מרכז חיפאי בין אוניברסיטאי

לבינה מלאכותית ומדעי הרוח

אתיקה - חברה - סביבה



הגשה למל"ג-ות"ת במסגרת קול קורא

ל "תכנית המאה" לקידום תחום מדעי הרוח

חיפה, יולי 2021



Photo by jesse orrico on Unsplash

 

מרכז חיפאי בין אוניברסיטאי

לבינה מלאכותית ומדעי הרוח:

אתיקה - חברה - סביבה

הגשה למל"ג-ות"ת במסגרת קול קורא

ל "תוכנית המאה" לקידום תחום מדעי הרוח

חיפה, יולי 2021

ועדה נכבדה,

אנו מתכבדים להגיש בזאת הצעה משותפת של אוניברסיטת חיפה ושל הטכניון להקמת מרכז לבינה מלאכותית ומדעי הרוח: אתיקה, חברה, סביבה - כמענה לקול קורא של המל"ג מיום 19 במאי, 2021. המרכז יהיה בינאוניברסיטאי ועל-פקולטטי, משותף לאוניברסיטת חיפה והטכניון. הוא יעסוק בנושאים מגוונים בתפר שבין בינה מלאכותית ומדעי הרוח, כולל השלכות בתחומי מדעי החברה והסביבה. המרכז ישלב הוראה ומחקר בינתחומיים מפקולטות שונות באוניברסיטת חיפה ובטכניון.

פניית המל"ג מגיעה בתקופה בה ברור לכל כי גובר הצורך בשילוב מדעי הרוח, ואתיקה בראשם, בתהליכים אנליטיים העוסקים בחיזוי, כריית מידע מתוך נתוני עתק, עיבוד מידע (טקסטואלי ואחר), כמו גם פיתוחים בתחומי עיבוד השפה הטבעית, חברה ומשפט. להערכתנו, צורך ממשי זה יחזק את התחומים השונים במדעי הרוח, יעלה את יוקרתם, יביא לדיאלוג בינתחומי של חוקרים בתחומי מדעי הרוח עם חוקרים מכלל המדעים, ירחיב ויעמיק את המחקר וההוראה במדעי הרוח, ויאפשר להטמיע את תחומי מדעי הרוח בתחומי הטכנולוגיה והמדעים השונים (טבע וחברה). אני צופים כי מהלך זה, שיש בו צורך ממשי שבא דווקא מתחומי ההנדסה והטכנולוגיה, יהפוך את מדעי הרוח לאטרקטיביים ורלוונטיים בחיי היום-יום ויביא למשיכה ולגידול משמעותי במספר הסטודנטים. מהלך זה יסייע בעקיפין גם להצלתם של תחומים נכחדים במדעי הרוח שאנו סבורים כי יש להצילם כמעט בכל מחיר.

פיתוח טכנולוגיות מיחשוב חדשניות של בינה מלאכותית באקדמיה ובתעשייה עומד במרכז תעשיית ההייטק כיום. ברור לכל כי יש בכוחן של טכנולוגיות אלה לשפר ולהיטיב עם הפרט והחברה. עם זאת, ברור גם כי ללא ההכנה הבינתחומית הנדרשת, קיים בתחומים אלה פוטנציאל ממשי להשפעה מזיקה על כל תחומי החיים האנושיים על פני כדור הארץ: פיתוח אמצעי שליטה ריכוזית, מעקב, חדירה לפרטיות, פגיעה בחופש הפרט, הטיות חברתיות והגדלת הפערים החברתיים כל אלה הם רק חלק מהנזקים הסביבתיים הפוטנציאלים. לעניות דעתנו, יש להטמיע את האחריות החברתית ואת הזהירות המתבקשת בפיתוח הבינה המלאכותית ויישומיה השונים.

לאור כל זאת, אוניברסיטת חיפה והטכניון רואים בהקמת מרכז לבינה מלאכותית ומדעי הרוח, שיתמקד בהשלכות אתיות, חברתיות וסביבתיות של פיתוח ויישום בינה מלאכותית, אתגר ראשון במעלה, ממשי ומתבקש, בעל השלכות עמוקות ביותר, חברתיות וקהילתיות, כמו גם אקדמיות טהורות.

**בברכה,**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**המשנה הבכיר לנשיא הטכניון: רקטור אוניברסיטת חיפה: פרופסור עודד רבינוביץ פרופסור גור אלרואי**

[**רקע**](#_heading=h.8zasl6bf6xu8)4

[**האתגר**](#_heading=h.tyjcwt)7

[**מטרות**](#_heading=h.z337ya)8

[**עקרונות**](#_heading=h.3j2qqm3)9

[**מגמות בינלאומיות**](#_heading=h.1y810tw)10

[**יעדים ותוכנית עבודה**](#_heading=h.p8my3hqe7dm4)12

[**תוכניות לימודים במרכז: תואר שני**](#_heading=h.h6fpdogbie9t)14

[**תנאי קבלה לתוכניות הלימודים**](#_heading=h.xm0cyg8rbory)18

[**מבנה הלימודים**](#_heading=h.r6niy38iyvrw)18

[**תוכנית מס. 1: בינה מלאכותית במדעי הרוח (AI and humanities)**](#_heading=h.u7ddw75gy0sx)22

[**תוכנית מס. 2: בינה מלאכותית ואתיקה (AI and ethics)**](#_heading=h.f1g3c3tt6cx0)22

[**תוכנית מס. 3: בינה מלאכותית וקיימות: חברה וסביבה ( AI for society and environment)**](#_heading=h.tkqqo05yp1ob)22

[**תוכנית מס. 4: בינה מלאכותית וקבלת החלטות בממשל ובעסקים (AI and Decision Making in Law and Business)**](#_heading=h.n2c964xnu8ya)23

[**תוכניות מחקר - תקצירים**](#_heading=h.k31huzqu7jek)39

[**חברי הסגל במרכז כולל הפרופיל המחקרי שלהם והדגשים הבין תחומיים**](#_heading=h.t6d66494bmmy)48

[**תוכניות עתידיות**](#_heading=h.gnl3eorqhtna)58

# 

# רקע

המהפכה התעשייתית הרביעית מאופיינת על ידי פיתוח מערכות בינה מלאכותית המחקות את האינטליגנציה האנושית. המהפכה מתרחשת בקצב אקספוננציאלי חסר תקדים היסטורי. למערכות בינה מלאכותית מתפתחות קיימת זיקה עמוקה לתחומים רבים במדע, בטכנולוגיה ובהנדסה. היישומים מצריכים שינוי תודעתי-מוסרי. על פי דוח מקנזי [[1]](#footnote-1)(McKinsey, 2021) עיקר השינוי הוא תודעתי, כאשר הבינה המלאכותית עוברת מהמעבדה אל מקום העבודה, עם השלכות עמוקות על התעשייה והחברה.

התפתחות מערכות בינה מלאכותית בכל תחומי החיים מתורגמת גם לצמיחה כלכלית עולמית. מחקר בבינה מלאכותית נוטה להפריד באופן אנליטי בין סוגיות טכנולוגיות וחברתיות, ורואה בבינה מלאכותית אובייקט טכני שרק מאוחר יותר, לאחר יישומה, עשויות להיות לו השלכות חברתיות[[2]](#footnote-2). אחת מנבואות הזעם הקשורות לעתיד התפתחות הבינה המלאכותית מתייחסת להשלכות החברתיות השליליות, ומנבאת את העמקת הפער הכלכלי-חברתי בשל דרישה גוברת והולכת בתעשייה ובמחקר למיומנויות גבוהות ומתגמלות. האתגר של האקדמיה הוא ביצירת מרחב לשיתוף פעולה משגשג למחקר עתידי בבינה מלאכותית המשלב היבטים אתיים. המרכז יקנה לסטודנטים במדעי הרוח יכולות במדעים דיגיטליים, ואילו לסטודנטים מתחומים הנדסיים - יקנה ידע במדעי הרוח ו'שפה הומניסטית'. המרכז 'לבינה מלאכותית ואתיקה' יקדם שיתופי פעולה בחינוך ובמחקר בבינה מלאכותית במטרה לגשר על הפער שבין פיתוחים טכנולוגיים לבין ההשלכות החברתיות הנובעות מהן. נבואת הזעם המיוחסת לעתיד עושה כבר כעת את צעדיה הראשונים. בשל כך, גוברת והולכת הדרישה מהאקדמיה להכשיר בוגרים למהפכה התעשייתית הרביעית. ההכשרה תבוא לידי ביטוי בפיתוח מערכות מבוקרות ובהטמעת האחריות האתית והחברתית הראויה כמענה לצורכי התעשייה והמחקר.

חוקרים ומפתחים העוסקים בפיתוח מערכות העושות שימוש באלגוריתמים של בינה מלאכותית נוטים להכיר יותר ויותר בחשיבותם של מדעי הרוח והחברה, ובראשם האתיקה, ביישום ופיתוח של מערכות לומדות. בנוסף, גוברת ההכרה כי תחומים רבים ושונים המהווים את התשתית לעיסוק בבינה מלאכותית, למידת מכונה, למידה עמוקה וביג דאטה מהווים אתגר גדול לאנושות. תחום הבינה המלאכותית מעצב מחדש שאלות יסודיות במדעי הרוח, כך, למשל, בהקשר של פרטיות, נגישות למידע, פערים חברתיים, שוויון ושוויון הזדמנויות - שאלות שיש להן כמובן גם השלכות חברתיות, כלכליות ופוליטיות, ועל כן הן רלוונטיות גם למדעי החברה, למשפטים, ולשאר התחומים. הבינה המלאכותית מאתגרת את יסודות הקיום האנושי, את המדענים והחוקרים, בתחומי המדע בכלל, ובמדעי הרוח בפרט. ניתן אף לומר שנחוץ סוג חדש של מדענים חוצי-תחומי מחקר, שיובילו את ההשתלבות העתידית של בינה מלאכותית בחיי האדם באחריות הראויה. תפקידו של המרכז המתוכנן בשתי הרמות - הוראה ומחקר - הוא ליצור את התשתית הראויה לפיתוחים עכשוויים ועתידיים, של בינה מלאכותית שיהיו אחראיים מוסרית וחברתית לרווחת החברה האנושית, וזאת מבלי לעצור את ההתפתחות של תחום הבינה המלאכותית, שלה חשיבות גוברת והולכת בכל תחומי החיים.

צורך זה הוא הבסיס לשיתוף פעולה בין אוניברסיטת חיפה והטכניון, שבוגריו הם חוד החנית של תעשיית ההיי-טק בארץ ובעולם. ייחודה של אוניברסיטת חיפה הוא בהיותה אוניברסיטת מחקר חוצת תחומים, וכור היתוך המאחד תחומי ידע שונים. בה נוצרים תחומי המחקר וההוראה של המחר, ובדרכה מגלמת אוניברסיטת חיפה את המרקם החברתי המקיים שורה ארוכה של שיתופי פעולה חוץ ופנים מוסדיים. שיתוף פעולה של אוניברסיטת חיפה עם הטכניון הינו חיבור מנצח, שכן יביא לידי ביטוי את מעמדו של הטכניון הממוקם במקום הראשון באירופה ובמקום ה-15 בעולם בתחום הבינה המלאכותית[[3]](#footnote-3), עם חוזקותיה של האוניברסיטה במדעי הרוח ובלימודים רב-תחומיים.

אוניברסיטת חיפה יחד עם הטכניון מבקשים בהצעה זו לכונן מרחב אקדמי משותף לאיגום המחקר ולהוראה בתואר מתקדם, שבו תתאפשר הזנה הדדית של חוקרים וסטודנטים ממדעי הרוח ושל חוקרים וסטודנטים העוסקים בבינה מלאכותית. מרחב זה ימצב את מדעי הרוח בכלל, ואת תחום האתיקה בפרט, כתשתית חיונית לעשייה המחקרית והיישומית-טכנולוגית של הבינה המלאכותית. צעד זה יגשר על הפער שנוצר בין ההאצה הטכנולוגית שאנו חווים מאז המחצית השנייה של המאה ה-20, לבין אי-הוודאות העמוקה באשר להשלכות החברתיות, המוסריות והסביבתיות של השימוש בבינה מלאכותית באספקטים השונים של החיים האנושיים. כדי לגשר על פער זה הכרחי להבין לעומק מהי מערכת אלגוריתמית, כיצד היא 'חושבת ומקבלת החלטות', מהן המגבלות של מערכות קיימות, וכיצד להטמיע שיקולים אנושיים רלוונטיים בפיתוח אלגוריתמים שאמורים לסייע בקבלת החלטות בתחומי הדיון של מדעי הרוח והחברה, הרווחה והבריאות חקר ההיבטים השונים של המעבר לשימוש בבינה מלאכותית, יאפשר, אם כן, להתמודד גם עם הסכנות הטמונות בשיפורים הטכנולוגיים המהותיים שמאפיינים עד כה את המאה ה-21, והצפויים להשפיע על חיינו בעתיד הנראה לעין. בין הסכנות ניתן למנות הטיות ואפליות שכבר מתגלות במערכות טכנולוגיות שונות שנכנסו לשימוש רחב, כפועל יוצא של היבטים אתיים שלא באו לידי ביטוי במהלך פיתוחן. ריכוז כוח בידי מיעוט גורמים, השלכות על דרכי עיצוב התודעה ושאלות שעוסקות בהגדרת מרחב הפרט בחברה המקוונת הן דוגמאות נוספות. מכיוון שהטכנולוגיה לא ניתנת לעצירה, יש לפתח בדחיפות תחומי מחקר והוראה שייעודם לגשר על הפער בין תחומי ההנדסה לבין התחומים ההומניים על ידי העמקת האינטראקציה הבינתחומית ובניית תוכניות הוראה משולבות. דוגמה מובהקת לכך היא הוראת האתיקה בכל התוכניות הנרקמות.

חבירה זו היא צורך מתבקש לאור כל המתואר לעיל, על מנת להכשיר בוגרים במספר מסלולים המשלבים את שני העולמות. התעשייה משוועת לבוגרים רחבי אופקים, המכירים את מגוון ההיבטים הנדרשים לתעשייה שצומחת ומשתנה בתחום בו מתחוללים שינויים עצומים של בינה מלאכותית, מכונות לומדות ומאגרי נתונים עצומים. הקמת מרחב הלימודים לתואר מוסמך תביא לשילוב תוכניות ליבה מתחומים שונים במדעי הרוח, אתיקה, אמנויות, משפט וחברה כהכנה לפתרונות הנדסיים בעולמות הבינה המלאכותית, פתרונות אשר יהיו ישימים בשדות חברתיים מרכזיים, בדגש על עולמות הסביבה והקיימות, הממשל והעסקים. חבירה כזאת תכשיר בוגרים שיהוו את חוד החנית ואת המנוף העיקרי והתשתיתי לתעשייה בתחום הבינה המלאכותית ונגזרותיה. תארים אלה והמסלולים המשולבים בהם, הם צו השעה וצורך קיומי להמשך התפתחות תעשייה חשובה זו.

# האתגר

החל מהמחצית השנייה של המאה ה-20 מתחוללת מהפכה טכנולוגית שבבסיסה תקשורת ונתונים. השלב הנוכחי מתמקד בכלי בינה מלאכותית (AI), אשר תופסים כיום מקום מרכזי בחיינו האישיים והחברתיים, והשפעתם על חיינו הולכת וגדלה. מערכות שאוספות, אוגרות ומנתחות נתונים, ולאחר מכן תרות אחר מאפיינים  דומים ושונים, משליכות מהן, ומעצבות באמצעותם את ההתנהגות האנושית מבחינה חברתית, סביבתית, כלכלית, בריאותית, משפטית וכמובן אתית. החיבור בין התפתחות טכנולוגית לבין אתיקה מתבטא בחדירה של בינה מלאכותית לאורחות החיים ובהשלכות מרחיקות לכת על ההתנהגות האנושית, הן לחיוב והן לשלילה. היעדר הבנה וכישורי ניתוח של יכולות הבינה המלאכותית לצד התעצמות היכולות ההולכות וגוברות של מאגרי הנתונים, ובפרט הגברת מורכבות הנתונים, מהווים כיום את האתגרים המובילים אוטומציה של תהליכי עבודה במישורים השונים. אלה הם תהליכים הרלוונטיים לפיתוח הטכנולוגיה עצמה, ובכלל זה ניהול אבטחת רשת, ותהליכי הטמעת טכנולוגיה והשימושים האפשריים לה על ידי גורמי ממשל, תעשייה והמשתמש הבודד. מכאן, הצורך האקוטי במרחב אקדמי פלורליסטי, רב תחומי, מגוון והטרוגני אשר ינתח היבטים שונים של מהפכת מידע זו. המרכז יהווה מקום מפגש בין תחומי מחקר שונים שישתפו פעולה ויתרמו למאמץ האקדמי לזיהוי דילמות חברתיות שונות הקשורות לבינה מלאכותית, מאמץ שהשלכותיו על התעשייה יפורטו להלן. הצטלבות הצירים תקדם התפתחות של תחומים חדשים בחזית המדע. תוכניות הלימוד תוכננו לגשר על הפער הבינתחומי, ויהוו אבן שואבת לחוקרים ולסטודנטים לתארים מתקדמים.

מבין אתגרי התוכנית לגשר על הפער הבינתחומי וליצור את הצטלבות הצירים מצויה התמודדות עם המודעות ההולכת וגוברת לחשיבות בינה מלאכותית אמינה, שקופה ואחראית.  התמודדות עם האתגר הפדגוגי משלבת למידה טכנולוגית לצד פיתוח יכולות חשיבה ביקורתית על מגמות ההתפתחות הטכנולוגית והפעלת מנגנוני ההגנה מפגיעה ומהטיה אפשרית.

# מטרות

מטרת התוכניות המוצעות ללימודים במרכז היא ליצור פלטפורמה המהווה צומת, הצטלבות בין התחומים השונים לצורך הקניית שפה משותפת לאנשי מדעי הרוח (ולצידם חוקרים מתחומי החברה השונים), מפתחי בינה מלאכותית ואנשי תעשייה. שפה משותפת זו, ייעודה לפתוח ערוץ תקשורת ולאפשר שיח שבו ניתן יהיה לבנות פתרונות יעילים לכל בעיה באזורים ההנדסיים ו/או ההומניים באמצעות חשיבה בינתחומית משותפת. הקניית השפה המשותפת תאפשר דיון פתוח וביקורתי בהיבטים הטכניים, לצד מושגי יסוד בעבודה משותפת, תחבר ותפרה את השיח בין התחומים, ותוביל לפתרונות וליישומים יצירתיים ופורצי דרך בהתמודדות עם בעיות שונות במרחב הקיומי המשותף.

גמישות מחשבתית זו היא סוד ההצלחה של תופעה שהפכה זה מכבר למגמה עולמית. סיעור מוחות ושילוב צוותי עבודה מגוונים מאפשרים חשיבה רב ממדית המקיפה נושאים בעלי השלכות הדדיות. חשיפה לתכני מדעי הרוח והחברה תאיר אספקטים שהיו זרים עד כה לעולם ההנדסי והטכנולוגי. מאידך, בוגרי מדעי הרוח והחברה ייחשפו להיבטים הטכניים ולמשמעות ההנדסית של השינויים ברמת הפרויקט או התוצר של כל גוף טכנולוגי. כך, בדומה לתהליך ה-'ריעיון'[[4]](#footnote-4), שבו בעיה נבחנת על ידי צוות משתתפים מגוון, רב תחומי. ובהשאלה מעולם היזמות שהוכיח את היתרונות של עבודה בקבוצות מגוונות המחברות אינטליגנציות, צוות מגוון ומרובה אינטליגנציות יגשר על הפערים ויוביל לחשיבה יצירתית פורצת דרך.

אחת ממטרות המחקר בתחום הבינה המלאכותית היא פיתוח אלגוריתמים שיוכלו לתת ניבויים הלימים אמפירית בתחומי מדעי הרוח והחברה - למשל בתחומים של חלוקת משאבים וצדק חברתי, שימור סביבה, משפט וממשל, כדוגמת האלגוריתמים הקיימים לניבוי ההסתברות לפשיעה חוזרת בקרב עבריינים מורשעים. המטרה היא לפתח אלגוריתמים שיקבלו החלטות "נכונות". דהיינו, החלטות שהמומחים הטובים ביותר בתחום היו מקבלים בתחומי מדעי הרוח והחברה השונים - למשל, אלגוריתם המדמה שופט חסר הטיות מוסריות. תחת הנחות אלה, למדעי הרוח יש תפקיד בסיסי מהותי בניתוח, זיהוי והבנה לעומק של מערכות השיקולים והמאפיינים הרלוונטיים  לפיתוח אלגוריתמים ומערכות לומדות מסוג זה.

# עקרונות

מרכזים, ברוח זה המוצע כאן, הוקמו במסגרות תאגידיות שונות. העיקרון המנחה המשותף לכלל היוזמות הללו, על-פי מחקרי השוק של גרטנר[[5]](#footnote-5) - אופטימיזציה של ביצועי עובדים אינה מספקת כיום. צוותי עבודה המשלבים אינטליגנציות שונות, נוטים לחדש מהר יותר, להצביע על בעיות מהר יותר, ולהציע פתרונות טובים יותר. כך עולה ממחקר[[6]](#footnote-6) המדגיש את איכות צוות העבודה בתעשייה כיחידה המתפקדת היטב יחד. עובדה זו מחזקת פעם נוספת את הצורך לא רק בצוותי עבודה בעלי אינטליגנציות שונות, אלא גם, ואף בעיקר, את היכולת לשתף פעולה וליצור ביטחון אישי-פסיכולוגי בקרב חברי הצוות[[7]](#footnote-7). בכך גם טמון הערך המוסף של שפה משותפת הנרכשת באקדמיה, הפותחת אופקים חדשים עבור הסטודנט, והמעניקה לחברות בתעשייה ערך משמעותי נוסף.

אחת הדוגמאות המפורסמות היא פרויקט 'אריסטו' של גוגל  (Google)[[8]](#footnote-8) משנת 2012. הפרויקט בחן את הגורמים המגדירים קבוצת אנשים כצוות. מטרת הפרויקט לבחון צוותי עבודה במטרה לזהות את הגורמים המבדילים בין צוותי עבודה מצליחים לבין צוותים כושלים. בין הגורמים המובילים שזוהו נמנו תכונות כגון אמינות בין חברי הצוות, שיח שוויוני, ורגישות חברתית גבוהה. כל אלה ורבים אחרים מתורגמים לביטחון אישי-פסיכולוגי של חברי הצוות

מכאן, חשיבות תוכניות הלימוד מתבטאת בהקניית שפה משותפת המאפשרת את השיח השוויוני בעבודת צוות רב-תחומי בתעשיה הקולטת, והמעלה את בטחונו האישי של העובד. התוכנית תענה על הצורך הגובר בשילוב בוגרי מדעי הרוח במרחב הטכנולוגי העתידי ובהטמעת תהליכים ומגמות בשוק העבודה הישראלי והעולמי. פלטפורמה זו תענה על הצורך הגובר בתעשייה לשילוב חשיבה רב מערכתית, מרובת אינטליגנציות, המגדירה מחדש את גבולות האחריות הציבורית של חברות בתעשייה.

מגמה כלל עולמית זו שהתחילה בשילוב בוגרי  מדעי הרוח נבעה בעיקרה ממצוקת כוח אדם וממחסור של תקנים בצוותים הנדסיים, ואולם מעז יצא מתוק ותרומתם של בוגרי מדעי הרוח שינתה את אופן העבודה בצוותים הנדסיים שפעלו בעיקר לפי סרגל נוסחאות חד ממדי. תרומתם של בוגרי מדעי הרוח שיכללה במידה רבה את צורת החשיבה של אותם צוותים מגוונים.

# מגמות בינלאומיות

בעולם מתפתחת מגמה נרחבת של שילובי בינה מלאכותית במגוון תחומי יישום. בד בבד עם החדירה הנרחבת של הבינה המלאכותית לכל תחומי חיינו, מתעוררות שאלות אתיות רבות לגבי השפעתם של כלים טכנולוגיים מבוססי בינה מלאכותית על האנושות. מכונות לומדות שפעולתן האוטונומית מוגדרת כאטומה ו/או בלתי מוסברת, לא רק מאתרות אנומליה אלא יוצרות מודל למידה חדש. כמענה לכך, הולכות ומתפתחות תוכניות מחקר ולימוד רבות המשלבות היבטים אתיים בכל הקשור לפיתוח יישומי בינה מלאכותית.

אחת התוכניות המובילות באוניברסיטת הרווארד היא תוכנית המשלבת לימודי אתיקה בשיעורים הנדסיים.  התוכנית מיושמת בהרווארד בפקולטה למדעי המחשב וזו גם משמעות הדגש על האותיות CS Embedded)[[9]](#footnote-9) EthiCS). בתעשייה, כמו באקדמיה, מתקיים תהליך הפנמת החשיבות הרבה של ההשלכות האתיות והחברתיות של בינה מלאכותית והצורך לקחת אותן בחשבון במשולב עם הפיתוח.

בהרווארד מציגים את טכנולוגיות המחשוב החדשניות המשפרות והמיטיבות לעתים קרובות, לצד יכולתן להשפיע גם בדרכים מזיקות או לא הוגנות. כך, למשל, מערכת COMPASS המשמשת ככלי עזר לשופטים בארה"ב לחזות את מידת המסוכנות של עבריינים העומדים בפני שחרור על תנאי, הוכחה כמפלה לרעה עבריינים שחורים. תוכניות לזיהוי פנים עשויות להאיץ ולקדם את ערוצי האבטחה, אך גם להפלות אוכלוסיות מסוימות - ולראיה, תוכנת זיהוי הפנים של גוגל שזיהתה אפרו-אמריקאים כגורילות. כלי רכב אוטונומיים עשויים לשחרר אותנו מטעות אנוש, אך גם להוות סיכונים חדשים לנהגים אנושיים ולהולכי רגל, ולראיה כתאונות לא מעטות המלוות את הפיתוחים של טסלה. בהרווארד מדגישים את הצורך לפיו מדעני המחשב של ימינו חייבים ללמוד לעצב מערכות אחראיות מבחינה מוסרית וחברתית. ואכן, התוכנית פועלת מתוך התחייבות שיתופית של הפקולטה למדעי המחשב והפקולטה לפילוסופיה בהרווארד. התוכנית  מלמדת סטודנטים למדעי המחשב כיצד לחשוב על השלכות האתיות והחברתיות של עבודתם וכיצד להתמודד איתן. התוכנית מיושמת היום בטכניון והיא פועלת במתכונת המחברת את כל הפקולטות ההנדסיות והמשלבת אתיקה בשיעורי הנדסה שונים.

כפי שניתן ללמוד מיעדי הפיתוח של SDG[[10]](#footnote-10) ותחזיותיו לשנת 2030, בעשור האחרון החברות התעשייתיות בעולם מחויבות מתוקף הסכמים בינלאומיים לאתיקה, לקיימות סביבתית וכלל עולמית. מגמה זו מתעצמת לאור ההשלכות של העולם התעשייתי על שינויי האקלים וסכנת הקיום העולמית. סכנה זו הפכה מוחשית ומיידית, ומתייחסת  ליקום  כאל משאב מתכלה שיש לשמרו למען הדורות הבאים.

ההצהרה של האגודה המדעית למחשוב ACM[[11]](#footnote-11), נועדה לתמוך ביתרונות של קבלת החלטות אלגוריתמיות תוך התייחסות לעקרונות שהאגודה מפרסמת. ביניהם: מודעות, הסבר, אימות, בדיקה ופיצוי במקרה הצורך. יותר ויותר אלגוריתמים מיישמים תהליכי קבלת החלטות מוסדיות המבוססות על ניתוח, הכולל גילוי, פרשנות ותקשורת של דפוסים משמעותיים בנתונים. ניתוח בתחומים עתירי נתונים מסתמך על יישום בו-זמני של סטטיסטיקה, תכנות מחשבים ומחקר תפעולי על מנת לכמת ביצועים. ישנן ראיות הולכות וגוברות לכך שהאלגוריתמים וניתוחים מסוימים הם כה סבוכים עד שבלתי ניתן יהיה להתחקות אחר מנגנון קבלת ההחלטות שלהם, ולכן לא ניתן יהיה לקבוע מתי התפוקות עשויות להיות מוטות או שגויות. באחריותם של המודלים המוסדיים שנבנים, לדאוג למענה לפרמטרים הנזכרים בהצהרה. כך, בכל שלב של פיתוח ופריסה של המערכת, במטרה למזער נזקים.

לאור מגמה עולמית זו ורבות אחרות, מתחזק הצורך בתוכניות לימוד רב-תחומיות המשלבות תחומי למידה הנדסיים והומניים, המשלימים זה את זה, כמתבקש מהעולם האקדמי הצופה את פני המחר. תוכנית הלימודים, הקורסים והמגמות במרחב שיוקם יכשירו את השילוב של דור העתיד, בוגרי הנדסה ומדעי הרוח, לאתגרי העתיד ויהוו תשתית לתעשייה בעולם הומני ואתי כיאה למאה ה-21.

# 

# יעדים ותוכנית עבודה

אוניברסיטת חיפה והטכניון במשותף רואים במרחב החדש הזדמנות לחולל שינוי חשוב, ואף הכרחי, פריצת דרך אקדמית במחקר ותוכניות לימודים המגשרות בין לימודי הנדסה ומדעים לבין מדעי הרוח.

השימוש בבינה מלאכותית מעלה שאלות עמוקות בנושאים של אתיקה, חברה, משפט, וממשל, ומאידך, בהכוונה והבנה נכונה, מתמודד עם האתגרים של איכות חיים, קבלת החלטות וקיימות. המרכז יתמקד במחקר והוראה רב-תחומיים ובינתחומיים של השימוש העכשווי ו/או העתידי בבינה מלאכותית: בתהליכי קבלת החלטות, בתכנון וניתוח יישומיים, בהשלכות אתיות וחברתיות של בינה מלאכותית, באתגרים רגולטוריים שהשיומש בבינה מלאכותית מעלה, ובהשלכות על הסביבה והחיים על פני כדור הארץ.

מטרת המרכז היא לחקור וללמד נושאים ושאלות פתוחות הנמצאים היום בחזית המחקר בבינה מלאכותית. המרכז יחשוף סטודנטים לתואר שני במדעי הרוח וההנדסה לפרקטיקה הקיימת והעתידית של בינה מלאכותית: השלכות, דילמות ואתגרים אתיים, חברתיים וסביבתיים המתעוררים בהקשר של בינה מלאכותית, שימוש במערכות חכמות בתחומי היצירה השונים, ועוד.

בין יעדי המרח"ב המשותף:

* אנשי סגל בכיר משני המוסדות ילמדו בתכניות המרכז. ההוראה תהיה בשילוב מכסת ההוראה הרגילה בפקולטות, ככל האפשר, בתלות בצרכי ההוראה בחוגים או בפקולטות.
* מלגות לפוסט-דוקטורנטים, לדוקטורנטים ולסטודנטים מצטיינים לתואר שני - יתוקצבו.
* פעילויות המחקר במרכז יכללו, מעבר למחקר השוטף, ימי עיון, סדנאות וכנסים.
* פעילויות שוטפות מיוחדות כגון סמינרים מחקריים משותפים יתקיימו במטרה לחזק את שיתוף הפעולה המחקרי בין חוקרי האוניברסיטה וחוקרי הטכניון.  למשל: על מנת להעמיק את האינטראקציה הבינתחומית. בכל אחת מתוכניות הלימוד מתוכננים סמינרים שכוללים הוראה כפולה או משולבת של מרצה מתחום ההנדסה ומרצה מתחום מדעי הרוח (או מדעי החברה / המשפטים / האמנויות).
* המרכז יעודד אנשי סגל בכירים לפתח קורסים חדשים, באמצעות תקציבים ייעודיים ובאמצעות הקלות בהוראה.
* הפיתוח הסופי של מרבית השיעורים החדשים יושלם במהלך 3-4 השנים הראשונות. הפיתוח יימשך ככל שיידרש גם מעבר לשנים אלה.
* תוכניות הלימודים תיפתחנה בסמסטר א' תשפ"ג. התארים (MA בינה מלאכותית ומדעי הרוח, כולל ההתמחויות) יהיו משותפים לאוניברסיטת חיפה ולטכניון. התארים יובאו לאישור הגורמים המוסמכים באוניברסיטת חיפה, הטכניון והמל"ג במהלך שנה"ל תשפ"א.
* צוותי המחקר יחלו בעבודתם לכל המאוחר בתחילת שנת הלימודים תשפ"ב. צורכי המחקר, גישה לספריות, מאגרי מידע ומעבדות יסופקו על ידי האוניברסיטה והטכניון בהתאם לתוכנית התקציבית. המגמה היא לגייס כבר במהלך תשפ"ב דוקטורנטים ופוסט-דוקטורנטים שיצטרפו לקבוצות המחקר.
* תוכניות עתידיות לפיתוח: תוכנית לתואר שני: בינה מלאכותית ואמנויות יצירה, תכנית דו-חוגית לתואר ראשון בבינה מלאכותית ואמנויות יצירה; תוכנית לימודים בינלאומית לתואר שני (הוראה באנגלית); למידה מקוונת בינ-מוסדית ובינלאומית.
* המרכז יפתח את שעריו לקהילה. כנסים והרצאות שיתקיימו במרכז יפורסמו בערוצי התקשורת של הקהילה ויזמינו את הקהל הרחב להרחבת אופקים.
* במרכז יתקיימו פעילויות שמטרתן לאפשר אינטראקציה ותקשורת בלתי אמצעית בין החוקרים.
* קשרי קהילה: המרכז יקיים אירועים משותפים עם הספריות העירוניות ועם המוזאונים בעיר חיפה ובסביבתה.
* מיקומו הפיזי של המרכז בטכניון ו/או באוניברסיטה ייקבע בעתיד. הרכבל החדש מקצר את הדרך שבין הקמפוסים לכדי דקות ספורות.

מבנה ארגוני

המרכז יהיה בינ-אוניברסיטאי משותף לאוניברסיטת חיפה והטכניון, ויוקם בהתאם לכללים שיידרשו על ידי המל"ג-ות"ת והוועדה לקידום מדעי הרוח.

# 

# תוכניות לימודים במרכז: תואר שני

מטרת תוכניות הלימוד של המרכז היא להקנות שפה משותפת לכל התחומים, ולהעניק חשיבה אתית כמיומנות חיונית עבור מהנדסים בכל תחומי ההנדסה. הנגשת מושגים תסייע לסטודנטים לסגל חשיבה אתית וחברתית בעבודתם.

המרכז יציע בשלב ראשון ארבע תוכניות לימודים לתואר שני. התוכניות תהיינה פתוחות לבוגרי כל המדעים השונים - מדעי הרוח, החברה, והמדעים המדויקים, ומתוך כך תייצרנה מפגש ושיח בינתחומי. ראו תנאי קבלה ודרישות השלמה ותוכניות עתידיות בהמשך.

1. בינה מלאכותית במדעי הרוח **(AI and humanities)**

כלי בינה מלאכותית ולמידת מכונה מוצאים את דרכם אל התחומים הקלאסיים של מדעי הרוח, בהם הם מסייעים בדיגיטציה של כתבי יד, סיוע בתיקון אוטומטי של שגיאות, תעתוק באמצעות כלי עיבוד שפה טבעית, ניתוח תכנים של כתבי יד, חיפוש ציטוטים ומובאות, יצירת היררכיות של כתבי יד על סמך זיהוי מקורות מוקדמים המצוטטים בכתבי יד מאוחרים, זיהוי אוטומטי של אלמנטים רטוריים בשירה, זיהוי ואימות מחברים על סמך סגנונות כתיבה.

היבט אחר הקשור לראייה ממוחשבת הוא חיבור חלקים: שברים של כלי חרס, פיסות של פפירוס תוך שילוב של ראיית מכונה, ואף כריית טקסט (בחיבור קרעי מסמכים), שחזור מבנים עתיקים מתוך הריסותיהם ועוד. התועלת הרבה של הטכנולוגיה, האמורה להקל על החוקר, מציבה אתגרים חדשים - עד כמה ניתן לסמוך על תוצאות התהליכים החישוביים המבוצעים על ידי הקופסה השחורה? עד כמה אמין זיהוי המחבר, והאם ניתן להסתמך על כך בקביעת מקורו של כתב יד? עד כמה ניתן לסמוך על הצעת שחזור מבנה שמציע המחשב? על החוקרים העושים שימוש בכלים הטכנולוגיים להכיר את מגבלותיהם, להבין אותם ולדעת להתמודד איתם.

מטרת התוכנית כפולה - מצד אחד, לחשוף בפני אנשי מדעי הרוח כלים טכנולוגיים שיכולים לסייע במחקרם, ומצד שני - לפתח יכולת ביקורתית מתוך מודעות למגבלותיהם ולהתמודד עם האתגרים החדשים שהם מעלים.

התוכנית תדגיש אתגרים אתיים ומתודולוגיים תוך שימוש בכלי בינה מלאכותית בחקר הרוח האנושית ובתחום היצירתיות החישובית: מודעות למתח המצוי בין חשיבה כמותית לבין תקשורת מילולית והמבט הייחודי של החוקר; שאלת עקרון המוצא והביסוס של הידע העומדת במרכז השיח במדעי הרוח הנתפסת בדרך כלל כשקופה עבור מנועי בינה מלאכותית גנרית; הצורך בהנגשת מאות שנות מסורת של לימוד מחקרי קלאסי מן הרנסאנס ועד ימינו, אשר כלי הבינה המלאכותית נוטים 'לדלג' על תוצריהן, לעיתים בשל היעדר היכרות והבנה מספקת, ולעיתים בשל העובדה שאלו אינם זמינים לשימוש חישובי. התוכנית תעסוק, למשל, במתחים שבין תהליכי הבניית מידע מבוססי פרשנות אנושית לבין כלי למידה עמוקה; בבניית מודעות לצורך שבתרגום של מאות שנים של תשתית ידע וחשיבה בתחומי מדעי הרוח והאמנויות; בשאלות יסוד בזיקה שבין יצירתיות חישובית לבין מודל האמן האינדיבידואל; ובשאלות יסוד בתחום הניצול האפשרי לרעה של יצירתיות חישובית בכלכלה  ובפוליטיקה.

1. בינה מלאכותית ואתיקה (AI and Ethics)

מערכות אלגוריתמיות הופכות לחלק מחיי היומיום שלנו. עקב ההשפעה המשמעותית שלהן על חיינו, נושא השקיפות וההוגנות של מערכות אלו מושך לאחרונה תשומת לב מחקרית רבה. גוברת ההכרה שהיבטים אתיים, כגון הוגנות, שקיפות או פרטיות, והגורמים המשפיעים עליהם, צריכים להיות חלק בלתי נפרד מתהליכי החשיבה והפיתוח של מערכות העושות שימוש בכלי בינה מלאכותית ליישומים יומיומיים.

התוכנית תתמקד בהצגת היבטים אתיים הכרוכים בפיתוח מערכות העושות שימוש בבינה מלאכותית, בסיכונים הקיימים בהן, ובכלים ובשיטות המאפשרים מניעה וגילוי הטיות במערכות אלגוריתמיות. בוגרי התוכנית ייחשפו ויבחנו לעומק את האתגרים האתיים המתעוררים בפיתוח מערכות חכמות של בינה מלאכותית, בעקבות שימוש רחב בהן, ובניתוח נתוני עתק. השילוב של סטודנטים בעלי רקע מגוון, מתחומי הטכנולוגיה והבינה המלאכותית, וממדעי הרוח והחברה, יאפשר לדון באתגרים הללו מפרספקטיבות רלוונטיות שונות, להבין ולהכיר את השאלות ואת ההיבטים השונים המחייבים פתרון. כמו כן, הם יוכלו לפתח בעבודת צוות מערכות אלגוריתמיות, תוך מתן מענה לאתגרים השונים, כאשר המטרה היא להיטיב עם החברה.

1. בינה מלאכותית וקיימות: חברה וסביבה (AI for Society and Environment**)**

תחום הבינה המלאכותית תופס תפקיד מרכזי במימוש של יעדים לפיתוח בר-קיימא. בינה מלאכותית משמשת כיום במגוון יישומים סביבתיים, כמו תכנון עירוני למיזעור פליטות פחמן, זיהוי דליפות נפט בים, חיזוי שינויים במפלס מי תהום, מיפוי ארוזיה של קרקעות ועוד. עם זאת, ברוב המקרים מדובר בפיתוחים מחקריים ראשוניים ולא בשימוש סטנדרטי / מסחרי. בעתיד, צפוי שנושאים כגון הגנת סביבה, פיתוח תשתיות אנרגיה ירוקה, התמודדות עם שינויי אקלים, כמו גם נושאים חברתיים מרכזיים כגון מניעת עוני, שיפור החינוך, קידום צמיחה ועידוד חדשנות - כל אלה יהיו תלויים ביכולות הבינה המלאכותית לשמור על אמינות, שקיפות ובטיחות בתהליכי הפיתוח. התוכנית תתמקד בקשר שבין קיימות וטכנולוגיה, ובעיקר באתגרים שמציבה הבינה המלאכותית לפיתוח חברתי וסביבתי בהיבטים פילוסופיים שונים.

מטרת התוכנית היא להקנות לבוגריה הבנה עמוקה של כלי בינה מלאכותית, יכולת פיתוח של יישומים חברתיים / סביבתיים באמצעות בינה מלאכותית, וכן לאפשר לבוגריה לשמש כסוכנים חברתיים בעלי השפעה חיובית על מדיניות השימוש בבינה מלאכותית. בוגרי התוכנית יוכלו להשתלב ברשתות שיתוף פעולה בינלאומיות, כמו גם בתעשיות שונות המקדמות פיתוח בינה מלאכותית, ולתרום לקידום יעדי פיתוח בר-קיימא שונים וליישומם, וכן לפתח מערכות טכנולוגיות שונות המוכוונות להתמודדות עם אתגרי קיימות שונים. לכל סטודנט/ית בתוכנית יהיו שני מנחים מתחומי ידע שונים, ומתוך מגמה - גם משני המוסדות המשתתפים בתוכנית, וזאת על מנת לקדם את הרב-תחומיות ההכרחית בתוכנית זו.

1. בינה מלאכותית וקבלת החלטות בממשל ובעסקים  (AI and Decision Making in Law and Business)

התמורות הטכנולוגיות בתחום הבינה המלאכותית לא פסחו על  תחומי המשפט, הממשל והעסקים. כפי שתואר למעלה, המגזר העסקי מניע תהליכי אלגורתימזציה בתחומים רבים, ולצד התועלות הטמונות באיסוף ובשימוש מושכל במידע, צפים סיכונים. מחקרים רבים בהובלת וביוזמת מדינות באיחוד האירופי נועדו להתמודד עם סיכונים אלו, ובעקבותיהם מאומצות מסגרות רגולציה המטילות חובה על עסקים להעריך את הסיכונים הכרוכים בפיתוח ובשימוש באלגוריתמים, עוד בטרם מסחור המוצרים והשירותים. במאי 2018 הוביל האיחוד האירופי רגולציה שמטרתה להתמודד עם בעיית הפרטיות, ה-DGPR, אשר משליכה על תחום האיסוף, האגירה והניתוח של נתונים, ולכן גם על תחום הבינה המלאכותית. בנוסף, הצעת הרגולציה של האיחוד האירופי בנושא בינה מלאכותית, אשר נמצאת בשלבי קבלה, תחייב כל עסק להסביר בפירוט את תפקוד הבינה המלאכותית, ולהוכיח שהוא אינו מסכן את האינטרסים ואת זכויות הציבור על פי חוקי האיחוד האירופי. לצד אלו, הבינה המלאכותית מעלה שאלות של זכויות יוצרים הנוגעות למידע ולאלגוריתמים שיפותחו, וגם בתנאים להפקעת זכויות אלו או לשיתוף במידע - שיתוף אשר מתבקש על מנת שניתן יהיה לבקר את דרכי קבלת ההחלטות, וכן להתגונן מפני איומים ושימוש לרעה.

משטרים משפטיים רבים אחרים מושפעים מפיתוח הבינה המלאכותית - החל ממשטרי הנזיקין (מהו האלגוריתם הסביר?), וכן בסוגיות מתחומי הענישה, הראיות, החוזים והמשפט הציבורי, אשר בכולם מוצאים אנו כיום שימוש באלגוריתמים. מבחינה תורת-המשפט, אחת השאלות המהותיות לדיון זה היא: האם מכונה יכולה להיות סוכן מוסרי מושלם [AMA - artificial moral agents]? ומכאן: כיצד יש לחשוב על משטרים משפטיים שמבוססים על הסכמה או רצון?

בתקופה זו רגולטורים שונים ניצבים בפני אתגרים הנובעים מאחריותם לפתח שיטות ומשטרים משפטיים שבאמצעותם ניתן לאמת את כלי הבינה המלאכותית המיושמים על ידי התאגידים ואף על ידי הרשויות. כבר כעת מפותחות ומוטמעות מערכות שנועדו לסייע בקבלת החלטות בתחומי ממשל רבים, הן ברשות המחוקקת, הן ברשות המבצעת, והן ברשות השופטת, ולא מן הנמנע שחלק מהמערכות הללו תתפתחנה לכדי מערכות שמצריכות מעורבות אנושית שולית או פיקוחית בלבד. כך, למשל, מתקיימים כעת דיונים באשר לאפשרות של החלפת תפקיד השופט המסורתי במכונה שמנגנון קבלת ההחלטות שלה אינו נהיר, לצד דיון בהטיות בתהליך איסוף וניתוח המידע אשר מוביל לפגיעה בשוויון, וכל זאת במקביל לדיון  בכלים שתכליתם להבטיח תיעוד ובקרה, על מנת למזער הטיות ושימוש לרעה.

מעבר לרובד העסקי והרגולטורי, תעסוק התוכנית גם באתגרים שמציבה הבינה המלאכותית למשטר הדמוקרטי עצמו, לאור המניפולציות שבינה מלאכותית מאפשרת, הן בהיבטי השליטה ב"חדרי מידע", והן בהיבטי ריכוז הכוח (הדאטה) במוקדים ספציפיים.

בזירה הבינלאומית התפתחות הבינה המלאכותית הובילה לתובנות המשליכות על מאזן האימה, המתייחס לפיתוח מערכות נשק אוטונומיות שמשנות את פני הזירה הבינלאומית, ומזמנות דיונים בערכאות בינלאומיות. קובעי מדיניות יאלצו לתחזק מערכות אנליטיות, הווה אומר: מנגנונים לניתוח נתונים של תהליכי קבלת החלטות המדמות את סטנדרט תהליך קבלת ההחלטות האנושי, המסורתי.

**תכניות עתידיות לפיתוח (בשנתיים-שלוש הראשונות להרצת המרכז):** בכוונתנו לפתח בשנתיים-שלוש הראשונות של המרכז שתי תכניות לימוד: האחת, תוכנית ייעודית נוספת לתואר שני בנושא של בינה מלאכותית ואמנויות יצירה: אמנות פלסטית, תיאטרון, מוסיקה, ספרות, שירה. השניה, תוכנית דו-חוגית לתואר **ראשון** בבינה מלאכותית ואמנויות יצירה. בהקשר זה, של בינה מלאכותית ואמנויות, מתעוררות שאלות ספציפיות וביתר חדות (יחסית לתחומי התוכניות האחרות שלעיל), באשר לטבעה של החשיבה האנושית לעומת טבעה האלגוריתמי של החשיבה המלאכותית. ברור כי לחשיבה האנושית שני מאפיינים חשובים: מצד אחד, חשיבה דדוקטיבית-לוגית, ומצד שני חשיבה אינדוקטיבית-יצירתית. המרכיב הדדוקטיבי הראשון הוא מטבעו אלגוריתמי וניתן באופן עקרוני לשחזור וחיקוי מלא על ידי מערכת מלאכותית (קורט גדל, כמו גם רוג'ר פנרוז ואחרים חולקים על קביעה זו משיקולים הקשורים בהוכחות אי-השלמות של גדל). לעומת זאת, המרכיב השני האינדוקטיבי-נורמטיבי (שקיים גם בהקשר של אתיקה, למשל) דורש הכללה מאוסף מקרים פרטיים אל הכלל, והכללה מסוג זה באופן עקרוני אינה ניתנת לשחזור מלא על ידי חישוב, לפחות ללא קביעה מראש של הפרמטרים לפיהן ההכללה צריכה להתבצע (עקרונית קיימים אינסוף פרמטרים הניתנים להכללה מתוך אוסף סופי של מקרים- זו הסיבה העיקרית לכשלונה של מערכת פורמלית אינדוקטיבית! ראה, למשל, פרדוקס אבני הברקת של נלסון גודמן). אולם, קביעה זו אינה יכולה להתקבל כתוצאה של ה"חישוב" האינדוקטיבי, מכיוון שהיא עצמה קובעת את החישוב האינדוקטיבי. אמנם, ייתכן כי מערכת אלגוריתמית יכולה לחקות במידה טובה של הצלחה הכללות אינדוקטיביות שגורות המאפיינות את החשיבה האנושית הסטנדרטית בהקשרים הרגילים, למשל אלה המוכרים לנו מחיי היום-יום. אולם חיקוי מסוג זה תלוי בידיעה מראש של ההקשרים הרלוונטיים, בעוד (סביר להניח) שדווקא האספקט היצירתי של החשיבה האנושית אינו מוגבל למקרים בהם ההקשרים הרלוונטיים ידועים מראש. בנוסף לכך, האספקט היצירתי של החשיבה האנושית, אספקט שבא לידי ביטוי באמנויות היצירה, כמו גם באופן מעט שונה במדעים, עדיין לא מובן כמעט במלואו: האם מדובר באינטואיציה, סוג של רגישות (או אולי העדר-רגישות?) לסביבה החברתית, או סוג של חישה, חוש שישי? האם אינטואיציה זו אינה אלא סוג של ניחוש? אם כן, האם הניחוש שרירותי (במובן זה, גם מערכת אלגוריתמית יכולה "לנחש") או שמא הניחוש עצמו מונחה על ידי כללים (אינדוקטיביים או אחרים?), ואם כן, מה קובע אותם? האם ניתן להטמיע כללים כאלה במערכת אלגוריתמית לומדת? שתי תכניות הלימודים בהקשר של בינה מלאכותית ואמנויות (לתואר ראשון ולתואר שני) יעסקו בשאלות מסוג זה, ברמות שונות, בהבדלים, כמו גם בקשרי הגומלין, בין חשיבה אלגוריתמית לבין חשיבה יצירתית.

# 

# תנאי קבלה לתוכניות הלימודים

תנאי קבלה לתוכניות הלימודים במרכז משותפות לכל התוכניות.

לתוכניות יתקבלו בוגרי תואר ראשון בכל התחומים, שסיימו תואר ראשון בממוצע ציונים התואם את דרישות הקבלה למסלול אותו הם בוחרים (מסלול מחקרי עם תזה או מסלול לא-מחקרי ללא תזה, ראו למטה). מועמדים שאינם עומדים בקריטריונים של הקבלה יזומנו לדיון בוועדת הקבלה. הישגים מקצועיים של בעלי ניסיון מעשי ומכתבי המלצה של המועמדים יילקחו בחשבון בעת הדיון על הקבלה.

# 

# מבנה הלימודים

לימודי התואר השני מיועדים להעמיק את ידיעותיו של התלמיד, ולאפשר התמחות באחת מתוכניות הלימודים ובמסלול נבחר (מתוך שניים): מסלול לימודים הכולל כתיבת עבודת גמר (מסלול א') או מסלול לימודים הכולל פרויקט גמר, כגון מבחן או עבודה (מסלול ב').

תנאי קבלה

**מסלול א' (עם תזה)**: תואר ראשון ממוסד מוכר להשכלה גבוהה בארץ או בחו"ל, בציון סופי של 85 לפחות, ציון סופי של 80 לפחות בחוג השני, ואישור הוועדה החוגית לתואר שני.

**מסלול ב' (ללא תזה)**: תואר ראשון ממוסד מוכר להשכלה גבוהה בארץ או בחו"ל, בציון סופי של 80 לפחות, וציון סופי של 76 לפחות בחוג השני.

המועמדים יגישו בקשה לקבלה במזכירות החוג.

מועמדים בעלי תואר ראשון ממוסד מוכר להשכלה גבוהה, ששפת ההוראה בו אינה עברית, חייבים לעמוד בבחינת ידע בעברית ברמה הנדרשת על-ידי האוניברסיטה כתנאי קבלה.

מועמדים בוגרי אוניברסיטאות מוכרות מחו"ל, וכן בוגרי אוניברסיטאות בעלות שלוחות בישראל, אשר פועלות באישור המועצה להשכלה גבוהה, יחויבו לעמוד בהצלחה במבחן GRE.

תוכנית לימודים לתואר שני במסלול עם תזה

**סה"כ נ"ז 28 + הצעת מחקר + תזה:**

**היקף השיעורים:** התלמידים ילמדו שיעורים בהיקף כולל של 28 שש"ס, מתוכן לכל היותר 6 שש"ס בדרג 3, והיתרה בדרג 4. סדנת עבודות הגמר היא שיעור חובה. ניתן לקחת שיעור רלוונטי, עד 4 שש"ס, מחוץ לחוג (באישור יו"ר הוועדה החוגית לתואר שני). ניתן להשתתף בקולוקוויום החוגי במשך שנה אחת (2 שש"ס).

הוועדה החוגית לתואר שני יכולה לחייב את התלמידים בלימודי שפה זרה שנייה, בהתאם לנושא עבודת הגמר המחקרית.

**עבודות סמינריוניות:** בנוסף לתנאי המעבר של כל שיעור, התלמידים יגישו שתי עבודות סמינריוניות אצל מורים שונים, בסמינריונים מדרג 4.

**תנאי מעבר משנה א' לשנה ב':** התלמידים חייבים לצבור 16 שש"ס בשנה א', בציון ממוצע של לפחות 90, ולמצוא מדריך לתזה. תלמידים שלא יעמדו בתנאים אלה, יעברו בשנה ב' ללימודים אל מסלול ב'. במקרים מיוחדים, הוועדה החוגית לתואר שני רשאית לחרוג מכללים אלה בהתאם לשיקול דעתה.

**עבודת גמר מחקרית:** התלמידים נדרשים לכתוב עבודת גמר מחקרית בהדרכת חבר סגל.

על התלמידים למצוא מדריך לעבודתם ביוזמתם. עליהם להגיש את הצעת המחקר לעבודה לאישור המדריך, ואחר כך לאישור הוועדה החוגית לתואר שני, עד תום הסמסטר הראשון בשנה ב'.

שקלול ציון סופי:

50% - תזה, 20% - שתי העבודות הסמינריוניות (במשקל שווה), ו-30% - ממוצע משוקלל של שיעורי החובה והבחירה.

תוכנית לימודים לתואר שני במסלול ללא תזה:

סה"כ נ.ז - 32 + פרויקט גמר

היקף השיעורים: התלמידים ילמדו שיעורים בהיקף כולל של 32 שש"ס, מתוכן לכל היותר 6 שש"ס בדרג 3, והיתרה בדרג 4. ניתן לקחת שיעור רלוונטי, עד 4 שש"ס, מחוץ לחוג (באישור יו"ר הוועדה החוגית לתואר שני). ניתן להשתתף בקולוקוויום החוגי במשך שנה אחת (2 שש"ס).

עבודות סמינריוניות: בנוסף לתנאי המעבר של כל שיעור, התלמידים יגישו שלוש עבודות סמינריוניות אצל מורים שונים, בסמינריונים מדרג 4.

תנאי מעבר משנה א' לשנה ב': התלמידים חייבים לצבור 16 שש"ס בשנה א', בציון ממוצע של לפחות 80.

במהלך שנה ב' ללימודי התואר, התלמידים יכתבו פרוייקט גמר בהדרכת מדריך מצוות התכנית, לאחר שקיבלו אישור לפרוייקט מראש המסלול.

**שקלול ציון סופי:**

30% - שלוש עבודות סמינריוניות (במשקל שווה), 50% - ממוצע משוקלל של שיעורי החובה והבחירה; 20% - פרוייקט גמר

תוכניות לימודים

שיעורי השלמה לכל התוכניות (שאינם לתואר):

(א) מבוא לאתיקה קיים

(ב) אתיקה ופרטיות חדש

(ג) זרמים מרכזיים במדעי הרוח והחברה מותאם

(ד) חשיבה כמותית ואלגוריתמית חדש

שיעורי חובה משותפים לכל התוכניות:

(א) שקיפות אלגוריתמית - סמינר קיים

(ב) האם מכונות יכולות לחשוב - סמינר מותאם

(ג) בין מידע מובנה לבין למידה עמוקה - סמינר חדש

(ד) סמינר לכתיבת תזה במסלול מחקרי מותאם

(ה) סמינר מחקר - מרצים אורחים, נושאים משתנים - שנתי חדש

|  |
| --- |
| **מקרא: קורס** חדש **- קורס חדש (לפיתוח) שאינו קיים**  **קורס** קיים **- קורס קיים ומתאים לתוכנית**  **קורס** מותאם **- קורס קיים מותאם לתוכנית ( לתואר שני)**  **1 נק"ז =1 שש"ס** |

# תוכנית מס. 1: בינה מלאכותית במדעי הרוח (**AI and humanities)**

קורסי חובה בתוכניות:

(א) יסודות מדעי הרוח הדיגיטליים: כלי ניתוח והמחשה קיים

(ב) יצירתיות חישובית חדש

(ג) קורס פרויקט גמר חדש

קורסי בחירה בתוכניות:

(ד) אתיקה של נתונים חדש

(ה) חקר חישובי של התנהגות אנושית מותאם

(ו) מבוא ללוגיקה חדש

(ז) יסודות התכנות בשפת פייתון למדעי הרוח והחברה חדש

(ח) מבוא ליישומי בינה מלאכותית במדעי הרוח חדש

(ט) יישומי ראיה ממוחשבת במדעי הרוח הדיגיטליים מותאם

(י) יישומי כריית טקסט במדעי הרוח הדיגיטליים מותאם

(יא) למידה עמוקה קיים

(יב) מבוא למדעי הנתונים מותאם

# תוכנית מס. 2: בינה מלאכותית ואתיקה (**AI and ethics)**

קורסי חובה בתוכניות:

(א) אתיקה למהנדסים חדש

(ב) בינה מלאכותית, חברה וכוח חדש

(ג) קורס פרויקט גמר חדש

קורסי בחירה בתוכניות:

(ד) אתיקת מכונה חדש

(ה) תקשורת/אינטראקציית אדם ומכונה חדש

(ו) מבוא למדעי הנתונים חדש

(ז) תכנות בסיסי לחקר נתונים מותאם

(ח) כלים ושיטות למניעת וגילוי הטיות במערכות אלגוריתמיות מותאם

(ט) אתיקה של מכונות חדש

(י) שילוב אתיקה בתהליך פיתוח מוצר חדש

(יא) תכנות בסיסי לחקר נתונים מותאם

# 

# תוכנית מס. 3: בינה מלאכותית וקיימות: חברה **וסביבה ( AI for society and environment)**

קורסי חובה בתוכניות:

(א) הסביבה בעידן הטכנולוגי חדש

(ב) אקו-פילוסופיה: חשיבה וסביבה מותאם

(ג) קורס פרויקט גמר חדש

קורסי בחירה בתוכניות:

(ד) הרשת כסביבה לימודית קיים

(ה)  אתיקה בפעולה חדש

(ו) משבר האקלים: סיבות, משמעויות ופתרון קיים

(ז) בינה מלאכותית וקיימות חברתית חדש

(ח) מדעי חברה חישוביים מותאם

(ט) בינה אנושית ומלאכותית - המודל ההיברידי חדש

(י) בין איכות חיים רמת חיים חדש

(יא) מבוא למדעי הנתונים מותאם

(יב) תכנות בסיסי לחקר נתונים חדש

# תוכנית מס. 4: בינה מלאכותית וקבלת החלטות בממשל ובעסקים **(AI and Decision Making in Law and Business)**

קורסי חובה בתוכניות:

(א) בינה מלאכותית: חוק, משפט ורגולציה מותאם

(ב) מקרי מבחן: כלי בינה מלאכותית בשדה המשפט והעסקי מותאם

(ג) קורס פרויקט גמר חדש

קורסי בחירה בתוכניות:

(ד) כלכלת המידע מותאם

(ה) מערכות בינה מלאכותית למשפטנים מותאם

(ו) טכנולוגיה, משפט, וצדק דיגיטלי מותאם

(ז) שוויון ואפליה בממשקי אדם-מכונה מותאם

(ח) בינה מלאכותית במשפט הבינלאומי מותאם

(ט)  בינה מלאכותית וסייבר קיים

(י) הבינה המלאכותית והמשטר הדמוקרטי מותאם

(יא) אתיקה בעסקים מותאם

(יב) כלים טכנולוגיים במשפט הפלילי מותאם

(יג) בינה מלאכותית אתיקה ומשפט מותאם

תוכנית לימודים: שיעורי השלמה

(א) מבוא לאתיקה. 4 נק"ז

מבוא היסטורי לתורת המוסר, מהפילוסופיה היוונית ועד לפילוסופיה של ה-20. הקורס ידון בגישותיהם המוסריות של אפלטון, אריסטו, הובס, רוסו, יום, קאנט, מיל, ניטשה ורולס. באמצעות העיון בשיטות אלה נכיר את ההבחנות והמושגים הבסיסיים בפילוסופיה של המוסר.

(ב) אתיקה ופרטיות. 4 נק"ז

בתחום האתיקה יוצגו מושגי יסוד כגון זכויות ומקורן וטיעונים לביסוס הזכות לפרטיות. בהיבט התיאורטי תבוסס זכות זו בשני מודלים - המודל הסוציולוגי ותורת המשחקים. בהיבט המעשי יוצג מגוון גדול של מקרי מבחן שינותחו בכיתה.

בתחום הפרטיות, הסטודנטים ילמדו רעיונות ומודלים סביב מנגנונים טכנולוגיים לשמירה על פרטיות. במהלך הקורס ייחשפו הסטודנטים למגוון כלים ואלגוריתמים שנועדו לשמור על פרטיות המשתמשים (תוך התחשבות בהיבטים של אבטחת מידע). במהלך הקורס יידרשו הסטודנטים להתנסות בכלים שונים וליישם אלגוריתמים של פרטיות.

(ג) זרמים מרכזיים במדעי הרוח והחברה (לבוגרי המדעים המדויקים). 4 נק"ז

מדעי הרוח והחברה עוסקים בכל ההיבטים שהופכים אותנו לאנושיים - היסטוריה, פילוסופיה, סוציולוגיה, פסיכולוגיה, אמנות, בלשנות ועוד. הקורס יעסוק בשאלות גדולות ובמגמות מרכזיות, בחשיבה במדעי הרוח והחברה היום ובגלגוליהם ההיסטוריים.

(ד) חשיבה כמותית ואלגוריתמית. 4 נק"ז

הקורס עוסק במושגי יסוד בפיתוח וניתוח של אלגוריתמים קומבינטוריים מתוך נקודת מבט תיאורטית ודגש רעיוני. נושאים שחלקם יילמדו: אלגוריתמי מיון, אלגוריתמים בסיסיים בגרפים: סריקות, מסלולים קצרים ביותר, עץ פורש מינימום, זרימה ברשתות, עצי החלטה ורגרסיה. שיטות חישוביות לייצוג ידע, מערכות לומדות. מטרת הקורס להקנות הבנה בסיסית בעקרונות יסוד מרכזיים של חשיבה כמותית ואלגוריתמית גם לחסרי רקע בתחום.

סמינרים משותפים לכל התוכניות

חובה לכל התוכנית:

(א) שקיפות אלגוריתמית – סמינר. 4: נק"ז

הקורס הוא סמינר מתקדם העוסק באחת השאלות המטרידות הן את הקהל הרחב, הן את אנשי המקצוע והן את מקבלי ההחלטות: כיצד להפוך את טכנולוגיות המחשוב, ובעיקר את אלו המבוססות על בינה מלאכותית, להוגנות, ידידותיות ומובנות יותר למשתמשיהן? במהלך המחצית הראשונה של הקורס נסקור ונלמד מחקרים בתחומים הקשורים להוגנות של מערכות ממוחשבות ול"שקיפות אלגוריתמית", בהמשך יתבקשו הסטודנטים להכין ולהציג סקירה מעמיקה של נושא אחד מתחומי הלימוד. כמו כן יתבקשו הסטודנטים לאתר מאמרים הקשורים לכל אחד מן הנושאים הנלמדים בקורס, כהכנה לשיעור שבועי.

(ב) האם מכונות יכולות לחשוב - סמינר. 4 נק"ז

מאז שהוצבה על ידי אלן טיורינג, שאלה זו מעסיקה הוגי דעות, חוקרים, סופרים ואנשי אמנות, כמו גם מפתחי בינות מלאכותיות. סמינר מקיף זה יעסוק בשאלות אודות טבעה של החשיבה האנושית, כגון: מהי היכולת לחשוב? האם החשיבה האנושית ניתנת לחיקוי על ידי מחשב? טיעון גדל, לוקאס ופנרוז, המשיב בשלילה לשאלה זו. האם מחשבים יכולים לייצר משמעות? בהקשר זה: מבחן טיורינג ופרדוקס "החדר הסיני". כיצד החשיבה האנושית מתפתחת? במה דומה ההתפתחות של הפעילות החישובית של בינה מלאכותית לתהליכי התפתחות של ילדים?

(ג) בין מידע מובנה ולמידה עמוקה - סמינר. 4 נק"ז

הקורס יורכב מהרצאות בהן נלמד נושאים עיקריים בלמידה עמוקה ותרגול שיעסוק במימוש של רשתות ורכישת ניסיון בספריות ללמידה עמוקה Pytorch. הקורס יתחיל במושגים ובאלגוריתמים בסיסיים, בפרט שיטות אופטימיזציה מתקדמות, רשתות עמוקות לרצפים של נתונים ועוד.

(ד) סמינר לכתיבת תזה במסלול המחקרי / לחילופין: סדנת פרויקט גמר למסלול לא מחקרי: מרעיון לפרסום מדעי: מיומנויות מידע לתלמידי מחקר. 2 נק"ז

קורס זה עונה על צורך שחוצה את כל תחומי הלימוד והמחקר באוניברסיטה. הצורך להשתמש במידע אקדמי בצורה מושכלת ואתית. מסיבה זו הקורס מפותח על ידי הספרייה והיחידה להוראה מתוקשבת בשיתוף חברי סגל מפקולטות שונות. על בסיס קורס זה יכול כל חוג לפתח תכנים ייחודיים, וכן להזמין תכנים נוספים בספרייה.

(ה) סמינר מחקר - מרצים אורחים, נושאים משתנים - שנתי

שיעורי התמחות לפי תוכנית

(1) בינה מלאכותית במדעי הרוח: **(Digital Humanities**)

קורסי חובה בתוכנית:

(א) יסודות מדעי הרוח הדיגיטליים: כלי ניתוח והמחשה. 212.8331 4 נק"ז

מדעי הרוח הדיגיטליים הם התחומים המתפתחים במדעי הרוח, והנותנים כלים מחקריים ובסיסיים לכל חוקר במדעי הרוח. בקורס נכיר את הנושאים בהם עוסקים מדעי הרוח הדיגיטליים, נפגוש פרויקטים מרכזיים, ואת המושגים המרכזיים בתחום. נפגוש כלי ניתוח והמחשה בסיסיים בתחום חקר הטקסט וחקר הרשתות, ונתרגל את השימוש בהם. מתוך כך, נדון בשאלה: כיצד החשיבה הכמותית ממשיכה גישות קיימות במדעי הרוח מחד, אך גם נותנת פרספקטיבה חדשה למבט על מדעי הרוח ככללם. ונשאל את השאלה – האם החיבור בין מדעי הרוח ודיגיטליות יכול גם לשנות תפיסות של אנשים העוסקים בעולם הדיגיטלי. הקורס יעניק לסטודנט כלים בסיסיים להכרת התחום הדיגיטלי.

(ב) יצירתיות חישובית. 4 נק"ז

יצירתיות חישובית הוא תחום חדש העוסק בשימוש בכלי בינה מלאכותית בתחומי היצירה השונים: כתיבה, אמנות פלסטית, מוזיקה, תיאטרון, קולנוע וטלוויזיה. הקורס יקנה יסודות מעשיים ותיאורטיים בתחום, יקנה לתלמידים יכולות שימוש במוצרי מדף בתחום וניתוח תוצרים של יצירתיות חישובית. הקורס מדגיש מודים שונים של שימוש ביצירתיות חישובית ויעמוד על האבחנות בין תפיסות של המכונה כחלופה לאדם לבין תפיסות של המכונה ככלי בידי היוצר האנושי, כמו גם בין השימוש הנסייני אמנותי ביצירתיות חישובית לבין השימוש התעשייתי והכלכלי ביצירתיות חישובית, אפשרויות ניצול לרעה בשדות החברה, הכלכלה והפוליטיקה והיחסים הבין אישיים.

(ג) קורס פרויקט גמר. 4 נק"ז

קורס פרויקט הגמר במסלול זה יהווה מרחב לימודי ייחודי שייצור סינרגיה בין אנשי הרוח והחברה לבין בוגרי המסלולים הטכנולוגיים. הפרויקטים יבוצעו בצוותים הטרוגניים ויונחו במשותף על ידי חוקרים מן התחום הטכנולוגי וחוקרים ממדעי הרוח / פרופסור אמן מחוגי האמנות, מוסיקה ותיאטרון במתכונת PBL - Project Based Learning.

קורסי בחירה בתוכנית:

(ד) אתיקה של נתונים. 4 נק"ז

הקורס יעסוק בסוגיות ייחודיות וכבדות משקל המשלבות בין עולמות האתיקה לבין עולם הטכנולוגיה. על מנת לזהות סוגיות אלו, לנתחן ולהציע להן פתרונות, נדרשת היכרות עם שני סוגי העולמות. מטרת העל של הקורס הינה לייצר דיאלוג בין שני העולמות, באופן שיניח את הבסיס ליצירת שפה משותפת.

(ה) חקר חישובי של התנהגות אנושית (מדמ"ח) . 4 נק"ז

קורס מעבדה בו סטודנטים יבצעו פרויקט מחקרי בתחום חקר חישובי של התנהגות חישובית. הפרויקט יכלול הקמת מערכת ושימוש בטכנולוגיות והרצת פיילוט ניסויי עם ניתוח תוצאות. חלק מהפרויקטים יהיו בשיתוף עם חוקרים אחרים באוניברסיטה (פסיכולוגיה, פיזיותרפיה, חינוך ועוד) ויתכן ויתבצעו בשטח המעבדה שלהם. דוגמאות לפרויקטים: מעקב אחר תנועות פנים, ידיים, ניתוח תנועות גופניות, מעקב אחר תנועות עיניים בעת ביצוע משימה ועוד.

(ו) מבוא ללוגיקה. 4 נק"ז

הקורס יטפל בנושאי יסוד בלוגיקה: תחשיב הפסוקים, תחשיב הפרדיקטים, סינטקס וסמנטיקה, שפה, נוסחאות נורמליות, תורת הרברנד, מבנים של שפה מסדר ראשון, הוכחות פורמליות בעזרת רזולוציות, משפט הקומפקטיות ושימושיו, משפט והשלמות, כריעות וחישוביות.

(ז) יסודות התכנות בשפת פייתון למדעי הרוח והחברה. 4 נק"ז

הקורס יקנה יסודות תכנות בשפת פייתון לבוגרי מדעי הרוח והחברה. הקורס מיועד לכלל הסטודנטים בתוכנית. מבנה המחשב. חומרה ותוכנה. ייצוג נתונים. מושגי יסוד ותכנות בשפת פייתון: ערכים, משתנים, ביטויים, טיפוסי נתונים בפייתון, התניות, חוגים (לולאות), פונקציות, רקורסיה, קלט/פלט (טקסטואלי וגרפי). סביבת הפיתוח של פייתון: שטח העבודה, הרצת תוכניות, ניפוי שגיאות. פיתוח מובנה של תוכניות, ארגון מידע במחשב. אלגוריתמים: פיתוח מונחה נכונות וסיבוכיות בסיסית. פתרונות אלגוריתמים למיון, חיפוש ובעיות במספרים שלמים.

(ח) מבוא ליישומי בינה מלאכותית במדעי הרוח. 4 נק"ז

הקורס מיועד לבוגרי מדעי הרוח והחברה בתוכנית. הקורס יבהיר את עקרונות הבינה המלאכותית והלמידה העמוקה, ידגים וינתח פרויקטים מחקריים בתחום, יקנה יכולת שימוש בסיסי בכלי מדף בתחום. מהבחינה התיאורטית, יתמקד התחום בשאלות מתודולוגיות יסודיות בדבר ההבדל בין דגמי חשיבה כמותית לחשיבה מילולית ויישום המסורת הביקורתית האקדמית של תחומי הרוח בעת מימוש פרויקטים מבוססי בינה מלאכותית.

(ט) יישומי ראיה ממוחשבת במדעי הרוח הדיגיטליים. 4 נק"ז

הקורס יתמקד ביישומי ראיה ממוחשבת במדעי הרוח הדיגיטליים, בעיקר בשדות של ארכיאולוגיה ותולדות האמנות, כגון מיפוי ממוחשב של חפירות, דימות ממצאים ארכיאולוגיים ושיטות למחקר השוואתי של מידע ויזואלי מדומת, סיווג ואפיון מאפייני חומר, מצע, פרקטיקות יצירה, נרטיב ביצירות אמנות, ניתוח גיאוגרפי כרונולוגי התפתחותי של מידע ויזואלי וכיוצא בזה.

(י) יישומי כריית טקסט במדעי הרוח הדיגיטליים. 4 נק"ז

הקורס יסקור באופן תיאורטי ומעשי מושגי בסיס בעיבוד דאטה טקסטואלי. נתייחס לשלבים השונים הכרוכים בעיבוד וניתוח סטטיסטי של טקסט בקנה מידה גדול. לצורך זה, נסקור ייצוגים שונים של טקסט, נבצע הסרת רעשים, והפרדה למשפטים ומילים. כמו כן, נעסוק בתיוג הטקסט במימדים של מבנה תחבירי ומידע סמנטי. הקורס יקנה מיומנות ביישומים שונים בהם שימוש בלקסיקונים, למשל לצורך ביצוע הסקה של ביטויים רגשיים (sentiment analysis), ואיתור מבוסס-חוקים של פרטי מידע רלבנטיים בטקסט (information extraction). נושאים מתקדמים שייכללו, במידת האפשר, הם איתור של ביטויים אופיינים בטקסט בשיטות סטטיסטיות, וייצוג תמציתי של טקסט באמצעות נושאים סמנטיים. במסגרת הקורס נעשה שימוש פעיל בשפת פייתון הפופולרית לעיבוד טקסט. ידע רלבנטי בשפה זו יוקנה במסגרת הקורס ואינו נדרש מראש. הקורס אינו עוסק בנושאים של עיבוד טקסט בשיטות של למידה. עם זאת, שילוב של קורס זה עם קורסים בנושאי למידה ממוכנת יאפשר שימוש מושכל בשיטות אלו.

(יא) למידה עמוקה, 203.3834. 4 נק"ז

הקורס יורכב מהרצאות בהן נלמד נושאים עיקריים בלמידה עמוקה ותרגול שיעסוק במימוש של רשתות ורכישת ניסיון בספריות ללמידה עמוקה Pytorch. הקורס יתחיל במושגים אלגוריתמים בסיסיים.

(יב) מבוא למדעי הנתונים. 4 נק"ז

הקורס יקנה היכרות עם עולם מדעי הנתונים וטכנולוגיית המחשוב כידע בסיסי הנדרש לסטודנטים מחוגים שאינם טכנולוגיים, ואשר יכשיר אותם לעשות שימוש בכלים של מדעי הנתונים. תינתן עזרה ייעודית בכל הקשור להתקנת כלי תוכנה רלוונטיים הנדרשים לקורס.

# 

(2) בינה מלאכותית ואתיקה **(AI and ethics)**

קורסי חובה בתוכנית:

(א) אתיקה למהנדסים. 4 נק"ז

הקורס יציג את תחום האתיקה ההנדסית מהיבטים רב-תרבותיים ורב-תחומיים שונים. הסוגיות התיאורטיות יעסקו באתיקה ובפילוסופיה של ההנדסה, תוך השוואה בין עמדות תרבותיות שונות על תחומים אלה. הלומדים בקורס יחשפו מתודולוגיות מחקר רלוונטיות ויבחנו מקרי מחקר המדגימים סוגיות תיאורטיות שונות כגון פתרון בעיות אתיות, כמו גם זיהוין של בעיות אתיות בתרבויות שונות, תוך עיון בטקסטים היסטוריים, ספרותיים ואמנותיים שונים המשליכים אור על תחום האתיקה למהנדסים.

(ב) בינה מלאכותית, חברה וכוח - סוגיות מתקדמות באתיקה למעצבי טכנולוגיות. 4 נק"ז

הקורס יעסוק באתגרים האתיים השונים שמציבה הבינה המלאכותית בפני החברה האנושית, כמו גם בדרכים שבהן מעצבת החברה את הטכנולוגיה. הלומדים בקורס יכירו את התהליכים ואת הסוגיות השונות המלוות את פיתוח הקוד האתי של מהנדסי טכנולוגיות דיגיטליות, ויעסקו בהשפעות של עקרונות אתיים אלה על אוכלוסיות שונות בעולם ועל מערכות היחסים בין החברה לפרט.

(ג) קורס פרויקט גמר מחקרי.

קורסי בחירה בתוכנית:

(ד) אתיקת מכונה. 4 נק"ז

הקורס יציג את התחום של אתיקת מכונה, תחום העוסק בדילמות האתיות והמוסריות שאיתן צריכות להתמודד טכנולוגיות בינה מלאכותית, כמו גם בשאלות של זהות, רצון חופשי ומודעות מכונה. הלומדים בקורס יכירו את הסיבות לשילובם של מימדים אתיים שונים בתהליכים החישוביים של מכונות הפועלות באופן עצמאי, ויתמודדו עם האתגרים הפילוסופיים והמעשיים הכרוכים בפיתוח אתיקת מכונה, כגון המוסריות והאתיות של מכונות מחד, והזכויות של מכונות מאידך.

(ה) תקשורת/אינטראקציית אדם ומכונה. 4 נק"ז

הקורס יעסוק בסוגיות שונות בקשר אדם-מכונה מההיבט התקשורתי: מהי תקשורת? כיצד מתנהלת תקשורת בין שותף אנושי למכונה? מהי יצירתיות חישובית? מה מאפשר האנשה של מכונות, ומהם רובוטים חברתיים? כיצד מכונות יכולות להשפיע על תגובות של בני אדם, וכיצד הן יכולות לשכנע בני אדם לפעול? האם מכונות נתפסות כאמינות? כמו כן, הקורס יעסוק באתגרים השונים העומדים בפני תחום התקשורת בהקשר של בינה מלאכותית, ובייחוד באופן שבו התקשורת עם מכונות משפיעה על התקשורת האנושית.

(ו) מבוא למדעי הנתונים. 4 נק"ז

הקורס יקנה היכרות עם עולם מדעי הנתונים וטכנולוגיית המחשוב כידע בסיסי הנדרש לסטודנטים מחוגים שאינם טכנולוגיים, ואשר יכשיר אותם לעשות שימוש בכלים של מדעי הנתונים. תינתן עזרה ייעודית בכל הקשור להתקנת כלי תוכנה רלוונטיים הנדרשים לקורס.

(ז) תכנות בסיסי לחקר נתונים. 4 נק"ז

הקורס יקנה ידע בסיסי בתכנות לצורך הכרת נושא התכנות וכמבוא לקורסים מתקדמים יותר אשר דורשים הבנה זו, ולעתים יכולת תכנות מוגבלת. תינתן עזרה ייעודית בכל הקשור להתקנת כלי תוכנה רלוונטיים הנדרשים לקורס.

(ח) כלים ושיטות למניעת וגילוי הטיות במערכות אלגוריתמיות. 4 נק"ז

מטרת הקורס ללמד את העקרונות ההנדסיים של תכנון מערכות תוכנה מבוזרות. הקורס יעסוק בתיאוריות הרלוונטיות והדגמתן באמצעות פלטפורמות תוכנה מעשיות. בסיום הקורס הסטודנטים ירכשו את המידע הדרוש לתכנון מערכות תוכנה מבוזרות על ידי שימוש בטכניקות ובתיאוריות המובילות בתחום.

(ט) אתיקה של מכונות. 4 נק"ז

מטרת הקורס היא לדון בהיבטים הפרקטיים של המוסר באמצעות דיון בסוגיות פילוסופיות מוסריות. הדיונים יתמקדו בענפי האתיקה היישומית ובסוגיות הנוגעות לצד האפל של פריצות דרך טכנולוגיות, בהקשר האישי והציבורי. בין הנושאים שיידונו: סחר וחופש המידע, אבטחת מידע, עיבוד מידע מוטה מטרה, השלכות אתיות של בינה מלאכותית, וההשלכות של הצטברות מאגרי הדאטה על האינדיבידואל ופרטיותו.

(י) שילוב אתיקה בתהליך פיתוח מוצר. 4 נק"ז

הקורס מיועד לבוגרי מדעי הרוח והוא יעסוק במושגי יסוד בתחום ההנדסה. יילמדו מושגים בסיסיים בתהליך פיתוח מוצר, אסטרטגיות המצאת מוצר וההתפתחות הטכנולוגית של המוצרים. במחזור חיי המוצר משולבים תהליכי סקר, תכן ואנליזה. אלה הם תהליכים רבי שלבים המצריכים מעורבות אנשי מקצוע מתחומים שונים. התזה של הקורס תטען לנחיצות של כללי אתיקה בתהליכים, החל בהגדרת מדויקת של הבעיה שהמוצר נועד לפתור ועד לפיתוח מוצר המצאתי, לתועלות שעדיין לא הוגדרו כצרכים. מטרת הקורס ללמד את תהליכי הפיתוח של המוצר לצד ההשלכות האתיות של תהליכי התכן והאנליזה בהתפתחות חיי המוצר במטרה להבטיח אמינות ובטיחות.

(יא) תכנות בסיסי לחקר נתונים. 4 נק"ז

הקורס יקנה ידע בסיסי בתכנות לצורך הכרת נושא התכנות וכמבוא לקורסים מתקדמים יותר, אשר דורשים הבנה זו, ולעתים יכולת תכנות מוגבלת. תינתן עזרה ייעודית בכל הקשור להתקנת כלי תוכנה רלוונטיים הנדרשים לקורס.

(3) בינה מלאכותית וקיימות: חברה וסביבה: (**AI for society and environment)**

קורסי חובה בתוכנית:

(א) הסביבה בעידן הטכנולוגי. 4 נק"ז

הקורס מיועד לכלל הסטודנטים בטכניון, ומטרתו להקנות ידע מקיף ויכולת חשיבה ביקורתית בנושאי הסביבה בארץ ובעולם. הקורס יעסוק בהשפעות העידן הטכנולוגי על הסביבה מצד אחד, ושימוש בטכנולוגיות מתקדמות לטיפול בבעיות הסביבתיות שנוצרו מצד שני. הקורס יכלול התייחסות ספציפית לחמישה משאבים מרכזיים: מים, אוויר, מזון, אנרגיה והשטחים הפתוחים. כל משאב ייבחן בהיבט של הבעיה הסביבתית העכשווית, ובהיבט של טכנולוגיות זמינות / עתידיות לטיפול בבעיה.

(ב) אקו-פילוסופיה: חשיבה וסביבה. 4 נק"ז

מטרת הקורס היא להעניק את הכלים המחשבתיים לדיון באופן מושכל בשאלות הדוחקות הנוגעות למשבר הסביבתי ולערוך היכרות עם העמדות השונות בנוגע אליהן. זאת על-ידי היכרות עם המושגים, הרעיונות והטיעונים המרכזיים שהתפתחו במחשבה הסביבתית, כגון משאבי החינם של הטבע, טביעת הרגל האקולוגית, נוסחת I שווה PAT, והטרגדיה של הכלל.

(ג) קורס פרויקט גמר.

קורסי בחירה בתוכנית:

(ד) הרשת כסביבה לימודית. 4 נק"ז

היכרות עם התפתחויות ברשת העולמית וסוגיות הקשורות להוראה ולמידה בעזרתה. דיון בנושאים בחזית החינוך המדעי-טכנולוגי, כגון: מיומנויות המאה ה-21, עקרונות פדגוגיים של סביבות למידה מקוונות, למידה מעורבת והיברידית, למידה א-סינכרונית וסינכרונית מרחוק, למידה מבוססת מיקום, הכיתה ההפוכה וקורסים מקוונים מרובי משתתפים (MOOC).

(ה) אתיקה בפעולה. 4 נק"ז

לצמיחה הכלכלית הגלובלית יש גם צד שלילי: השמדה של יערות, הרס המוני של אדמות ובתי גידול ימיים, ירידה במגוון הביולוגי ופליטה מסיבית של גזי חממה המערערים את היציבות באקלים העולמי. האנושות מתמודדת עם בעיות אקולוגיות חדשות, ועם מחלות זיהומיות חדשות המאיימות על העולם. מהי אחריות המהנדס? נדון בקוד ההתנהגות והשלכותיו על הסביבה.

הקורס יעסוק בהשלכות האתיות של שינויים במערכות האקולוגיות, לצד הבנה של תהליכים רבים ותפקודים שונים הנצפים בנוף הסביבתי. כמו כן, הקורס יעסוק בהסבר לחשיבות שיקול קנה המידה וההיררכיה במחקרים אקולוגיים, בצורך לנתח נתונים כדי לבחון את השפעת קנה המידה במחקרים אקולוגיים, בשינויים שבייעודי קרקע, בתזוזה של מפעלים, ובהשלכות על מערכות אקולוגיות לטווח הרחוק.

(ו) משבר האקלים: סיבות, משמעויות ופתרון. 4 נק"ז

מטרת הקורס היא להעמיק את ההיכרות עם משבר האקלים. הקורס עוסק בשאלות: מדוע קהילת המדע טוענת שמשבר האקלים כל-כך חמור, מהו המתווה שקהילת המדע מציעה בכדי למנוע חלק משמעותי ממשבר האקלים, מהו 'תקציב הפחמן', מה יש בהסכם האקלים של פריז. תכני הקורס מציגים ממצאים של דו"חות מדעיים עדכניים, וכן עשרות קטעי וידיאו שבהם מדענים/ות ואנשי/נשות ארגונים אזרחיים מסבירים/ות נושאים מרכזיים.

(ז) בינה מלאכותית וקיימות חברתית - הטוב, הרע, היפה והמכוער. 4 נק"ז

ההשפעה של בינה מלאכותית על קיימות חברתית יכולה להיות חיובית, אך יכולה להיות בעלת השלכות שליליות מרחיקות לכת. הקורס יעסוק בנושאים הקשורים לבינה מלאכותית ובשאלות של כוח חברתי. ההשפעות המועילות כמו גם המזיקות של בינה מלאכותית בתחומים כגון: פינטק, כלכלה ומסחר, וההשלכות של אלה על סוגיות כגון ריבוד חברתי. בקורס נבחן הצעות שונות למדידת התרומה של בינה מלאכותית לקיימות חברתית ונדון בשאלות הקשורות להשפעות ארוכות הטווח של השימוש בבינה מלאכותית.

(ח) מדעי החברה החישובים. 4 נק"ז

בעידן הטכנולוגי, מרבית האינטראקציות החברתיות מותירות טביעת אצבע דיגיטלית שאיגומן יוצר מאגרים רבים של נתוני ענק. מאגרים אלו כוללים תיעוד של אזרחים על ידי רשויות המדינה החל מלידה ועד פטירה, רישומים של כניסות לאתרי אינטרנט, תיעוד של שיחות טלפון, ומעקב של מיקום הפרט ושל חפציו בכל זמן. הזמינות של נתוני ענק אלו מאפשרת לממשלות, לחברות עסקיות ולחוקרים באקדמיה ללמוד ולחקור באופן נרחב ועמוק את ההתנהגות האנושית בכלל, ואינטראקציות חברתיות בפרט, באמצעות כלים וטכניקות לעיבוד נתונים הלקוחים ממדעי המחשב. המחקר האקדמי המתקיים במסגרת מדעי החברה החישוביים הוא במהותו רב-תחומי, ומשלב תחומי דעת דיסציפלינאריים שונים במדעי החברה יחד עם מומחים ממדעי המחשב ומערכות לניהול נתונים. הזמינות של נתוני ענק בשילוב עם יכולות חישוביות מתקדמות ומומחים דיסציפלינאריים המשתפים פעולה הובילו לממצאים ולתובנות חדשות בתחומים רבים, כגון אי שוויון, חיזוי אופן וקצב התפשטותן של מגפות, וגם על הקשר שבין התפשטותן של מגפות ואי שוויון.

(ט) בינה אנושית ומלאכותית - המודל ההיברידי. 4 נק"ז

תחום האינטליגנציה ההיברידית (HI) הוא אחד התחומים הצומחים בתחום הבינה המלאכותית, העתיד להוות קרש קפיצה לעתיד המשלב בינה אנושית עם יכולות של בינה מלאכותית. בניגוד לתחזיות האימה בהן המכונה תחליף את האדם על יכולותיו, תחום האינטליגנציה ההיברידית מבקש להרחיב את היכולות האנושיות ולא לצמצם או להעלים אותן. היתרון הגדול של המודל ההיברידי הוא בחיבור בין יכולות המכונה לבין הערך המוסף האנושי, המתבטא במערכת ערכים. הרחבת היכולת האנושית משמרת מחד את הייחודיות האנושית האתית, ומאידך - המכונה מרחיבה ומשפרת את מודל קבלת ההחלטות האנושי.

(י) בין איכות חיים לרמת חיים: מבט היסטורי. 4 נק"ז

הקורס עוסק בייצור ועיצוב, ובאידיאולוגיות השונות לגבי אלה, לאורך המאות ה-19 וה-20. הציר המרכזי סביבו נע הקורס הוא המתח שבין הפן הפרקטי – קרי, הפונקציונליות של האובייקט, הרצון לייעל את תהליך הייצור עבור היצרנית ולהוזיל את המוצר עבור הצרכן (׳רמת חיים׳), לבין הפן הכרוך בהשפעות עקיפות יותר של המוצר, בעיקר על חוויית החיים של המשתמשים, אך גם של בעלת המלאכה – פן שמתבטא בחומריות של המוצר, האסתטיקה, אופן השימוש, תהליך הייצור עצמו, כלי העבודה וכדומה (׳איכות חיים׳).

סוגיות אלה ועוד ייבחנו דרך היכרות עם טקסטים קאנוניים על ייצור ומוצרים, שעיצבו את המחשבה על מלאכות ועיצוב, ובתוך כך גם את אופני המלאכה עצמם, וכמובן את המוצרים. על מנת לעמוד על התפתחות הרעיונות האלה, יילמדו הטקסטים לפי סדר כרונולוגי.

הגישות המרכזיות יהיו: מטריאליזם, תנועת ה-Arts and Crafts, אר נובו, מודרניזם מוקדם, אסתטיקת מכונה, מודרניזם, באוהאוס, תנועת העיצוב הטוב, פוסטמודרניזם, עיצוב בעידן טכנולוגיית המידע, חומריות ואינטליגנציה חומרית.

(יא) מבוא למדעי הנתונים. 4 נק"ז

הקורס יקנה היכרות עם עולם מדעי הנתונים וטכנולוגיית המחשוב כידע בסיסי הנדרש לסטודנטים מחוגים שאינם טכנולוגיים, אשר יכשיר אותם לעשות שימוש בכלים של מדעי הנתונים. תינתן עזרה ייעודית בכל הקשור להתקנת כלי תוכנה רלוונטיים הנדרשים לקורס.

(יב) תכנות בסיסי לחקר נתונים. 4 נק"ז

הקורס יקנה ידע בסיסי בתכנות לצורך הכרת נושא התכנות וכמבוא לקורסים מתקדמים יותר אשר דורשים הבנה זו, ולעתים יכולת תכנות מוגבלת. תינתן עזרה ייעודית בכל הקשור להתקנת כלי תוכנה רלוונטיים הנדרשים לקורס.

**(4)** בינה מלאכותית וקבלת החלטות בממשל, משפט ועסקים**: (AI and decision making in government, law and business)**

## 

קורסי חובה בתוכנית:

(א) בינה מלאכותית: חוק, משפט ורגולציה. 4 נק"ז

טכנולוגיות מידע מחוללות שינויים מרחיקי לכת בסביבה העסקית ומהוות כר להתפתחותם של מודלים עסקיים חדשים. הרשת יוצרת גם מציאות פוליטית חדשה: מאפשרת היווצרות קהילות חדשות, מעצבת את יחסי הפרט והקהילה, ומחוללת שינויים בתפקידה של המדינה. המרחב הקיברנטי, המהווה עורק מרכזי של חיי המסחר, הכלכלה, החברה והפוליטיקה, חווה בשנים האחרונות קליטה של מערכות בינה מלאכותיות, המשנות את מהלך החיים, והמייצרות הזדמנויות וסיכונים - החל ממתקפות סייבר וכלה בשימוש לרעה באלגוריתמים. הקורס ייתן את התשתית הבסיסית לעולם הרגולציה, כשבמוקד יעמדו תיאוריות של רגולציה, היחס שבין רגולציה משפטית לסוגי הסדרה חברתיים ומוסריים אחרים, ויישום התובנות בעולמות הבינה המלאכותית.

(ב) מקרי מבחן: כלי בינה מלאכותית בשדה המשפט ובשדה העסקי. 4 נק"ז

קורס זה יעסוק בסוגיות ספציפיות שעולות בשדה המשפט ובשדה העסקי, מתוך התבוננות על הממשק שבין כלי הבינה המלאכותית והמסגרת המשפטית שבתוכה הם מופעלים. הסטודנטים יעסקו באתגרים ספציפיים, שלהם יידרשו פתרונות משדה הבינה המלאכותית ומעולם המשפט. בין האתגרים שעל הפרק: ענישה או שחרור בערבות, חוזים חכמים, אחריות נזיקית של כלים אוטונומיים, ביטויי שנאה ותמיכה בטרור, חדשות כזב, פרופילים לפרסומות ולתמחור, אלגוריתמים לחיזוי שוק והשקעות, ואבטחת מידע.

(ג) קורס פרויקט גמר מחקרי.

## 

קורסי בחירה בתוכנית:

(ד) כלכלת **המידע - Information Economics:** בינה מלאכותית, השקעות, יזמות והסדרת השוק. 4 נק"ז

מידע הפך בשנים האחרונות להיות קטר כלכלי בכל המגזרים במשק. דאטה וכלי ניתוח אלגוריתמיים מייצרים תועלות עסקיות בעלות ערך כלכלי. עם זאת, כלכלת המידע מייצרת אתגרים שעלולים לפגוע בתחרות ובחדשנות. קורס זה יעסוק בניתוח קשרי הגומלין בין מידע, טכנולוגיה ואדם, במטרה להבין תהליכים להם אנו עדים, כגון צמיחה מהירה של חברות, יצירת מונופולים בסביבה תחרותית, ומתווכים מסוגים חדשים. באמצעות הבנה של עקרונות יסודיים של מידע כמוצר פרטי או ציבורי נפתח יכולת לנתח את הפעילות העסקית סביב מידע. נעסוק בגישות הכלכליות השונות שבאמצעותן ניתן להבין ולשכלל את השוק האלגוריתמי ונתוודע למגוון של מודלים עסקיים. בין השאר נעסוק בריכוזיות ובהגבלים עסקיים, בהגנת הצרכן והמשתמש, ובמבנה התמריצים של החברות האוספות, האוגרות והמנתחות מידע. נתייחס לכלים כלכליים להתמודדות עם האתגרים, לרבות משטרי רישוי, מיסוי וביטוח, וכן למשטרי הקניין הרוחני ולכללי המסחר בחברות נסחרות, אשר עשויים להשליך על מבנה התמריצים לפיתוח השוק. הכלים שירכשו הסטודנטים יאפשרו להם להציע יוזמות ליישום הידע הנלמד בארגונים שונים.

(ה) מערכות בינה מלאכותית למשפטנים. 4 נק"ז

הקורס מעלה סוגיות ייחודיות וכבדות משקל אשר משלבות בין עולמות המשפט והאתיקה לבין עולם הטכנולוגיה. על מנת לזהות סוגיות אלו, וכמובן על מנת לנתחן ולהציע להן פתרונות, נדרשת היכרות עם שני סוגי העולמות. מטבע הדברים, להרבה מאנשי מדעי הנתונים אין היכרות מעמיקה עם תחום המשפט, ולהרבה מאנשי המשפט אין רקע טכנולוגי. אופן זה מקשה על דיאלוג בין הצדדים. כתוצאה מכך, עלולים אנשי מדעי הנתונים לפתח טכנולוגיות שמעוררות קשיים אתיים וחוקיים שניתן היה למזער מראש עם הבנה מעמיקה יותר של עולם המשפט, ומנגד - אנשי המשפטים עלולים לצפות לפתרונות אשר אינם מתיישבים עם הכלים הקיימים במציאות. מטרת העל של הקורס היא לייצר דיאלוג בין שני העולמות, באופן שיניח את הבסיס ליצירת שפה משותפת בה הצדדים אינם מתמצאים עדיין בנבכי העולם האחר, אך מסוגלים לזהות בשלב מוקדם את נקודות המפגש הפוטנציאליות ביניהם.

(ו) טכנולוגיה, משפט, וצדק דיגיטלי. 4 נק"ז

הקורס יעסוק בגישות הבסיסיות לצדק דיגיטלי ולדרכים ליישוב סכסוכים בכלים דיגיטליים. השימוש בכלי הטכנולוגיה, בדגש על בינה מלאכותית, מחייב התייחסות לתיאוריות צדק בסיסיות, ובהן צדק מתקן, צדק גמולי, צדק חלוקתי, צדק הליכי וצדק מאחה. זאת, מכיוון שהשימוש בכלי הבינה המלאכותית עלול לאתגר תפיסות צדק אלו, אך גם יש ביכולתה של הבינה המלאכותית לקדם יישום של תפיסות צדק כאלה ואחרות (by design). בנוסף, השימוש בכלי הבינה המלאכותית מאפשר להבין את האופן שבו מחלוקות נתפסות (באמצעות התבוננות על דרכי ארגומנטציה), ואף למסד דרכים ליישוב סכסוכים (למשל, באמצעות מערכות שמוטמעות בפלטפורמות לרכישה מקוונת, כגון אמזון ואיביי).

(ז) שוויון ואפליה בממשקי אדם-מכונה. 4 נק"ז

הקורס יעסוק בהיבטים שונים של הזכות לשוויון ושוויון זכויות, בהקשרים טכנולוגיים. בין השאר נעסוק באפליה על רקע גזע, מין, לאום, דת, נטייה מינית, גיל, מצב משפחתי ומוגבלות, כפי שהיא באה לידי ביטוי בסביבה העסקית-משפטית שבה שירותים ומוצרים מסופקים, בתנאים שונים, על בסיס מידע שנאסף על ידי סנסורים שונים (בידי המגזר העסקי ובידי רשויות השלטון). דגש יושם במשפט השוואתי ובדרכים שבהן מדינות שונות מתמודדות עם האופן שבו אנו מבינים פגיעה בזכות לשוויון, ובדרכים משפטיות להתמודד עם הסיכון שמתקיימת אפליה, או עם תוצאות האפליה, לכשזו מוכחת. את הקורס תלווה השאלה: תחת אלו תנאים ממשק אדם-מכונה הוא חלק מהבעיה או חלק מהפתרון?

(ח) בינה מלאכותית במשפט הבינלאומי. 4 נק"ז

בקורס זה נעסוק בצמתים שבהם המשפט הבינלאומי נפגש עם הטכנולוגיות החדשות. בין השאר נדון בהסדרת כלי נשק אוטונומיים ודיני הלוחמה באופן כללי, כלים לניהול משאבי סביבה בינ-מדינתיים, כלי להתמודדות עם משברים בינלאומיים (בדגש על משברים כלכליים), לרבות כלים להתמודדות עם גלי הגירה ופליטים. חטיבה במסגרת הקורס תעסוק במסגרות הבינלאומיות המסדירות השקעות במדינות מתפתחות ומפותחות, מתוך הכרה בחשיבות שבהתמודדות עם הפער הדיגיטלי בין מדינות הדרום לצפון.

(ט) בינה מלאכותית וסייבר. 4 נק"ז

מרחב הסייבר נתון בשנים האחרונות להתקפות המאיימות על מהלך החיים התקין בתחום המסחר, החברה, הכלכלה, והביטחון האישי והלאומי. הצורך לאבטח את הסייבר ולהגן עליו מפני מתקפות עשוי לחייב ניטור מסיבי ומתמשך של משתמשי הרשת, שימוש בטכנולוגיות מעקב פולשניות, כריית מידע וניתוח נתונים (big data), וכן שיתוף פעולה רחב היקף בין גורמי אכיפת החוק לבין גופים מסחריים, כמו ספקי תקשורת, רשתות חברתיות ומנועי חיפוש. הקורס יעסוק בסוגיות מתחום אבטחת המידע (לרבות כלים התקפיים, כלים להגנה מפני מתקפות, וכלי מעקב), והאופן שבו סוגיות אלו מוסדרות מבחינה משפטית. דגש יושם בסביבה הכלכלית-חברתית (אקו-סיסטם) שבתוכה מעוררות הדילמות, שבפניהן ניצבים ארגונים ומדינות בבואן להתגונן (וכן לתקוף, או לעקוב) במרחב הווירטואלי. הקורס יערוך היכרות עם משטרים שונים שחולשים על הגנת הסייבר ועל מתקפות הסייבר, לרבות תביעות משפטיות שהתעוררו בהקשרים אלו, ולרבות תקנים שונים שגובשו, ומוסדות שונים שהוקמו.

(י) הבינה המלאכותית והמשטר הדמוקרטי. 4 נק"ז

כלי הבינה המלאכותית הגיעו לפוליטיקה: מפלגות, מועמדים ואף מדינות זרות עושים שימוש בכלים שונים מתחום הבינה המלאכותית, באופן שמאתגר הנחות יסוד מתחום התיאוריה הפוליטית ופרקטיקות רבות-שנים במדינות דמוקרטיות שונות. נתבונן באופנים שבהם כלי הבינה המלאכותית (לרבות האופנים שבהם נעשה שימוש בכלים אלו בקהיליות פוליטיות מסוימות), משפיעים על הבנת זכויות דמוקרטיות בסיסיות, כגון חופש הביטוי, חופש המחאה וההפגנה, חופש ההתאגדות, הזכות לבחור ולהיבחר ופרטיות.

(יא) אתיקה בעסקים. 4 נק"ז

החל מתקופת יוון העתיקה דנו פילוסופים בשאלות מוסריות - האופן בו ראוי שבני אדם יחיו את חייהם,. במחצית השנייה של המאה ה-20 החלה התעניינות בשאלות של מוסר גם בהקשר של ארגונים. ארגונים פועלים בשוק דינמי ותחרותי, בתנאי אי ודאות, כאשר מנהלים ועובדים מונעים על ידי רצון להצליח באופן אישי ולהשיג את המטרות הארגוניות.

לפיכך, קבלת החלטות בארגונים מתבססת לעתים גם על שיקולים לא מוסריים. התוצאה היא עוולות מוסריות כגון פגיעה באיכות הסביבה, הפרת זכויות עובדים וחוסר הוגנות כלפי לקוחות. על מנת לצמצם את הנזקים האפשריים, מנהלים ועובדים בארגונים נקראים לפעול תוך התחשבות בבעלי עניין שונים בארגון. החלת עקרונות אתיים על פעילות של ארגונים מונעת גם מטענה רווחת כי "אתיקה משתלמת". קרי, יש הצדקה כלכלית להתנהגות אתית של ארגונים.

(יב) כלים טכנולוגיים במשפט הפלילי: פורנזיקה, חיפוש והצפנה. 4 נק"ז

במאה ה-21 עולם המשפט הפלילי שזור בטכנולוגיה עילית: כלים טכנולוגיים מתקדמים, לרבות כלי בינה מלאכותית, מבצעים חקר פורנזי ומשמשים לחיפוש ומעקב, הן במרחב הפיזי, והן במרחב הווירטואלי. משכך, תחום הפורנזיקה הדיגיטלית מצוי בחזית המחקר הטכנו-משפטי, ומצריך עיון שרגיש לחשיבה מוסרית ומדעית כאחד. כך גם שדה פלילי נוסף - שימוש בכלי בינה מלאכותית כמסייע לגורמי האכיפה בגיבוש מדיניות וביישומה במקרים קונקרטיים על דרך בניית פרופילים שונים, אשר נועדו לחזות התנהגות עבריינית או רסידיביזם. לצד אלו, המשפט עצמו מבנה עבירות מחשב ייחודיות, כגון חדירה למחשב, ובמקביל נדרש לעדכן עבירות "פיזיות", כגון עבירות מין, למרחב הווירטואלי, שאינו כולל מגע. לבסוף, הטכנולוגיה מאפשרת כיום הצפנה, אשר לצד התועלת שבשמירה על פרטיות מהווה אתגר לרשויות החוק. הקורס יפרוש בפני התלמידים את הגישות התיאורתיות-פילוסופיות להפללה ולענישה פלילית, וכן לשימוש באמצעי עזר כגון חיפוש ומעקב, ולצידן את מערכות הדין, התובנות הקרימינולוגיות והגישות הטכנולוגיות החולשות על הפרקטיקות הנהוגות במקטע חברתי זה.

(יג) בינה מלאכותית, אתיקה ומשפט. 4 נק"ז

טכנולוגיות של למידת מכונה ובינה מלאכותית חודרות לתחומי חיים שונים ומשפיעות על מנעד רחב של פעולות בחיי הפרט, החברה והממשל. מערכות אלה מעוררות שורה של אתגרים חדשים בתחום האתיקה והמשפט. הקורס נועד לפתח כלים אתיים, טכנולוגיים ומשפטיים לטיפול בהיבטים האתיים והמשפטיים הקשורים לפיתוח ושימוש בטכנולוגיות אלה. הקורס מכוון פרויקטים מעשיים בשיתוף מדעי הנתונים בטכניון.

# תוכניות מחקר - תקצירים

1. מרכז למחקר משותף של אוניברסיטת חיפה והטכניון: מרכז לחקר האינטראקציה בין אינטליגנציה אנושית לאינטיליגנציה ממוחשבת:

רשימת חוקרים:א**.** פרופ' אלדד יחיעם, הנדסה תעשיה וניהול, הטכניון.

ב. פרופ' ברוך עיטם, פסיכולוגיה, אוניברסיטת חיפה.

ג. ד"ר עפרה עמיר, הנדסה תעשיה וניהול, הטכניון.

ד. פרופ' עדו ערב, מדעי ההתנהגות, הטכניון.

ה. פרופ' מור פלג, מערכות מידע, אוניברסיטת חיפה.

ו. ד"ר ארנון קרן, פילוסופיה, אוניברסיטת חיפה.

תיאור המרכז:

הרבה מהיכולות הקוגנטיביות שאותן מעריכים מבחני אינטליגנציה פופולריים נראות חשובות פחות היום משהיו כאשר פותחו מבחנים אלה, משום שבימינו פלאפון שעלותו 100$ יכול להשלימן ולבצע הרבה מפעולות החשיבה מהר יותר וטוב יותר מבני אדם. לעומת זאת, יכולות אחרות, שאותן לא נראה שמבחני אינטליגנציה פופולריים מעריכים, הופכות בימינו להיות יותר ויותר חשובות. האינטראקציה בין אינטליגנציה אנושית לאינטליגנציה ממוחשבת מערערת לכן על תוקפם של מבחני אינטליגנציה מקובלים. אך זאת רק אימפליקציה אחת של שיתוף הפעולה הזה: לאור היכולות הגדלות של אינטליגנציה ממוחשבת, ולאור האינטראקציה הגוברת בין בני אדם למכונות חכמות, מאוד ייתכן שהיכולות הקוגניטיביות של בני אדם ישתנו בדרכים שעדיין אינן ברורות דיין.

האינטראקציה הגוברת בין אינטליגנציה אנושית לאינטליגנציה ממוחשבת אף מולידה סיכונים חברתיים משמעותיים, לצד הזדמנויות להתמודדות משופרת עם בעיות חברתיות, קיימות וחדשות. כך, השימוש והתלות הגוברת באינטליגנציה ממוחשבת עלולים להגדיל באופן דרמטי צורות של אי-שוויון, באשר פרטים וקבוצות בעלי אינטליגנציה-טכנולוגית גבוהה יזכו בנתח גדל והולך של משאבים והשפעה חברתית; ולאור השימוש הגובר בכלים חדשים להפצה וליצור של דיסאינפורמציה, והספק לגבי היכולת הקוגניטיבית האנושית להתמודד עמן, מוסדות חברתיים שמבוססים על יכולתם של פרטים לקבל החלטות מבוססות-ידע נתקלים באיומים חדשים. מנגד, מהפכת הביג-דאטה מספקת לבני אנוש אפשרויות חדשות ליהנות ממכונות חכמות, ולעבוד ולשתף עמן פעולה בתחבורה, חינוך, בריאות, ועוד. אילו יכולות קוגניטיביות נדרשות כדי לשגשג בעולם כזה, שנמצא ממש מעבר לפינה?

מרכז המחקר על אינטליגנציה 2.0 יתמקד בהזדמנויות ובסיכונים אלה. מטרתנו היא לברר את התנאים שבהם השימוש באינטליגנציה ממוחשבת תורם או פוגע באינטליגנציה אנושית או בהתנהגות אינטליגנטית. אגב כך, המחקר יוכל לתרום לפיתוחם של כלים אינטליגנטים ממוחשבים שהשפעתם החיובית גדולה, והשלילית נמוכה. עבודתם של ששת החוקרים השותפים במרכז מתייחסת למטרה זאת מפרספקטיבות שונות, אך ביניהן קשרים חשובים, המאפשרים את שיתוף הפעולה המחקרי ביניהם. להלן תיאור קצר של שישה פרויקטים שייחקרו במסגרת המרכז, המקובצים תחת שתי מסגרות רחבות: ההערכה והשיפור של אינטליגנציה 2.0.

ההערכה של אינטליגנציה 2.0: הסתמכות חכמה על מערכות מידע חדשות. מוביל: ד"ר ארנון קרן, פילוסופיה:

התפתחות המתמדת של טכנולוגיות חדשות המספקות מצד אחד מידע ומצד שני מיסאינפורמציה מעוררת ספק לגבי יכולתנו להסתמך עליהן בצורה מושכלת, ולגבי עצם הבנתנו את ההבחנה בין הסתמכות אינטליגנטית יותר ופחות על מקורות מידע חדשים. מכיוון שרבים מאיתנו חסרים את המשאבים הדרושים לבחירה מושכלת וחכמה לגבי הסתמכות על מקורות מידע חדשים אלה, נראה שאנו עומדים בפני בחירה בין אמון עיוור לבין דחיה חסרת בסיס של מקורות מידע חדשים אלה. נבחן תגובות אפשריות לאתגר זה לגבי עצם האפשרות להסתמכות מושכלת על טכנולוגיות מידע חדשות, תוך בחינת קווי הדמיון בין אתגר זה לבין אתגר בעל דמיון מבני הנוגע ליכולתם של הדיוטות להסתמך באופן חכם ומושכל על מומחים. נבחן את המידה שבה דרכי התמודדות שהציעו אפיסטמולוגים ביחס לאתגר ההסתמכות על המומחיות יכולות לעזור להתמודד עם ההמשגה של אתגר ההסתמכות האינטיליגנטית על טכנולוגיות מידע חדשות. אנו משערים שכפי שנמצא במקרה של בעיית המומחה-הדיוט (Goldman 2001; Anderson 2011), אמון במערכות מידע חדשות אינו צריך להיות עיוור; עם זאת, אנשים שאינם מומחים לעיתים קרובות לא מנצלים באופן מיטבי מקורות מידע חשובים לגבי אמינות מקורות המידע (Keren 2018; Erev, Roth & Sonsino, under review) בחירה בין מקורות שונים של מידע נרחב. מוביל: פרופ' אלדד יחיעם, כלכלה התנהגותית, מדעי הקוגניציה לאור המספר העצום של אפשרויות זמינות לקבלת מידע באופן מיידי, הבדלים אישיים בקבלת ההחלטות ביחס לבחירה בין החלופות, בהיקף החקירה, ובהתגברות על נקודות אופטימום מקומיות, הופכים לחשובים יותר ויותר עבור האינטיליגנציה החדשה. ההערכה, האמידה, וההבנה של הבדלים בין אישיים אלה היא אתגר חשוב, המצריך שינוי גישה מזה שמאפיין מבחני אינטליגנציה קיימים. אנו נבחן כיצד ניתן להעריך "אינטליגנציית בחירה" שכזאת בעזרת דילמות קוגנטיביות מובנות. נפתח משימות גלישה מובנות, שיאפשרו לאפיין את הרכיבים העיקריים של סגנונות חיפוש מידע, כמו רוחב החיפוש וההתגברות על נקודות אופטימום מקומיות (התמקדות בתוצאות בעלות רלוונטיות חלקית); כך, במחקר כזה בחנו כיצד פרטים על הספקטרום האוטיסטי מבצעים חיפושי רשת וזיהינו שימוש באסטרטגיה של חיפוש רחב יותר (Yechiam & Yom-Tov, 2021). פרויקט נוסף יחקור משימת שני-מומחים, שבמסגרתה פרטים יצטרכו להתמודד עם מצבים בהם הם נתקלים בעמדות מומחים שונות זו מזו, אך שבהם לכל אחד משני המומחים מידע רלבנטי, שיכול לתרום לתובנות. באמצעות פרדיגמות אלה נבקש ליצור קשרים בין השטחים השונים של קבלת החלטות (כלכלה התהנגותית), חיפוש מידע, ואינטליגנציה.

פעלנות אנושית (Human Agency) בהקשר של אינטנליגנציה ממוכנת. מוביל: פרופ' ברוך עיטם, פסיכולוגיה:

מחקר בפסיכולוגיה קוגנטיבית ומדעי המוח מן העשור האחרון הראה שחוויות של פעלנות (agency) ושל אפקטיביות (שהפעולה שלי השפיעה על הסביבה) חשובה ללמידה אנושית Haggard & Eitam, 2015;; Hemed et al., 2019)). סביבות שמספקות חוויות כאלה מקיימות מוטיבציה אנושית ומכווינות אותה, ומרחיבות חקירה ורכישה של ידע בסיסי כיצד לפעול בעולם. המחקר שלנו יבחן כיצד חווית הפעלנות מושפעת מן השימוש בכלי אינטיליגנציה ממוכנת. מתי ההגדלה במידת ההשפעה האובייקטיבית בזכות השימוש בכלים מתורגמת לעליה בתחושת הפעלנות, ומתי היא מובילה דווקא לירידה – למשל, בגלל חוסר וודאות לגבי ייחוס האחריות והקרדיט לפעולה, כאשר הכלי עצמו נתפס כפועל. עוד נבחן הבדלים בין-אישיים בתפיסת פעלנות בהקשר של פעולה הנעזרת בכלי אינטיליגנציה מלאכותית.

שיפור אינטליגנציה 2.0: מה נדרש עבור עבודת צוות מעורב (אנושי-ממוכן) מיטבית. מובילה: ד"ר עפרה עמיר, בינה מלאכותית, ואינטראקציה בין אדם למחשב:

אחד האתגרים הגדולים של ימינו הוא המורכבות העולה ועומס המידע הגדל שמקורם במערכות שבהם אינטראקציות הולכות וגוברות, המשלבות בין אדם ומכונה. נבחן את ההתמודדות עם האתגר הזה בשני הקשרים: ההקשר של עבודת צוות אנושית, וההקשר של שיתוף פעולה בין בני אנוש לסוכן-ממוחשב. בעבודת צוות אנושית, חברי הצוות צריכים לשתף ביניהם מידע באופן יעיל (Amir et al., 2019). בשיתוף פעולה בין אנשים לסוכן ממוחשב, אנשים צריכים לחזות את התנהגותם של סוכנים עם מנגנוני קבלת החלטות מורכבים (Amir, Doshi-Velez & Sarne, 2019). המחקר שלנו יבחן את ביצועיהם של אנשים במצבים כאלה ויפתח שיטות חדשניות ש"מזקקות" את המידע החשוב והרלוונטי ביותר בתרחישים מורכבים ומציגות אותו בפני אנשים באופן יעיל התומך בהבנה ובקבלת ההחלטות טובה. תחום עיקרי שבו נתרכז הוא תחום הטיפול הרפואי, שבו צפויים להשתלב באופן הולך וגובר סוכנים ממוחשבים חכמים, שייעצו לצוות ולמטופלים.

אינטליגנציה רפואית (אנושית וממוחשבת). מובילה: פרופ' מור פלג, מערכות מידע:

בשדה הרפואי, בינה מלאכותית תומכת במטפלים ובמטופלים ומסייעת בפרשנות ואבחון נתונים ובהגדרת יעדים, תכנון טיפול, וניהול טיפול מותאמים אישית ומבוססי ראיות (Peleg et al., 2017a, b). מחקרים קודמים מדגישים שלושה חסרונות של האינטליגנציה הרפואית של השילוב בין אדם למכונה: אי-זיהוי המצב הפסיכו-סוציאלי של המטופלים, היעדר ראיות בנוגע לטיפולים אפקטיביים מותאמים אישית (Peleg et al., 2017c), וחוסר מעורבות ודבקות של המטופל. אנו נחקור את ההשפעה המשותפת של שלושת ההיבטים האלה של אינטראקצית אדם-מכונה בשדה הרפואי על ידי (א) איסוף נתוני מטופלים מתיקים רפואיים, חיישנים, ודיווח עצמי; (ב) פיתוח ויישום שיטות למידת מכונה כדי להעריך את ההקשר הנוכחי של המטופל ולחזות את תהליך הטיפול היעיל ביותר; (ג) מתן שירותי אימון מותאמים אישית שעושים שימוש בתיאוריות פסיכולוגיות (Peleg et al., 2018; Erev & Roth, 2014; Eitam et al ,, 2013) והוראה חכמה (Michalowski et al., 2020). אנו משערים כי מערכת המשלבת שירותים כאלה באפליקציה סלולרית תביא להעצמה של המטופלים ותאפשר להם להבין טוב יותר, ולהיות מעורבים יותר בבריאותם ובטיפול בהם.

צמצום מלכודות למידה. מוביל: פרופ' עדו ערב, כלכלה התנהגותית, מדעי הקוגניציה:

אחד הגורמים העיקריים התורמים להשפעה השלילית של טכנולוגיות חדישות הוא ההתכנסות של משתמשים מסוימים למלכודות למידה (learning traps) התורמות להתפתחות "חוסר אונים נרכש" (learned helplessness). מלכודות אלה מצמצמות את התועלת שמפיקים משתמשים אלה מטכנולוגיות חדישות, ותורמות לעליה באי-שיוויון. מחקרים קודמים (Erev & Roth 2014) מצביעים על שלושה סוגים של גורמים שתורמים למלכודות אלה: תחושת העדר שליטה, סבירות נמוכה להצלחה ובעיות קואורדינציה (שיווי משקל לא יעיל). נחקור את ההשפעה המשותפת של גורמים אלה על ידי איסוף נתונים במגוון רחב של תנאים, פיתוח מודלים תיאוריים, וארגון תחרויות חיזוי בחירה (Erev et al., 2017). אנו משערים כי ניתן לחזות את ההשפעה של שלושת הגורמים באמצעות מודל יחיד של למידה ברשתות החברתיות. בנוסף נעריך את ההשלכות של התוצאות במחקרי התערבות.

1. מחקר משותף של אוניברסיטת חיפה והטכניון: פרופ' עפרה עמיר וד"ר גיא אבני - מחקר בבינה מלאכותית לפיתוח שיטות שיתופיות ממוקדות משתמש להסברת התנהגותן של מערכות המקבלות החלטות לאורך זמן

ממכוניות אוטונומיות למערכות הממליצות על טיפול רפואי. מערכות מבוססות בינה מלאכותית (AI) הופכות נפוצות יותר ויותר. למערכות אלו יש פוטנציאל להיטיב עם החברה בתחומים כגון תחבורה, בריאות וחינוך. חשוב לציין, כי מערכות אלו אינן פועלות בוואקום - המערכות מופעלות ומתקשרות עם אנשים במגוון רחב של דרכים. על מנת להשתמש ולהסתמך על מערכות לומדות, אנשים צריכים להיות מסוגלים לצפות ולהבין את התנהגותן. לדוגמה, נהג של רכב אוטונומי יצטרך לצפות מצבים שבהם המכונית נכשלת ומעבירה את השליטה, בעוד שרופאה תצטרך להבין את הטיפול המומלץ על ידי המערכת כדי לקבוע אם הוא עולה בקנה אחד עם העדפות המטופל. התחום Explainable AI (הסברי מערכות בינה מלאכותית) שואף להבהיר למשתמשים איך מערכות לומדות מקבלות החלטות ובכך להפוך את ההתנהגות של מערכות אלו לשקופה יותר. ישנם כמה היבטים בהם יש חוסרים בכלים הקיימים של Explainable AI:

ראשית, רוב השיטות הקיימות מתמקדות במתן הסברים 'מקומיים' להחלטות חד פעמיות של מודלים של למידת מכונה. הן אינן מתאימות לתיאור התנהגותן של מערכות הפועלות לאורך זמן. שנית, רוב השיטות הקיימות אינן מתחשבות בהקשר שבו ניתנים הסברים, ואינן מתאימות את ההסברים לצרכים ולמאפיינים הספציפיים של המשתמשים. לבסוף, רוב השיטות אינן אינטראקטיביות, מה שמגביל את היכולת של המשתמשים להשיג הבנה מעמיקה של המערכת.

בפרויקט מחקר זה, אנו שואפים לפתח שיטות שיתופיות ממוקדות משתמש להסברת התנהגותן של מערכות המקבלות החלטות לאורך זמן. שיטות ההסבר ישתמשו בכלים, ברעיונות ובמתודולוגיות של בינה מלאכותית, אימות פורמלי ואינטראקציה בין אדם למחשב. המחקר עוסק בפיתוח אלגוריתמים שיקבעו איזה מידע על התנהגות המערכת ישותף עם משתמשים, יוצעו הסברים המותאמים לצרכים ולמאפיינים של המשתמשים, וממשקים המאפשרים למשתמשים לחקור באופן יזום את היכולות של המערכות.

1. מחקר משותף של אוניברסיטת חיפה והטכניון: פרופ' אורי פלונסקי וד"ר יפים רוט:

עבור קובעי מדיניות, אחד האתגרים הגדולים במהלך מגפת COVID-19 עוסק בשאלה כיצד לעודד את הציבור לעמוד בהנחיות הריחוק הפיזי ודומות להן. מכיוון שציות להנחיות מצריך מאמץ חוזר ונשנה לאורך זמן בסביבה חברתית, המחקר בודק החלטות חוזרות במשחקים חברתיים שנועדו לחקות את סביבת קבלת ההחלטות במגיפה. במספר מחקרי מעבדה אנו מראים כי ההתנהגות הנפוצה במגפה צפויה להיות שאננות ופזיזות. המחקר בוחן מספר פתרונות פוטנציאליים המבוססים על אפליקציות ניטור לטלפון חכם על מנת להראות אפשרות להפחתה משמעותית בהתנהגויות פזיזות כאשר האלטרנטיבה, התנהגות אחראית, מובילה לתגמול טוב יותר רוב הזמן. לבסוף, אנו מדגימים את הפוטנציאל של גישה זו במחקר שנערך בקרב עובדי בית אבות בשיא המגפה. ההתערבות הגדילה את הציות להנחיות מ -20% ל -75%.

1. ארז פירט, דני סטטמן, החוג לפילוסופיה, אוניברסיטת חיפה: האם עלינו להתאים את האתיקה של סוכנים מוסריים מלאכותיים לזו שלנו?

בעתיד הקרוב, דרגת האוטונומיה של מערכות מלאכותיות נפוצות תגיע לנקודה שבה נהיה חייבים להרשות להן לקבל החלטות מוסריות כחלק מתיפקודן היומיומי – מכוניות אוטונומיות, שרתים אישיים, רובוטי משק בית, בוטי מסחר, מערכות נשק אוטונומיות ועוד, הן דוגמאות לסוגי המערכות שיתמודדו עם מצבים מוסריים מורכבים המצריכים שיפוט מוסרי ברמה הקרובה לרמה אנושית. בעשרות השנים הקרובות, חוזים המומחים כי למכונות מסוג זה יהיו יכולות קוגניטיביות דומות לשלנו, מה שידרוש מאיתנו לשלוט עליהן למען ביטחוננו. הניסיון להבטיח שפעולותיהן של מכונות עוצמתיות אלה לא מובילות לתוצאות לא רצויות ואף הרסניות עבור בני האדם, מכונה 'בעיית השליטה' (The AI Control Problem), והגישה הבולטת העכשווית להתמודדות עם בעיה זו, בה מתמקד מחקר זה, נקראת 'התאמת ערכים' (AI Alignment). כשאנו חושבים על התאמת ערכיהן של מערכות בינה מלאכותית לערכים אנושיים, אנו שואלים את עצמנו כיצד נוכל לוודא שהערכים (המוסריים) או העקרונות המנחים את נוהלי ההחלטה של מערכות עתידיות אלו מותאמים לערכים אנושיים.

במחקר זה, אנו מציעים לחקור את הישימות הטכנולוגית והנאותות המוסרית של גישות מסוימות להתאמת ערכים, לסוגים מסוימים של סוכנים מוסריים מלאכותיים (סמ"מ) עתידיים: גישת 'מקסום התועלת' (Utility Maximization Approach), הגורסת שסוכנים המונחים על ידי פונקציית תועלת כלשהי יכולים להיות מותאמים למוסר אנושי (יהיה אשר יהיה), וגישת 'למידת הערכים' (Value Learning Approach), הגורסת כי אפשר לאמן מערכות בינה מלאכותית בעזרת שיטות למידה של 'חיזוק הפוך' (Inverse Reinforcement Learning), על מנת לחלץ העדפות אנושיות על ידי התבוננות בהתנהגות אנושית.

בנאותות מוסרית הכוונה היא לשאול האם נכון מוסרית להתאים את ערכיהם של סוגים שונים של סמ"מ, בעלי רמות שונות (עד רמה אנושית ויותר) של סוכנות מוסרית. במחקר זה נתרכז בסמ"מ פונקציונלי (functional) ובסמ"מ מלא (full-blown). את הסוג הראשון ניתן להשיג על ידי יישום גישות top-down כגון גישת ה–rule based, גישות bottom-up כגון למידת מכונה או גישות אבולוציוניות, או על ידי גישות היברידיות, שמשלבות את הגישות שהוזכרו, ומביאות לידי מראית עין של מוסריות. הגישה השנייה ספוקולטיבית יותר, מאחר ואינה נטועה בטכנולוגיה עכשווית, אלא בהערכות מומחים לגבי עתיד הטכנולוגיה בעוד עשרות שנים. היא משערת את קיומן של מערכות בעלות אינטליגנציה אנושית ומעלה ויכולות חישה מתקדמות, הנקראות 'מכונות סופר אינטליגנטיות' (Super-intelligent machines).

בהסתמך על הנאמר לעיל, אנו מעלים את שאלת המחקר העיקרית שלנו: האם הגישות המוזכרות לפתרון בעיית התאמת הערכים, דהיינו מקסום התועלת ולמידת הערכים, הן ישימות ונאותות מוסרית, כאשר מדובר בסוגי הסמ"מ שאנו חוזים בעתיד הקרוב והלא-כל-כך-קרוב? במילים אחרות, האם אנו רוצים וצריכים להתאים את האתיקה של סמ"מ עתידיים לזו האנושית?

1. פרופ' מירי ברק, הפקולטה למדע ולטכנולוגיה, טכניון: האתיקה של המחקר - חינוך לאתיקה בבינה מלאכותית: הבניית ידע ומודעות להתנהגות אתית במחקר מדעי-הנדסי

בינה מלאכותית מציבה אתגר לא רק בתחום הפיתוח ההנדסי אלא גם בהיבט האתי והחינוך לאתיקה. החל משימוש בטכנולוגיה לזיהוי פנים ויצירת 'פרופיל בוחרים', דרך ממשקי מוח-מכונה ועד מזל"טים חמושים, כל אלו עלולים להוות איום על חופש הפרט, החברה והסביבה בה אנו חיים. חינוך לאתיקה מהווה חלק חיוני בהכשרתם של סטודנטים ברחבי העולם. הספרות מראה שקורסים רבים בתחום האתיקה מבוססים על הגישה הפרטנית שכוללת קריאה של יחידות לימוד וביצוע מבחנים. גישה זו נמצאה כמעודדת למידה שטחית, מאחר ואינה מאפשרת שיח בין לומדים וחשיפה למגוון דעות ועמדות. המחקר הנוכחי יתבסס על הגישה ההבנייתית-חברתית (Social constructivism) בפיתוח והפעלה של קורסים. על פי גישה זו, הבניית ידע והתפתחות קוגניטיבית מושגים בעזרת שיח ואינטראקציות בינ-אישיות. מכיוון שבינה מלאכותית וחינוך לאתיקה הינם תחומים חוצי-דיסציפלינות, המחקר הנוכחי שם למטרה לבחון את יעילות ההוראה וחווית הלמידה מנקודת מבטם של המרצים והסטודנטים, מפקולטות שונות. בהיבט ההוראה, המחקר יבחן השקפות של מרצים לגבי חינוך לאתיקה בבינה המלאכותית ואת הנושאים שנמצאים בחזית הדיון האתי. בהיבט הלמידה, המחקר יבחן ידע ומודעות להתנהגות אתית ומיומנויות הדרושות בתהליכי מחקר ופיתוח הנדסי. בהתבסס על מתודת השיטות המעורבות, הנתונים ייאספו באמצעות שאלונים וריאיונות חצי-מובנים וינותחו באופן כמותי ואיכותני. חשיבות המחקר היא בזיהוי עקרונות עיצוב ומרכיבי תכנון של קורסים המקדמים פרקטיקה אתית בבינה מלאכותית. בהתבסס על מסקנות המחקר, תוצגנה המלצות אופרטיביות לפיתוח קורסים המעודדים למידה משמעותית, חווייתית ושיתופית.

1. פרופ' צבי קופליק, החוג למערכות מידע, אוניברסיטת חיפה: שקיפות אלגוריתמית

עם התפשטות השימוש במערכות המבוססות על בינה מלאכותית לכל תחומי חיינו, מתעוררות שאלות רבות לגביהן. מערכות אלו נמצאות בכלי הרכב שלנו ומסייעות לנו בשמירת מרחק, שמירה על נתיב הנסיעה ומניעת התנגשות. הן מגינות עלינו מפני הונאות בכרטיסי האשראי שלנו, הן ממליצות לנו אילו חדשות לקרוא ועוד. מערכות אלו נתפסות כ"קופסאות שחורות". לא ברור כיצד הן פועלות ולא ברור כיצד ועל סמך מה הן מגיעות למסקנות. יתרה מזו, הוכח לא אחת שקיים סיכון רב להטיות ואפליות במערכות אלו, בפרט כאשר הן פועלות בסביבה שלא הותאמו מראש אליה, עקב מאפייניהן. הן מבוססות על לימוד תופעה מתוך דוגמאות מאפיינות האמורות לייצג את העולם בו הן יפעלו, אלא שלא תמיד אוסף הדוגמאות המשמש לאימון מייצג את העולם האמיתי. אחת השאלות המטרידות יותר ויותר את הקהל הרחב, כמו גם את אנשי המקצוע ואת מקבלי ההחלטות: כיצד להפוך את טכנולוגיות המחשוב, בעיקר אלו המבוססות על בינה מלאכותית, להוגנות, ידידותיות ומובנות יותר למשתמשי הקצה? המחקר מתמקד בבחינת האתגרים השונים - החל בגורמים לבעיה, בין אם אלו הנתונים או האלגוריתם או הכיוונון של המשתמשים, דרך דרכים למנוע הטיות/חוסר הגינות ועד לזיהוי תקלות. בשלב הנוכחי מתמקד המחקר בתפיסת המשתמשים את הוגנותן של מערכות אלגוריתמיות ובתפקידם של הסברים בהפיכת המערכות לשקופות, הגברת האמון ושכנוע המשתמשים לבטוח בהן. כיוון מחקר עתידי מתמקד בהטמעת הוגנות ושקיפות בתהליכי הפיתוח המקובלים של מערכות מידע - החל משלב הגדרת דרישות המערכת והתייחסות לדרישות ההוגנות והשקיפות, דרך התוכן, המימוש והבדיקה לבחינת העמידה בדרישות.

1. פרופ' טל ז'רסקי, הפקולטה למשפטים, אוניברסיטת חיפה: קבלת החלטות אוטומטיות/אוטונומיות

מחקרים בנושאים הנוגעים להחלטות אוטומטיות/אוטונומיות ובכללן בינה מלאכותית: מחקר בהגדרת ההוגנות בהחלטות אלו - תוך הבנה שחלק מההגדרות להוגנות סותרות ודורשות איזון. מחקר נוסף, בוחן דרכים חדשניות לזיהוי חוסר הוגנות בקבלת ההחלטות. המחקר עוסק באסדרה בפועל של החלטות אלו במדינות שונות, בדרישת השקיפות ובאופיה של טענת האפליה בהחלטות מעין אלו.

1. פרופ' מרגריטה (ריטה) אוסדצ'י, החוג למדעי המחשב, באוניברסיטת חיפה: למידת מכונה ולמידה עמוקה

תחומי המחקר המרכזיים הם: למידה עמוקה פרקטית - רשתות עמוקות מאפשרות לפתור בעיות רבות, אולם דורשות הרבה משאבים, גם מבחינת דוגמאות למידה (שדורשות תיוג ידני) וגם מבחינת משאבים חישוביים. המחקרים תרמו בשני אופנים: לגבי דוגמאות הלמידה, פותחו שיטות חדשניות ללמידה מדוגמאות מעטות וללמידה מרצפים מתחלפים של דאטה. לגבי המשאבים החישוביים, פותחה דרך להפחית משאבים חישוביים של רשתות על ידי שיטת צמצום שעבורה ניתן להוכיח מתמטית את רמת הרובוסטיות של הרשת המצומצמת מבלי להתפשר על רמת הביצועים שלה.

הבטחת פרטיות: המחקר מתמקד בשמירת הפרטיות של מידע פרטי של אנשים. הוא כולל שיטות לשמירה על פרטיות במערכות לזיהוי ביומטרי. העבודה שעוסקת בחישוב בטוח של זיהוי פנים זכתה בפרס BEST PAPER בכנס מרכזי של אבטחת מידע. עבודות רבות בעולם התבססו על המאמר הזה.

פרשנות של רשתות עמוקות: 1) שימוש בלמידה עמוקה למידול של קבלת החלטות של אנשים ושימוש במודלים קוגניטיביים כדי לפרש את החיזוי של רשתות עמוקות אלו. 2) פיתוח מודלים חדשניים לשקיפות של רשתות סיווג עמוקות בתחומי מדעי הטבע (כימיה, פיזיקה וביולוגיה).

1. ד"ר אורן זלצמן, הפקולטה למדעי המחשב, טכניון: אלגוריתמיקה של תכנון ותנועה רובוטית

המחקרים מתמקדים בבחינה מחודשת של אלגוריתמים קלאסיים של מדעי המחשב, כלים ופרדיגמות, על מנת להתמודד עם האתגרים החישוביים המתעוררים בעת תכנון תנועה של רובוטים.

באמצעות שילוב טכניקות מתחומים מגוונים כגון גיאומטריה חישובית, תורת הגרפים ולמידה חישובית, השאיפה היא לספק אלגוריתמים יעילים מערכות רובוטיות מורכבות.

המעבדה לרובוטיקה חישובית מהווה מרכז למחקר בסיסי עבור בעיות תכנון רובוטיות. במעבדה מתמקדים במערכות רובוטיות מורכבות כגון אלו הנמצאים במרכזים לוגיסטיים ובאפליקציות רפואיות. לאחר זיהוי האתגרים החישוביים המאפיינים את המערכת, במעבדה מפתחים אלגוריתמים העונים על האתגרים הללו על מנת לספק לקהילה הרובוטית כלים פרקטיים המעוגנים בתיאוריה.

# חברי הסגל במרכז כולל הפרופיל המחקרי שלהם והדגשים הבין תחומיים

|  |
| --- |
| חברי סגל חוקרים : שיתופי פעולה אוניברסיטת חיפה והטכניון: |
| 1. **פרופ' מרכז מחקר משותף: אינטליגנציה 2.0: מרכז לחקר האינטראקציה בין אינטליגנציה אנושית לאינטליגנציה ממוחשבת:** 2. **פרופ' אלדד יחיעם, הנדסה תעשיה וניהול, טכניון:**   [**https://ie.technion.ac.il/~yeldad/**](https://ie.technion.ac.il/~yeldad/)   1. **פרופ' ברוך עיטם, פסיכולוגיה, אוניברסיטת חיפה:**   [**https://www.researchgate.net/profile/Baruch-Eitam**](https://www.researchgate.net/profile/Baruch-Eitam)   1. **פרופ' עפרה עמיר, טכניון:**   [**https://scholar.harvard.edu/oamir/home**](https://scholar.harvard.edu/oamir/home)  [**Ofra Amir | The Faculty of Industrial Engineering & Management (technion.ac.il)**](https://web.iem.technion.ac.il/site/academicstaff/ofra-amir/)   1. **פרופ' עדו ערב, מדעי ההתנהגות, טכניון:**   **https://web.iem.technion.ac.il/site/academicstaff/ido-erev/**   1. **פרופ' מור פלג, מערכות מידע, אוניברסיטת חיפה:**   [**https://sites.google.com/hevra.haifa.ac.il/mpeleg**](https://sites.google.com/hevra.haifa.ac.il/mpeleg)   1. **ד"ר ארנון קרן, פילוסופיה, אוניברסיטת חיפה:**   [**https://www.researchgate.net/profile/Arnon-Keren**](https://www.researchgate.net/profile/Arnon-Keren) |
| 1. **פרופ' עפרה עמיר, טכניון וד"ר גיא אבני, אוניברסיטת חיפה: מחקר משותף בבינה מלאכותית לפיתוח שיטות שיתופיות ממוקדות משתמש להסברת התנהגותן של מערכות המקבלות החלטות לאורך זמן.** 2. **ד"ר גיא אבני:**   [**https://sites.google.com/view/gavni**](https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fsites.google.com%2Fview%2Fgavni&data=04%7C01%7Cdina.raphael%40technion.ac.il%7C91365b6946b043ff0b5c08d9475331fc%7Cf1502c4cee2e411c9715c855f6753b84%7C1%7C0%7C637619244922112295%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=bcsFKZZ2%2BcR7Kn1qFSIJx7iql1gFCq0Pq9fTmarvHlE%3D&reserved=0) |
| 1. **פרופ' אורי פלונסקי, טכניון וד"ר יפים רוט, אוניברסיטת חיפה: מחקר משותף בתחום קבלת החלטות התנהגותיות, תהליכי למידה אנושיים, מידול אופן פעולה וחיזוי התנהגות, כלכלה התנהגותית ויישומים של קבלת החלטות התנהגותיות למדיניות ציבורית.** 2. **פרופ' אורי פלונסקי:**   [**https://oriplonsky.com**](https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Foriplonsky.com%2F&data=04%7C01%7Cdina.raphael%40technion.ac.il%7C23d8c88f43b840b0356408d94a953c6a%7Cf1502c4cee2e411c9715c855f6753b84%7C1%7C0%7C637622827092504995%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=GvXnn4YTz7IXhA1cAdec4CFFoNuZX422mC85NhUezzY%3D&reserved=0)   1. **ד"ר יפים רוט:**   [**https://hw.haifa.ac.il/en/people/human/yroth1**](https://hw.haifa.ac.il/en/people/human/yroth1) |
| 1. **פרופ' רקפת אקרמן וד"ר תומר שגיא: המחקר נערך ביחד עם עמית נוסף מהטכניון, פרופ' אבי גל, ועם סטודנט משותף לדוקטורט, ד"ר רועי שרגא. המחקר עוסק בשימוש בתובנות מתוך מחקר בסיסי במדעי החברה לתועלת של אלוגריתמים אוטומטיים למיזוג מסדי נתונים. במחקר שפורסם בשנת 2019, אלגוריתמים ואנשים ביצעו את אותה מטלה. על פי ממצאים קודמים בשני התחומים, שילבנו את החוזקות של האלגוריתמים בנקודות החולשה של האנשים ולהיפך. התוצאה היתה מיזוג מסדי נתונים טוב מכל התהליכים הידועים עד לאותה עבודה.**      1. **רקפת אקרמן:**   [**Rakefet Ackerman | The Faculty of Industrial Engineering and Management, Technion**](https://ackerman.net.technion.ac.il/)   1. **פרופ' אביגדור גל:**   [**http://ie.technion.ac.il/~avigal**](http://ie.technion.ac.il/~avigal)  [**https://agp.iem.technion.ac.il/avigal/**](https://agp.iem.technion.ac.il/avigal/)  [**/**](https://agp.iem.technion.ac.il/avigal/)   1. **ד"ר תומר שגיא:**   [**https://is-web.hevra.haifa.ac.il/index.php/en/staff-4/faculty/18-english/staff/236-dr-tomer-sagi**](https://is-web.hevra.haifa.ac.il/index.php/en/staff-4/faculty/18-english/staff/236-dr-tomer-sagi)   1. **ד"ר רועי שרגא:**   [**https://web.iem.technion.ac.il/site/he/three-doctoral-students-and-a-master-student-winning-excellence-scholarships/**](https://web.iem.technion.ac.il/site/he/three-doctoral-students-and-a-master-student-winning-excellence-scholarships/) |
| 1. **פרופ' רקפת אקרמן ופרופ' אשר קוריאט: היבטים מטא-קוגניטיביים של אינטראקציה בין אדם למחשב: האתגרים העומדים בפני אנשים בעת עבודה ולמידה בסביבות מקוונות. כלים לניתוח שימושיות של יישומים ממוחשבים.**   **פרופ' רקפת אקרמן ופרופ' אשר קוריאט: מחקר משותף נערך בענף שעסק בהתפתחות השיפוט העצמי של מידת ההצלחה הצפויה וההחלטות לגבי ויסות מאמצי חשיבה המתבססות על השיפוט הזה.**   1. **פרופ' אשר קוריאט:**   [**Prof. Asher Koriat - Publications (haifa.ac.il)**](https://iipdm.haifa.ac.il/index.php/koriat-publications) |
| 1. **פרופ' רקפת אקרמן ופרופ' מוריס גולדשמיט: שיתוף פעולה מחקרי בשני קוי מחקר.**   **האחד כלל את המחקר הראשון בתוך קו מחקר שעסק בהשוואת ויסות מאמצי חשיבה כאשר המטלה מוצגת על מסך המחשב לעומת הצגה של אותה מטלה על דף נייר.**  **קו המחקר השני, עסק באתגר של מענה לשאלות ידע קשות במצבים חברתיים. במחקר נמצא שאנשים מתקשים מאד לענות "לא יודע" והם ישתמשו בה רק כאשר הם לא מוצאים שום אפשרות תשובה בעלת סיכוי מספק להיות נכונה וגם כוללת תרומה משמעותית של מידע לאדם ששאלה את השאלה. פרופ' אקרמן המשיכה את שני קוי המחקר הללו עם סטודנטים בטכניון.**   1. **פרופ' מוריס גולדשמיט:**   [**Prof. Morris Goldsmith (haifa.ac.il)**](https://iipdm.haifa.ac.il/index.php/2014-10-01-08-42-51/faculty/4-prof-morris-goldsmith) |

|  |
| --- |
| חברי סגל מהטכניון: |
| 1. **פרופ' יוחאי כרמל: הנדסת הסביבה בטכניון. פיתוח כלים ומתודולוגיה לחקר המגוון הביולוגי עתיר הנתונים.**   [**Carmel’s Ecology Lab | Technion**](https://ecologylab.net.technion.ac.il/)  [**Biodiversity Informatics | Carmel’s Ecology Lab (technion.ac.il)**](https://ecologylab.net.technion.ac.il/research-3/research-2/biodiversity-informatics/) |
| 1. **פרופ' מרים זקסנהויז: למידת מכונה לבקרה רובוטית ממשקי מחשב מוח פולשניים (**BMIs**) ממשקי מחשב מוח לא פולשניים (**BCIs**) בקרה מוטורית חישובית מתנדים עצביים ושליטה בתנועות קצביות שליטה בהשראה ביולוגית של הליכה דינמית של רובוטים דו-קוטביים גנרטורים תבנית מרכזית (**CPGs**).**   [**https://meeng.technion.ac.il/members/miriam-zacksenhouse/**](https://meeng.technion.ac.il/members/miriam-zacksenhouse/) |
| 1. **פרופ' עזרי טרזי: חשיבה עיצובית גנרטיבית באמצעות בינה מלאכותית ויצירה פיסולית מורפולוגית בהדפסת תלת מימד.**   [**https://designtech.net.technion.ac.il**](https://designtech.net.technion.ac.il) |
| 1. **פרופ' אסף שוסטר: למידה עמוקה מבוזרת ומדרגית; למידה עמוקה לרפואה אישית; אקראיות בלמידה עמוקה; ניתוח של זרמי נתונים מהירים; עיבוד אירועים מורכבים (**CEP**); אינטרנט של דברים ומערכות חכמות; פרטיות; אבטחת סייבר; ניהול ענן.**   [**MLIS - The Machine Learning and Intelligent Systems | Relevant Projects (interia.website)**](http://mlis.interia.website/relevant-projects/?id=802) |
| 1. **פרופ' מירי ברק: יישומי ענן בחינוך המדעי-הנדסי.**   Google scholar profile: [**https://scholar.google.co.il/citations?user=P7QmUhsAAAAJ&hl=en**](https://scholar.google.co.il/citations?user=P7QmUhsAAAAJ&hl=en)  Personal Website: [**http://barakmiri.net.technion.ac.il**](http://barakmiri.net.technion.ac.il/)  Research Group Website: [**http://sltgroup.wix.com/Technion**](http://sltgroup.wix.com/Technion) |
| 1. **ד"ר אורן זלצמן: רובוטיקה, תכנון תנועה אלגוריתמי, יסודות תכנון רובוטי, אתגרים חישוביים ברובוטיקה.**   [**http://orensalzman.com/**](http://orensalzman.com/) |
| 1. **ד"ר יונתן בלינקוב: עיבוד שפה טבעית, כתת-תחום של בינה מלאכותית. שימוש בעיבוד שפה טבעית לטובת מדעי הרוח ואף פרסמתי בנושא מדעי הרוח הדיגיטליים בעבר.**   [**https://www.cs.technion.ac.il/~belinkov/**](https://www.cs.technion.ac.il/~belinkov/)  [**Educational Neuroimaging Center (ENIC) | Neurocognitive Center for Language, Literacy, and Reading Development (technion.ac.il)**](http://neuroimaging-center.technion.ac.il/) |
| 1. **פרופ' דב דורי: תחומי המחקר כוללים הנדסת מערכות מבוססת מודלים, מידול קונספטואלי של מערכות מורכבות, ארכיטקטורת מערכות ועיצוב, הנדסת תוכנה ומערכות וביולוגיה של מערכות. פרופ' דורי המציא ופיתח מתודולוגיית תהליך אובייקט** (OPM**), שאומצה לאחרונה כ-** ISO **19450.**   [**https://web.iem.technion.ac.il/site/academicstaff/dov-dori/**](https://web.iem.technion.ac.il/site/academicstaff/dov-dori/) |
| 1. **ד"ר יועד קנת: חקר המורכבות של קוגניציה ברמה גבוהה יותר באוכלוסייה טיפוסית וקלינית, תוך התמקדות ביצירתיות, מחשבה אסוציאטיבית, ידע וחיפוש זיכרון. כדי לחקור נושאים אלה, הוא מיישם שיטות חישוביות, התנהגותיות ועצביות.**   [**The Cognitive Complexity Lab | Faculty of Industrial Engineering & Management (technion.ac.il)**](https://cognitive-complexity.net.technion.ac.il/) |
| 1. **ד"ר שרה קרן: תכן זיהוי מטרות (Goal recognition design) היא גישה אלגוריתמית חדשנית המשפרת את יכולת ההיסק לגבי מטרות משתמשים (סוכנים) הפועלים במערכת.**   [**Sarah Keren Academics (wixsite.com)**](https://sarahkeren.wixsite.com/sarahkeren-academics) |
| 1. **ד"ר אליק פלמן: פילוסופיה של המדע, השפה והאתיקה של הסביבה.**   [**Alik Pelman | Technion Israel Institute of Technology - Academia.edu**](https://technion.academia.edu/AlikPelman) |
| 1. **פרופ' ראובן כץ: ממשקי מחשב למוח (BCI) לשיקום; תכנון מכני ותכנון אופטו-מכני; מטרולוגיה אלקטרו-אופטית ובדיקה; רטט של מכונות.**   [**https://meeng.technion.ac.il/he/members/reuven-katz**](https://meeng.technion.ac.il/he/members/reuven-katz/) |
| 1. **פרופ' יעקב בן חיים: קבלת החלטות בתנאים של אי-וודאות.**   [**https://yakovbh.net.technion.ac.il/**](https://yakovbh.net.technion.ac.il/)  [**https://info-gap.technion.ac.il/**](https://info-gap.technion.ac.il/) |
| 1. **פרופ' שגיא דליות: כרייה ואינטגרציה של נתונים טמפורליים-מרחביים (מדע אזרחי ומיפוי השתתפותי), שמטרתם להעשיר ולהגדיל את בסיס הידע הסביבתי הקיים ולפתח שירותים חדשים מבוססי מיקום. המחקר עוסק בעיקר באופטימיזציה של נתיבים, הכרה סביבתית, העשרת מערכות ניטור סביבתי, שירותים מבוססי מיקום וזיהוי דפוסים והתנהגות של אובייקטים נעים.**   [**http://ecsl.net.technion.ac.il/**](http://ecsl.net.technion.ac.il/)  [**https://mlis.technion.ac.il/people/**](https://mlis.technion.ac.il/people/) |
| 1. **פרופ' תומר טולדו: מידול וניתוח מערכות תחבורה, כולל מודלים של סימולציית תנועה וכיול ואימות שלהם, התנהגות נהיגה ונסיעות ומערכות תחבורה חכמות. אנא עיינו בדף המחקר שלי לקבלת פרטים נוספים.**   [**https://toledo.net.technion.ac.il/**](https://toledo.net.technion.ac.il/) |
| 1. **פרופ' אורית חזן: מדעי המחשב, הנדסת תוכנה וחינוך למדעי הנתונים.**   [**Orit Hazzan (technion.ac.il)**](https://orithazzan.net.technion.ac.il/) |
| 1. **פרופ' קרל מאטינס: כריית נתונים ולמידת מכונה במחקר המחבר תחבורה וצדק.**   [**Fair Transport Lab – Exploring justice of transport systems (technion.ac.il)**](https://fairtransport.technion.ac.il/)  [**Fair Transport Lab - Faculty of Architecture and Town Planning, IIT (technion.ac.il)**](https://architecture.technion.ac.il/research/labs/fair-transport-lab/) |
| 1. **ד"ר ליאת לבונטין מחקר משותף עם עפרה עמיר: מחקר בנושא ממשקי אדם-מחשב, והוא מציע תובנות על המנגנונים הפסיכולוגיים המכוונים את בחירותיהם של צרכנים פוטנציאליים.**   [**The Faculty of Industrial Engineering & Management | (technion.ac.il)**](https://web.iem.technion.ac.il/site/) |
| 1. **פרופ' איתן יעקובי: חקר מידע ותורת הקידוד עם יישומים לזיכרונות לא נדיפים, זיכרונות אסוציאטיביים, אחסון DNA, אחסון ואחזור נתונים ואחזור מידע פרטי.**   [**Eitan Yaakobi | Computer Science Department Technion**](https://yaakobi.net.technion.ac.il/) |
| 1. **פרופ' מירילה בן כהן: הגיאומטריה של צורות. גיאומטריה דיפרנציאלית דיסקרטית, אופטימיזציה מספרית וניתוח הרמוני, עבור יישומים כגון, אוספי צורות גלישה, אנימציה וסימולציה נוזלית. בהינתן שתי צורות אנושיות הדרך לעצב אלגוריתם שמוצא נקודות מתאימות כמייצג את "הצורה האנושית" מבחינה חישובית.**   [**https://mirela.net.technion.ac.il/**](https://mirela.net.technion.ac.il/) |
| 1. **פרופ' ענבל טלגם-כהן: תכנון וניתוח של אלגוריתמים עם יישומים כלכליים וחברתיים.**   [**Inbal Talgam-Cohen's Website**](http://www.inbaltalgam.com/) |
| 1. **ד"ר תמיר חזן: ההיבטים התיאורטיים והמעשיים של למידת מכונה. המחקר מתמקד בפתרונות מתמטיים לבעיות מודרניות בחיים האמיתיים הממחישות התנהגות סטטיסטית לא מסורתית.**   [**https://ie.technion.ac.il/~tamir.hazan/tamir.html**](https://ie.technion.ac.il/~tamir.hazan/tamir.html) |
| 1. **ד"ר יבגני ברזק: מחקר העוסק בצומת של עיבוד שפה טבעית (NLP) ומדע קוגניטיבי. כיצד ניתן להשתמש בתנועות עיניים במהלך הקריאה כדי להסיק את מאפייני הקורא ואת המצב הקוגניטיבי, ולשפר את NLP. תחומי עניין קשורים אחרים כוללים רב לשוניות, מתודולוגיית הערכה ורכישת שפה מעוגנת.**   [**Yevgeni Berzak Homepage (mit.edu)**](http://people.csail.mit.edu/berzak/index.html) |
| 1. **ד"ר רשף מאיר: המחקר עוסק בהצטלבות של חישוב וכלכלה, המכונה לעתים תורת המשחקים החישובית. בפרט, מנגנוני התנהגות ועיצוב אסטרטגיים המקדמים שיתוף פעולה, יציבות ורווחה חברתית בקרב סוכנים רציונליים או מוגבלים-רציונליים.**   [**Reshef Meir (technion.ac.il)**](https://reshef.net.technion.ac.il/) |
| 1. **פרופ' רועי רייכארד: מחקרים בצומת של עיבוד שפה חישובית, תקשורת אנושית (למשל שיחות, מדיה חברתית), מדעי החברה, ההתנהגות והבריאות (כלכלה, תורת המשחקים, פסיכולוגיה ופסיכיאטריה) וסיבתיות ופרשנות מודל.**   [**Roi Reichart Homepage (technion.ac.il)**](https://ie.technion.ac.il/~roiri/) |
| 1. **ד"ר אור אלכסנדרוביץ': מדע וטכנולוגיה של הבנייה, פיסיקה של הבנייה, מיקרואקלים עירוני, היסטוריה עירונית, היסטוריה של הארכיטקטורה ושל טכנולוגיות בנייה, מדעי הרוח והחברה הדיגיטליים.**   [**https://oraleks.net.technion.ac.il/**](https://oraleks.net.technion.ac.il/) |
| 1. **פרופ' יעל אלואיל: היסטוריה של המרחב הישראלי-פלסטיני, מתודולוגיות עיצוב של מגורים כאדריכלות-עתק, מגורים והעיר, ביטויים של אומנות ,סוגיית הארכיון בתחומים של היסטוריה ויזואלית ומרחבית.**   [**Yael Allweil – Faculty of Architecture and Town Planning (technion.ac.il)**](https://allweil.net.technion.ac.il/) |
| 1. **ד"ר יאיר גולדברג: מודלים מתמטיים בחקר מגיפות.**   [**https://yairgo.net.technion.ac.il/mathematical-models-in-epidemic-research**](https://yairgo.net.technion.ac.il/mathematical-models-in-epidemic-research/) |
| 1. **ד"ר דינה רפאל מארק-רייך: אתיקה של מכונות ויזמות.**   [**https://t-hub.technion.ac.il/team/**](https://t-hub.technion.ac.il/team/) |

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

|  |
| --- |
| חברי סגל חוקרים : מאוניברסיטת חיפה: |
| 1. **פרופ' אפרים לב: דיקן הפקולטה למדעי הרוח, אוניברסיטת חיפה.**   [**http://hcc.haifa.ac.il/images/CV-Efraim\_Lev1.10.20.pdf**](http://hcc.haifa.ac.il/images/CV-Efraim_Lev1.10.20.pdf)  [**https://genizah.haifa.ac.il/index.php?option=com\_content&view=article&id=12:2014-10-21-08-07-59&catid=14&Itemid=127&lang=he**](https://genizah.haifa.ac.il/index.php?option=com_content&view=article&id=12:2014-10-21-08-07-59&catid=14&Itemid=127&lang=he) |
| 1. **פרופ' מאיר חמו: פילוסופיה של הפיזיקה, פילוסופיה של המדע, פילוסופיה של הנפש וקוגניציה.**   [**https://scholar.google.com/citations?user=DNgP7zwAAAAJ&hl=en**](https://scholar.google.com/citations?user=DNgP7zwAAAAJ&hl=en) |
| 1. **פרופ' דני סטטמן: מתמחה באתיקה, פילוסופיה פוליטית ופסיכולוגיה מוסרית.**   [**https://www.danielstatman.com**](https://www.danielstatman.com) |
| 1. **פרופ' שאול סמילנסקי: מתמחה באתיקה, החוג לפילוסופיה.**   [**https://www.saulsmilansky.com**](https://www.saulsmilansky.com/) |
| 1. **פרופ' טל ז'רסקי: מחקר בהיבטיים משפטיים, רגולטורים ואתיים של מערכות החלטה אוטונומיות ואוטומטיות.**   [**https://law.haifa.ac.il/index.php/he/tzarsky**](https://law.haifa.ac.il/index.php/he/tzarsky) |
| 1. **ד"ר ברכה ניר: מתמחה בהיבטים שונים של תקשורת אנושית (כולל התפתחות התקשורת והשפה), בקשר בין שפה, כח וחברה, כמו גם בשימוש במערכות חישוביות לניתוח נתונים במדעי הרוח.**   [**https://hw.haifa.ac.il/en/people/csd/bnir**](https://hw.haifa.ac.il/en/people/csd/bnir) |
| 1. **פרופ' צבי קופליק: מומחה בממשקי אדם-מחשב חכמים, עוסק במחקר ב"שקיפות אלגוריתמית" - הפיכת מערכות ממוחשבות למובנות והוגנות למשתמשיהן.**   [**https://tsvikak.hevra.haifa.ac.il/**](https://tsvikak.hevra.haifa.ac.il/) |
| 1. **פרופ' אמנון רייכמן: משפט ציבורי, רגולציה, זכויות אדם, משפט במצבי חירום, סייבר, משפט חברה ותרבות, פורמליזם.**   [**https://law.haifa.ac.il/index.php/he/reichman**](https://law.haifa.ac.il/index.php/he/reichman) |
| 1. **פרופ' מיכל גל: משפט בינלאומי. הסדרת התנהגותם של גופים עסקיים שכוחם בשוק מתבסס על זכויות קניין רוחני, שהוא תולדה צפויה של מגמות עכשוויות ובראשן המידע.**   [**פרופ' מיכל גל (haifa.ac.il)**](https://law.haifa.ac.il/index.php/he/mgal#tab-2) |
| 1. **פרופ' דפנה רבן: כלכלת מידע, חברת המידע, אוריינות מידע, שיתוף מידע, ביבליומטריה, משחקים רציניים.**   [**https://scholar.google.com/citations?user=z-Wxh-8AAAAJ&hl=en&inst=11183788154206457931**](https://scholar.google.com/citations?user=z-Wxh-8AAAAJ&hl=en&inst=11183788154206457931) |
| 1. **פרופ׳ עירית הדר: תהליכים קוגניטיביים של מפתחי אלגוריתמים ומערכות, היבטים קוגניטיביים, ארגוניים ותרבותיים בהגנה על פרטיות המידע, שילוב חשיבה עיצובית בפיתוח טכנולוגיות מידע.**   [**https://sites.google.com/hevra.haifa.ac.il/dr-irit-hadar/home**](https://sites.google.com/hevra.haifa.ac.il/dr-irit-hadar/home) |
| 1. **פרופ' ריטה אוסדצ'י: למידה עמוקה ולמידת מכונה, ראייה ממוחשבת, פרטיות בנתונים ביומטריים, אבטחה ופרטיות בלמידת מכונה ומדעי הנתונים**   [**http://rita.osadchy.net/**](http://rita.osadchy.net/) |
| 1. **פרופ' מאיר יעיש: ריבוד ואי-שוויון חברתי, סוציולוגיה של החינוך, שוק העבודה, תאוריית הבחירה הרציונלית.**   [**https://soc.hevra.haifa.ac.il/~yaish/**](https://soc.hevra.haifa.ac.il/~yaish/) |
| 1. **ד"ר דן קוטליאר: חוקר אלגוריתמים של ביג דאטה ובינה מלאכותית והשפעתם על עיצוב מחדש של מושגי יסוד סוציולוגיים, כגון זהות, בחירה וכוח. עבודתו מבוססת על מחקר אמפירי בסצנת ההייטק הישראלית ובעמק הסיליקון, והיא חושפת את הדרמה החברתית שמאחורי פעולת האלגוריתמים. מחקריו מראים עוד איך ידע אלגוריתמי משתלב עם סוגי ידיעה מסורתיים יותר.**   [**https://behevrat-haadam.org**](https://behevrat-haadam.org/%D7%91%D7%A8%D7%9B%D7%95%D7%AA-%D7%9C%D7%93%D7%9F-%D7%A7%D7%95%D7%98%D7%9C%D7%99%D7%90%D7%A8-%D7%A9%D7%94%D7%AA%D7%A7%D7%91%D7%9C-%D7%9B%D7%97%D7%91%D7%A8-%D7%A1%D7%92%D7%9C-%D7%97%D7%93%D7%A9-%D7%91/)  [**http://isracom.org.il/.upload/media\_frames\_11\_Part3.pdf**](http://isracom.org.il/.upload/media_frames_11_Part3.pdf) |
| 1. **ד"ר קרני שגל פפרקורן: היא פוסט דוקטורנטית באוניברסיטת אוטווה העוסקת ברגולציה הגלובאלית שחלה על מערכות** AI**. תחומי העניין של קרני כוללים גם את דיני הנזיקין ותחולתם על מערכות** AI**, וכן את אידיאל ה-**Responsible AI **והתהליכים באמצעותם מפתחי מערכות יוכלו להתקדם לעברו.**   [**Karni Chagal-Feferkorn | Centre for Law, Technology and Society | University of Ottawa (uottawa.ca)**](https://techlaw.uottawa.ca/aisociety/people/chagal-feferkorn-karni) |
| 1. **ד"ר משה בלידשטיין: מחקר באש ובאור בדתות העולם העתיק,** [**פרויקט דיגיטלי**](http://tiresias.haifa.ac.il/) **שמטרתו לבנות ולהנגיש טקסטים מהעולם העתיק בחיפוש נושאי.**   [**https://scholar.google.com/citations?user=ZXT6tbcAAAAJ&hl=iw**](https://scholar.google.com/citations?user=ZXT6tbcAAAAJ&hl=iw)  [**https://haifa.academia.edu/MosheBlidstein**](https://haifa.academia.edu/MosheBlidstein) |
| 1. **ד"ר שי סרוגו: מחקר בקבוצות המתמקד בתיעוש שבין תחילת ה-20 לבין מלחמת העולם השנייה, תקופת עידן המכונות והתיעוש, הקולוניאליזם האירופי, והופעת מדינת הלאום.**   [**Shai Srougo | University of Haifa - Academia.edu**](https://haifa.academia.edu/SSrougo) |
| 1. **ד"ר ארז פירט: חוקר אתיקה ובינה מלאכותית, פילוסופיה של המדעים.**   [**https://wolffund.org.il/2020/03/25/%D7%90%D7%A8%D7%96-%D7%A4%D7%99%D7%A8%D7%98/**](https://wolffund.org.il/2020/03/25/%D7%90%D7%A8%D7%96-%D7%A4%D7%99%D7%A8%D7%98/) |
| 1. **ד"ר אלכס אלטשולר: חוקר בינתחומי המתמחה במחשבה חברתית ובמגוון סוגיות הקשורות להיערכות והתמודדות עם סיכונים, משברים ומצבי חירום.**   [**https://ash.harvard.edu/people/alex-altshuler**](https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fash.harvard.edu%2Fpeople%2Falex-altshuler&data=04%7C01%7Cdina.raphael%40technion.ac.il%7Ca44e067d688f4cd529ab08d94c76445e%7Cf1502c4cee2e411c9715c855f6753b84%7C1%7C0%7C637624893114191157%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=qL7iAKJW50B1lyFcqqZyj9NnyvXtxnv9B109etKtN3E%3D&reserved=0) |

# תוכניות עתידיות

* **פיתוח תכנית לימודים לתואר שני בנושא של בינה מלאכותית ואמנויות יצירה (ראו לעיל).**
* **פיתוח של תכנית דו-חוגית לתואר ראשון אמנויות יצירה ובינה מלאכותית (ראו לעיל).**
* **בנייה ומיסוד של מעבדת עבודה לבניית דיאלוג, עדכון והפרייה בינתחומיים.**
* **שיתופי פעולה מחקריים והוראתיים חוצי פקולטות;**
* **ליבת הוראה משותפת לדוקטורנטים ופוסט דוקטורנטים (סדנאות וסמינרים).**
* **פיתוח תוכנית לימודים בינלאומית; למידה מקוונת בינמוסדית ובינלאומית;**

1. [Artificial intelligence articles | McKinsey & Company](https://www.mckinsey.com/featured-insights/artificial-intelligence): [Artificial intelligence articles | McKinsey & Company](https://www.mckinsey.com/featured-insights/artificial-intelligence) [↑](#footnote-ref-1)
2. Mind the gap! On the future of AI research: [Mind the gap! On the future of AI research | Humanities and Social Sciences Communications (nature.com)](https://www.nature.com/articles/s41599-021-00750-9) [↑](#footnote-ref-2)
3. CSRankings: Computer Science Rankings: [CSRankings: Computer Science Rankings](http://csrankings.org/#/fromyear/2016/toyear/2020/index?ai&vision&mlmining&nlp&ir&europe) [↑](#footnote-ref-3)
4. Ideation [↑](#footnote-ref-4)
5. GARTNER ISRAEL ADVISORY LTD: <https://www.gartner.com/en> [↑](#footnote-ref-5)
6. Prof. Amy C. Admondson: Harvard Business School: Teaming: How organizations learn, innovate and compete in the knowledge economy. <https://www.hbs.edu/faculty/Pages/profile.aspx?facId=6451>: [↑](#footnote-ref-6)
7. Prof. Amy C. Admondson: ‘Psychological Safety’: How to turn a group of strangers into a team. [↑](#footnote-ref-7)
8. <https://rework.withgoogle.com/print/guides/5721312655835136/> [↑](#footnote-ref-8)
9. <https://embeddedethics.seas.harvard.edu/about> [↑](#footnote-ref-9)
10. Sustainable Development Goals: <https://www.unsdsn.org/sdg-index-and-monitoring?gclid=CjwKCAjwq7aGBhADEiwA6uGZpxevYUu_Sve7wSg1bF2AV4JBDRzYwIOezA158rTk185F2dIqXjAPOxoCm6kQAvD_BwE> [↑](#footnote-ref-10)
11. <https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf> [↑](#footnote-ref-11)