

# שם הקורס: אמינות, הוגנות ועקביות של מערכות לומדות

מספר הקורס: 236802

סמסטר: אביב תשפ"א

מרצה:	יניב רומנו
שעות הרצאה:	
שעת תרגול:	
דרישות קדם:	1. 044202, 094412 הסתברות (או קורסים דומים) 2. 046195, 236756 מערכות לומדות
	<b>מקצועות קדם מומלצים:</b>
	1. 094423 מבוא סטטיסטיקה (או קורסים דומים, כגון 046202 עיבוד וניתוח מידע)
	2. 236781, 046211 למידה עמוקה
אתר הקורס:	

## תאור הקורס

### סילבוס בעברית

בקורס זה נלמד על שיטות להבטחת אמינות, עמידות, הוגנות, והישנות של מסקנות המתקבלות ממערכות למידה מורכבות, כגון למידה עמוקה. נחשף לבעיות של אלגוריתמי למידה מודרניים ונכיר שיטות מעשיות הנסמכות על תיאוריה עשירה להתמודד עם אתגרים אלו. הקורס יינתן במתכונת סמינר וכסה מאמרים בחזית התחום של מדעי הנתונים. נתרכז בשיערוך אי ודאות בחיזוי, כמו כן זיהוי ומניעת הטייה מגדרית של מערכות לומדות. בנוסף, נסקור שיטות מתקדמות לבדיקות השערות מרובות ככלי לזיהוי תגליות מדעיות ממאגרי מידע ולאיינטרפרטציה של אלגוריתמי למידה.

### תוצאות למידה

1. חשיפה לאתגרים בשימוש במערכות לומדות ביישומים שבהם האמינות היא דרישה הכרחית.
2. למידת מתודולוגיות להבטחת אמינות המסקנות המתקבלות ממערכות לומדות.

### 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.

### דרישות הקורס

הציון יתבסס השתתפות פעילה בכיתה, הצגת מאמר ופרויקט.

### רשימת ספרות

מאמרים מעיתונים ומכנסים מובילים בתחום של למידה חישובית וסטטיסטיקה. דוגמאות למאמרים שיכוסו בסמינר:

1. *A tutorial on conformal prediction*. Journal of Machine Learning Research, G. Shafer, and V. Vovk, 2008.
2. *Predictive inference with the jackknife+*. R.F. Barber, E. J. Candes, A. Ramdas, and R. J. Tibshirani, arXiv:1905.02928, 2019.
3. *Equality of opportunity in supervised learning*. M. Hardt, E. Price, and N. Srebro. NeurIPS, 2016.
4. *Panning for gold: Model-X knockoffs for high dimensional controlled variable selection*. E. J. Candes, Y. Fan, L. Janson, L. Lv. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2018.