# Seminar on Failures Modes in Machine Learning (FAILML)
# (2360839, Spring 2025)

**Time**: Mondays 13:30-15:30

**Place**: Taub 3 (may change - check before semester starts)

**Syllabus**: https://cs.technion.ac.il/courses/all/224/236839.pdf

**Course website**: https://webcourse.cs.technion.ac.il/02360839

*** please register on webcourse to make sure you get all course-related announcements ***

**Outline**:

- Every week we will meet to hear about papers related to how machine learning can fail, and how it can be fixed.
- In each session, two students will present two papers (45 min. each), and two students will present a short, critical discussion on the papers (5 min. each).
- Both papers will be on the same (or closely related) topic, and in each week we will explore a different topic.
- Each student in the course must sign up for two papers (spanning two different meetings): one as a presenter, and one as a discussant. The requirements of each role are described next.

**Presenters**:

- Your task here is to choose a paper to present, read it thoroughly, and present it to the class in a clear and didactic manner.
- The available papers vary in their focus, difficulty, length, and technical depth. Your grade will be relative to the difficulty of your chosen paper. For example, if you choose an `easy' paper, you must be exceptional when presenting; if you choose a `hard' paper, you'll be given some slack.
- Regardless of the paper's difficulty - fitting any full length paper into a 45 min. talk is challenging. Your primary goal (and an important factor in your grade) is to be able to pick out and present the most relevant parts of the paper, in a way that is a coherent and complete summary of the paper. It will be your job to "separate the wheat from the chaff".
- Please practice your presentation to make sure it is streamlined, clear, and engaging.
- Note: some papers have multiple versions available online - please use the link provided in the table below.

**Discussants**:

- In addition to the main paper presentations, in each week we will also have two students (one per paper) take on the role of discussants (or 'critics') whose role is to present a short, critical, alternative take on the papers presented.
- Show critical thinking! Identify implicit assumptions, hidden weaknesses, and lurking limitations. Consider the paper within a broader, realistic setting. Think about the bigger picture. In any case - present your case in a respectful and constructive manner.
- Be precise and to the point - you only have 5 minutes!
- More details will be given in our first meeting.

**Signing up for papers:**

- The dates, topics, and paper titles are given in the table below.
- Sign-up begins **Sunday 23/3/25 at 10:00**, and is on a **first-come first serve basis**. At that time, the table will become editable, and you will be able to sketch yourself in.
- Needless to say: DO NOT EDIT A TABLE ENTRY IF IT HAS ALREADY BEEN TAKEN.
- Be sure to go over the papers until that time, and choose with care.
- Students presenting in the same meeting can choose to work on and present the two papers **jointly**, or **individually**. You will need to make this decision when signing up to papers (more details below), so be sure to coordinate with your partner first!
- Don't forget to sign up to **two papers** - once as a **presenter**, and once as a **discussant**.
- **Note**: In extreme cases, there is a chance that I may have to reassign you (but if this happens, it is for good reason). So until you receive confirmation, please know that signing up to a paper does not guarantee that this is the paper you'll be presenting.
- Extra slack will be given to the students presenting first!

**Dates and papers:** (*notice special time and/or day)

| # | date | topic | paper | presenter | discussant |
|---|---|---|---|---|---|
| 1 | 31/3 | | Intro lecture | | |
| 2.1 | 7/4 | domain adaptation | Domain-Adversarial Neural Networks | Adi Grauer | Lior Dvir |
| 2.2 | | domain adaptation | Understanding Robust Learning through the Lens of Representation Similarities | Lior Dvir | |
| 3.1 | 21/4 | domain adaptation | Joint distribution optimal transportation for domain adaptation | Omer Iedovnik | Omer Ben Laish |

| | | | | | |
|---|---|---|---|---|---|
| 3.2 | | out-of-distribution | Accuracy on the Line: On the Strong Correlation Between Out-of-Distribution and In-Distribution Generalization | Tal Haklay | Yara Shamshoum |
| 4.1 | 28/4* | out-of-distribution | Fishr: Invariant Gradient Variances for Out-of-Distribution Generalization | | Adi Grauer |
| 4.2 | | test-time adaptation | Test-Time Training with Self-Supervision for Generalization under Distribution Shifts | | Moriya Menachem |
| 5.1 | 5/5 | worst-group | Why does Throwing Away Data Improve Worst-Group Error? | Amit Ben Yossef | Dovid Parnas |
| 5.2 | | distributionally robust optimization | Distributionally Robust Neural Networks For Group Shifts: On The Importance Of Regularization For Worst-Case Generalization | Tomer Ashuach | Tal Haklay |
| 6.1 | 12/5 | distributionally robust optimization | Just Train Twice: Improving Group Robustness without Training Group Information | Ram Binshtock | Yazan Othman |
| 6.2 | | distributionally robust optimization | Modeling the Second Player in Distributionally Robust Optimization | Liran Cohen | Tomer Bitan |
| 7.1 | 19/5 | shortcut learning | On the Foundations of Shortcut Learning | Elizabet Khaimov | Roy Maor Lotan |
| 7.2 | | memorization | Learning and Memorization | Yara Shamshoum | Amit Ben Yossef |
| 8.1 | 26/5 | invariance | Invariant Representations without Adversarial Training | Roy Maor Lotan | Arkadi Piven |
| 8.2 | | invariance | Discovering Environments with XRM | Moriya Menachem | |

| | | | | | |
|---|---|---|---|---|---|
| 9.1 | 9/6 | adversarial | MMA Training: Direct Input Space Margin Maximization Through Adversarial Training | Omer Ben Laish | Ram Binshtock |
| 9.2 | | adversarial | Attacks Which Do Not Kill Training Make Adversarial Learning Stronger | Yazan Othman | 1 Liran Cohen |
| 10.1 | 16/6 | adversarial | Theoretically Principled Trade-off between Robustness and Accuracy | Dovid Parnas | Neta Katz |
| 10.2 | | adversarial | A Closer Look at Accuracy vs. Robustness | Neta Katz | Elizabet Khaimov |
| 11.1 | 23/6 | strategic classification | Strategic Classification Made Practical | Dan Kalifa | Bana Sadi |
| 11.2 | | strategic classification | Generalized Strategic Classification and the Case of Aligned Incentives | Arkadi Piven | Omer Iedovnik |
| 12.1 | 30/6 | self-selection | Classification Under Strategic Self-Selection | Bana Sadi | Dan Kalifa |
| 12.2 | | strategic classification | Adversaries With Incentives: A Strategic Alternative to Adversarial Robustness | Tomer Bitan | Tomer Ashuach |
| 13.1 | 7/7 | performative prediction | Outside the Echo Chamber: Optimizing the Performative Risk | | |
| 13.2 | | performative prediction | Optimal Classification under Performative Distribution Shift | | |