

# סילבוס - תוכנית הוראה לקורס היבטים אתיים בבינה מלאכותית

ד"ר אהרונה פפר | המחלקה לניהול  
Artificial Intelligence Ethics | 55-047-02 on

שיעור	סוג הקורס:
2	היקף נ"ז:
תשפ"ז	שנת לימודים:
קיץ	סמסטר:
	יום ושעה
בתיאום מראש	שעת קבלה:
<a href="mailto:aharonap@biu.ac.il">aharonap@biu.ac.il</a>	מייל מרצה:
_____	קישור לאתר למדה:



## תיאור הקורס ומטרות למידה

### תקציר הקורס

הקורס עוסק בזיהוי, ניתוח ומציאת פתרונות לסוגיות אתיות, חברתיות ומוסריות ביישום טכנולוגיות של בינה מלאכותית. הקורס מפנה זרקור אל "הצד האפל" של התחום ומציג כיצד לצד ההזדמנויות והיתרונות המשמעותיים שמביאים עמם יישומי בינה מלאכותית, צצות מגוון שאלות אתיות, כגון חשש לפגיעה בפרטיות, באמון הלקוחות, בהוגנות, ולעיתים גם אפלייה ופגיעה במיעוטים.

במסגרת הקורס נבחן טכנולוגיות בינה מלאכותית שונות דרך הפריזמה של הגישות האתיות המסורתיות ותוך הצגת דוגמאות עדכניות ממגוון תחומים ותעשיות (לרבות היטק, פיננסים, בריאות ורפואה, משפט, מסחר אלקטרוני, שיווק, פרסום דיגיטלי, משאבי אנוש ועוד), ותוך דיון במחקרים אקדמיים, תחקירים עיתונאיים וסרטונים.

נלמד לזהות את האתגרים והסוגיות הבעייתיות וגם נכיר את הדרכים המתאימות להתמודדות עימן: פתרונות נורמטיביים, טכניים, או במידת הצורך גם רגולטוריים. במקרים מסוימים נראה כי דווקא "הבעיות" האתיות עשויות לייצר עבור עסקים הזדמנויות חדשות.

### מטרות/תוצרי הלמידה

#### ידע

1. הלומדים יגדירו את הפרספקטיבות האתיות לבחינת טכנולוגיות קיימות וחדשות
2. הלומדים יבחנו את ההיבטים האתיים הקשורים לטכנולוגיות בינה מלאכותית הרווחות היום
3. הלומדים יסקרו את הרגולציה הקיימת והמתהווה בתחום, בארץ ובעולם

#### מיומנויות

1. הלומדים ינתחו את השיטות והכלים הפרקטיים המשמשים לשמירה על פרטיות משתמשים
2. הלומדים יעריכו את הכלים הפרקטיים המסייעים לצמצום הטיית וחסור הוגנות באלגוריתמים



## למידה פעילה - תכנון מהלך השיעורים:

מס' השיעור	נושא השיעור	למידה פעילה	קריאה/ צפיה נדרשת	הערכה תהליכית/מעצבת
1	מבוא. העולם בעידן הבינה המלאכותית	The moral machine experiment – אירוע כיתה	1	השיעורים ישולבו בדיונים בקבוצות, תרגולים אינטראקטיביים ועוד.
2	פילוסופיה של המוסר ואתיקה של טכנולוגיה		2	
3	פרטיות – מהי פרטיות והזכות לפרטיות ומהי חשיבותה. כיצד מתעררת הזכות לפרטיות בעידן ה-AI	דיון כיתה – מכוניות חכמות והערכות סיכון	3	
4	פרטיות בנתונים – הקושי לשמר פרטיות במאגרי נתונים וסקירת כלים משמרי פרטיות (שיטות ודרגות להתממה והצפנה, פרטיות דיפרנציאלית, הנדסת פרטיות)		4	
5	הרגולציה על השימוש במידע האישי בארץ ובעולם – חוק הגנת הפרטיות, תקנות ה-GDPR ועוד.	דיון כיתה – הפרות פרטיות ב-Clearview		
6	הטיות בקבלת החלטות אלגוריתמית – דוגמאות מתחומים שונים		5	
7	העקרונות של למידת מכונה והגורמים להטיות שהיא יוצרת באלגוריתמים		6	
8	כלים לצמצום הטיות וחוסר הוגנות באלגוריתמים; הרגולציה על תחום הבינה המלאכותית	Courtroom algorithm game: Can you make AI fairer than a judge? MIT Technology Review.	7	
9	תחום התעסוקה והמשאב האנושי בעידן ה-AI – קבלה לעבודה, מעקב אחר עובדים ושימוש בנתוני עובדים			
10	אתיקה בניסויים דיגיטליים ואפליה בפרסום דיגיטלי		8,9	
11	מצגות סטודנטים			
12	מצגות סטודנטים			
13	סיכום			

\*ייתכנו שינויים בסילבוס בהתאם לקצב ההתקדמות ואפקטיביות הלמידה

משקל בציון הסופי	תיאור התוצר
20% מהציון הסופי	מצגות כיתה בצוותים
80% מהציון הסופי	מבחן מסכם

## דרישות הקורס

הצגת מצגת, בצוותים, בשיעורים האחרונים של הסמסטר. קבלת ציון עובר בבחינה. נוכחות חובה בשיעורי מצגות הסטודנטים וב- 80% מסך השיעורים. מומלץ להגיע לכל השיעורים ולהשתתף באופן פעיל בדיוני הכיתה, בניסויים ובפעילויות.

## ביבליוגרפיה: תכנים לקריאה, צפיה והאזנה

### חובה:

1. Tristan Harris, (2025) "**Why AI is our ultimate test and greatest invitation**" TED2025 [https://www.ted.com/talks/tristan\\_harris\\_why\\_ai\\_is\\_our\\_ultimate\\_test\\_and\\_greatest\\_invitation](https://www.ted.com/talks/tristan_harris_why_ai_is_our_ultimate_test_and_greatest_invitation)
2. Vallor, S., Raicu I., and Green B. (2022). "**Technology and Engineering Practice: Ethical Lenses to Look Through**". *The Markkula Center for Applied Ethics at Santa Clara University*. <https://www.scu.edu/ethics-in-technology-practice/ethical-lenses/>
3. Kosinski, M. "**The End of Privacy**", *Talks at Google* <https://www.youtube.com/watch?v=VUwBcTgztU>
4. Gregory Vial, Julien Crowe, and Patrick Mesana (2024) "**Managing Data Privacy Risk in Advanced Analytics**" *MIT Sloan Management Review Special Report "Overcoming the hard problems to advance AI practice"* 18-22.
5. Cathy O'neil "**Weapons of math destruction**" Talks at Google: <https://www.youtube.com/watch?v=TQHs8SA1qpk>
6. Hao K. (2019) "**This is how AI bias really happens—and why it's so hard to fix**", *MIT Technology Review* <https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/>
7. "**Algorithmic Fairness**" (2025), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/algorithmic-fairness/>

8. Kramer A., Guillory J., and Hancock J. (2014) "**Experimental evidence of massive-scale emotional contagion through social networks**" *PNAS* 111(24)  
<https://doi.org/10.1073/pnas.1320040111>
9. Basileal I., Korolova A., and Heidemann J. (2021) "**Auditing for Discrimination in Algorithms Delivering Job Ads**". *Proceedings of the Web Conference '21, ACM*  
<https://doi.org/10.1145/3442381.3450077>

## העשרה:

1. Jules Polonetsky, Omer Tene, and Kelsey Finch, "**Shades of Gray: Seeing the Full Spectrum of Practical Data De-identification**", 56 *Santa Clara L. Rev.* 593 (2016). Available at: <http://digitalcommons.law.scu.edu/lawreview/vol56/iss3/3>
2. בירנהק מ. "פרטיות בעיר בעידן הדיגיטלי" פרק 3 בספר: חתוקה ט. (עורכת) העיר בעידן הדיגיטלי. אוניברסיטת תל אביב 2018.
3. Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, (2016) "**Machine Bias**: There's software used across the country to predict future criminals. And it's biased against blacks." *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
4. Cathy O'Neil, Jake Appel, and Sam Tyner-Monroe (2024) "**Auditing Algorithmic Risk**" *MIT Sloan Management Review Special Report "Overcoming the hard problems to advance AI practice"* 1-8.
5. Marcin Lukasz Bartosiak and Artur Modlinski (2022) "**Fired by an algorithm? Exploration of conformism with biased intelligent decision support systems** in the context of workplace discipline" *Career Development International* 27 601-615
6. **EU AI Act** – High-level summary <https://artificialintelligenceact.eu/high-level-summary/>
7. גאמס נ. (2022) "עבודה מרחוק במעקב קרוב" מגזין דה-מרקר, פברואר 2022
8. Tiidenberg, K. (2017) "**Ethics in digital research**", in *Handbook of qualitative data collection*, Sage. 466–481