Oversight and Investigations Majority Staff, 2013, 'A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes', *Staff Report for Chairman Rockefeller*, 18 December, United States Senate, Committee on Commerce, Science and Transportation, https://www.commerce.senate.gov/services/files/0d2b3642-6221-4888-a631-08f2f255b577.
- Radin, Joanna, 2017, '"Digital Natives": How Medical and Indigenous Histories Matter for Big Data', *Osiris*, 32 (1), 43-64.
- Snowberg, Erik and Leeat Yariv, 2018, 'Testing The Waters: Behavior Across Participant Pools', *NBER Working Paper*, No. 24781, http://www.nber.org/papers/w24781.
- Various. Please read and consider the tweets related to 'Black in the ivory', beginning with: https://twitter.com/thePhDandMe/status/1272990673214398464 within the context of Bergis Jules, Ed Summers and Vernon Mitchell, 2018, 'Documenting The Now: Ethical Considerations for Archiving Social Media Content Generated by Contemporary Social Movements: Challenges, Opportunities, and Recommendations', White Paper, DocNow, https://www.docnow.io/docs/docnow-whitepaper-2018.pdf.

Technical

- Review APIs, scraping websites, and parsing PDFs (Part 2 of 'Telling Stories with Data') to address any shortcomings.

Week 3 - Women and gender

Ethical

Core:

- D'Ignazio, Catherine, and Lauren F. Klein, 2020 *Data Feminism*.
- Gebru, Timnit, 2020, 'Race and Gender', *The Oxford Handbook of Ethics of AI*, Chapter 13, Oxford University Press.

Additional:

- Borgerson, Janet L., 2007, 'On the Harmony of Feminist Ethics and Business Ethics', *Business and Society Review*, 112 (4):477-509.
- D'Ignazio, Catherine, and Lauren F. Klein, 'Feminist data visualization', *Workshop on Visualization for the Digital Humanities* (VIS4DH), Baltimore. IEEE. 2016.
- Hill, Kashmir, 2017, 'What Happens When You Tell the Internet You're Pregnant', *Jezebel*, 27 July, https://jezebel.com/what-happens-when-you-tell-the-internet-youre-pregnant-1794398989.
- Keyes, Os, 2018, 'The misgendering machines: Trans/HCI implications of automatic gender recognition', *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1-22, https://dl.acm.org/doi/pdf/10.1145/3274357.
- Quintin, Cooper, 2017, 'Pregnancy Panopticon', *DEFCON* 25, https://www.eff.org/files/2017/07/27/the_pregnancy_panopticon.pdf.
- Woods, Heather Suzanne, 2018, 'Asking more of Siri and Alexa: feminine persona in service of surveillance capitalism', *Critical Studies in Media Communication*, 35.4, pp. 334-349.

Technical

- Review the essentials of Bayesian models by going through McElreath, 2020, *Statistical Rethinking*, 2nd Edition, (at least chapters 1, 2, 4, 7, 9, 11, 12, and 13) to address any shortcomings.

Week 4 - Race

Tom Davidson, Assistant Professor, Sociology, Rutgers University, will be invited to join the discussion briefly this week.

Ethical

Core:

- Noble, Safiya Umoja, 2018, *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press.

Additional:

- Buolamwini, Joy and Timnit Gebru, 2018, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', *Proceedings of Machine Learning Research Conference on Fairness, Accountability, and Transparency*, 81: pp. 1–15, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf
- Davidson, Thomas, Debasmita Bhattacharya, and Ingmar Weber, 2019, 'Racial bias in hate speech and abusive language detection datasets', *arXiv*, https://arxiv.org/abs/1905.12516.
- Kwet, Michael, 2019, 'Digital colonialism: US empire and the new imperialism in the Global South', *Race & Class* 60.4, 3-26.
- Scheuerman, M. K., Wade, K., Lustig, C., and Brubaker, J. R., 2020, 'How We've Taught Algorithms to See Identity: Constructing Race and Gender in Image Databases for Facial Analysis', *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), 1-35.
- Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, 2019, 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science*, Vol. 366, Issue 6464, pp. 447-453, DOI: 10.1126/science.aax2342, https://science.sciencemag.org/content/366/6464/447/tab-pdf

Technical

- Pick a project from The Markup's Show Your Work section (https://themarkup.org/series/show-your-work) and reproduce it, writing your own code. You may pick whatever language you are comfortable in.

Week 5 - Natural Language Processing

Xanda Schofield, Assistant Professor, Computer Science, Harvey Mudd College, will be invited to join the discussion briefly this week.

Ethical

Core:

- Hovy, Dirk and Shannon L. Spruit, 2016, 'The Social Impact of Natural Language Processing', *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 591–598, https://aclweb.org/anthology/P16-2096.pdf.
- Prabhumoye, Shrimai, Elijah Mayfield, and Alan W Black, 2019, 'Principled Frameworks for Evaluating Ethics in NLP Systems', Proceedings of the 2019 Workshop on Widening NLP, https://aclweb.org/anthology/W19-3637/.

Additional:

- Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama and Adam T. Kalai, 2016, 'Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings', Advances in Neural Information Processing Systems 29 (NIPS 2016), http://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-d.
- Chang, Kai-Wei, Vinod Prabhakaran, and Vicente Ordonez, 2019, 'Bias and Fairness in Natural Language Processing', *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): Tutorial Abstracts*, https://aclweb.org/anthology/D19-2004/.
- Hutchinson, Ben, Vinodkumar Prabhakaran, Emily Denton, Kellie Webster, Yu Zhong, and Stephen Denuyl, 2020, 'Social Biases in NLP Models as Barriers for Persons with Disabilities', *arXiv*, https://arxiv.org/abs/2005.00813.
- Solaiman, Irene, Miles Brundage, Jack Clark, Amanda Askell, Ariel Herbert-Voss, Jeff Wu, Alec Radford, Gretchen Krueger, Jong Wook Kim, Sarah Kreps, Miles McCain, Alex Newhouse, Jason Blazakis, Kris McGuffie, Jasmine Wang, 2019, 'Release Strategies and the Social Impacts of Language Models', *arXiv*, https://arxiv.org/abs/1908.09203.
- Tatman, Rachel, 2020, 'What I Won't Build', *Widening NLP Workshop 2020*, Keynote address, 5 July, https://slideslive.com/38929585/what-i-wont-build and http://www.rctatman.com/talks/what-i-wont-build.
- Zhao, Jieyu, Tianlu Wang, Mark Yatskar, Vicente Ordonez and Kai-Wei Chang, 2017, 'Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints', *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,* pp. 2979–2989, https://aclweb.org/anthology/D17-1323.pdf.
- (Optional/fun/horrifying) Hao, Karen, 2020, 'The messy, secretive reality behind OpenAI's bid to save the world', *MIT Review*, 17 February, https://www.technologyreview.com/2020/02/17/844721/ai-openai-moonshot-elon-musk-sam-altman-greg-brockman-messy-secretive-reality/.

Technical

- Implement a NLP model via Hugging Face or Spacy, depending on your language preference.

Week 6 - AI Ethics

Jamie Duncan, Junior Policy Analyst, Artificial Intelligence Hub, Innovation, Science and Economic Development Canada, will be invited to join the discussion briefly this week.

Ethical

Core:

- Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff, Bobby Filar, Hyrum Anderson, Heather Roff, Gregory C. Allen, Jacob Steinhardt, Carrick Flynn, Seán Ó hÉigeartaigh, Simon Beard, Haydn Belfield, Sebastian Farquhar, Clare Lyle, Rebecca Crootof, Owain Evans, Michael Page, Joanna Bryson, Roman Yampolskiy, Dario Amodei, 2019, 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation', *arXiv*, https://arxiv.org/abs/1802.07228.
- Jobin, A., Ienca, M., and Vayena, E, 2019, 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence*, 1(9), pp. 389-399. https://www.nature.com/articles/s42256-019-0088-2.

Additional:

- Australian Human Rights Commission, 2019, 'Human Rights and Technology Discussion Paper', December, https://tech.humanrights.gov.au/sites/default/files/2019-12/TechRights_2019_DiscussionPaper.pdf.
- Crawford, Kate, Amba Kak and Jason Schultz, 2020, 'Submission to the Australian Human Rights Commission Human Rights & Technology Discussion Paper', AI Now Institute, New York University, 13 March.
- Kaplan, Andreas, Michael Haenlein, 2019, 'Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence', *Business Horizons*, Volume 62, Issue 1, pp. 15-25.
- Leslie, David, 2019, 'Understanding Artificial Intelligence Ethics and Safety: A guide for the responsible design and implementation of AI systems in the public sector', Alan Turing Institute.
- Luciano, Floridi, and Cowls Josh, 2019, 'A Unified Framework of Five Principles for AI in Society', *Harvard Data Science Review*, 1 July, https://hdsr.mitpress.mit.edu/pub/l0jsh9d1.
- Paglioni, Vincent, 2015, 'The Ethics of Intelligent Machines', Investment Management Consultants Association, https://investmentsandwealth.org/getattachment/f3614756-1e1d-49c7-a201-29dbc22d8fbf/IWM15NovDec-EthicsIntelligentMachines.pdf
- Winfield, Alan F., Katina Michael, Jeremy Pitt, Vanessa Evers, 2019, 'Machine Ethics: the Design and Governance of Ethical AI and Autonomous Systems', *Proceeding of IEEE*, Volume 107, Issue 3, pp. 509-517.

- Use RASA (https://rasa.com/) to build a chatbot, or OpenAI's GPT-2 or GPT-3 to generate text.

Week 7 - Privacy

Jonathan A. Obar, Assistant Professor, Department of Communication Studies, York University, will be invited to join the discussion briefly this week.

Ethical

Core:

- Hyunghoon Cho, Daphne Ippolito, Yun William Yu, 2020, 'Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs', *arXiv*, https://arxiv.org/abs/2003.11511.
- Obar, Jonathan A. and Oeldorf-Hirsch, Anne, 2018, 'The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services' *TPRC 44: The 44th Research Conference on Communication, Information and Internet Policy*, http://dx.doi.org/10.2139/ssrn.2757465.

Additional:

- Blumberg, Andrew J. and Peter Eskersley, 2009, 'On Locational Privacy, and How to Avoid Losing it Forever', https://www.eff.org/wp/locational-privacy.
- de Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel, 2013, 'Unique in the Crowd: The privacy bounds of human mobility', *Scientific Reports*, vol 3, https://doi.org/10.1038/srep01376.
- Obar, Jonathan A., and Anne Oeldorf-Hirsch, 2018, 'The clickwrap: A political economic mechanism for manufacturing consent on social media', *Social Media+ Society*, 4.3, 2056305118784770
- Solove, Daniel J, 2007, '"I've Got Nothing to Hide" and Other Misunderstandings of Privacy', *San Diego Law Review*, Vol. 44, p. 745-772.
- Zimmeck, Sebastian, Story, Peter, Smullen, Daniel, Ravichander, Abhilasha, Wang, Ziqi, Reidenberg, Joel, Cameron Russell, N., & Sadeh, Norman, 2019, 'MAPS: Scaling Privacy Compliance Analysis to a Million Apps', *Proceedings on Privacy Enhancing Technologies*, Volume 3, pp. 66-86.
- Zimmer, Michael, Priya Kumar, Jessica Vitak, Yuting Liao and Katie Chamberlain Kritikos, 2018, '"There's nothing really they can do with this information': unpacking how users manage privacy boundaries for personal fitness information", *Information, Communication & Society*, Vol 23, Issue 7, pp. 1020-1037.

Technical

- Find or generate a dataset, then implement differential privacy on it. Examine and discuss the results.
  - Oberski, Daniel, and Frauke Kreuter, 2020, 'Differential Privacy and Social Science: An Urgent Puzzle', Harvard Data Science Review, https://doi.org/10.1162/99608f92.63a22079.
  - Rubinstein, Benjamin I. P. and Francesco Alda, 2017, 'diffpriv: An R Package for Easy Differential Privacy', *Journal of Machine Learning Research*, 18, pp. 1-5.

Week 8 - Images/video with particular reference to facial recognition

Jeffrey Knockel, Research Associate, Citizen Lab, University of Toronto, will be invited to join the discussion briefly this week.

Ethical

Core:
- Buolamwini, Joy, Vicente Ordóñez, Jamie Morgenstern, and Learned-Miller, Erik, 2020, 'Facial recognition technologies: A primer', *Algorithmic Justice League*, 29 May.
- Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton, 2020, 'Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing', *In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20)*. Association for Computing Machinery, New York, NY, USA, 145–151. DOI:https://doi.org/10.1145/3375627.3375820.

Additional:
- Hill, Kashmir, 2020, 'Wrongfully Accused by an Algorithm', *New York Times*, 24 June, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.
- Hill, Kashmir, 2020, 'The Secretive Company That Might End Privacy as We Know It', *New York Times*, 18 January, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html.
- Knockel, Jeffrey, and Ruohan Xiong, 2019, '(Can't) Picture This 2: An Analysis of WeChat's Realtime Image Filtering in Chats', *Citizen Lab*, 15 July,

https://citizenlab.ca/2019/07/cant-picture-this-2-an-analysis-of-wechats-realtime-image-filtering-in-chats/.

- Learned-Miller, Erik, Vicente Ordóñez, Jamie Morgenstern, and Joy Buolamwini, 2020, 'Facial recognition technologies in the wild: A call for a federal office', *Algorithmic Justice League*, 29 May,

Technical

- Chollet, Francois, and J. J. Allaire, 2018, *Deep Learning with R*, Chapter 5 'Deep learning for computer vision'.

Week 9 - Corporate Surveillance

Ethical

Core:

- Zuboff, Shoshana, 2019, *The Age of Surveillance Capitalism*, and watch related interview: https://www.youtube.com/watch?v=hIXhnWUmMvw
- Zuboff, Shoshana, 2019, 'Written Testimony Submitted to The International Grand Committee on Big Data, Privacy, and Democracy', 28 May, Ottawa, https://www.ourcommons.ca/Content/Committee/421/ETHI/Brief/BR1057373725/br-external/ZuboffShoshana-e.pdf and watch related video https://youtu.be/6N2kJNwGgUg?t=4869.

Additional:

- Bennett Cyphers and Gennie Gebhart, "Behind One-Way Mirror: A Deep Dive Into the Technology of Corporate Surveillance", https://www.eff.org/files/2019/12/11/behind_the_one-way_mirror-a_deep_dive_into_the_technology_of_corporate_surveillance.pdf
- Marczak, Bill and John Scott-Railton, 2020, 'Move Fast and Roll Your Own Crypto: A Quick Look at the Confidentiality of Zoom Meetings', *Citizen Lab*, 3 April, https://citizenlab.ca/2020/04/move-fast-roll-your-own-crypto-a-quick-look-at-the-confidentiality-of-zoom-meetings/.
- Parsons, Christopher, Andrew Hilts, and Masashi Crete-Nishihata, 2017, 'Approaching Access: A comparative analysis of company responses to data access requests in Canada', *Citizen Lab*, Research Brief No. 106. Available at: https://citizenlab.ca/wp-content/uploads/2018/02/approaching_access.pdf.
- (Optional/fun) Duhigg, Charles, 2012, 'How Companies Learn Your Secrets', *New York Times*, 19 February, https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

- Create a datasheet or model card for an open source dataset or model.
    - Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford, 2018, 'Datasheets for Datasets', *arXiv*, https://arxiv.org/abs/1803.09010
    - Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji and Timnit Gebru, 2019, 'Model Cards for Model Reporting', *FAT '19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 220–229 https://doi.org/10.1145/3287560.3287596.

Week 10 - Privacy and surveillance in Canada and other countries

Lisa Austin, Professor, Law, University of Toronto, will be invited to join the discussion briefly this week.

Ethical

Core:

- Khoo, Cynthia, Kate Robertson, and Ronald Deibert, 2019, 'Installing Fear: A Canadian Legal and Policy Analysis of Using, Developing, and Selling Smartphone Spyware and Stalkerware Applications,' *Citizen Lab*, Research Report No. 120, University of Toronto, June, https://tspace.library.utoronto.ca/bitstream/1807/96321/1/stalkerware-legal.pdf.
- Obar, Jonathan A., 2017, 'Keeping Internet Users in the Know or in the Dark? The Data Privacy Transparency of Canadian Internet Carriers: A Third Report', *IXMaps*, https://ixmaps.ca/docs/DataPrivacyTransparencyCanadianCarriers-2017.pdf
- Ruan, Lotus, Crete-Nishihata, Masashi, Knockel, Jeffrey, Xiong, Ruohan and Dalek, Jakub, 2020, 'The Intermingling of State and Private Companies: Analysing Censorship of the 19th National Communist Party Congress on WeChat,' *The China Quarterly*, pp. 1–30. doi: 10.1017/S0305741020000491.

Additional:

- Austin, Lisa, and David Lie, 2019, 'Safe Sharing Sites', New York University Law Review, Vol. 94, No. 4, pp. 581 - 623.
- Knockel, Jeffrey, Christopher Parsons, Lotus Ruan, Ruohan Xiong, Jedidiah Crandall, and Ron Deibert, 2020, 'We Chat, They Watch: How International Users Unwittingly Build up WeChat's Chinese Censorship Apparatus,' *Citizen Lab*, Research Report No. 127, University of Toronto, May,

https://tspace.library.utoronto.ca/bitstream/1807/101395/1/Report%23127--wechattheywatch-web.pdf.

- Obar, Jonathan A., and Brenda McPhail, 2018, 'Preventing Big Data Discrimination in Canada: Addressing design, consent and sovereignty challenges', Centre for International Governance Innovation (CIGI), https://www.cigionline.org/articles/preventing-big-data-discrimination-canada-addressing-design-consent-and-sovereignty.
- Parsons, Christopher, Adam Molnar, Jakub Dalek, Jeffrey Knockel, Miles Kenyon, Bennett Haselton, Cynthia Khoo, and Ron Deibert, 2019, 'The Predator in Your Pocket: A Multidisciplinary Assessment of the Stalkerware Application Industry,' *Citizen Lab*, Research Report, No. 119, University of Toronto, June, https://tspace.library.utoronto.ca/bitstream/1807/96320/1/stalkerware-holistic.pdf.
- Scott, James C., 1998, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed.*
- Various, 'GDPR Checklist', https://gdpr.eu/checklist/.
- Various, 'Summary of privacy laws in Canada', Office of the Privacy Commissioner, https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/02_05_d_15/.

Technical

- TBD based on student interest.

Week 11 - Algorithmic decision-making

Ethical

Core:
- Spiegelhalter, David, 2020, 'Should We Trust Algorithms?', *Harvard Data Science Review*, 31 January, https://doi.org/10.1162/99608f92.cb91a35a.
- Molnar, Petra and Lex Gill, 2018, 'Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System,' *Citizen Lab and International Human Rights Program*, Faculty of Law, University of Toronto, Research Report No. 114, University of Toronto, September, https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf.

Additional:
- De-Arteaga, Maria, Riccardo Fogliato, and Alexandra Chouldechova, 'A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores'. https://arxiv.org/abs/2002.08035.
- Mitchell, Shira, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum, 2018, 'Prediction-Based Decisions and Fairness: A Catalogue of Choices,

Assumptions, and Definitions', *arXiv*, 1811.07867.
https://arxiv.org/abs/1811.07867.

- Rudin, Cynthia, Caroline Wang, and Beau Coker, 'The Age of Secrecy and Unfairness in Recidivism Prediction', *Harvard Data Science Review*, https://hdsr.mitpress.mit.edu/pub/7z10o269.
- Suresh, Harini, Natalie Lao, and Ilaria Liccardi, 'Misplaced Trust: Measuring the Interference of Machine Learning in Human Decision-Making', https://arxiv.org/pdf/2005.10960.pdf

Technical

- McElreath says that researchers use point estimates to describe posterior distributions, not to support particular decisions. But this isn't always viable. Using a post from the Stan Case Study (https://mc-stan.org/users/documentation/case-studies.html) as a guide, please develop a Bayesian hierarchical model in Stan. Please post-process your model to support/recommend a decision, and justify your choices.

Week 12 - History of ethical concerns broadly, and domain-specific ethical practices

Ethical

(Please pick two.)
- Medicine:
    - Parker, Michael, J A Muir Gray, 2001, 'What is the role of clinical ethics support in the era of e-medicine?', *Journal of Medical Ethics*, 27 suppl I:i33–i35
    https://jme.bmj.com/content/medethics/27/suppl_1/i33.full.pdf
    - Chancellor, S., Baumer, E. P., & De Choudhury, M. (2019). Who is the" Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1-32.
    https://doi.org/10.1145/3359249
    - Vayena, Effy, and Alessandro Blasimme, 2020, 'The Ethics of AI in Biomedical Research, Medicine and Public Health', *The Oxford Handbook of Ethics of AI*, Chapter 37, Oxford University Press.
- Engineering
    - Davis, Michael, 1991, 'Thinking Like an Engineer: The Place of a Code of Ethics in the Practice of a Profession',
    https://www.jstor.org/stable/pdf/2265293.pdf?refreqid=excelsior%3A9 4aaba1458bc97cf0563cf7d16861188

- ○ Michaelson, Christopher, 2014, 'The Competition for the Tallest Skyscraper: Implications for Global Ethics and Economics', *CTBUH Journal*, Issue IV, [https://www.jstor.org/stable/pdf/24192831.pdf?ab_segments=0%252F basic_SYC-5187%252Ftest&refreqid=excelsior%3A9bf439c8785e93d0 09d8e42608e6b425](https://www.jstor.org/stable/pdf/24192831.pdf?ab_segments=0%252Fbasic_SYC-5187%252Ftest&refreqid=excelsior%3A9bf439c8785e93d009d8e42608e6b425)
  - ○ Millar, Jason, 2020, 'Engineering', *The Oxford Handbook of Ethics of AI*, Chapter 23, Oxford University Press.
  - ○
- Statistics
  - ○ Wells, Martin, 2020, 'Statistics', *The Oxford Handbook of Ethics of AI*, Chapter 26, Oxford University Press.
- Law
  - ○ Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner, 2016, 'Machine Bias', *ProPublica*, [https://www.propublica.org/article/machine-bias-risk-assessments-in-c riminal-sentencing](https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing)
  - ○ Eubanks, Virginia, 2014, 'How Big Data Could Undo Our Civil-Rights Law', https://prospect.org/justice/big-data-undo-civil-rights-laws/
  - ○ Surden, Harry, 2020, 'Law: Basic Questions', *The Oxford Handbook of Ethics of AI*, Chapter 38, Oxford University Press.
- Finances
  - ○ Geslevich Packin, Nizan, Yafit Lev Aretz, 2015, 'Big Data and Social Netbanks: Are You Ready to Replace Your Bank?', https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2567135
- Education
  - ○ Mayfield, E., Madaio, M., Prabhumoye, S., Gerritsen, D., McLaughlin, B., Dixon-Román, E., & Black, A. W. (2019, August). Equity beyond bias in language technologies for education. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 444-460). https://doi.org/10.1177/2053951720913064
  - ○ Rubel, A., & Jones, K. M. (2016). Student privacy in learning analytics: An information ethics perspective. *The information society*, *32*(2), 143-159. https://doi.org/10.1080/01972243.2016.1130502
  - ○ Zeide, Elana, 2020, 'Education', *The Oxford Handbook of Ethics of AI*, Chapter 42, Oxford University Press.
- General non-computational
  - ○ Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59-68). https://dl.acm.org/doi/pdf/10.1145/3287560.3287598
- Earlier calls for ethics in computing

- ○ Agre, Philip E., 1997, 'Towards a critical technical practice: Lessons learned from trying to reform AI', *Social science, technical systems, and cooperative work: Beyond the great divide*, Ed. by Geoffrey C. Bowker, Susan Leigh Star, Will Turner, and Les Gasser. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 131–158. URL: https://web.archive.org/web/20040203070641/http://polaris.gseis.ucla.edu/pagre/critical.html.
- ○ Friedman, Batya, and Helen Nissenbaum, 1996, 'Bias in computer systems', *ACM Trans. Inf. Syst,* 14, 3 (July 1996), 330–347. DOI: https://doi.org/10.1145/230538.230561.

Technical

- TBD based on student interest.

Assessment

Four ethical and technical blog posts (30 per cent)

Over the course of the term, you are expected to submit four blog posts that each comprise two aspects: 1) ethical and 2) technical. These two aspects should be related to each other. Only your best three blog posts will count, that is each blog post will account for 10 per cent of your overall mark.

For the first aspect (ethical), you are expected to write a moderate length discussion (think a paper of about two to three pages), of a reading, or set of readings, that we have covered over the past two weeks. Strong submissions will not limit themselves to reviewing a reading but will draw in larger issues and detail their own point of view.

For the second aspect (technical), you are expected to implement some small related technical aspect of what we have covered in the past two weeks. For instance, if we covered natural language processing then you may critically review a paper, and put together a chat bot.

To be clear, these two aspects should be related, tied together, and should be in the one blog post.

You should submit your blog post by emailing me (rohan.alexander@utoronto.ca) a link to the relevant blog post on your website.

The proposed specific list of deadlines is:
- Blog post 1: midnight, Sunday 24 January, 2021.

- Blog post 2: midnight, Sunday 7 February, 2021.
- Blog post 3: midnight, Sunday 7 March, 2021.
- Blog post 4: midnight, Sunday 21 March, 2021.

In Week 1 we will discuss how these dates fit in with your other commitments and finalise them at that point. The instructor reserves the right to alter these due dates, if circumstances warrant it.

The instructor will make the marking guide available at least a week before the submission deadline.

Paper 1 (30 per cent)

Task

Please gather and clean data on UofT salaries from the Sunshine List. Then conduct a Bayesian statistical analysis of your dataset to discuss the extent to which gender has an effect on salary. Finally please prepare a paper of around 10 pages that discusses your analysis. (Hint: gender is not explicitly part of the Sunshine List, you will need to grapple with what to do.)

Background

You should make appropriate use of appendices for additional and supporting material, and thoroughly reference your paper, but neither the appendices nor the reference list count toward your page limit. Your paper should have an appropriate title, author, date, abstract, and introduction. It should document and overview your dataset. It should clearly specify your model, and then discuss the results of your analysis and any weaknesses. Your analysis should be fully reproducible, with code and data hosted on a public GitHub repo. Additionally, you should include a thorough discussion of ethical considerations relevant to your analysis. This would likely take at least three pages, but you are welcome to write as much as is needed to make the points you would like to make. Likely the best way to do this is to include a brief overview of the ethical points that you would like to discuss, and then include the rest of the discussion in an appendix. I understand that Bayesian analysis may be new to you. I will assist you with putting together the model, but it is up to you to understand and interpret the output.

Submission

To submit your paper you should email me (rohan.alexander@utoronto.ca) a link to a public GitHub repo. That repo should contain your paper in PDF format and all supporting code and data. **Please send this email by midnight, Sunday, 14 February,**

2021. Please do not make any changes to the repo after this. I will make the marking guide available at least a week before the submission deadline.

Paper 2 (40 per cent)

Task

In consultation with me, please identify an appropriate research question and data source that, like the requirement for Paper 1, combines both ethical and technical aspects. Please prepare a paper that represents your best attempt to answer this question and shows off your ability to engage in thoughtful, ethical, critique. The paper should be as long as necessary, although all extraneous material should be included in appendices. The expectation is that this paper should make an original contribution, that could be published in an academic journal.

Background

Please see the background provided for Paper 1, as this applies for Paper 2 as well. You must send the email with the GitHub link to me by **midnight, Sunday, 23 April, 2020**. Please do not make any changes to the repo after this. I will make the marking guide available at least a week before the submission deadline. **No extensions are possible because of deadlines for instructors to submit grades.**

Miscellaneous

Accommodations

Students with diverse learning styles and needs are welcome in this course. If you have a disability or a health consideration that may require accommodations, please feel free to approach Student Services and/or the Accessibility Services Office http://www.studentlife.utoronto.ca/as as soon as possible. The Accessibility Services staff are available by appointment to assess needs, provide referrals and arrange appropriate accommodations. The sooner you let us know your needs, the quicker we can assist you in achieving your learning goals in this course.

Children in the classroom

We are in the middle of a pandemic. It is unlikely that we will meet in person. But if we do, babies (bottle-feeding, nursing, etc) are welcome in class as often as necessary. You are welcome to take breaks to feed your infant or express milk as needed, either in the classroom or elsewhere including:
https://familycare.utoronto.ca/childcare/breastfeeding-at-u-of-t/. A list of baby

change stations is also available:
https://familycare.utoronto.ca/childcare/baby-change-stations-at-u-of-t/. For older
children, I understand that unexpected disruptions in childcare can happen. You are
welcome to bring your child to class in order to cover unforeseeable gaps in childcare.

Unforeseen events

We are in the middle of a pandemic. The weighting of assessment could change slightly
(+/-5 percentage points) in response to changed circumstances. If there is a need for
more significant changes, then following university rules, students will be given at least
one week's opportunity to comment on the change, and vote, if it is at all reasonably
possible.

Grading

Please consult:
- the Faculty of Information's Grade Interpretation Guidelines:
  https://ischool.utoronto.ca/wp-content/uploads/2016/11/grade-interpretation
  .pdf
- The University Assessment and Grading Practices Policy:
  http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+
  Assets/Policies/PDF/grading.pdf
- The Guidelines on the Use of INC, SDF, & WDR:
  http://www.sgs.utoronto.ca/facultyandstaff/Pages/INC-SDF-WDR.aspx

These documents will form the basis for grading in the course.

Writing support

As stated in the Faculty of Information's Grade Interpretation Guidelines, 'work that is
not well written and grammatically correct will not generally be considered eligible for
a grade in the A range, regardless of its quality in other respects.' With this in mind,
please make use of the writing support provided to graduate students by the SGS
Graduate Centre for Academic Communication. The services are designed to target
the needs of both native and non-native speakers and all programs are free. Please
consult the current workshop schedule
http://www.sgs.utoronto.ca/currentstudents/Pages/Current-Years-Courses.aspx for
more information.

Academic integrity

Please consult the University's site on Academic Integrity
http://academicintegrity.utoronto.ca/. The Faculty of Information has a zero-tolerance
policy on plagiarism as defined in section B.I.1.(d) of the University's Code of Behaviour

on Academic Matters
http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/ppjun011995.pdf. You should acquaint yourself with the Code. Please review the material in Cite it Right and if you require further clarification, consult the site How Not to Plagiarize
http://advice.writing.utoronto.ca/wp-content/uploads/sites/2/how-not-to-plagiarize.pdf.

Cite it Right covers relevant parts of the U of T Code of Behaviour on Academic Matters (1995). It is expected that all Faculty of Information students take the Cite it Right workshop and the online quiz. Completion of the online Cite it Right quiz should be made prior to the second week of classes as the workshop is now interactive. To review and complete the workshop, visit the Orientation Workshop portion of the Inforum site: https://inforum.library.utoronto.ca/workshops/orientation.