



Detailed Syllabus

Course Name		
Trustworthy Machine Learning (0368-4075)		
Instructor		
Dr. Mahmood Sharif		
Semester		
Spring ("Bet")		
Prerequisites		
Some background in machine learning		
Grading		
Project (65%) Homework assignments (20%) Paper reviews (10%) Participation (5%)		
Schedule		
<i>Date</i>	<i>Lec. num.</i>	<i>Subject and content (assignments, readings, etc.)</i>
02/24	1	Introduction and general background
03/03	2	General background • Assignment 1 - release
03/10	3	Inference-time integrity - attacks (part 1) • Project - list release
03/24	4	Inference-time integrity - attacks (part 2) • Assignment 1 - submission • Project - forming teams and selecting projects • Paper reviews - review 1 submission
03/31	5	Inference-time integrity - defenses • Assignment 2 - release
04/07	6	Training-time integrity - attacks • Paper reviews - review 2 submission
04/11	7	Training-time integrity - defenses • Assignment 2 - submission
04/28	8	Data privacy • Assignment 3 - release • Project – intermediate report • Paper reviews - review 3 submission
05/12	9	Model privacy



Detailed Syllabus

05/19	10	Differential privacy <ul style="list-style-type: none"> • Assignment 3 – submission • Paper reviews - review 4 submission
05/26	11	Explainability <ul style="list-style-type: none"> • Assignment 4 - release
06/02	12	Fairness <ul style="list-style-type: none"> • Paper reviews - review 5 submission
06/09	13	Project - (almost) final presentation <ul style="list-style-type: none"> • Assignment 4 - submission
08/04	-	Project - final submission

Notes

- Late days: Each student will be allocated four grace credits. Each grace credit can be used at the student's convenience to extend an assignment-submission deadline by 24 hours.
- Students will work on the semester-long projects in pairs or (preferably) teams of three.
- Depending on the number of registered students, there may be a final exam. In such a case, the exam's weight in the final grade will be 20%, and the project's weight will be decreased to 45%.



סילבוס מפורט

שם הקורס		
למידת מכונה אמינה (0368-4075)		
מרצה		
ד"ר מחמוד שריף		
סמסטר		
ב'		
דרישות הקורס		
רקע בלמידת מכונה		
הרכב הציון הסופי		
פרויקט (65%) תרגילי בית (20%) קריאת מאמרים (10%) השתתפות (5%)		
מבנה הקורס		
נושא השיעור ותכני השיעור (מטלות, רשימת קריאה, משימות וכיו"ב)	מס' שיעור	תאריך שיעור
מבוא ורקע כללי	1	02/24
רקע כללי • תרגיל בית 1 - פרסום	2	03/03
תקיפות בשלב ההיסק (חלק 1) • פרויקט - פרסום רשימת פרויקטים	3	03/10
תקיפות בשלב ההיסק (חלק 2) • תרגיל בית 1 - הגשה • פרויקט - חלוקה לקבוצות ובחירת פרויקט • קריאת מאמרים - הגשת דו"ח 1	4	03/24
הגנות בשלב ההיסק • תרגיל בית 2 - פרסום	5	03/31
תקיפות בשלב האימון • קריאת מאמרים - הגשת דו"ח 2	6	04/07
הגנות בשלב האימון • תרגיל בית 2 - הגשה	7	04/11
פרטיות הנתונים • תרגיל בית 3 - פרסום • פרויקט - הגשת דו"ח ביניים • קריאת מאמרים - הגשת דו"ח 3	8	04/28
פרטיות המודל	9	05/12



סילבוס מפורט

פרטיות דיפרנציאלית • תרגיל בית 3 - הגשה • קריאת מאמרים - הגשת דו"ח 4	10	05/19
הסברת מודלים • תרגיל בית 4 - פרסום	11	05/26
הגינות • קריאת מאמרים - הגשת דו"ח 5	12	06/02
פרויקט - הצגה (כמעט) סופית בכיתה • תרגיל בית 4 - הגשה	13	06/09
פרויקט - הגשה סופית	-	08/04

הערות

- לכל סטודנט תוקצב מכסה של ארבעה ימים (סה"כ) לדחיית הגשת תרגילי הבית אשר יוכלו להשתמש בהם כרצונם.
- העבודה על הפרויקטים הינה בקבוצות של שלושה סטודנטים (אפשרות מועדפת) או בזוגות.
- בהתאם למספר הסטודנטים הרשומים, ייתכן שתתקיים בחינה סופית. במקרה כזה, יוקצה לבחינה 20% משקל בעת חישוב הציון הסופי, ולפרויקט יוקצה 45%.