EE 8227: Secure Machine Learning

Instructor Information

Name: Dr. Reza Samavi
Office location: ENG457
Office hours: by appointment – please email.
E-mail: samavi@torontomu.ca
Website: https://www.ee.torontomu.ca/~samavi

Email Policy

Email is the main way to contact the instructor. In accordance with the Policy on TMU Student E-mail Accounts (Policy 157), TMU requires that any electronic communication by students to TMU's faculty or staff be sent from their official TMU email account. All students in full and part-time graduate degree programs are required to activate and maintain their TMU online identity in order to regularly access TMU's e-mail, RAMSS, and my.torontomu.ca portal.

Course Description

With the extensive proliferation of machine learning models (ML), specifically Generative models (GM), Large Language models (LLM) and ML for safety-critical systems, it is long overdue to study the security and trustworthiness of these models. To address this gap, this course focuses on the theories and techniques for enhancing the security and robustness of machine learning algorithms. Modern machine learning and deep learning models are shown to be vulnerable to a slight perturbation of input queries or training datasets. A number of machine learning algorithms can also memorize and expose private information about individuals. Disclosure of sensitive data not only leads to privacy breaches but also could result in discrimination or issues of fairness. This course is designed to fill this gap and specifically covers the following topics: privacy-preserving statistics and machine learning; adversarial machine learning; certified robustness; poisoning attacks and countermeasures; accountability, transparency and interpretability in machine learning, federated learning to support privacy; and considerations for trustworthy machine learning.

Course Details

Prerequisites / co-requisites

The course is open to interested engineering and science graduate students with a solid undergraduate-level mathematical background. Undergraduate-level knowledge of probability, statistics, algorithms and data structures, and machine learning is assumed.

Course Website

https://courses.torontomu.ca/d2l/home/917957

Course objectives and intended learning outcomes

This course provides a platform for students to strengthen their knowledge at the intersection of security and machine learning. At the end of this course, students will be able to:

1. Analyze and formulate machine learning attack surface, adversarial capabilities and goals, and develop algorithms to counteract the attacks.
2. Analyze, design, and develop algorithms to ensure privacy in data science and machine learning.
3. Explore new research directions in the interplay between security, privacy, and robustness in machine learning models.

Texts and readings

No textbooks are required for this class. All relevant materials will be made available online on the course website on D2L. The materials mainly come from seminal and recent papers in the field including (but not limited to) Neurips, ICML, ICLR, CCS and USENIX.

Teaching Methods

1. Students' participation and interaction is a major component of this course. Therefore, lectures are accompanied by student presentations and participation.
2. Notes/slides from the class lectures will be posted on D2L.
3. If the university decides the courses to be delivered virtually, students are NOT required to turn ON their cameras during lectures. However, when a student presents a seminar topic, screen sharing and at minimum audio communication is required. The University has issued a minimum technology requirement for remote learning. Details can be found at: https://www.torontomu.ca/covid-19/students/ Please ensure you meet the minimum technology requirements as specified in this link.

Topics and Course Schedule*

Schedules and contents are tentative and subject to modifications before the semester starts or as we make our way through the course and based on the students' feedback.

General Topics	Week	Detailed Description
ML: attacks on Privacy	1	Course review, Introduction to Machine Learning (ML) and Security
	2	**Attacks on ML**: **Privacy**, Reidentification and reconstruction attacks – **Tutorial #1-ML**
	3	**Attacks on ML**: **Privacy**, membership attacks- **Tutorial #2-OpenDP**
Defense against privacy attacks	4	**Foundations of Differential Privacy (DP) 1**: Randomized response
	5	**Foundations of Differential Privacy (DP) 2**: Global Sensitivity and Laplace mechanism
	6	DP Composition, Group Privacy
	7	Gaussian mechanisms, DP and synthetic data generation
	8	**Real-world DP applications:** Healthcare, Social media, Financial- **Project progress**
Attacks on machine learning integrity and ML robustness	9	**Attacks on ML**: **Integrity** - Poisoning attacks and Adversarial examples – **Tutorial #3**
	10	**Machine learning robustness** – Adversarial training (Empirical methods)
	11	**Machine learning robustness** – Certified methods
	12	**ML Robustness and real-world applications:** Image processing, Robotics
Research Directions	13	Open research problems in Secure Machine Learning – **Final Project presentation**

* Note – Any changes and additions to this schedule will be communicated in class and posted on the D2L site. The hours for each lecture is approximated and also includes time for students' presentations.

Evaluation*

No.	Title	Value	Detail / general description
1	Participation & Presentation	25%	Students present an assigned paper or through the project presentations lead the class discussion. The presenters will be evaluated on the critical content of the subject matter and communication skills. For participation, which would not be more than 5% of the course grade, one or two easy questions will be asked, just to make sure you're following up the content of the course.
2	Project/ Research paper	35%	Students will complete an individual (or group) research project. The focus of the projects is on developing new model, theory, or algorithms in one of the subtopics of secure machine learning. The project can also be an implementation of known algorithms on a new application domain (a domain of interest to the student such as financial sector, robotics, healthcare). For high-quality projects the instructor will help the student to publish the work in one of the reputable machine learning conferences.
3	Assignments	40%	There will be two take-home assignments one during the semester and one at the end of the semester in lieu of the final exam.

* All evaluation items will be completed individually. For the project/research paper depending on the topic and the scope of the project, the instructor may approve the project to be completed by two students. Further information about project/research ideas and more details on the delivery of the exam (and access to graded items) will be provided by the instructor on the course website.

University Academic Policies and Additional Information

Students are reminded that they are required to adhere to all relevant university policies found in their online course shell in D2L and/or on the following URL: http://ryerson.ca/senate/course-outline-policies It is student's responsibility to familiarise themselves with all relevant University academic policies.

The most relevant policies

For information on academic policies pertaining to issues such as course management, grading practices, and appeals, students are to refer to the TMU Senate Policies: Policy 164 – Graduate Status, Enrolment, and Evaluation, Policy 166 – Course Management Policy , and Policy 152 – Graduate Student Academic Considerations and Appeals

Academic Accommodation Support
Students are required to immediately inform their instructors of any situation which arises during the semester, which may have an adverse effect upon their academic performance, and must request any considerations and accommodations according to the relevant policies and well in advance. Failure to do so will jeopardize any academic appeals.

Academic Accommodation Support (AAS) is the university's disability services office. AAS works directly with incoming and returning students looking for help with their academic accommodations. AAS works with any student who requires academic accommodation regardless of program or course load.

- Learn more about Academic Accommodation Support
- Learn how to register with AAS

Academic Accommodations (for students with disabilities) and Academic Consideration (for students faced with extenuating circumstances that can include short-term health issues) are governed by two different university policies. Learn more about Academic Accommodations versus Academic Consideration and how to access each.

Accessibility

- Please study this accessibility statement. The instructor will do every effort to improve the accessibility of this course.
- In case of remote teaching, any technologies used in this course and any known accessibility features or barriers (if applicable) will be communicated ahead of time with the students.
- The students should email the instructor as soon as they discover an accessibility barrier with any course materials or technologies.

Turnitin

- Turnitin.com is a plagiarism prevention and detection service to which TMU subscribes. It is a tool to assist instructors in determining the similarity between students' work and the work of other students who have submitted papers to the site (at any university), internet sources, and a wide range of books, journals and other publications. While it does not contain all possible sources, it gives instructors some assurance that students' work is their own. No decisions are made by the service; it generates an "originality report," which instructors will evaluate to judge if something is plagiarized.

- Students agree by taking this course that their written work will be subject to submission for textual similarity review to Turnitin.com. Instructors can opt to have student's papers included in the Turnitin.com database or not. Use of the Turnitin.com service is subject to the terms-of-use agreement posted on the Turnitin.com website. Students who do not want their work submitted to this plagiarism detection service must, by the end of the second week of class, consult with their instructor to make alternate arrangements.

- Even when an instructor has not indicated that a plagiarism detection service will be used, or when a student has opted out of the plagiarism detection service, if the instructor has reason to suspect that an individual piece of work has been plagiarized, the instructor is permitted to submit that work in a non-identifying way to any plagiarism detection service.

Academic Integrity and Plagiarism

- TMU's Policy 60 - Academic Integrity policy, applies to all students at the University. Forms of academic misconduct include plagiarism, cheating, supplying false information to the University, and other acts. The most common form of academic misconduct is plagiarism. Plagiarism is a serious academic offence and penalties can be severe. In any academic exercise, plagiarism occurs when one offers as one's own work the words, data, ideas, arguments, calculations, designs or productions of another without appropriate attribution or when one allows one's work to be copied.

- All academic work must be submitted using the citation style approved by the instructor. Students may refer to the TMU Library's list of Citations and Style Guides for more information.

- **It is assumed that all examinations and work submitted for evaluation and course credit will be the product of individual effort, except in the case of group projects arranged for and approved by the course instructor. Submitting the same work to more than one course, without instructor approval, is also considered a form of plagiarism.**

- Furthermore, the unauthorized use of the intellectual property of others, including your professor, for distribution, sale, or profit is expressly prohibited. Intellectual property includes, but is not limited to: slides, lecture notes, presentation materials used in and outside of class, lab manuals, course packs, exams, etc.

- Students are advised that suspicions of academic misconduct may be referred to the Academic Integrity Office (AIO). Graduate students who are found to have committed academic misconduct will have a Disciplinary Notation (DN) placed and remain on their academic record, **which will exclude them to be eligible for any scholarships and/or awards**. In addition, they could be assigned one or more of the penalties ranging from a **grade of "zero" (0) on the**

work, a grade of "F" in the course, to DA (Disciplinary action), DA-S (Disciplinary action with suspension), (DW) Disciplinary withdrawal, up to an expulsion or even revocation of a degree.

- For more detailed information on these issues, please refer to the full online text for the Ryerson Senate Policy 60: Academic Integrity. For more information on how to avoid academic misconduct situations, for clues and tips, visit the Academic Integrity website.

Important Resources Available at Toronto Metropolitan University

- **The Library** provides research workshops and individual assistance. If the University is open, there is a Research Help desk on the second floor of the library, or students can use the Library's virtual research help service at https://library.ryerson.ca/ask/ to speak with a librarian.

- **Student Life and Learning Support** offers group-based and individual help with writing, math, study skills, and transition support, as well as resources and checklists to support students as online learners.

- You can submit an **Academic Consideration Request** when an extenuating circumstance has occurred that has significantly impacted your ability to fulfill an academic requirement. You may always visit the Senate website and select the blue radial button on the top right hand side entitled: Academic Consideration Request (ACR) to submit this request).

  Please note that the Provost/ Vice President Academic and Dean's approved a COVID-19 statement for Fall 2022 related to academic consideration. This statement will be built into the Online Academic Consideration System and will also be on the Senate website (www.ryerson.ca/senate) in time for the Fall term:

  Policy 167: Academic Consideration for Fall 2022 due to COVID-19: Students who miss an assessment due to cold or flu-like symptoms, or due to self-isolation, are required to provide a health certificate. All absences must follow Senate Policy 167: Academic Consideration.

  Also NOTE: Policy 167: Academic Consideration does allow for a once per term academic consideration request without supporting documentation if the absence is less than 3 days in duration and is not for a final exam/final assessment. If the absence is more than 3 days in duration and/or is for a final exam/final assessment, documentation is required. For more information please see Senate Policy 167: Academic Consideration.
- TMU COVID-19 Information and Updates for Students summarizes the variety of resources available to students during the pandemic.

- TMU COVID-19 Vaccination Policy

- In we switch to remote teaching, familiarize yourself with the tools you will need to use for remote learning. The Remote Learning guide for students includes guides to completing quizzes or exams in D2L Brightspace, with or without Respondus LockDown Browser and Monitor, using D2L Brightspace, joining online meetings or lectures, and collaborating with the Google Suite.

- Information on Copyright for Faculty and students.

Wellbeing Support
At Toronto Metropolitan University (TMU), we recognize that things can come up throughout the term that may interfere with a student's ability to succeed in their coursework. These circumstances are outside of one's control and can have a serious impact on physical and mental well-being. Seeking help can be a challenge, especially in those times of crisis.
If you are experiencing a mental health crisis, please call 911 and go to the nearest hospital emergency room. You can also access these outside resources at anytime:

- Distress Line: 24/7 line for if you are in crisis, feeling suicidal or in need of emotional support (phone: 416–408–4357)
- Good2Talk: 24/7-hour line for postsecondary students (phone: 1-866-925-5454)
- Keep.meSAFE: 24/7 access to confidential support through counsellors via My SSP app or 1-844-451-9700

If non-crisis support is needed, you can access these campus resources:

- Centre for Student Development and Counselling: 416-979-5195 or email csdc@ryerson.ca
- Consent Comes First – Office of Sexual Violence Support and Education: 416-919-5000 ext: 553596 or email osvse@ryerson.ca
- Medical Centre: call (416) 979-5070 to book an appointment

We encourage all Toronto Metropolitan University community members to access available resources to ensure support is reachable. You can find more resources available through the Toronto Metropolitan University Mental Health and Wellbeing website.