Abuses, Bias and Blessings of Data

Theoretical and practical look at the contemporary issues of using data driven and intelligent systems in challenging contexts.

The Course Structure

Lectures

Theory & Applications in context

Research Report (50%)

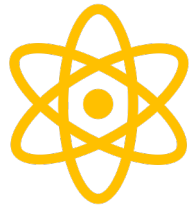
10% Report Plan

40 Final Report

Debate Seminars

Seminars

Debates (25%)



Computing Lab Sessions

Practical experience

Labs (25%)

The Course Focus and Content

Topics

Introduction and Setting the Scene	• Data, representation learning,
Bias in Algorithms / Models	• Reducing, Detecting, Measuring, Mitigating.
Explaining Blackbox Algorithms	• Algorithms, use case, limitations, insights.
Fair ML	• Concepts, Metrics, principles, frameworks.
Human-ML interaction	• Trust, Autonomy, Efficacy, Responsibility

Skills

- Debating
- Implementation & Practical Experience
- Critical Evaluation

Practical Labs

CSCM21 – Assignment 25%

Practical Lab Rubric

	%	Levels of Performance				
Criteria		Fail (0 -39)	Fail (small margin) (40-49)	Pass (50 -59)	Merit (60 -69)	Distinction (70-100)
1. Code Completion: Coding task completion to support the learning objectives and exploration of the core concepts.	30	Some code completed but one or more full sections of the workbook not completed.	Most code completed some small tasks incomplete.	All coding tasks completed satisfactorily.	All coding tasks completed including evidence of additional exploration of relevant concepts.	All coding tasks completed extensive evidence of additional exploration of almost all relevant concepts.
2. Code Review Evidence of practical coding skills and understanding	20	Little evidence of understanding the code.	Some evidence of understanding the code implemented in the lab.	Satisfactory evidence of understanding the code implemented in the lab	Evidence of understanding all aspects of code in the lab notebook including that which was not implemented in the lab. Can include some minor misunderstandings.	Evidence of understanding all aspects of code in the lab notebook including that which was not implemented in the lab.
3. Discussion: Depth of understanding of core concepts covered in the lab.	50	Little to no evidence of concept or method understanding gained throughout the lab. Factual explanation on implementation only.	A few relevant insights made regarding concepts and methods. Some gaps in concept understanding or terminology use.	Discussion contributions display an understanding of all the underlying concepts including correct use of terminology.	Discussion contributions display an understanding of the underlying concepts including correct use of terminology. Supported by empirical evidence generated through investigations of the methods used in the lab.	Discussion contributions display a deep understanding of the underlying concepts including correct use of terminology. Supported by large amount of varied empirical evidence generated through investigations of the methods used in the lab. Includes reasoned critical analysis of the methods used. E.g. good practical use case and/or limitations.

Debate Seminars

CSCM21 – Assignment 25%

Debate Seminar Rubric

	%	Levels of Performance				
Criteria		0- 39 (fail)	40-49 (fail by small margin)	50 – 59 (Pass)	60 -69 (Merit)	70 – 100 (Distinction)
Individual Scores 75%						
1. Organization and Clarity: viewpoints and responses are outlined both clearly and orderly.	20	Entirely unclear in most parts, total lack of augment being tied to an idea (reasoned premise)	Unclear in most parts, lack of many augments being tied to an idea (reasoned premise)	Clear in some parts but not overall, some arguments were tied to an idea (reasoned premise)	Mostly clear and orderly presentation and many arguments were tied to an idea (reasoned premise)	Completely clear and orderly presentation and all arguments were tied to an idea (reasoned premise) in a tight logical fashion
2. Presentation Style: tone of voice, use of gestures, and level of enthusiasm are convincing and respectful to the audience.	20	No style features were used. Statements, responses and/or body language was not consistently respectful	Few style features were used; not convincingly. Statements, responses and/or body language was not consistently respectful	Few style features were used but they were used convincingly. Respectful at all times.	All style features were used, most convincingly. Respectful at all times.	All style features were used convincingly. Enthusiasm empathy connection with audience using anecdotes and metaphors. Respectful at all times.
3. Depth of Topic Knowledge Breadth and depth of topic understanding in debate	20	No evidence that they learned / researched the debate topic. No variance in topic knowledge presented.	Little evidence that they learned / researched the debate topic. Little variance in topic knowledge presented.	Varied argument presented based on some evidence of understanding / research of the topic.	Sufficient evidence of knowledge demonstrated by augments and supporting from understanding / research on the topic.	Deep comprehensive understanding of topic evident in many varied arguments, evidence and response.
4. Use of Arguments (rebuttals): Major points are given to support viewpoint with supporting logical reasoning adopting - induction – deduction – equivalency. (responded directly to other team arguments and dealt with effectively)	20	No relevant reasons given. Arguments and their points were irrelevant and/or unclear.	Few relevant reasons given. Arguments and their points were largely irrelevant and/or unclear.	Some relevant reasons given Few convincing arguments and most points were irrelevant and/or unclear	Many reasons given: several convincing arguments and most points were on topic and clear.	Many convincing arguments and all points were on topic and very clear.
5. Use of Examples and Facts: examples and facts are given to support reasoning, with references	20	No relevant supporting examples/facts and no sources	Some relevant supporting examples/facts	Major points adequately supported by examples / facts given with sources	Major points well supported with examples and/or facts given with robust sources	Major points well supported with examples and/or relevant facts / statistics given, with robust and scientific sources
Team Score (25%):						
1. Depth of Topic Knowledge Breadth and depth of topic understanding in debate notes	50	No evidence that they learned / researched the debate topic. No variance in topic knowledge presented.	Little evidence that they learned / researched the debate topic. Little variance in topic knowledge presented.	Varied argument presented based on some evidence of understanding / research of the topic.	Sufficient evidence of knowledge demonstrated by augments and supporting from understanding / research on the topic.	Deep comprehensive understanding of topic evident in many varied arguments, evidence and response.
2. Score from Audience Rating	50		Dissuaded	Neutral	Informed	Persuaded

CSCM21 –
Assignment 50%

Research Report

Deliverables

Part 1: Definition & Scope

Title & Study Type

- Short statement (Hypothesis or Conjecture) on what is the subject of the report.
- **(Main section)**
  - In this section try to highlight the **precise focus** of the study – be specific rather than general and abstract with your language.

Name and briefly outline the experimental method(s) that will be used to carry out the study. Include:

- any data to be used – emphasises on scope size of data, number of data sets etc.
- Concepts to be discussed - emphasises on scope number of concepts.
- Assessed 10%

Part 2:Full Research Report

Part 2: Full Research Report (In the Style of a Paper)

- Assessed 40%

Topic of Study - Self Identified

- Bias in Algorithms / Models
  - Reducing, Detecting, Measuring, Mitigating.
- Explaining Blackbox Algorithms
  - Algorithms, use case, limitations, insights.
- Fair ML
  - Concepts, Metrics, principles, frameworks.

Styles of Study

CRITICAL CONCEPTS
DISCUSSION

EMPIRICAL
EXPERIMENTAL STUDY

LARGE SCALE RESEARCH
STUDY DESIGN

Critical Concepts Discussion

As study that critically apprises current state-of-the-art approaches to: fair ML, Bias or Explainability through a multidisciplinary perspective(s)

- Purpose of the Study
  - Hypothesis / Conjecture
  - Motivated by the debate seminars
- Background & Literature Review
  - Clarity of concepts: definitions and their origin (social science, legal studies)
- Study Design
  - Structured Reviews
  - Thematic Analysis
- Discussion & Further Work
  - Critical discussion on current state-of-the-art philosophical and empirical use of the concepts in formalized ML / data driven approaches
  - Further Work

Critical Concepts Discussion Inspiration

- [1] Google, “Perspectives on Issues in AI Governance,” 2019.
- [2] EUROPEAN COMMISSION, “High-Level Expert Group on Artificial Intelligence,” pp. 2– 36, 2019.
- [3] S. Amershi *et al.*, “Guidelines for human-AI interaction,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2019.
- [4] M. J. Kusner and J. R. Loftus, “fairer algorithms.”
- [5] R. Binns, “On the apparent conflict between individual and group fairness,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 514–524.
- [6] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, D. Pedreschi, and F. Giannotti, “A Survey Of Methods For Explaining Black Box Models,” *ACM Comput. Surv.*, Feb. 2018.
- [7] C. Rudin, “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead.”

Empirical Experimental Study

Must include the application of method(s) to data set(s)

- Purpose of the Study
  - What would you like to investigate?
  - Empirical exploration of a method or methods to address a challenge.
  - Motivated by the lab(s) – Extension and write-up
- Background & Literature Review
  - Origins of methods used and explored in the study.
  - Applications and limitations.
- Study Design
  - Explanation of experimental design and approach to explore the purpose of the study.
  - Empirical Results Presented
- Discussion & Further Work
  - Critical analysis of the results
  - Conclusions & limitations
  - Further study proposals

Empirical Experimental Study Inspiration

- [8] T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, and A. Kalai, “Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings,” Jul. 2016.
- [9] J. Krause, A. Perer, and K. Ng, “Interacting with Predictions,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI ’16*, 2016, pp. 5686–5697.
- [10] V. Petsiuk, A. Das, and K. Saenko, “RisE: Randomized input sampling for explanation of black-box models,” *Br. Mach. Vis. Conf. 2018, BMVC 2018*, vol. 1, 2019.
- [11] M. Ngan and P. Grother, “Face Recognition Vendor Test (FRVT) - Performance of Automated Gender Classification Algorithms,” 2015.
- [12] J. Buolamwini, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification *,” 2018.
- [13] E. Black, S. Yeom, and M. Fredrikson, “FlipTest,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 111–121.
- [14] C. Sweeney and M. Najafian, “Reducing Sentiment Polarity for Demographic Attributes in Word Embeddings using Adversarial Learning,” 2020.

Large Scale Research Study Design

A study that would need large resources to carry out e.g. 100s or 1000s of participants o r one with a critical Longitudinal component.

- Purpose of the Study
  - Hypothesis
  - Motivated by real world problem that has not be invested in the literature to date.
- Background & Literature Review
  - Contextual Setting Explanation
  - Importance, prevalence or significance of study
  - Survey of similar studies
- Study Design
  - Explanation of experimental design & method
  - Technical architecture
  - Participant journey, expectation, engagement,
- Discussion & Further Work
  - Most Critical risks difficulties with study
  - Limitations of conclusions
  - Potential challenges to study

Large Scale Research Study Design Inspiration

- [15] M. Katell *et al.*, “Toward situated interventions for algorithmic equity,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 45– 55.
- [16] A. Datta, M. C. Tschantz, and A. Datta, “Automated Experiments on Ad Privacy Settings,” *Proc. Priv. Enhancing Technol.*, vol. 2015, no. 1, pp. 92–112, 2015.
- [17] M. Sendak *et al.*, “‘The human body is a black box,’” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 99–109.
- [18] A. Lucic, H. Haned, and M. de Rijke, “Why does my model fail?,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 90–98.
- [19] A. Coston, A. Mishler, E. H. Kennedy, and A. Chouldechova, “Counterfactual risk assessments, evaluation, and fairness,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 582–593.
- [20] U. Bhatt *et al.*, “Explainable Machine Learning in Deployment,” 2020.

Reading and Disscussion

- Importance
- **Rigor**
- Contribution

- Purpose of the Study
- Background & Literature Review
- Study Design
- Discussion & Further Work