# Course: Reliability, Fairness and Robustness of Learning Systems

**Course Code:** 0601

**Semester:** 2/2021

| Instructor | Tel.: 046195, 236756
|-------------|--------------------------|
| 1. 044202, 094412
| 2. 236781, 046211

**Course Content:**

- The course will cover methods for ensuring reliability, stability, fairness, and robustness of the conclusions derived from complex learning systems, such as deep learning.
- We will be exposed to modern learning algorithms and familiarize ourselves with practical methods based on rich theory to deal with these challenges.

**Course Format:** Seminar format and articles at the forefront of the field of data science. Focus on reducing uncertainty in prediction, as well as identifying and preventing gender biases in learning systems. Additionally, we will review advanced methods for testing many hypotheses as a tool for identifying scientific discoveries from data and for interpreting learning algorithms.

---

**Course Requirements:**

- Required courses:
  - 094423, Introduction to Statistics (or similar courses)
  - 046211, Deep Learning

- Recommended courses:
  - 044202, Data Processing and Analysis
  - 236756, Learning Systems

---

**Course Information:**

- Interested in enrolling in the course?
  - Please fill out the form at the following link:
    - [https://forms.gle/hd1f5mJtnhEJxzy19](https://forms.gle/hd1f5mJtnhEJxzy19)
  - It is important to note that registering through the link above does not constitute enrollment in the course. Due to space constraints, only a portion of the students who complete the form may be able to participate.
  - You can also send an email to the course instructor at yromano@technion.ac.il

---

**Course Information:**

- Comments:
  - The course will focus on modern learning algorithms and their applications in complex systems, such as deep learning.
  - We will be exposed to practical methods based on rich theory to address these challenges.
  - The course format will include seminar discussions and articles at the forefront of data science.
  - Focus will be on reducing uncertainty in predictions, as well as identifying and preventing gender biases in learning systems. Additionally, we will review advanced methods for testing many hypotheses as a tool for identifying scientific discoveries from data and for interpreting learning algorithms.

---

**Course Information:**

- Additional comments:
  - The course will cover methods for ensuring reliability, stability, fairness, and robustness of the conclusions derived from complex learning systems, such as deep learning.
  - We will be exposed to modern learning algorithms and familiarize ourselves with practical methods based on rich theory to deal with these challenges.
  - The course format will include seminar discussions and articles at the forefront of data science.
  - Focus will be on reducing uncertainty in predictions, as well as identifying and preventing gender biases in learning systems. Additionally, we will review advanced methods for testing many hypotheses as a tool for identifying scientific discoveries from data and for interpreting learning algorithms.
English Syllabus

In this course, we will study methodologies to guarantee the reliability, robustness, equity, and reproducibility of advanced machine learning systems, such as deep neural nets. We will discuss recent concerns about modern machine learning algorithms and will tackle these by introducing flexible tools that are supported by theoretical guarantees. This course will be given as a seminar and will cover influential papers in the field of data science. We will focus on prediction uncertainty estimation as well as mitigation of discrimination against minorities. We will also explore frameworks for multiple hypothesis testing as powerful tools for making new scientific discoveries and for interpreting complex learning systems.

Learning Outcomes

1. Familiarize with challenges in using modern machine learning algorithms in high-stakes applications.
2. Learn flexible methodologies to produce data-driven inferences that can be trusted.

רשימה של מאמריםمواובלים בתחום של למידה חישובית וסטטיסטיקה.