

**ועדת משנה של המיזם
הלאומי למערכות נבונות
בנושא אתיקה ורגולציה של
בינה מלאכותית**

דין וחשבון

נובמבר , 2019

חברי הוועדה:

יו"ר הוועדה: פרופ' קרין נהון, המרכז הבינתחומי הרצליה ונשיאת איגוד האינטרנט הישראלי
מזכירת הוועדה: ד"ר דלית קן דרור פלדמן, מנחה משפטית, הקליניקה למשפט, לטכנולוגיה ולסייבר, אוניברסיטת חיפה ומרצה במכללה האקדמית צפת

חברי הוועדה על פי סדר הא' ב':

- מר יואב אבנשטיין, נציג ישראל בוועדת התקינה הבינלאומית בעניין בינה מלאכותית ובוועדת האתיקה האירופאית לבינה מלאכותית
- מר אורי אליאביב, יועץ עיסקי ומייסד קהילת Machine & Deep Learning Israel
- פרופ' ניבה אלקין קורן, ראשת המרכז למשפט וטכנולוגיה והמרכז לסייבר, משפט ומדיניות, הפקולטה למשפטים, אוניברסיטת חיפה
- עו"ד עמית אשכנזי, היועץ המשפטי, מערך הסייבר הלאומי, משרד רוה"מ
- עו"ד יערה בן שחר, היועצת המשפטית, רשות התקשוב הממשלתית, משרד רוה"מ
- מר אולג ברודט, מנהל החדשנות של Cyber@BGU ומנהל מו"פ במעבדות החדשנות של דויטשה טלקום בישראל
- ד"ר רן גלעד-בכרך, מחלקת המחקר Microsoft Research, מיקרוסופט
- מר שלומי הוד, מדען נתונים
- עו"ד יורם הכהן, מנכ"ל איגוד האינטרנט הישראלי (ע"ר)
- גבי ענבר נאור, חברת טאבולה (Taboola)
- מר שי פלצ'י, יועץ לנושאי למידה חישובית ומדע המידע
- עו"ד צפירי נוימן, היועץ המשפטי, הלשכה המשפטית, הרשות הלאומית לחדשנות טכנולוגית
- ד"ר ניל נמיר, יועץ מדיניות מו"פ, מועצה לאומית למחקר ופיתוח
- מר דן קוטליאר, המחלקה לסוציולוגיה ואנתרופולוגיה, האוניברסיטה העברית
- מר עמית קרן, מנכ"ל דויטשה טלקום ישראל
- ד"ר תהילה שוורץ אלטשולר, ראש התוכנית לדמוקרטיה בעידן המידע, המכון הישראלי לדמוקרטיה
- ד"ר רועי שיינדורף, המשנה ליועץ המשפטי לממשלה (משפט בין-לאומי)
- מר עידן שער, יועץ בתחום הסייבר והבינה המלאכותית

משקיפים:

- גבי שירה ריבנאי בהיר, המחלקה לסוציולוגיה ואנתרופולוגיה, אוניברסיטת בן-גוריון
 - עו"ד לימור שמרלינג מגזניק, מנהלת המכון הישראלי למדיניות טכנולוגיה
- צוות כתיבת הדו"ח (לפי סדר הא-ב):** עו"ד עמית אשכנזי, ד"ר רן גלעד-בכרך, פרופ' קרין נהון, ד"ר דלית קן דרור פלדמן, מר עמית קרן, ד"ר תהילה שוורץ אלטשולר

גילויים נאותים:

העמדות המובעות בדו"ח אינן בהכרח מייצגות את עמדות המוסד או הארגון אליהם שייכים חברי הוועדה.

הדו"ח להלן נכתב ע"י צוות האתיקה והרגולציה שעבד במסגרת המיזם הלאומי למערכות נבונות בראשות פרופ' בן ישראל ופרופ' מתניה. הדו"ח הסופי של המיזם ומסקנותיו נמצאים בשלבי עבודה סופיים, ויכללו גם חלק מהמסקנות של הדו"ח להלן, לאחר התאמות בראייה כוללת.

תקציר מנהלים והמלצות הוועדה

ועדה זו הינה ועדת משנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית. על הוועדה הוטל לדון בהיבטים האתיים המשפטיים והרגולטוריים של תחום הבינה המלאכותית. בנוסף הוטל על הוועדה להמליץ על עקרונות מנחים בהקשר הישראלי אשר ילקחו בחשבון כחלק מהתכנית הלאומית, שתוגש לרוה"מ, להפיכת ישראל למובילה בתחום.

המסמך נכתב עבור שני קהלי-יעד שונים: ראשית, ככלי עזר למקבלי החלטות בתחום ועבור מפתחי מערכות בינה מלאכותית. הדוח עוסק בטכנולוגיות ובשימושים שכיחים בתחום הבינה המלאכותית כיום ובשנים הקרובות ועל כן התמקד בהיבטים של בינה מלאכותית צרה ובעיקר למידת מכונה. הנחה מרכזית שליוותה את כתיבת הדוח היא ששילובם של שיקולים אתיים בשלבי הפיתוח והתחזוקה של מוצר משולב בינה מלאכותית, לטווח הארוך, יעניקו יתרון לחברות ולמדינות שפועלות לפיהן. בנוסף, הוועדה העדיפה לנקוט בגישה של התערבות מידתית, ותוך הנחה כי יש לתת מענה מראש למקרים בעייתיים אפשריים ולא רק בדיעבד.

יוער כבר עתה, הוועדה לא עסקה בשאלות של שיקולים רגולטוריים ואתיים בתחום דיניי הלחימה.

מבנה הדוח

הפרק הראשון בדוח דן במאפיינים הייחודיים של טכנולוגיית הבינה המלאכותית, אשר רלוונטיים לדיון בנושא המשילות האתית והמשפטית בתחום הבינה המלאכותית.

הפרק השני הפרק השני בדוח הציג את המלצות הוועדה ביחס לעקרונות הערכיים והאתיים שיש להחיל על פעילות פיתוח ושימוש בטכנולוגיית בינה מלאכותית. המלצות אלה גובשו בהתבסס על ההיבטים הייחודיים שאותרו בידי הוועדה, בזיקה לדיונים דומים המתנהלים בגופים בינלאומיים ציבוריים ופרטיים.

בנוסף הוועדה פיתחה כלי עזר למקבלי החלטות. כלי העזר בנוי מקבוצת שאלות שמקבלי החלטות צריכים לענות עליהם בבואם לבדוק כשלים אתיים אפשריים, וממפת שכיחות, שמראה את השכיחות של תופעות אתיות לאור שרשרת הפיתוח של המוצר.

הפרק השלישי הפרק השלישי מציג את האפשרויות השונות בתחום האסדרה של בינה מלאכותית, כחלק מהדיון אודות התפקיד של המשפט ומוסדות משפטיים בקידום ודאות, מימוש עקרונות ערכיים ותמיכה בחדשנות של בינה מלאכותית.

פרק ראשון: מה ייחודי בטכנולוגיית הבינה המלאכותית:

1. מערכות בינה מלאכותית נוטות להקצין יחסי גומלין קיימים במערכות חברתיות.
2. האתגר "להנדס" ערכים בייחוד במערכות בינה מלאכותית הפועלות בתחומים שבהם ההחלטות שהופעלו באופן מסורתי בידי גורם אנושי.
3. סיכונים לפרטיות בשל העובדה שמערכות בינה מלאכותית מבוססות על עיבודי מידע רב, ממקורות שונים.
4. מורכבות מוצרים שלובי בינה מלאכותית מעוררים קושי בהשגת אמון הציבור.
5. שינויים בתחרותיות בשוק והגדלת הפערים בין חברות גדולות וקטנות.
6. שינוי בקטגוריות מוכרות לאחריות של יצרני מוצרים, ספקי שירותים, ועסקים הנשענים עליהם.

פרק שני: אתיקה ובינה מלאכותית

- הוועדה קוראת לבנות תכניות של ידע והכשרה לאתיקה של אנשי המקצוע העוסקים בפיתוח מערכות בינה מלאכותית.
- הוועדה מאמינה כי שמירה על מערכת עקרונות אתיים על פיהם יפעלו ארגונים פרטיים והסקטור הציבורי, הוא קריטי לטווח ארוך.
- הוועדה זיהתה את העקרונות האתיים הבאים כחשובים וכאלה שמקבלי החלטות צריכים לשים לב אליהם:
 1. הוגנות – חתירה לשוויון מהותי, מניעת הטיות (במידע, בתהליך, בתוצר), מניעת הפליה, והימנעות מהגדלת פערים חברתיים, כלכליים והשכלתיים.
 2. אחריותיות (Accountability)

- א. שקיפות –הנגשה, בהתאם להקשר ולנסיבות, של מידע על התהליך עצמו ועל דרך קבלת ההחלטה.
- ב. הסברותיות – הסבר הפעולה וההחלטה - היכולת להסביר את תהליך קבלת ההחלטה של המערכת (ברמת המשתמשים כפרטים, גם ברמת הכלל אם המערכת משפיעה על קבוצות, וגם עבור מפעילי המערכת עצמם).
- ג. אחריות וניהול סיכונים – חלוקת האחריות האתית והמשפטית לגורמים הרלבנטיים בשרשרת הערך, וחלוקת ניהול הסיכונים ביניהם. קביעת האחריות ליצירת כללים לאמצעים סבירים למניעת הסיכון בהתאם להקשר ולעוצמת הסיכון הצפוי, לניהול הסיכונים ולמינויו של גורם אחראי לניהול הסיכונים.
3. כיבוד זכויות אדם והגנה עליהן – כיבוד כלל זכויות האדם ומודעות לצורך הגובר בהגנה על זכויות מסוימות שעלולות להיפגע יותר בעידן הבינה המלאכותית וביניהן:
- א. שמירה על שלמות הגוף - הגנה מפני פגיעה בחייו או בגופו של אדם.
- ב. פרטיות – הגנה על פרטיות, לרבות מניעת פגיעה בפרטיות באמצעות איסוף המידע, ניתוח ועיבוד המידע, שיתוף המידע ובשימושים אחרים וחדשים בו.
- ג. השמירה על האוטונומיה - שמירה על יכולתו של אדם לקבל החלטות תבוניות, ובין השאר, מניעת השפעה שאינה הוגנת או אינה מודעת על התנהגות היחיד.
- ד. זכויות אזרחיות ופוליטיות – לרבות הזכות לבחור, הזכות לחופש הביטוי והזכות לחופש הדת והמצפון.
4. הגנת סייבר ואבטחת מידע – שמירה על התפקוד התקין של המערכות, הגנה על המידע המשמש אותן ומניעת שימוש לרעה בידי גורם זדוני.
5. בטיחות - מניעת סכנה (לפרט, לחברה) וצמצום גרימת נזק.
- א. בטיחות "פנימית" התמקדות בבטיחות תהליך הפיתוח של כלי הבינה המלאכותית.
- ב. בטיחות חיצונית – התמקדות בסכנה והנזק עבור הסביבה והלקוחות, כתוצאה מהשימוש.
6. קיום שוק תחרותי – שמירה על שוק תחרותי ועל כללי משחק מאפשרים.
- הוועדה ממליצה לראות בכל העוסקים בתחום הבינה המלאכותית כאחראים לפעול באופן חוקי ואתי. על העוסקים בתחום להישאר מעודכנים בנושא ואי עמידה בתנאים אלו עלולה להיחשב כרשלנות.
 - הוועדה פיתחה כלי לבחינת האתגרים האתיים עבור מקבלי החלטות. הכלי בנוי משני חלקים:
 - חלק א' מורכב מקבוצת שאלות ראשוניות שיש לשאול בעת פיתוח מוצר בינה מלאכותית והבחנות את מידת ההשפעה של מערכת הבינה המלאכותית. שאלות אלה מיועדות למפתחי המוצרים לאורך כל שרשרת הפיתוח והייצור, ויש לעדכן אותן מעת לעת.
 - 1. מהי עוצמת הפגיעה הפוטנציאלית בפרט?
 - 2. מהי עוצמת ההשפעה הפוטנציאלית על התודעה?
 - 3. מהי עוצמת הפגיעה הפוטנציאלית בציבור?
 - 4. האם ישנה השפעה על הקצאת משאבים לציבור, ומהי?
 - 5. האם צוות הפיתוח מגוון דיו?
 - 6. מהי עוצמת הפגיעה הצפויה אם יעשו שימוש לרעה במוצר או שהמוצר יצא משליטה?
 - 7. האם יש דרך מהירה לזיהוי של תקלות אתיות במערכת שלא נצפו מראש?
- חלק ב' של הכלי הוא שימוש ותחזוקה של מפת שכחיות, המאפשרת לאתר את המוקדים האתגריים מבחינת הטמעת הערכים האתיים בשלבי הפיתוח של המערכת (לאו דווקא את עוצמת הנזק או הפגיעה, כי אם את שכחיות הפגיעה). מפת השכחיות המוצגת בדוח בוחנת עשרה מקרי-בוחן מייצגים שאירעו, אותם הוועדה בחרה, ומאתרת את הנקודות בהן עלולים להתעורר אתגרים אתיים לפי ציר הזמן ושלבי הפיתוח.
- עשרת מקרי הבוחן שנבדקו ועל פיהם נבנתה מפת השכחיות:
1. מערכת בינה מלאכותית לסינון מועמדי עבודה.
 2. שימוש בבינה מלאכותית להשפעה על תודעת האנשים.
 3. חיזוי סכנות ממחלות.
 4. מערכת להערכת מסוכנות של עצירים.
 5. משחק בו שחקנים וירטואליים מונחי בינה מלאכותית צברו נשקים שוברי שוויון.

6. בוט שהפך לגזען.
7. בוט שמתחזה לאדם.
8. הטיות בזיהוי פנים.
9. מכוניות אוטונומיות דרסה למוות הולכת רגל.
10. מערכות המלצה של תוכן מציגות מידע שונה לאוכלוסיות שונות.

מפת שכיחות אתגרים אתיים בשלבי פיתוח מערכת בינה מלאכותית

אקוסיסטם	מעקב אחרי ביצועים	הפצה	הערכת המודל	בניית מודל	ארגון נתונים	איסוף נתונים	הבנת הצורך העסקי	
1,4,9	1,3,4,9	1,4,9	1,3,4	3,4		1,3,4		הוגנות
4				3,4			4	שקיפות
9				3,4			4	הסבריות
2,5,6,9	3,4,5,6	3,4	1,3,4				1,2,3,4	אחריות
2,9			1	1	1,2	1	2,9	פרטיות
6,7,10	10						7,10	חופש בחירה
2,9					2			אבטחת מידע
4,9	4	4		4			4,9	זכויות אדם
3,5,8	3,4,5	3,4,5,8	3,4	3,4		3,4	3,4	בטיחות
	5	5,6					5	שוק תחרותי

מקרא:

1	מקרה בוחן 1 מערכת בינה מלאכותית לסינון מועמדים לעבודה
2	מקרה בוחן 2 שימוש בבינה מלאכותית להשפעה על תודעת האנשים
3	מקרה בוחן 3 חיזוי סיכון ממחלות
4	מקרה בוחן 4 מערכת להערכת מסוכנות של עצירים
5	מקרה בוחן 5 משחק בו שחקנים וירטואלים מונחי בינה מלאכותית צברו נשקים שוברי שיוויון
6	מקרה בוחן 6 בוט שהפך לגזען
7	מקרה בוחן 7 בוט שמתחזה לאדם
8	מקרה בוחן 8 מכוניות אוטונומיות דרסה למוות הולכת רגל
9	מקרה בוחן 9 הטיות בזיהוי פנים
10	מקרה בוחן 10 מערכות המלצה של תוכן מציגות מידע שונה לאוכלוסיות שונות

שכיחות המקרים הבעייתיים הייתה נמוכה (מקרה בודד)
שכיחות המקרים הבעייתיים הייתה בינונית (2 מקרים)
שכיחות המקרים הבעייתיים הייתה גבוהה (3 מקרים ומעלה)

פרק שלישי : בינה מלאכותית ואסדרה

המטרה הכללית של אסדרה היא להביא להפנמה ולשינוי התנהגות של אנשים. תחום הבינה המלאכותית הוא דינאמי, גלובלי ומושפע מפיתוחים חוצי-גבולות, מתפתח במהרה, עוסק בנתוני עתק, וכזה שהמגזר הפרטי מניע חלק ניכר מההתקדמות הטכנולוגית בו.

הוועדה זיהתה כי קיימות מספר אפשרויות של אסדרה להתמודדות עם אתגרים של בינה מלאכותית :

1. אסדרה באמצעות חקיקה באמצעות דין ייעודי.
2. אסדרה באמצעות פיתוח שיפוטי.
3. אסדרה באמצעות תקינה מקצועית.
4. אסדרה-עצמית לא משפטית.

לכל אחת משיטות האסדרה יתרונות וחסרונות, באיזון שבין ניהול סיכונים, ודאות, תופעות לוואי, גמישות וחדשנות. בחלק זה וועדה מתארת את היתרונות והחסרונות של סוגי האסדרה השונים, וההתאמה שלהם לאזורי סיכון שונים.

אפשרויות אסדרה בתחום הבינה המלאכותית

סוג האסדרה	מאפיינים	יתרונות	חסרונות	המלצות הוועדה
אסדרה באמצעות חקיקה באמצעות דין ייעודי	חוק ייעודי או תיקון לחוקים אחרים שנאכף בידי רשות מדינתית קיימת או בידי גורמים פרטים	<ul style="list-style-type: none"> • מגביר את הבהירות לגבי הערכים המוגנים • מאפשר פיתוח שיפוטי קונקרטי על בסיסי קווים מנחים של המחוקק • גמישות חלקית 	<ul style="list-style-type: none"> • העדר התמחות מקצועית בארגון אחד • אכיפה לאחר מעשה בלבד. • עלול להגביר את חוסר ודאות • לא מתערב במערכת יחסי הכוחות הקיימים שבו יתכן יתרון לגורמים מסוימים 	מתאים בעיקר לאזורי סיכון בינוני וגבוה
אסדרה באמצעות פיתוח שיפוטי	ללא חוק ספציפי	<ul style="list-style-type: none"> • ללא חיכוך רגולטורי או משפטי ישיר • גמישות • מאפשר פיתוח שיפוטי 	<ul style="list-style-type: none"> • בדרך כלל פיתוח שיפוטי יהיה במקרים מובהקים יותר של נזק – ולכן לא ברור שיאפשר מענה הולם לכל טווח הסיכונים לנזק • העדר התמחות מקצועית בארגון אחד • חוסר ודאות • יתרון לשחקנים חזקים 	מתאים למצבים של סיכון בינוני ואי בהירות לגבי כיווני הפיתוח
אסדרה באמצעות תקינה מקצועית או של קהילה מקצועית/אתית	מאפשר ייבוא בידי המשפט בעתיד	<ul style="list-style-type: none"> • גמישות • לגיטימיות גבוהה בקהילה המקצועית • תהליך שיתופי 	<ul style="list-style-type: none"> • סיכון להדרה של המשפט וערכיו • תלות במשפט לקבלת תוקף מחייב, פיקוח ואכיפה • יתרון לשחקנים חזקים 	מתאים לאזורי סיכון בינוני ונמוך, ומתאים כמסגרת לפיתוח ובחינה של החלת העקרונות האתיים
אסדרה לא-משפטית	ללא נורמה משפטית (למשל החלת עקרונות אתיים)	<ul style="list-style-type: none"> • גמישות • לגיטימיות גבוהה בקהילה המקצועית 	<ul style="list-style-type: none"> • סיכון להדרה של המשפט וערכי היסוד שלו (שוויון, הגינות, זכויות אדם) • תלות בקהילה המקצועית לפיתוח • חסר מנגנון אכיפה אמין • יתרון לשחקנים חזקים 	מתאים לאזורי סיכון נמוך, שם מספיקה אסדרה לא-משפטית, ואזורי סיכון גבוה, שם ההתפתחות הטכנולוגית מהירה יחסית לאפיק המשפטי

- הוועדה רואה חשיבות להתאים את האסדרה הנבחרת לחקיקה, למדיניות ולתקינה מקובלות במדינות המפותחות, על מנת שמדינת ישראל תוכל להישאר בחוד החנית של התחום.
- הוועדה ממליצה לייצר עמדה ממשלתית ברורה ואחידה של מדינת ישראל שתוצג במנות הבינלאומיות החשובות.

- הוועדה ממליצה למפות את השחקנים השונים המצויים בשרשרת הערך ובאקוסיסטם של המוצרים והשרותים, כדי ליצור מסגרת מותאמת של אחריות ומערכת תמריצים כדי שיפנימו את העקרונות האתיים שנסקרו בפרק השני של הדוח. במיפוי יש לכלול גם את גורמי המחקר הבסיסי, המחקר היישומי, המחקר התעשייתי, הגורמים המיישמים וכד'.
- אחריות הנהלת הארגון לבחון באופן קונקרטי את הסיכונים הנובעים לערך המוגן מהפעילות המוצעת, ולהתאים לה מסגרת ניהול סיכונים. על ארגונים לקיים "מרחב ניסוי" ו- "מעגלי בקרה" בטרם מימוש בפועל של הטכנולוגיה. מובן כי רמת ועומק הניסוי הנדרש הינו בזיקה לעקרונות האתיים שהוצעו בדוח זה.
- הוועדה סבורה כי קידום רגולציה-מסייעת, בשלבים הראשוניים של פיתוח מערכות בינה מלאכותית, יכולה לסייע רבות בהטמעת העקרונות והערכים האתיים.
- הוועדה ממליצה לרשויות האסדרה לשלב עיקרון תהליכי שיאפשר ניסוי מבוקר של המדיניות הרגולטורית המיושמת.
- הוועדה ממליצה לבחון את אפשרות התפיסה של סביבה רגולטורית מבוקרת (regulatory sandbox) בדומה להמלצות המתגבשות ב-OECD.
- הוועדה ממליצה להתחיל באיתור הרשויות המסדירות את נושא משאבי מידע המשמשים לעיבודי טכנולוגית בינה מלאכותית, וכן את אותן רשויות אסדרה האחראיות על שוק המוצר שיווצר. איתור רשויות אסדרה אלה יסייע בהסרת חסמים לשימוש בבינה מלאכותית ולהתאמה של המעטפת הרגולטורית לסיכונים הקיימים מהמערכת.
- הוועדה ממליצה כי בכל הקשור למימוש בינה מלאכותית הכרוך בעיבוד מידע אישי, או שיש לו השלכות על קבלת החלטות מבוססות מידע אישי המוסדרות באמצעות דיני הפרטיות, רשות האסדרה המובילה תהיה רשות הגנת הפרטיות, אשר תפעל בתיאום עם רשויות האסדרה הייעודיות, ככל שיהיו.
- א. בהתאם לכך הוועדה ממליצה כי רשות הגנת הפרטיות תבחן את העקרונות המנחים שהוצגו לעיל ותגבש תכנית מימוש בתחומי המידע האישי בהקשר של בינה מלאכותית.
- ב. בפרט, נראה כי יש לוודא כי לרשות הגנת הפרטיות משאבים טכנולוגיים וחומריים לפתח מסגרת טכנולוגית ומשפטית מעודכנת לתחום אנונימיזציה של מידע. היכולת לבצע אנונימיזציה של מידע אישי, ברמת סמך סבירה, היא יכולת תשתיתית לפיתוח וקידום בינה מלאכותית.
- הוועדה ממליצה כי הרשות לתחרות תגבש דרכי התמודדות שמטרתן שמירה על תחרות הוגנת בתחום; שמירה על צרכנים ועל נגישות הטכנולוגיה; ומניעת גלגול סיכוני הטכנולוגיה או עלויותיה באופן שאינו יעיל לשחקנים חלשים יותר במורד "שרשרת הערך".
- הוועדה ממליצה להקים מנגנון תיאום פנים ממשלתי על-משרדי. כל זאת, על מנת ליצור מדיניות אחידה, ברורה וקוהרנטית, בין כלל משרדי הממשלה. מכיוון שלכל החלטה בתחום הבינה המלאכותית השפעות רוחביות מידיות, יש ליצור לבחון את כלל ההיבטים הרלוונטיים של כל החלטה על הרגולטורים השונים ועל החברה בכללותה.
- בנוסף לתיאום הכולל והמנחה, הוועדה צופה שיעלה הצורך במקרים מסוימים בתיאום בפרויקטים רוחביים חוצי-משרדים. בהתאם לכך מוצע להקים מנגנון תיאום ממשלתי בפורמט אפקטיבי כדי להעביר מידע מקצועי ורגולטורי בין הגורמים השונים ולאפשר להם קבלת החלטות מבוססות תמונה רחבה יותר.
- הוועדה ממליצה כי יוקם מנגנון תיאום ומוקד ידע ממשלתי כדי לסייע לממשלה ולרשויות האסדרה בקידום הנושא באופן קוהרנטי. בדיוני הוועדה עלה חשש מפני אי תיאום או סתירות שעלולות להיגרם כפועל יוצא של ההצעה להתבסס על רשויות האסדרה המגזריות. הוועדה לא דנה במאפייני המנגנון אלא ממליצה על הצורך בקיום פונקציה כזו במסגרת ההיערכות האסדרתית.

- הוועדה ממליצה כי רשויות האחראיות על משאבי מידע כחלק מתחום הליבה שלהם, יידרשו לבצע בחינה בהתאם לעקרונות העל האתיים. עליהן לבחון האם בעת פריסה של טכנולוגיות בינה מלאכותית או שימוש בהם בתחומי הפעילות המאוסדרים על ידם, יש צורך בהתאמה של המסגרת החלה כדי לקדם את ההגנה על האינטרסים המאוסדרים.

פרק 1 – הקדמה

ועדה זו הינה ועדת משנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית. על הוועדה הוטל לדון בהיבטים האתיים המשפטיים והרגולטוריים של תחום הבינה המלאכותית. בנוסף הוטל על הוועדה להמליץ על עקרונות מנחים בהקשר הישראלי אשר ילקחו בחשבון כחלק מהתכנית הלאומית, שתוגש לרוה"מ, להפיכת ישראל למובילה בתחום.

השיח על אודות בינה מלאכותית, אתיקה, משפט ורגולציה מצוי בעלייה בשנים האחרונות. זאת, לאור ההתקדמויות הטכנולוגיות, המייצרות שאלות חדשות אודות סיכונים, אחריות, ותועלות חברתיות הנובעות מפיתוח ושימוש בבינה ממלאכותית בהקשרים שונים. להיבטים אתיים משפטים ורגולטורים יש השפעה על מרחב הפעולה של שחקנים שונים המפתחים ומשתמשים בבינה מלאכותית. להיבטים הנ"ל ישנה אף השפעה על הגנה על ערכים חברתיים מוכרים ובהם האוטונומיה של הפרט, הזכות לפרטיות, חדשנות והוגנות. במסגרת זאת יש גם ציפייה לרתימת האפשרויות הטכנולוגיות לקפיצת מדרגה ברווחה האנושית הכוללת.

הנחות עבודה של הוועדה

1. המסמך נכתב עבור שני קהלי-יעד שונים: ראשית, ככלי עזר למקבלי החלטות בתחום האתי והמשפטי מתוך ראייה משפטית, אתית וטכנולוגית. שנית, עבור מפתחי מערכות בינה מלאכותית, הנוקטים לקבל החלטות בעלות השלכות אתיות ומשפטיות.
2. הדוח הנוכחי עוסק בטכנולוגיות ובשימושים שכיחים בתחום הבינה המלאכותית כיום ובשנים הקרובות. על כן הדוח התמקד בעיקר בהיבטים אתיים ורגולטוריים של בינה מלאכותית צרה (Narrow AI)¹, כזו שממוקדת במשימות מוגדרות, וב-machine learning.
3. הנחה מרכזית שליוותה את כתיבת הדוח היא ששילובם של שיקולים אתיים בשלבי הפיתוח והתחזוקה של מוצר משולב בינה מלאכותית, לטווח הארוך, יעניקו יתרון לחברות ולמדינות שפועלות לפיהן.
4. הוועדה העדיפה לנקוט בגישה של **התערבות מידתית**, לפיה ההתערבות אמורה להיות במידה הנדרשת בלבד, תוך בחינת כלל השיקולים הרלוונטיים (זכויות פרט, טובת הציבור, שיקולים מסחריים ותחרותיים, חדשנות וכדומה). בנוסף, הוועדה יצאה מנקודת הנחה כי יש לתת מענה מראש ולא רק בדיעבד למקרים בעייתיים אפשריים.

במה הוועדה לא עסקה?

הוועדה לא עסקה בשאלות של שיקולים רגולטוריים ואתיים בתחום דיני החימה. שאלת שילובם של מכוונות בינה מלאכותית אוטונומיות בשדה הקרב הינה שאלה שחוקרים וארגונים רבים מנסים לבחון בימים אלה. ישנה מחלוקת בנושא. מצד אחד, ישנם מדינות ומומחים בולטים בתחום דיני לחימה הסבורים כי בתנאים מסוימים לא תהיה מניעה בעתיד לשלב כלים אוטונומיים, ככל שיפותחו כאלה, למלחמה בשדה הקרב, וכי דווקא שילובם בשדה הקרב יכול לתרום לצמצום הנפגעים שאינם אמורים להיפגע ולקיומם של הדינים הבינלאומיים.² אל מול עמדה זו, מומחים שונים ובין השאר, הקליניקה המשפטית לזכויות אדם בינלאומיות באוניברסיטת הרווארד בשיתוף עם משמר זכויות האדם (Rights Watch Human), טוענים כי לאור הדינים הבינלאומיים השונים, אין לשלב טכנולוגיות של מכוונות הרג אוטונומיות בשדה הקרב.³ נראה כי בנושא זה עוד מוקדם להסיק מסקנות ויהיה נכון להמשיך את הדיון בפורומים ייעודיים לנושא.

¹ הגדרות של בינה מלאכותית צרה ומנוחים נוספים ניתן לראות בנספח ב' של הדוח.
² ראו למשל Schmitt N. Michael and Thurnher Jeffrey, 2013, Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict, Harvard National Security Journal 231.
³ ראו <https://www.hrw.org/report/2018/08/21/heed-call/moral-and-legal-imperative-ban-killer-robots> וכן ראו דיונים קודמים <https://www.hrw.org/report/2016/12/09/making-case/dangers-killer-robots-and-need-preemptive-ban> <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots> וגם <http://hrp.law.harvard.edu/staff/clinic-and-hrw-killer-robots-fail-key-moral-legal-test/>

פעילות הוועדה ומבנה הדוח

מטרת הדוח היא להבין מהן המערכות הרגולטוריות והאתיות הקיימות שיכולות לתת מענה לשאלות בתחום הבינה המלאכותית, היכן ניתן להסב פתרונות קיימים, והיכן צריך לייצר מסגרת חדשה. מיפוי הייחודיות של בינה מלאכותית הוא קריטי לזיהוי האתגרים וההזדמנויות שעומדים לפתחנו בנושא.

בהתאם לכך, הפרק הראשון בדוח דן במאפיינים הייחודיים של טכנולוגיית הבינה המלאכותית אל מול טכנולוגיות אחרות שקיימות. בנוסף, הפרק דן בשאלות שטכנולוגיה שכזו מעלה בשלבים השונים של מחזור החיים שלה, אל מול הקשרים חברתיים, רגולטוריים וערכים אתיים שונים. במסגרת זאת, בשלב מוקדם של הדיונים בתת הוועדה, עלה כי יש חשיבות רבה לקיומו של שיח בינתחומי בנושא זה. במסגרת בחינה האם מדובר בתופעה חדשה המחייבת הסדר חדש, או שמא התאמה של כללים קיימים, יש חשיבות לגישור בין השדה הטכנולוגי, בין השדה של מדעי החברה, ובין השדה האתי-המשפטי בהתבוננות על השאלות.

העיסוק בהתמודדות עם המאפיינים הייחודיים של טכנולוגיית הבינה המלאכותית נחלק לשני פרקים.

הפרק השני בדוח, הוא פרק המציע סדרה של **עקרונות אתיים**, שמטרתם לתת ביטוי לערכים חברתיים ואינטרסים מוכרים בעת התמודדות עם המאפיינים המיוחדים של הטכנולוגיה. הפרק מציע כלי שימושי שיאפשר למקבלי החלטות להבין ולהחליט היכן יש לפעול בתחום האתי, וסדרת שאלות שמקבלי החלטות הטכנולוגיים צריכים להיות מסוגלים לענות עליהם.

הפרק השלישי בדוח מציע סדרה של **עקרונות אסדרתיים**, שהוועדה מציעה לכלול בפיתוח המודל האסדרתי. המודל המוצע מקדם ערכים ראויים להגנה תוך צמצום תופעות לוואי של אסדרה ואחריות משפטית, שעלולות לפגוע בחדשנות. על מנת להדגים את השיח העקרוני וככלי עזר לקוראי הדוח, הוא כולל הדגמות של מקרים שבהם שולבה טכנולוגיה ואופן ההתמודדות המוצע איתה.

מה חדש ומיוחד בטכנולוגיית בינה מלאכותית?

בדיוני הוועדה עלו העמדות והמסקנות הבאות אודות הייחודיות של מערכות הבינה המלאכותית לעומת טכנולוגיות מתקדמות אחרות:

1. הקצנה של יחסי גומלין שקיימים במערכות חברתיות

מערכות בינה מלאכותית נוטות להקצין יחסי גומלין קיימים במערכות חברתיות. למשל, אם קיים אי שוויון בין קבוצות שונות חברתיות, מערכות בינה מלאכותית עשויות לשעתק אי שוויון זה, ולעיתים אף להקצין אותו (הדבר נכון אף לאפליה, חיזוק סטריאוטיפים, פגיעה בזכויות, הקצנת דעות וכיו"ב. לשם הנוחות נדגים טענה זו על נושא אי השוויון). יש כמה סיבות מרכזיות לתופעה זו:

א. מכיוון שמערכות הבינה המלאכותית תלויות במידע המוזן להם, הנתונים המוזנים עשויים לשקף אי שוויון שקיים ממילא בין קבוצות, ובמידה ובמידע המוזן יש עיוות, המערכת תלמד ותעתק עיוות זה.

ב. מערכות הבינה המלאכותית הופכות לשכיחות יותר, ונפוצות בהקשרים חברתיים רבים יותר. על כן ההשפעה שלהן (וכפועל יוצא ההטיות שעלולות להיות בהן) מגיעה לציבור רחב יותר.

ג. יש נטייה מוטעית להתייחס לתוצר של מערכות בינה מלאכותית, המנתחות מידע בקנה מידה גדול ובמהירות, כאמת מדעית. כתוצאה מכך, הסכנה היא שלא יופעלו לגבי מערכות אלו מנגנוני בקרה המופעלים על החלטות אנושיות זהות, כאשר יש חשד לעיוות.

ד. בשל מורכבות המערכת, קשה לצפות ולתקף מראש את ההתנהגות והפעולות של מערכות בינה מלאכותית. עקב כך קשה לעיתים להפריד בין אבחון "אמיתי" המבוסס על תהליך בדיקה ובקרה תקף, כפי שהדבר נעשה באלגוריתמים קלאסיים, או בהפעלת שיקול דעת מדעי אנושי, לבין אבחון שמקורו בהטיה.

2. האתגר התהליכי "להנדס" ערכים

מערכות בינה מלאכותית פועלות גם בתחומים שבהם ההחלטות ושיקול הדעת הופעל באופן מסורתי בידי גורם אנושי. במקרים רבים אותו בעל התפקיד המקצועי, אשר מפעיל את שיקול הדעת ומקבל את ההחלטה, הינו מיומן או מוסמך גם בהפעלת שיקולים ערכיים בהתאם להקשר. בעת פיתוח מערכות טכנולוגיות משולבות בינה מלאכותית, המחליפות שיקול דעת זה, האחריות לביטוי השיקולים הערכיים הללו עוברת מבעלי התפקידים המקצועיים (כגון, רופאים, שופטים וכד') לאנשים מתחום ההנדסה (כמו מהנדסים, מדעני מידע וכד'), דבר שפחות מתרחש בעת העיסוק באלגוריתמים קלאסיים.

3. סיכונים לפרטיות

המונח "פרטיות" הינו מונח רחב הזוכה להגדרות שונות. המכנה המשותף להגדרות אלה קשור, בין היתר, בזכות האדם "להיעזב במנוחה", בכיבוד האוטונומיה שלו ויכולתו לקבוע כיצד לפעול בענייניו. "פרטיות במידע" עוסקת בזכותו של האדם לשלוט או להשפיע על השימושים במידע אודותיו. במקרים רבים הפגיעה בפרטיות, שהיא פגיעה בפני עצמה, עשויה להיות מבוא לפגיעה בזכויות אחרות, כגון אפליה. בעידן המידע, שבו גורמים שונים ורבים אוספים מידע באופן שגרתי אודות יחידים, עולים אתגרים חדשים להגנה אפקטיבית על הזכות לפרטיות. אתגרים אלה באים לביטוי ביכולת לאסוף מידע כמעט ללא הגבלה, לעבדו ולהפיק ממנו תובנות חדשות אודות אדם שאינן קשורות בהכרח להקשר המקורי שבגינן נאספו. בהמשך לכך אתגר נוסף הוא שכמויות מידע אלה מאפשרות יכולת ניתוח של מאפייני היחיד, באופן שמאפשר השפעה עליו, השפעה שעשויה להיות לא תמיד הוגנת.

מערכות בינה מלאכותית מבוססות על עיבודי נתוני-עתק (big data), ממקורות שונים, לרבות הצלבה וחיבור של מידע, הנעשים לעתים תוך ניתוק מהמטרה ומההקשר המקורים שלשמן נאסף המידע. בנוסף, מערכות בינה מלאכותית מאפשרות הסקת מסקנות וביצוע פעולות המבוססות על מידע, באופן שמייצר סיכונים בהיקף ובאיכות חדשים לזכות לפרטיות, וכך גם להגנה על האוטונומיה של הפרט.

4. מורכבות שפוגעת באמון הציבור

המורכבות והעמימות של מוצרים ושירותים שלובי בינה מלאכותית, מעוררים קושי ביצירת הבנה ואמון של קהלים רחבים הן בטכנולוגיה עצמה, בהבנה כיצד המערכת עובדת, והן בהשפעות של טכנולוגיה זו. חוסר הבהירות הזו מובילה במקרים מסוימים לחוסר אמון, אפילו בנושאים בהם מערכות הבינה המלאכותית מובילות ליתרון חברתי ועסקי. כך לדוגמא, בתחום המכוניות האוטונומיות ההערכה היא שכדי לרכוש את אמון הציבור במכוניות ללא נהג, כמות התאונות הקטלניות בהן מכוניות אלו מעורבות צריכה להיות פי 1000 קטנה יותר מאשר כמות התאונות הקטלניות בהן מעורבות מכוניות עם נהג.⁴

פגיעה באמון הציבור בשימוש מסוים של בינה מלאכותית עלולה להוביל לפגיעה נרחבת יותר בכל התחום. ניסיון העבר מראה שאכזבה הובילה לשנים ארוכות של "חורף" בהן ההתקדמות בתחום נעצרה כמעט לגמרי. קבוצת המומחים HLEG באיחוד האירופאי טענה שאם הציבור לא יוכל לתת אמון במערכות בינה מלאכותית, הוא לא יאמץ אותן, וכתוצאה מכך יהיה הפסד ערך מוסף רב שיש למערכות להציע.⁵

5. שינויים בתחרותיות בשוק והגדלת פערים בין חברות

אחד היתרונות של מערכות בינה מלאכותית היא היכולת שלהן להמשיך ולהשתפר עם השימוש בהם. לדוגמא מערכת להמלצה על תוכן מקבלת משוב מהמשתמשים על איכות התוכן עליו המליצה וכך היא יכולה לשפר המלצות עתידיות. כך יוצא, שחברה גדולה, המסוגלת לאסוף משוב ממספר רב של משתמשים, תשפר את המערכת שלה בקצב גבוה יותר מאשר חברה מתחרה בעלת נתח שוק קטן יותר. במילים אחרות, שחקנים בעלי כוח שוק, כאלה עם נתוני-עתק הנחוצים לפיתוח מערכות בינה מלאכותית, מנצלים כלכלת רשת על מנת לעצב את האופן בו שחקנים חדשים נכנסים לתחום. דבר זה עשוי להוביל למצב שבו הפער בין השחקנים הגדולים בשוק לבין מתחרים קטנים יותר גדל במהירות מה שמקטין את התחרותיות במשק. המצב אפילו מקצין כאשר בוחנים שחקנים חדשים הרוצים

Shalev-Shwartz Shai, Shanmmah Shaked and Shashua Amnon, 2018, On a Formal Model of Safe and Scalable Self-driving

Cars, ArXiv, <https://arxiv.org/pdf/1708.06374.pdf>

Ethical Guidelines for Trustworthy AI, The High-Level Expert Group on Artificial Intelligence, EU, 2019,⁵ <https://ec.europa.eu/futurium/en/ai-alliance-consultation>

להתחרות בשוק – בהעדר כמות גדולה מספיק של נתונים להתחיל איתם, חברה כזו חסומה מלהיכנס לשוק.

6. שינוי בקטגוריות מוכרות לאחריות של יצרני מוצרים, ספקי שירותים, ועסקים הנשענים עליהם

מוצרים רבים משלבים כיום יכולות איסוף ועיבוד מידע, וקישור בין המוצר הפיזי ליכולת עיבוד מרוחקת, וזאת במסגרת המגמה הידועה כאינטרנט של הדברים – internet of things. יכולות האיסוף והעיבוד של מידע באמצעות מוצרים, מאפשר פיתוח שירותים חדשים הקשורים למוצר אולם למעשה, מעלים שאלות חדשות אודות האחריות לשירותים אלה, וחלוקת האחריות בין יצרן המוצר המספק את יכולת האיסוף ואספקת השירות, לבין מי שמספק את השירות בפועל.

בהמשך לכך, מוצרים שלובי בינה מלאכותית כוללים הן רכיב של מוצר פיזי, והן רכיב של חישוביות ויכולת פעולה. בכך החלוקה הקלאסית, שבין מוצר לשירות והאחריות למוצרים אל מול האחריות לשירותים, השתנתה ומחייבת בחינה מחודשת. מורכבות נוספת עשויה לקרות כאשר נעשה שימוש במוצרים ושירותים אלה כדי לספק שירות של גורמים עסקיים אחרים. כך למשל, שימוש ברחפן אוטונומי בידי רשת מרכולים בכדי לספק משלוחים לבתי הקוחות. הרחפן מבוסס על יכולות תעופתיות, יכולות ניווטיות והתמודדות עם הסביבה. בנוסף, הרחפן מספק שירותי מיפוי וחיזוי. כל אלה נרכשים בידי רשת מרכולים, אשר השירות שהרשת מציעה הוא משלוח.

2. אתיקה ובינה מלאכותית:

לאור האתגרים וייחודיות של טכנולוגיות בינה מלאכותית קבוצות מומחים וחברות שונות פרסמו בשנים האחרונות מסמכי מדיניות והצעה לכללי אתיקה. כך למשל קבוצת המומחים באיחוד האירופאי, HLEG, קבעה כי לא ניתן להשיג אמון במערכות הבינה המלאכותית רק באמצעות רגולציה, וכי ניתן ליצור בינה מלאכותית אמינה על ידי בחירת המסגרת האתית המתאימה והטמעתה בצורה נכונה.⁶

לצד זאת, יש לזכור כי מדינות רבות נמצאות כיום במרוץ לבכורה בתחום הבינה המלאכותית.⁷ האתגר של שימור יכולת תחרותית של מדינות מערביות הפועלות במסגרת דמוקרטית, ערכית ואתית, אל למול מדינות שבהן המסגרת הערכית שונה, מחריפה בתחום הבינה המלאכותית. מהלכים אלה מקודמים גם בזירות בינלאומיות שונות.⁸ בישראל רשות החדשנות מצביעה על החשיבות בהשקעה תשתיתית בתחום זה, כמאפשרת צמיחה והובלה, וכן מודעות לחשיבות של מסגרת אתית ומשפטית ברורה שתמנע אפקט מצנן על החדשנות.⁹ לאור תהליכים עולמיים אלו:

הוועדה מאמינה כי שמירה על מערכת עקרונית אתיים על פיהם יפעלו ארגונים פרטיים והסקטור הציבורי, הוא קריטי לטווח ארוך.

כמו כן, הוועדה קוראת לבנות תכניות של ידע והכשרה לאתיקה של אנשי המקצוע העוסקים בפיתוח מערכות בינה מלאכותית. הרעיון הוא ללמד את בעלי התפקידים השונים בתחום את הדילמות האתיות על מנת שהם יוכלו להעלות את הבעיות האתיות עוד בשלבים המוקדמים של הפיתוח ולתת להם מענה. כלומר, להטמיע כבר בתכנון ובפיתוח המערכת את שיקולי האתיקה. נושא זה אינו בתחום העיסוק של הוועדה, ועל כן לא פירטנו עליו מעבר לקריאה להכשיר ולחנך.

Ibid. Ethical Guidelines for Trustworthy AI, 2019⁶

⁷ ראו למשל את הצו הנשיאותי של ארה"ב מה-2019.11.2 בנושא מובילות בתחום הבינה המלאכותית שקובע שיש לפתח וליישם תכנית פעולה על מנת לשמור על היתרון היחסי של ארה"ב אל מול מדינות אחרות בתחום. כמו כן, ראו הדוח של ממשל צרפת מ-2018 שנכתב ע"י סדריק וילאני - https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf, והדוח של ממשלת בריטניה משנת 2018 - <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>. גם מדינות לא-דמוקרטיות התייחסו לעניין, למשל סין <https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence-development-plan> ורוסיה <https://futureoflife.org/ai-policy-russia/>.

⁸ ראו את הדוח של ה-OECD המקדם המלצות למדינות החברות לקדם את הפיתוח והשימוש בטכנולוגיית בינה מלאכותית באופן שיקדם את הרווחה החברתית - <https://www.oecd.org/going-digital/ai/principles/> ודו"ח רשות החדשנות 2018-2019 <https://innovationisrael.org.il/InnovationRapport18>

על רקע הנעשה בעולם, ולנוכח האתגרים שמייצרת טכנולוגיית הבינה המלאכותית, הוועדה זיהתה את העקרונות הבאים כבסיסי האתיקה בתחום הבינה המלאכותית, כחלק מהכלי שנציע למקבלי החלטות בתחום לעיצוב הבינה מלאכותית ולשימוש בה. מובן כי נדרש שיח רציף בין אנשי הטכנולוגיה, מדעי החברה והמשפט לעדכון ולבחינת ההתאמה של המסגרת האתית. עקב כך יש להתייחס להמלצות מטה כמשקפות את ניסיונם של חברי הוועדה לסכם את הדברים כפי שהם מובנים בעת כתיבת הדוח.

2.1 עקרונות אתיים בבינה מלאכותית:

1. **הוגנות** – חתירה לשוויון מהותי, מניעת הטיות (במידע, בתהליך, בתוצר), מניעת הפליה, והימנעות מהגדלת פערים חברתיים, כלכליים והשכלתיים.
2. **אחריותיות** (Accountability)
 - א. **שקיפות** – הנגשה, בהתאם להקשר ולנסיבות, של מידע על התהליך עצמו ועל דרך קבלת ההחלטה.
 - ב. **הסברתיות** – הסבר הפעולה וההחלטה - היכולת להסביר את תהליך קבלת ההחלטה של המערכת (ברמת המשתמשים כפרטים, גם ברמת הכלל אם המערכת משפיעה על קבוצות, וגם עבור מפעילי המערכת עצמם).
 - ג. **אחריות וניהול סיכונים** – חלוקת האחריות האתית והמשפטית לגורמים הרלבנטיים בשרשרת הערך, וחלוקת ניהול הסיכונים ביניהם. קביעת האחריות ליצירת כללים לאמצעים סבירים למניעת הסיכון בהתאם להקשר ולעוצמת הסיכון הצפוי, לניהול הסיכונים ולמינויו של גורם אחראי לניהול הסיכונים.
3. **כיבוד זכויות אדם והגנה עליהן** – כיבוד כלל זכויות האדם ומודעות לצורך הגובר בהגנה על זכויות מסוימות שעלולות להיפגע יותר בעידן הבינה המלאכותית וביניהן:
 - ד. **שמירה על שלמות הגוף** - הגנה מפני פגיעה בחייו או בגופו של אדם.
 - ה. **פרטיות** – הגנה על פרטיות, לרבות מניעת פגיעה בפרטיות באמצעות איסוף המידע, ניתוח ועיבוד המידע, שיתוף המידע ובשימושים אחרים וחדשים בו.
 - ו. **השמירה על האוטונומיה** - שמירה על יכולתו של אדם לקבל החלטות תבוניות, ובין השאר, מניעת השפעה שאינה הוגנת או אינה מודעת על התנהגות היחיד.
 - ז. **זכויות אזרחיות ופוליטיות** – לרבות הזכות לבחור, הזכות לחופש הביטוי והזכות לחופש הדת והמצפון.
4. **הגנת סייבר ואבטחת מידע** – שמירה על התפקוד התקין של המערכות, הגנה על המידע המשמש אותן ומניעת שימוש לרעה בידי גורם זדוני.
5. **בטיחות** - מניעת סכנה (לפרט, לחברה) וצמצום גרימת נזק.
 - ח. בטיחות "פנימית" התמקדות בבטיחות תהליך הפיתוח של כלי הבינה המלאכותית.
 - ט. בטיחות חיצונית – התמקדות בסכנה והנזק עבור הסביבה והלקוחות, כתוצאה מהשימוש.
6. **קיום שוק תחרותי** – שמירה על שוק תחרותי ועל כללי משחק מאפשרים.

דוגמאות לסיכונים ערכיים אתיים במערכות:

1. הוגנות

- המערכת מקבלת החלטות על הקצאת משאבים כגון חלוקת כספים ומתן טיפול רפואי.
- המערכת מעריכה אנשים לצורך קבלה למקום עבודה או מוסד חינוכי, קידום וכו'.
- המערכת מעריכה אנשים למטרות ענישה או הקלה בענישה.
- המערכת מקבלת החלטות שפוגעות ברכוש ובאינטרסים כלכליים של משתמשים.

2. אחריותיות

- המערכת לא מנגישה מידע למשתמש על דרכי פעולתה מול המשתמש.
- חלוקת האחריות במקרה של תקלה לא מוסדרת.

3. כיבוד זכויות אדם והגנה עליהן

- המערכת אוספת מידע אישי עם או בלי הסכמת המשתמשים.
- נעשה שימוש במידע אישי בפיתוח המערכת.
- המערכת מייצרת מידע אישי (לדוגמה מערכת זיהוי פנים שמאפשרת מעקב אחר אנשים).
- המערכת יכולה להשפיע על תפיסת העולם של המשתמשים.
- המערכת מסננת מידע או מייצרת מידע שיותאם אישית אל המשתמש על בסיס נתוניו האישיים (targeting) או ש"ייתפר" באופן ייחודי עבור המשתמש מבלי שהוא יוכל לדעת מהו המידע שניתן לאחרים.

4. הגנת סייבר ואבטחת מידע

- אם גורמים זדוניים יצליחו לפרוץ למערכת הם יוכלו לגרום לנזק בגוף, נזק לנפש, נזק כלכלי או לפגוע בביטחון המדינה.

5. בטיחות

- המערכת יכולה לגרום למוות או פגיעה גופנית במשתמשים או באנשים אחרים.
- תקלה במערכת יכולה לגרום למוות או פגיעה גופנית במשתמשים או באנשים אחרים.

6. קיום שוק תחרותי

- המערכת מייצרת יתרון למתחרים בעלי נתוני ענק.
- המערכת מבוססת על מאגר מידע גדול שנגיש רק למספר קטן של שחקנים בשוק.
- המערכת מייצרת תוך כדי פעולתה מאגר מידע גדול וייחודי, שלא יהיה נגיש למתחרים.
- הסכם אי-תחרות ותיאום אוטומטי בין חברות מתחרות המבוסס על מערכות בינה מלאכותית

להלן נפרט בקצרה עקרונות אלה, כפי שעלו במסמכי מדיניות שונים הן בסקטור הציבורי והן בסקטור הפרטי. חלק מעקרונות אלה מפורטים במסגרת מסמכי מדיניות של יצרניות הטכנולוגיה עצמן, כחלק ממגבלות עצמיות שהטילו. האמור מלמד כי בשיח שבין ערכים לחדשנות חברות אלה מודעות לצורך בקיום איזון כאמור, ואף מציעות נקודות איזון מיוזמתן.

2.1.1. הוגנות

טכנולוגיה איננה ניטרלית, וזאת מפני שהיא מבוססת על עבודת תכנות אנושית ועל אינטרסים מסחריים שונים. כמו כן מערכות הבינה המלאכותית מבוססות על מידע הקשור בהתנהגות אנושית, העשוי לשקף ולהקצין הטיות חברתיות מסוגים שונים. מסמכי מדיניות שונים ניסו להגדיר מהי הוגנות בהקשר זה, וכיצד ניתן לשמור עליה.

כך, למשל, חברת מיקרוסופט מגדירה הוגנות כיחס הוגן ומאוזן לכל אדם, תוך מתן יחס דומה לקבוצות שונות.¹⁰ חוסר ההוגנות עלול להתרחש בשלבים שונים של הפיתוח. יש למשל לתת ייצוג לכל האוכלוסיות במסדי הנתונים ולהימנע מהטיות של מגדר או גזע. חוסר הגינות עלול להיות גם העובדה שהציבור לא מודע למגבלות הבינה המלאכותית ומניח כי החלטותיה תמיד טובות יותר ומבוססות יותר. היבט נוסף שניתן לכלול תחת עקרון ההגינות הינו שמערכות הבינה המלאכותית יעצימו את כלל הקבוצות באוכלוסייה ולא רק קבוצה מסוימת. מעבר לייצוגיות במסדי הנתונים, מיקרוסופט ממליצה אף לגוון את צוות הפיתוח והמחקר כדי לשתף מגוון רב של אוכלוסיות בתהליך עצמו.¹¹ בדוח של ממשל צרפת מצוין כי אין לתת לטכנולוגיות בינה מלאכותית ליצור דרך חדשה למידורן של אוכלוסיות שונות. זו היא לדעתם, בבחינת דרישה דמוקרטית שכן מערכות הבינה המלאכותית הן המפתח לעתיד ולכן יש לאפשר לכלל הציבור ליהנות ממערכות אלה. על מנת לנסות להימנע מהטיות מובנות, דוח זה ממליץ להשקיע בין השאר מאמצים בגיוון האוכלוסיות של מפתחי טכנולוגיה, ובתוך כך לקדם נשים רבות יותר לעסוק בתחום. בנוסף הדוח קורא לממשל לתמוך בפיתוח בינה מלאכותית לתועלת החברה.¹²

חברת IBM מזהירה כי ההטיות השונות הקיימות האנשים המפתחים את התוכנה עלולות להיכנס לעיצוב המערכות. החברה טוענת כי זו אחריות הצוות המפתח לדאוג למזעור ההטיות על ידי מחקר מתמשך ואיסוף מידע המייצג את כל סוגי האוכלוסיות. על הצוות לפתח את המערכת ללא הטיות מכוונות ולבחון את המערכת על מנת לאתר הטיות שאינן מכוונות. בעת שמגלים הטיה, על הצוות המפתח לחקור מדוע היא נוצרה ולתקן אותה. כמו כן מציעה החברה ליצור ערוץ תקשורת למשוב המשתמשים, בו יוכלו לדווח על הטיות שהתגלו במערכת.¹³

לסיכום, על מנת לצמצם תוצר של טכנולוגיה מפלה או מוטה, יש ללמוד באופן פעיל על אוכלוסיית היעד, ולזהות מבעוד מועד אוכלוסיות שעלולות להיפגע, או להיות מיוצגות באופן לא מספק על-ידי מפתחי המערכת. לאור זאת, הועדה רואה חשיבות רבה בהיוועצות עם מומחי תוכן (domain experts) המתמחים בספרות החברתיות המסוימות בהן מבקשים לשלב מערכת AI. בדומה, היוועצות עם אנשים מאוכלוסיית היעד עצמה (באמצעות מחקר קצר-היקף) יכולה לסייע ליצור מערכת הוגנת יותר. למשל, אם מדובר בשילוב מערכת AI בביטוח הלאומי, מומלץ לדבר עם פקידי הביטוח הלאומי, אך גם עם מומחים מהאקדמיה ומחוצה לה (עובדים סוציאליים, סוציולוגים, אנתרופולוגים), שמבינים בסעד, בעוני, באי שוויון ועוד. בדומה, כדאי גם להיוועץ במקבלי קצבאות שיושפעו בפועל מהטמעת המערכת. הרעיון הוא לא לשמוע מה דעתם אודות הטמעת המערכת, אלא להבין טוב יותר את הצרכים שלהם, ואת האתגרים שהטמעת מערכת כזו עלולה לעורר.

¹⁰ The Future Computed - Artificial Intelligence and its role in society, Microsoft, 2018, pp. 58-59
https://blogs.microsoft.com/uploads/2018/02/The-Future-Computed_2.8.18.pdf

¹¹ שם, עמ' 68

¹² https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

¹³ Cutler Adam, Pribic Milena and Humphrey Lawrence, 2018, Everyday Ethics for Artificial Intelligence, IBM.
<https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>

2.1.2. אחריותיות:

2.1.2.א. שקיפות

הוועדה רואה בשקיפות ערך אתי מרכזי בפיתוח טכנולוגיה בכלל, ובפיתוח מוצרים שלובי בינה מלאכותית בפרט. שקיפות היא ערך בפני עצמו ביחוד במשטר דמוקרטי ומרכיב של אחריותיות מצד העוסקים בתחום. בנוסף, השקיפות היא גם כלי להשגה ולשמירה על ערכים אתיים אחרים כגון הוגנות והעדר הפלייה בתהליך. יתר על כן, שקיפות היא מרכיב מהותי ברכישת האמון של הציבור. בדוח וועדת המומחים באיחוד האירופאי לא היתה תמימות דעים לגבי משמעות הערך. לעיתים התייחסו למונח שקיפות כחלק משלבי הפיתוח והוצע לחשוב על דרגות שקיפות כתלות בקהל היעד, ולעיתים התייחסו לשקיפות כהסבר וכפירוש בלבד. אנו מפרידים בין המונח שקיפות לבין המונח הסברתיות. בעוד ששקיפות מדברת על גישה למידע כיצד התהליך פועל ומהם מרכיביו השונים, ההסברתיות מסבירה את הפעולה עצמה ואת התהליך בו התקבלה.

בדוח האיחוד האירופאי המליצו על שלוש רמות של דיווחים: 1. רמת מיקרו של אנשים, חוקרים ומקצוענים; 2. רמת מאזן – של מוסדות ציבוריים, חברות ואוניברסיטאות; 3. רמת מאקרו – של פוליטיקאים ודרך הדיווח שלהם לאזרחים. הקבוצה גם המליצה לשקול שקיפות לקבוצות מיוחדות, כמו אנשים עם מוגבלויות, שיעשו שימוש רב בטכנולוגיות בינה מלאכותית¹⁴.

עפ"י דוח האיחוד האירופאי ההבנה כיצד המערכת פועלת (שקיפות והסברה) היא חשובה על מנת ליצור אמון עם מערכות הבינה המלאכותית. לאור זאת, לדעתם יש להשקיע בשקיפות, בחינוך ובהסברה. מתבקש כי מוצר משולב בינה מלאכותית ידגיש את היתרונות הרבים הטמונים בשימוש בבינה מלאכותית, וכי רמת ההסבר צריכה להיות דומה להסבר שאנשים מקבלים מנותני השירות הרלוונטיים בתחום. צריך להסביר מדוע המערכות בטוחות לשימוש – ולאחר שהציבור יבין כיצד טכנולוגיה זו פועלת, גם מבלי להבין לעומק את דרך הפעילות, ניתן יהיה לרכוש את אמון הציבור במוצר ובמפתחים. האמון צריך ללוות את מערכת הבינה המלאכותית לאורך כל מחזור החיים של המערכת ולא רק בשלב הייצור, מכיוון שהמערכת צפויה להשתנות עם הזמן מעצם טבעה.

2.1.2.ב. הסברתיות

הסברתיות עוסקת בהסבר הפעולה וההחלטה, דהיינו העלאת מודעות הציבור לדרך הפעולה של המערכת באמצעות מתן הסברים מדוע המערכת הכריעה כפי שהכריעה - גם ברמת הפרט וגם ברמת הכלל.

דוח חברת IBM גורס כי ככל שטכנולוגיית הבינה מלאכותית תשולב במערכות השונות של החיים יהיה צורך גובר ועולה להסביר את הליך קבלת ההחלטות של המערכת באופן שיהיה ברור למשתמשי הקצה. כך גם, המשתמש צריך לדעת כי יצר קשר עם מכונה ולא עם אדם, וכי יהיה ברור בכל שלב שהוא יכול לשאול את המערכת מדוע היא פועלת כפי שהיא פועלת. הממשק לברור השאלות צריך להיות נגיש למשתמש בכל שלב. הליך העיבוד של המידע על ידי המערכת צריך להיות ניתן לבחינה ולבקרה בעיקר אם מדובר בעיבוד מידע רגיש. בכל מקום בו המערכת מסייעת לאדם להחליט החלטות רגישות, על המערכת להיות מסוגלת להסביר את המלצתה, לבסס אותה על המידע הרלוונטי והרעיון שעומד מאחורי המלצה¹⁵. גם דוח חברת מיקרוסופט מצטרף להמלצות דומות¹⁶. בדוח האיחוד האירופאי דנו בנוסף בעקרון המידע המאוזן – לפיו יש לתת מידע מספיק ולא מפורט מדי. החשש הוא שמידע מפורט יהיה בלשון מקצועית, שאינה בהירה או ברורה לציבור הרחב, וכתוצאה מכך ישיג את התוצאה ההפוכה¹⁷. ולבסוף, גם חברת גוגל קבעה כי על מערכות הבינה המלאכותית להיות מעוצבות באופן בו תוכלנה לתת הסברים רלוונטיים ואף להיות פתוחות לערעור על ההחלטה. על פי הצהרה זו המערכות תהינה נתונות להדרכה ולפיקוח אנושיים¹⁸.

Ethical Guidelines for Trustworthy AI, The High-Level Expert Group on Artificial Intelligence, 2019, ¹⁴

<https://ec.europa.eu/futurium/en/ai-alliance-consultation>

Cutler Adam, Pribic Milena and Humphrey Lawrence, 2018, Everyday Ethics for Artificial Intelligence, IBM. ¹⁵

<https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>

The Future Computed - Artificial Intelligence and its role in society, Microsoft, 2018, pp. 58-59 ¹⁶

https://blogs.microsoft.com/uploads/2018/02/The-Future-Computed_2.8.18.pdf

Ibid. Ethical Guidelines for Trustworthy AI ¹⁷

Pichai Sundar, 2018, AI At Google: Our Principles, <https://www.blog.google/technology/ai/ai-principles/> ¹⁸

2.1.2.ג. אחריות

האחריות היא הצלע השלישית של אחריותיות, והיא דנה בהסדרת וחלוקת האחריות האתית והמשפטית לאור אמות מידה של סיכון או נזק. הסדרה שכזו תכיל בין השאר את הנסיבות שבהן תוטל אחריות ועל מי, בהתייחס לשלבי התכנון, הפיתוח, ההשקעה, השיווק, ההטמעה והשימוש בטכנולוגיה – וזאת על מנת שיהיה ברור מה הם גבולות האחריות של הגורמים השונים. נושא האחריות הוא מורכב וכולל בעלי עניין שונים: מערכות אוטונומיות (רובוטים למשל) המפתחים, המשתמשים, היצרנים ועוד.

מורכבות חלוקת האחריות מחריפה בייחוד כאשר הבינה המלאכותית מהווה כיועץ-שותף להכרעה. כך, למשל, בינה מלאכותית המשולבת בבתי חולים ומטרתה להחליף או לסייע לרופאים. כיצד אמורה האחריות המשפטית להתחלק בין אנשי המקצוע (domain-experts) לבין הבינה המלאכותית במצבים שונים¹⁹? על פי טיוטת העקרונות האתיים של האיחוד האירופאי, יש לעצב את המערכת בדרך בה ניתן יהיה לעקוב בצורה מבוקרת אחרי החלטות המתקבלות לאור הקלט, הנתונים, המודלים לפני הליך האימון וכיו"ב. יתר על כן, יש להגדיר שיטות הסברתיות של מערכת הבינה מלאכותית²⁰.

חברת IBM החליטה לנקוט בעמדה נוקבת, הגורסת שמפתחי בינה מלאכותית אחראים על עיצוב הטכנולוגיה, ועל תהליך קבלת ההחלטות ועל התוצאות. על כן, לדעת החברה, יש לקבוע מדיניות ברורה בנוגע לחלוקת האחריות. כך גם יש להבהיר היכן מסתיימת האחריות. יש גם לחייב תיעוד של הליך הפיתוח וקבלת ההחלטות שכן התיעוד מעודד את השמירה על הכללים המשפטיים והאתיים²¹. לדעת חברת מיקרוסופט, על מי שמפתח מערכת בינה מלאכותית להיות אחראי לדרך בה היא פועלת. מינויו של גוף בוחן פנימי יכול לבדוק את נושא האחריות מעת לעת ולהעניק הכוונה מתאימה לחברה²².

באיחוד האירופי הציעו לחשוב על קביעת סעדים ופיצויים ועל קביעת גורם שניתן יהיה להתלונן בפניו על בעיות. הנחת היסוד שלהם היא שהגדלת האמון תלויה בידיעה שאם מישהו לא יעמוד בכללים – הוא ייתן על כך את הדין²³.

2.1.3 כיבוד זכויות אדם והגנה עליהן

הוועדה רואה בהגנה על כלל זכויות האדם כדבר חשוב ביותר, ביחוד הגנה על זכויות שעלולות להיפגע יותר בעידן הבינה המלאכותית, ובעיקר פרטיות, שמירה על האוטונומיה, זכויות פוליטיות ואזרחיות. לטכנולוגיית הבינה מלאכותית עלולות להיות השלכות רבות על המרחב האישי, האוטונומיה של הפרט וזכויות האדם הקשורות בכך. חלקן אין ביכולתנו עדין לצפות. בצו הנשיאותי בארצות הברית, שניתן בפברואר 2019, הוצהר כי על ארצות הברית לשמר את אמון הציבור בבינה מלאכותית ואף לשמר את הביטחון של האנשים בטכנולוגיה זו. בתוך כך על טכנולוגיה זו להגן על חירויות אזרחיות, על פרטיות ועל ערכים נוספים של התרבות האמריקאית²⁴.

הדגש בחלק זה הינו החשיבות של המודעות לפוטנציאל ההשפעות של מוצרי בינה מלאכותית, וקיומו של שיח רציף, מונחה ערכים וסיכונים, בין מובילי הידע השונים הקשורים בקידום הנושא לבין מפתחי המוצרים. הוועדה סבורה שקיומו של שיח רציף כשלעצמו מקדם את ההיתכנות של הטמעת ערכים באופן מאוזן וכזה שעומד בדרישות האחריות הכלליות של החברה.

האיחוד האירופאי קבע כי על מנת שניתן יהיה לבטוח בבינה-מלאכותית, על מוצרים מבוססי בינה מלאכותית לכבד את זכויות היסוד, הרגולציות, הערכים החברתיים ועקרונות הליבה של החברה. כל זאת יש לעשות תוך שמירה מתמדת על כוונה אתית לעשות טוב, הימנעות מכוונה לשימוש לרעה ושמירה על אוטונומיה של האנשים, על הצדק ועל היכולת להסביר את הפעולות. בנוסף, על הטכנולוגיה להיות בשליטה, שכן גם כוונות טובות עלולות לגרום לנזק בלתי צפוי. האיחוד האירופאי טוען כי התייחסות לאדם, לזכויות

¹⁹ Karni Chagal, 2018, The Reasonable Algorithm, Journal of Law, Technology and Policy, Forthcoming.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3095436

²⁰ Ibid. Ethical Guidelines for Trustworthy AI, EU, 2018.

²¹ Ibid. Cutler, Pribic and Humphrey, 2018.

²² Ibid. The Future Computed, 2018, p.73.

²³ Ibid. Ethical Guidelines for Trustworthy AI, EU, 2018.

²⁴ Executive Order on Maintaining American Leadership in Artificial Intelligence, 2019, <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>

שלו חייבות להיות משולבות בפיתוח ובעיצוב מערכת הבינה המלאכותית מראשיתה, לרבות בעת בחירת צוות הפיתוח. כך גם יש לשים דגש על המצבים בהם מעורבות אוכלוסיות פגיעות כמו ילדים, אנשים עם מוגבלויות, מיעוטים או מצבים בהם יש פערי-כוחות בין כותבי המערכת למשתמשים שלה. יש תמיד לזכור, כי לצד היתרונות הגדולים הטמונים בטכנולוגיה עלולות להיות השלכות שליליות²⁵.

חברת IBM, למשל, הדגישה במדריך שפרסמה את הצורך לפתח את מערכת הבינה המלאכותית בהתאם לנורמות ולערכים של המשתמשים הפוטנציאליים. לשם כך, המליצה, שעל הצוות המפתח את המערכת להבין את התרבות ואת ערכיה, ועל כן המליצה להשתמש בצוותים אינטרדיסציפלינריים מתחומים שונים (לדוג' מהאקדמיה, מהעולם העסקי והטכנולוגי) וחוקרי עיצוב תוכנה על מנת לוודא כי אכן ערכיהם של המשתמשים משתקפים בתוכנה²⁶. במקביל, נתן מטיאס מה-MIT Media Lab הציע כי יש לבחון ביסודיות שימוש בטכנולוגיה העלולה להשפיע (לגרום לנזק או לגרום לתועלת) על חיים או חירות כלשהי, וכי תוצאות המבחנים יהיו פומביים לעיון הציבור²⁷.

חברת גוגל הדגישה את השימוש הדו-מהותי שניתן לעשות בטכנולוגיות בינה מלאכותית. על מנת למזער שימושים שאינם לטובת האנושות, הבטיחה החברה לקחת בחשבון את השיקולים הבאים בעת פיתוח טכנולוגיות בינה מלאכותית: (א) השימוש המרכזי והעיקרי של הטכנולוגיה החדשה; (ב) האם לטכנולוגיה השלכות נרחבות; (ג) האם הטכנולוגיה ייחודית או שכיחה. בנוסף, חברת גוגל התחייבה שלא לקחת חלק בפיתוח טכנולוגיות אשכולות נזק רחב היקף; מערכות נשק שמטרתן לפגוע או לייצר פגיעה באנשים; טכנולוגיות שאוספות או עושות שימוש במידע לצורכי מעקב, הפוגעות בנורמות בינלאומיות מקובלות; ולא לקחת חלק בפיתוח טכנולוגיות שנוגדות עקרונות מקובלים של הדין הבינלאומי וזכויות אדם. גוגל התחייבה לפתח טכנולוגיית בינה מלאכותית רק במקרים בהם התועלת המצרפית של הטכנולוגיה עולה על הסיכונים האפשריים והחסרונות. כך, התחייבה החברה, במקרים המתאימים, לבחון את טכנולוגיית הבינה מלאכותית בסביבה מבוקרת ומוגבלת ואף לעקוב אחר כל צעד בהטמעת המערכת בסביבה הפתוחה אחר מכן²⁸.

בתוך כך, יש גם לחשוב על יצירת מעבדות או סביבות מוגנות אחרות בהן ניתן לבחון את ההשלכות של הבינה מלאכותית, תוך מזעור סיכונים אפשריים, למשל במעבדות בהם ניתן לבחון בינה מלאכותית חדשנית כמו TeraLab. גם בצרפת המליצו להקים מגרש חול לחדשנות (Innovation Sandboxes), מתחם מוגן, בו ניתן יהיה להטמיע את מערכות הבינה מלאכותית הניסיוניות ולבחון את דרך פעולתן ואת השפעתן על המערכת האקולוגית בתנאים מבוקרים, על מנת לאפשר חדשנות בתחום²⁹. בדומה המליצו בדוח שמואל נאמן הראשון להקים אתרי מבחן Testbeds³⁰.

2.1.3. א. פרטיות

נושא הפרטיות נדון רבות בהקשר של מערכות בינה מלאכותית, וזאת משום היותן מבוססות, במידה רבה על מידע אודות אנשים או הסקת מסקנות אודות אנשים באמצעות מידע אישי. דיני הגנת הפרטיות מסדירים את אופן איסוף עיבוד ושימוש במידע אישי. מידע אישי מוגדר כמידע אודות אדם מזוהה או מידע שניתן לזהות ממנו את האדם. דיני הפרטיות מקנים זכות תביעה אישית למי שפרטיותו נפגעה, וכן הם כוללים עיסוק רגולטורי "מסדיר" מקיף בהתאם ל-"עקרונות עיבוד מידע" מקובלים בעולם כלפי מי שאוסף ומעבד מידע אישי. מטרתם של עקרונות אלה לאפשר שימוש במידע אישי למטרות ציבוריות ועסקיות לגיטימיות תוך צמצום הסיכון לפרטיות אדם כתוצאה מכך. עקרונות עיבוד מידע מקובלים מבוססים על מסמכים מקובלים של ה-OECD, אולם הם גם מושפעים מהסביבה החוקתית והחברתית החלה במדינה בה הם מיושמים.

בישראל הזכות לפרטיות היא זכות יסוד המעוגנת בחוק יסוד: כבוד האדם וחירותו, והיא אף נתפסת, כחלק מהזכות לאוטונומיה של אדם, בהיותה מגנה על המרחב האישי שלו. במסגרת זו לדיני הפרטיות מטרה לגון מפני שימושים אסורים במידע אישי.

²⁵ Ibid. Ethical Guidelines for Trustworthy AI, EU, 2018.
²⁶ Rossi Francesca, Anna Sekaran, Jim Spohrer, and Mike Monteiro. 2018. Everyday Ethics for Artificial Intelligence, IBM, <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>

²⁷ Matias Nathan, 2016, <https://medium.com/mit-media-lab/the-obligation-to-experiment-83092256c3e9>

²⁸ Ibid. Pichai, 2018

²⁹ Villani Cedric, 2018, For a Meaningful Artificial Intelligence, https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

³⁰ שם. גץ ואחרים, 2018, עמ' 91.

בנוסף, תפיסת המימוש של דיני הפרטיות בתחום עקרונות עיבוד המידע מצויה בהתפתחות ניכרת בשנים האחרונות, שבמרכזה המגמה האירופאית בחקיקה הידועה כ- GDPR (General Data Protection Regulation). עקרונות עיבוד מידע מקובלים כוללים את הצורך בבסיס חוקי לאיסוף ועיבוד מידע (בדרך כלל הסכמה או בסיס חוקי אחר), הגבלת מטרת השימוש למטרה לשמה התקבל המידע, זכות עיון ותיקון במידע הנאסף, שקיפות כלפי נושא המידע, וחובה להגן על המידע. במשטרים מתקדמים יותר, כגון ה-GDPR, עקרונות מפורטים יותר הכוללים את החובה לעצב את המערכות הטכנולוגיות על מנת להגן על הזכות לפרטיות, הזכות להעביר מידע אישי בין ספקים, הזכות למחוק מידע.

משטרי הגנת הפרטיות נדרשים להתמודד כיום עם האתגר המשמעותי הנובע מהפער שבין חשיבותה העקרונית של ההסכמה לאיסוף מידע ושימוש בו, לבין המציאות הטכנולוגית כלכלית שברוב המקרים ההסכמה מבוססת על "הסכמים אחידים" שיש חשש שאינם מממשים את תכלית ההסכמה. מורכבות זו משפיעה גם בתחום הבינה המלאכותית בשל היותו מבוסס עיבודי מידע, היתוכי מידע, והפקת מידע ותובנות על בסיס מידע אישי.

החברות הגדולות מתייחסות לפרטיות באופן נרחב בהקשר של בינה מלאכותית. כך, למשל, חברת גוגל הצהירה כי היא מתחייבת לפתח את מערכות הבינה מלאכותית שלה תוך שמירה על עקרונות הגנת הפרטיות. בתוך כך, הדגישה גוגל כי יושם דגש על הודעה והסכמה לפגיעה, יעודדו תשתיות עם שמירה מובנית של פרטיות וכן תהיה שקיפות ויכולת לשלוט על השימוש שנעשה במידע³¹. חברת אינטל מסבירה שקיים פרדוקס מסוים – על מנת שניתן יהיה לגלות הטיות או סיכונים סייבר ולשמור על פרטיות המידע, יש צורך שמערכת הבינה מלאכותית תעבד מידע אישי. כלומר, יש לתת את הדעת כיצד ניתן מצד אחד למזער את הפגיעה בפרטיות ומצד שני לא לפגוע בתוך כך יתר על המידה בהגנה על המידע³². בנייר מדיניות נוסף הביעה אינטל את הצורך בפיתוח מחקר בתחום בינה מלאכותית ובין היתר לחשוב מחדש על פרטיות בעידן הנוכחי ולבחון האם עקרונות העבר וההווה כמו פרטיות בעיצוב (Privacy by Design) רלוונטיים בעידן הבינה המלאכותית³³.

אף במדריך של חברת IBM מציינים כי על הצוות המפתח מוטלת האחריות להשאיר למשתמש את היכולת לשלוט על ההתקשרויות שלו ועל המידע שלו. במדריך מציינים כי כבר כיום, לאור סקרים שבוצעו במדינות שונות, אנשים מדגישים את החשיבות של השליטה במידע שלהם ואת החשיבות להגנת הפרטיות במידע שלהם. צורך זה רק יתעצם עם שילוב מערכות הבינה מלאכותית. בהתאם החברה התחייבה להבהיר למשתמשים את מדיניות השימוש במידע ומדיניות הבעלות במידע וכן אפשרות אם ירצו שלא יישמר המידע עליהם עוד במערכת. כך יוכלו האנשים להחליט מהו המידע שהם מבקשים שלא ייכלל במידע המועבר לניתוח ולעיבוד הבינה מלאכותית. המערכת צריכה להציע לאנשים את האפשרות לסרב לענות או לקבל שירות בתחילת הדברים או אף במהלכם. בנוסף, תנאי הגנת הפרטיות צריכים גם כן להיות נגישים וברורים כל העת, ובין השאר, צריך לתת מענה מקיף על השימוש שנעשה במידע האישי ועל שיתוף המידע עם אחרים. יש לעשות שימוש במידע לבניית מערכות בינה מלאכותית רק אם יש הסכמה לעשות שימוש במידע³⁴.

לבסוף, חברת מיקרוסופט מציינת כי יש לעמוד בכל חוקי הגנת הפרטיות הקיימים ולהיות שקופים ביחס לאיסוף המידע, לשימוש בו ולאחסונו. כמו כן, מציינת החברה, כי יש להשאיר את השליטה במידע בידי המשתמש. עם זאת מיקרוסופט מאמינה כי יעלו שאלות חדשות שלא נדונו קודם לכן ועל כן על התעשייה והממשל יחדיו להסדיר את הנושא כך שלא תהיה פגיעה בפרטיות, לצד מקסום השימוש בבינה מלאכותית לשיפור חיי האדם תוך התייחסות לאתגרים החדשים³⁵.

יש לשים לב כי תיתכן סתירה בין שמירה על פרטיות והגנות. כך לדוגמה אם חברים רבים בקבוצה מסוימת באוכלוסייה ימנעו מלשתף מערכות בינה מלאכותית במידע האישי שלהם, יתכן שאיכות השירות, שמערכות כאלה תספקנה לקבוצה זו, תפגע.

Ibid. Pichai, 2018 ³¹

Hofman, D., & Masucci, R, 2018, Intel's AI Privacy Policy White Paper - Protecting individuals' privacy and data in the artificial intelligence world. <https://blogs.intel.com/policy/files/2018/10/Intels-AI-Privacy-Policy-White-Paper-2018.pdf> ³²

Intel, 2017, Artificial Intelligence, <https://www.intel.ai/ai/wp-content/uploads/sites/69/Intel-AI-Public-Policy-WP-2017.pdf> ³³

Ibid. Rossi et al., 2018, IBM. ³⁴

Microsoft, 2018, The Future Computed - Artificial Intelligence and Its Role in Society, https://blogs.microsoft.com/uploads/2018/02/The-Future-Computed_2.8.18.pdf ³⁵

ככל הנראה, ישנם גם שינויים מהותיים בין הטכנולוגיה הקיימת לבין הבינה המלאכותית, לאורם יהיה מקום לבחון מעת לעת את תחום הפרטיות הן במישור האתי והן במישור הרגולטורי. הדבר נכון בכל הנוגע למידע ולשילובו עם בינה מלאכותית, שלא מכוסה בדין הקיים. חוק הגנת הפרטיות הישראלי אינו מותאם לעידן הנוכחי ויש להתאימו בהתאם למדיניות הרגולטורית בעולם המערבי³⁶.

ב.2.1.3. השמירה על האוטונומיה

האוטונומיה מבוססת לא רק על היכולת הפיזית של אדם לבחור בין אפשרויות, אלא גם על זמינות המידע המאפשר בחירה מושכלת ובחינת המהימנות של מידע זה. אין לקחת נושאים אלו כמובנים מאליהם בעידן הבינה המלאכותית. כך גם, היכולת לנתח מידע על אדם ולהבין כיצד הוא פועל ומהם מאפייני האישיות שלו, דבר המתאפשר באמצעות הבינה המלאכותית, מאפשר ליצור תהליכי שכנוע ברמת עומק וחודרנות גבוהים.

האוטונומיה מכוונת גם למגוון ההחלטות האנושיות הכרוכות במגע עם הטכנולוגיה, ואשר הטכנולוגיה עלולה להחליש. בהתאם לכך הכוונה היא להחלטות של יחידים כצרכנים, אזרחים, מנהלים, בעלי תפקידים ציבוריים ונבחרי ציבור. בכל אחד מאלה, יש לבחון האם יישום טכנולוגי משפיע על האוטונומיה, ובאיזה אופן. חשוב לציין כי במסגרת דיון זה ייתכנו מקרים שבהם האוטונומיה מלכתחילה צרה יותר (בשל מאפיינים חברתיים, כלכליים, ערכיים), או שאוטונומיה צרה יותר תיתפס כראויה ערכית. כפועל יוצא מכך לא בהכרח יידרשו פעולות מיוחדות לשמירה על חופש בחירה.

ישנן טכנולוגיות בינה מלאכותית (כגון deep fake) שמטרתן לייצר מידע לא מהימן באופן שיהיה קשה עד בלתי אפשרי להבחין בינו לבין מידע מהימן. לטכנולוגיות אלו יש פוטנציאל לפגוע ביכולת של הפרט להבין את המציאות ולקבל החלטות בצורה מושכלת ואוטונומית, לפגוע באמון בין אנשים ובאמון בין אזרחים למוסדות השלטון. לדוגמא, לא ירחק היום בו ניתן יהיה לייצר באופן מלאכותי סרטון בו נראה מנהיג מדינה מכריז על מלחמה ובכך לגרור תוצאות הרות אסון. הוועדה בדעה כי על המדינה לבחון את ההתמודדות עם טכנולוגיות אלה בנפרד מדוח זה.

כיוון נוסף שיש לתת עליו את הדעת הוא חדירה של כלי בינה מלאכותית לתחום הפקת המידע והחדשות וההשפעה על האוטונומיה של הפרט. ערוצי תקשורת רבים משתמשים בבינה מלאכותית כדי לייצר חדשות מותאמות באופן אישי. לכלי זה יתרונות רבים אבל הוא גם מציב סכנה של חשיפה סלקטיבית – קבוצות באוכלוסייה שחשופות למידע חד-גוני ואינן נחשפות לעדויות וטיעונים שאינם עולים בקנה אחד עם תפיסת עולמם ולא על-פי בחירתם המפורשת. במצב זה נשללת מאוכלוסייה שכזו הבחירה החופשית או החשיפה לדעות שונות. כמו-כן אוכלוסיות כאלו חשופות לניסיונות השפעה בלתי הוגנת על-ידי גורמים אינטרסנטיים. זאת מכיוון שמקורות המידע שלהם מצומצמים ולכן פגיעה במספר קטן ממקורות אלו יכולה לגרום השפעה לא פרופורציונית. נושא זה הועלה כדרך אפשרית בה גורמים זרים יכולים להשפיע על תוצאות הבחירות.

ג.2.1.3. זכויות אזרחיות ופוליטיות

פגיעה בזכויות ביטוי, בשוויון, ובחופש הדת והמצפון עשויה להתרחש כאשר נוצרת השפעה על השיח הציבורי באמצעים אוטומטיים, כגון הדהוד יתר של דעות מסוימות והשתקת אחרות, הקצנה של השיח ומתן לגיטימציה לדעות שיכולות להיות פוגעניות עבור קבוצות מסוימות, וכן השפעה על תהליכי שכנוע בדעות ורעיונות. כל אלה עשויים לפגוע ביכולת לממש זכויות אזרחיות ופוליטיות שונות ומעלים חשש מפני התפשטות של שקרים ופגיעה בהליך דמוקרטי תקין. אלה עשויים לפגוע ביכולת לממש זכויות אזרחיות ופוליטיות שונות ויש לנסות למנוע אותן.

³⁶ ארידור-הרשקוביץ רחל ואלטשולר שוורץ תהילה, 2019, הצעת חוק הגנת הפרטיות התשע"ט-2019, המכון הישראלי לדמוקרטיה <https://www.idi.org.il/books/27125>

2.1.4 הגנת הסייבר ואבטחת מידע

הגנת הסייבר ואבטחת מידע עוסקות בהגנה על מחשבים ותקשורת מפני שימוש לרעה בהם, או במידע השמור בהם. מכאן שהגנת הסייבר ואבטחת מידע הן דרישות תשתיות לקידום הבטוח והתקין של פיתוח והטמעה של טכנולוגיות בינה מלאכותית. המידע הוא הדלק המניע את הגל הנוכחי של הבינה המלאכותית. לכן, העוסקים בבינה מלאכותית אוספים כמויות גדולות של מידע הן בשלב בניית הכלים והן בזמן הפעלת המערכות. מידע כזה יכול לכלול מידע אישי, רפואי, כלכלי ומידע רגיש אחר. כמו-כן מידע לא רגיש יכול להפוך רגיש כאשר הוא מוצלב עם מקורות מידע אחרים. לפעמים, אפילו כמות המידע יכולה להפוך מידע לרגיש. לפיכך, על העוסקים בתחום הבינה המלאכותית להקפיד על אבטחת המידע על-מנת למנוע זליגה של מידע ולמנוע מגורמים זדוניים גישה למידע³⁷. המערכות האלגוריתמיות התומכות החלטות אנושיות או מחליפות אותן הופכות להיות בעצמן יעד להגנה על מנת למנוע מימוש תרחישי שימוש לרעה עקב היותן ממוחשבות.

2.1.5 בטיחות

מערכות בינה מלאכותית עשויות להיות משולבות כתומכות החלטה או מקבלות החלטה באופן שהן יכולות להשפיע על בטיחות באופן ישיר ועקיף. כך לדוגמה כבר נרשמו מספר מקרים בהן מכוניות אוטונומיות היו מעורבות בתאונות קטלניות. סכנות הבטיחות כמו גם הנוק, יכולים לנבוע ממספר גורמים. ראשית, אי זיהוי של עיוותים, או מגבלות אחרות של הנתונים, יכולות להביא את המערכת לקבלת החלטות שגויות (לדוגמה בנושא המלצות על טיפול בחולים בדלקת ריאות)³⁸. אתגר נוסף הוא הטיפול במקרי קצה או מקרי קיצון. במקרים אלו יש לפתח מנגנונים שיגבילו את מידת הנוק. כמו-כן יש לנטר את ביצועי המערכת, לזהות נקודות כשל ולתקן אותן במהירות³⁹.

נושא נוסף שיש לתת עליו את הדעת הוא בטיחות בשלב בניית מערכת הבינה המלאכותית. כפי שנאמר, מערכות אלו דורשות מידע רב ולכן יש לשים לב שהמרוץ להשיג מידע כזה אינו מוביל ללקיחת סיכונים מיותרים. כך לדוגמה, כדי לבנות מכונית אוטונומית יש לאסוף נתונים במצבי נהיגה מגוונים, חלקם מסוכנים, על-מנת שהמערכת תלמד איך להגיב במצבים אלו. לכן, עלול להיווצר מצב שבו בשלב הפיתוח מציבים אנשים במצבי סיכון ואפילו מעודדים אותם לקחת סיכונים על-מנת לאסוף מידע שיסייע למערכת.

2.1.6 קיום שוק פתוח, תחרותי והוגן

הוועדה רואה בתחרות ובתהליך התחרותי תנאים לחופש ולחרות ותורם לחדשנות ולרווחה חברתית. מניעת תחרות מונעת משחקנים חדשים ליזום ולהיכנס לשוק, והנוק לטווח ארוך למדינה הוא גבוה. בהתאם לכך שמירה על שוק חופשי שיש בו תחרות הוגנת תאפשר לכל השחקנים בשרשרת הערך, בעיקר חברות קטנות-בינוניות וסטרטאפים ליהנות ולהרוויח מהפעילות.

בתחום הבינה המלאכותית, קיימים שחקנים בעלי כוח שוק משמעותי במקטעים שונים של שרשרת הערך הרלבנטית לפיתוח הטמעה ושיווק של הטכנולוגיה. כוח שוק זה נובע בין היתר ממאפיינים של כלכלת רשת, שווקים דו צדיים, ונגישות נרחבת למידע המשמש לעיבודים. ריכוזי כח כלכלי עלול להוביל גם לריכוז כח פוליטי, ומאפשר לחברות לקבוע את כללי המשחק בשוק. החשש הוא השפעתם של מגה-שחקנים אלה על השוק עלולה לפגוע בכניסתן של טכנולוגיות חדשות או יישומים חדשים, ובחדשנות הגלומה בבינה מלאכותית. למשל העוצמה של פלטפורמות המידע תורגמה לעוצמה בתחום הפרסום (לדוג' גוגל ופייסבוק), ולאחר מכן קמעונאות (לדוג' אמזון), הרחבת העוצמה הקיימת עשויה להשפיע על תחומי כלכלה נוספים.

בהתאם לכך, יש לבחון האם דיני התחרות, החוזים האחידים והגנת הצרכן הקיימים ערוכים לתת מענה מהותי ומוסדי לאתגרים הצפויים. לכך יש להוסיף את האתגר הגלובאלי, הנובע מכך שחלק מהשחקנים המשמעותיים פועלים בארצות הברית. לדעת המומחים מהאיחוד האירופאי, יש לבחון האם יש לשנות את דיני התחרות על מנת לאפשר תחרות רבה יותר בתחום הבינה מלאכותית. בתוך כך, יש למפות את התעשייה על פי המוכנות שלה לקלוט בינה מלאכותית ולתת מענה שונה לכל תחום בהתאם ליכולות כאמור⁴⁰.

³⁷ Ibid. Microsoft, 2018.

³⁸ Ibid. Microsoft, 2018, pp. 64.

³⁹ Ibid. Microsoft, 2018, pp. 61-65.

⁴⁰ Ibid. Ethical Guidelines for Trustworthy AI, 2019.

בבריטניה המליצו בדוח משנת 2016 להגביל את הכוח של החברות הגדולות עתירות המידע באמצעות כלים מתחום ההגבלים העסקיים⁴¹.

בהקשר זה, אתגר מהותי בתחום הבינה המלאכותית הוא הצורך בנגישות למאגרי מידע גדולים, הנמצאים בידי גופים מעטים, לעיתים בידי חברות פרטיות. כך, חברות האוחזות במידע רב יותר עלולות למנוע מחברות קטנות להיכנס לשוק ולהתחרות בו ועלולות להיות מוטות באופן ברור לטובתן בשירות אותו הן מעניקות⁴². מדינות אחדות, כמו צרפת, קראו לממשלה, בדוח שפורסם, לממשלה ליצור מסדי נתונים פתוחים, כטובין ציבוריים ולעגן את הפתיחות של אותם מסדי נתונים בחוק זכות יוצרים (ראו למשל בישראל את ההסדר הקובע כי אין זכות יוצרים ביצירות אחדות המנויות בסעיף 6 לחוק זכות יוצרים, התשס"ח-2007)⁴³. המלצה זו נועדה למנוע הטיות ולשמור על הוגנות פעולת המערכת באמצעות אימון של מערכות הבינה מלאכותית על מסדי נתונים הוגנים, מקיפים, המייצגים את כלל האוכלוסייה בהתאמה וכיו"ב. בנוסף קריאה זו נועדה ועל מנת לאפשר תחרות פתוחה והוגנת בתחום בין חברות קטנות שאין ברשותן מידע רב לבין חברות גדולות, להן ישנו מידע רב עליו ניתן לאמן את מערכות הבינה מלאכותית.

כפתרון נוסף ניתן לחשוב על הכנת מסדי נתונים פתוחים והנגשתם לציבור באמצעות טכנולוגיית בינה מלאכותית. כך עשו במחקר בו יצרו פרצופים על ידי בינה מלאכותית⁴⁴. כך גם ניתן לדאוג שהמאגר ייתן ייצוג לכלל האוכלוסייה תוך שמירה על פרטיות האנשים ושימוש בתוכן שנוצר על ידי המחשב. בדוח שמואל נאמן ציינו את החשיבות ביצירת מסדי נתונים ציבוריים משותפים. גם מיקרוסופט קראה לפתיחת מסדי נתונים כמו מסדי נתונים במאגרים ציבוריים. עם זאת, מציינת החברה, יש לחשוב על דרך בה תישמר הפרטיות (אולי להשקיע במחקרים של הסרת הזהות)⁴⁵.

2.2 כלי למקבלי החלטות לבחינת האתגרים האתיים

הוועדה ממליצה לראות בכל העוסקים בתחום הבינה המלאכותית כאחראים לפעול באופן חוקי ואתי. יתרה מזאת, מכיוון שהתחום מתפתח במהירות, כמו גם ההבנה שלנו את הסכנות, האפשרויות ודרכי ההתמודדות איתן, על העוסקים בתחום להישאר מעודכנים בנושא ואי עמידה בתנאים אלו עלולה להיחשב כרשלנות.

כדי לאפשר לעוסקים בתחום הבינה המלאכותית לזהות סיכונים אתיים ולהגיב אליהם בצורה מתאימה פיתחנו כלי עזר שברצוננו להעמיד לרשות מקבלי החלטות בתחום והציבור. בפרק זה נדגים את הכלי המוצע, אולם היות והתחום מתפתח והינו דינמי, יהיה צורך לשמור על עדכונו השוטף של הכלי.

הכלי בנוי משני חלקים:

1. חלק אי מורכב מקבוצת שאלות ראשוניות שיש לשאול בעת פיתוח מוצר בינה מלאכותית והבוחנות את מידת ההשפעה של מערכת הבינה המלאכותית. שאלות אלה מיועדות למפתחי המוצרים לאורך כל שרשרת הפיתוח והייצור, ויש לעדכן אותן מעת לעת.
 2. חלק ב' הוא שימוש ותחזוקה של מפת שכיחויות, המאפשרת לאתר את המוקדים האתגריים מבחינת הטמעת הערכים האתיים בשלבי הפיתוח של המערכת (לאו דווקא את עוצמת הנזק או הפגיעה, כי אם את שכיחות הפגיעה). מפת השכיחויות המוצגת בדוח בוחנת 10 מקרי-בוחן מייצגים אותם בחרנו, ומאשרת את הנקודות בהן עלולים להתעורר אתגרים אתיים (ראה עמ' 13) לפי ציר הזמן ושלבי הפיתוח. בעזרת מקרי הבוחן, המפה מנסה להעלות למודעות את התחומים בהם היו בעיות לארגונים אחרים בעבר, והיכן צריכים מקבלי החלטות צריכים לשים לב, להיות זהירים יותר ולפעול. יש לדאוג להזין את המפה במקרי בוחן עדכניים, רלוונטיים וחשובים באופן שוטף, אחרת מפה זו תתייתר.
- בעמודים הבאים נדגים כיצד לעצב כלי שכזה עבור ארגון המפתח מוצר מבוסס בינה מלאכותית.

House of Commons Science and Technology Committee, 2016, Robotics and Artificial Intelligence, ⁴¹ <https://publications.parliament.uk/pa/cm201617/cmselect/cmstech/145/145.pdf>

⁴² שם. ג' ואחרים, 2018, עמ' 90.

⁴³ Ibid. Villani, 2018.

⁴⁴ Karras Tero, Samuli Laine and Aila Timo, 2018, "A Style-Based Generator Architecture for Generative Adversarial Networks", Neural and Evolutionary Computing, <https://arxiv.org/abs/1812.04948>

⁴⁵ Ibid. Microsoft, 2018, pp. 78.

2.2.1 חלק א' של הכלי - קבוצת שאלות ראשוניות להערכת מידת השפעה של המערכת:

בבואנו לעצב את קבוצת השאלות שעל מקבלי החלטות לענות עליהם כדי לבדוק כשלים אטיים אפשריים, אנו מתמקדים במאפייני השפעה ופרמטרים שונים⁴⁶.

- 1. מהי עוצמת הפגיעה הפוטנציאלית בפרט?**
ככל שעוצמת הפגיעה הפוטנציאלית בפרט עולה, יש לתת את הדעת יותר לערכים כמו הוגנות, זכויות האדם, אחריותיות, שמירה על אוטונומיה ואבטחת המידע. ככל שהחברה יש לה נתח שוק גדול יותר – העוצמה של הפגיעה, גם אם היא קטנה, במצטבר עלולה לפגוע בקהל לקוחות רחב מאוד. בתחום הפרטיות, למשל, אמנות שונות כיום בעולם דנות בשאלת ההסכמה לאיסוף המידע ולעיבודו. במידה שלא ניתן אישור כזה יש לגלות זהירות יתר לא רק בשלב איסוף המידע, המוסדר כיום בחקיקות השונות, אלא אף בשלב של השימוש במידע לצרכי אימון מערכות הבינה המלאכותית. כך גם לגבי קיומו של מידע אישי. גם נושא זה מוסדר כיום ברובו בחקיקה, אולם אם יש שימוש במידע אישי יש לגלות רגישות יתר הן בשלב האיסוף והן בשלב אימון המערכת או שיתוף המידע עם אחרים.
- 2. מהי עוצמת ההשפעה הפוטנציאלית על התודעה?**
סעיף זה גם מתייחס לעוצמת הפגיעה בפרט, אך לדעת הוועדה דורש התייחסות נפרדת וממוקדת. במידה שלמערכת יש השפעה אפשרית על התודעה יש לגלות רגישות יתר, במיוחד במקרים בהם ייתכן שיעשה שימוש לרעה במערכת הבינה המלאכותית, תוך שימוש במניפולציות על התודעה (למשל – Deep Fake).
- 3. מהי עוצמת הפגיעה הפוטנציאלית בציבור?**
כך, למשל, כמות לקוחות או כוח השוק של החברה בעלת הטכנולוגיה של הבינה המלאכותית יכולים להעיד על פוטנציאל הפגיעה. יש לבחון מהו היקף ההשפעה של המערכת מבחינת מספר משתמשים או השפעה על השוק. יש לבחון גם האם קבוצה מסוימת בציבור עלולה להיפגע. כלומר, אם המערכת מתוקף עבודתה, יוצרת אבחנה בין קבוצות, יש לגלות רגישות יתר.
- 4. האם ישנה השפעה על הקצאת משאבים לציבור, ומהי?**
גם סעיף זה מתייחס לעוצמת הפגיעה בציבור אך לדעת הוועדה דורש התייחסות נפרדת וממוקדת. האם המערכת עשויה להשפיע על חלוקה של משאבים כספיים, ציבוריים או אחרים? במידה שהמערכת יכולה להשפיע על חלוקת משאבים יש לגלות רגישות יתר.
- 5. האם צוות הפיתוח מגוון דיו?**
יש לבחון האם צוות הפיתוח כלל מגוון רחב של אנשים שמייצגים בדגש על נציגות של קבוצות באוכלוסייה העלולות להיפגע מהמערכת. במידה ולא, יש להגביר את הזהירות בבחינה הערכית של המערכת. לשם בחינה זאת יש גם לבדוק האם ניתן לזהות את הקהלים שנמצאים בסיכון.
- 6. מהי עוצמת הפגיעה הצפויה אם יעשו שימוש לרעה במוצר או שהמוצר יצא משליטה?**
יש לבחון מהי עוצמת הפגיעה הפוטנציאלית אם יעשו שימוש לרעה במערכת. ככל שעוצמת הפגיעה – במישור הפרט, הציבור או הכלכלה עלולה להיות גדולה, יש להיות זהירים יתר ולנסות למזער את הנזק כמו – היכולת להפסיק את פעילות המערכת בעת שהיא יוצאת משליטה או שעושים בה שימוש לרעה.
- 7. האם יש דרך מהירה לזיהוי של תקלות אתיות במערכת שלא נצפו מראש?**
מידה שיכול להצטבר נזק רב עד אשר תזוהה התקלה יש צורך ברף בדיקות גבוהה ביותר לפני תחילת השימוש במערכת.

⁴⁶ בנוסף, ממשלת קנדה פיתחה כלי-שאלונים (Algorithmic Impact Assessment) שנועד לעזור להעריך ולצמצם סיכונים בעת פיתוח והטמעה של מערכות בינה מלאכותית. <https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai/algorithmic-impact-assessment.html>

2.2.2 חלק ב' של הכלי – מפת שכיחות אתגרים בשלבי הפיתוח של מערכת בינה מלאכותית:

מפת השכיחות מראה את השכיחות של תופעות אתיות לאור שרשרת הפיתוח של המוצר.

המפה אמורה לתאר את הנקודות בהן נתגלו כשלים בעבר ומידת הנפיצות של כשלים כאלו. השכיחות של האירועים עשויה להשתנות עם הזמן, ואירועים חדשים עלולים להתגלות. על כן, אנו ממליצים לשמור על כלי זה ככלי חי ולעדכן אותו באופן תדיר ככל שהניסיון בעולם גדל ומקרי בוחן נוספים יתגלו. אי לכך, מפת השכיחות המובאת כאן, היא הדגמה של השיטה שבה צריך לפתח כלי זה ואת אופן השימוש בו.

על מנת לבנות את מפת השכיחות השתמשנו ב-10 מקרי בוחן שמייצגים אתגרים שונים. מקרי הבוחן נבחרו מתוך מקרים שכבר קרו בעבר. רשימה זו נבחרה כך שתדגים את כל העקרונות האתיים (ראה סעיף 2.1) בעת שלבי הפיתוח השונים.

טבלה 1: מקרי בוחן שמשמשים כאבי-טיפוס לאתגרים אתיים

מקרה בוחן 1	<u>מערכת בינה מלאכותית לסינון מועמדים לעבודה</u> חברות מסחריות מקבלות מספר רב של פניות של מועמדים לעבודה בחברה. על-מנת לברור מתוך פניות אלו את המועמדים הטובים ביותר פיתחו מספר חברות כלים מבוססי בינה מלאכותית שיבצעו סינון של המועמדים. כלים אלו אומנו על בסיס החלטות העבר של החברות. בבדיקת מערכת כזו שפותחה בחברת אמזון נתגלה שהמערכת מפלה נשים בקבלה לעבודה במשרות טכניות. ההערכה היא שהפליה זו נובעת מכך שבעבר מנהלים בחברה נטו להפלות נשים בקבלה לעבודה והמערכת למדה לחקות התנהגות זו ⁴⁷ .
מקרה בוחן 2	<u>שימוש בבינה מלאכותית להשפעה על תודעת האנשים</u> חברת קיימבידג' אנליטיקה אספה נתונים אישיים של מיליוני פרופילים בפייסבוק, ללא הסכמתם או ידיעתם ועשתה במידע זה שימוש על מנת להשפיע עליהם למטרות פוליטיות במניפולציות תודעתיות שונות. ככל הנראה היה שימוש בטכנולוגית בינה מלאכותית לצורך ניסיון ההשפעה על התודעה של האנשים. פעילות זו נמשכה במשך מספר שנים ⁴⁸ .
מקרה בוחן 3	<u>חיזוי סיכון ממחלות</u> במהלך שנות ה-90 מספר מרכזי מחקר חברו יחד כדי לפתח מערכת שתעריך עד כמה מחלת דלקת הריאות מהווה סיכון חיים עבור החולים בה. מטרת המערכת הייתה לאפשר לרופאים להחליט את איזה מהחולים יש לאשפז ובאיזה חולים ניתן לטפל בקהילה. מעט לפני תחילת השימוש במערכת התגלה שההמלצות של המערכת עבור חולי אסטמה עלולות לסכן את חייהם. הסיבה לכך היא שהמידע שבו נעשה שימוש כדי לבנות את המערכת היא מוטה מכיוון שחולי האסטמה שחלו בדלקת ריאות קיבלו טיפול אינטנסיבי קודם שהציל את חייהם. המערכת הסיקה מכך שמחלת דלקת הריאות אינה מהווה סיכון חיים עבור חולי אסטמה ⁴⁹ .
מקרה בוחן 4	<u>מערכת להערכת מסוכנות של עצירים</u> בעת החלטה על הארכת מעצר של עציר אחד השיקולים המובאים בחשבון הוא רמת המסוכנות של העציר. החלטה זו מבוססת על פרמטרים רבים כגון עברו הפלילי של העציר. מספר מחוזות בארצות הברית אימצו מערכת בינה מלאכותית בשם Compas על מנת לסייע לשופטים בהערכת המסוכנות. בדיקה של המערכת גילתה שהיא מוצאת עצירים לבנים כמסוכנים פחות מעצירים שחורים ⁵⁰ .

<https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>⁴⁷

<https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data>⁴⁸

<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>⁴⁹

<http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf>⁴⁹

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>⁵⁰

<p><u>משחק בו שחקנים וירטואליים מונחי בינה מלאכותית צברו נשקים שוברי שיוויון</u></p> <p>במשחק Elite Dangerous מתחרים שחקנים בני-אנוש מול שחקנים מונחי בינה מלאכותית. על-מנת להפוך את המשחק למעניין יותר שונו בגרסה 2.1 המגבלות על השחקנים הווירטואליים כדי לאפשר להם לעוף ולהילחם טוב יותר. מנגנוני הבינה המלאכותית מצאו דרך לנצל שינויים אלו כדי לאגד ביחד כלי נשק בצורה שלא אפשרה למשתמשים אנושיים להתמודד איתם.⁵¹</p>	<p>מקרה בוחן 5</p>
<p><u>בוט שהפך לגזען</u></p> <p>חברת מיקרוסופט שחררה בוט במטרה ללמד אותו להתכתב בשפה חופשית עם משתמשים בטוויטר. מטרתו הייתה לקיים שיחה ותוך כדי ללמוד ולהשתפר בשיחה. פחות מיממה מתחילת הפעלתו של בוט זה התגלה שמכיוון שהבוט לומד לחקות את המשתמשים בו, בחרו מספר משתמשים להפוך אותו לגזען על-ידי שהם השתמשו בעצמם בהערות גזעניות.⁵²</p>	<p>מקרה בוחן 6</p>
<p><u>בוט שמתחזה לאדם</u></p> <p>חברת גוגל פיתחה מערכת, Google Duplex, המאפשרת לבוט לנהל שיחה בצורה שהקשתה על הדוברים בצד השני לזהות האם מדובר באדם או מכונה שהם מדברים איתו. בניית כלי זה דרשה גישה לכמויות אדירות של נתונים הנמצאים ברשות מספר קטן ביותר של חברות עתירות ידע.⁵³</p>	<p>מקרה בוחן 7</p>
<p><u>מכוניות אוטונומיות דרסה למוות הולכת רגל</u></p> <p>הולכת רגל שחצתה את הכביש במקום חשוך באריזונה נהרגה על ידי מכונית אוטונומית של אובר. ככל הנראה הרכב זיהה כי עומד לפניו מכשול ויכול היה להימנע מלהתנגש בו. עם זאת, מכיוון שהמהנדסים הנמיכו את רגישות התוכנה לזיהוי החסמים המכונית לא עצרה והאישה נהרגה. נהגת אנושית שהייתה ברכב לא הייתה ערנית מספיק ולא הצליחה למנועה את התאונה.⁵⁴</p>	<p>מקרה בוחן 8</p>
<p><u>הטיות בזיהוי פנים</u></p> <p>חברת אמזון פיתחה כלי למהנדסים המאפשר להם להוסיף זיהוי פנים למערכת אותה אם מפתחים. המערכת נועדה לשימוש, בין היתר, על-יד גורמי אכיפת חוק, משטרת הגבולות ועוד. בדיקה גילתה שמהערכת מבצעת אחוז גדול בהרבה של שגיאות כאשר היא מופעלת על אנשים בעלי גוון עור כהה מאשר כאשר היא מופעלת על אנשים עם גוון עור בהיר.⁵⁵</p>	<p>מקרה בוחן 9</p>
<p><u>מערכות המלצה של תוכן מציגות מידע שונה לאוכלוסיות שונות</u></p> <p>כדי למצוא מידע רלוונטי יותר עבור משתמשים חברות שונות משתמשות בבינה מלאכותית כדי להתאים מידע למשתמשים. בדיקה מצאה שמערכת הפרסומות של חברת גוגל מציגה פרסומות לאיתור מידע פלילי כאשר משתמש מחפש מידע על שם שנפוץ יותר באוכלוסיות מיעוטים.⁵⁶</p>	<p>מקרה בוחן 10</p>

גורמים שונים מעורבים בפיתוח מערכת הבינה מלאכותית: מנהלי מוצר, מדעני נתונים, אנשי מכירות, קניינים של מערכות כאלה ועוד. הכלי הבא, מציג דוגמאות לסוגיות אתיות העולות בשלבים השונים של תהליך הפיתוח ומאפשרת לעקוב אחריהן. מטרת הטבלה היא לסייע באיתור השלבים, בציר הזמן, בהם עלולות להתעורר בעיות אתיות. לשם כך אנו מתארים תהליך אופייני של פיתוח בינה מלאכותית. יש לציין שתהליך הפיתוח יכול להשתנות ממוצר לאחד למשנהו.

⁵¹ <https://futurism.com/this-video-games-artificial-intelligence-turned-on-players-using-super-weapons>

⁵² <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>

⁵³ <https://www.androidcentral.com/google-duplex-will-let-people-know-its-not-human>

⁵⁴ <https://www.cbc.ca/news/business/uber-arizona-crash-1.4594939>

⁵⁵ <https://www.theverge.com/2019/1/25/18197137/amazon-rekognition-facial-recognition-bias-race-gender>

⁵⁶ <https://www.bostonglobe.com/business/2013/02/06/harvard-professor-spots-web-search-bias/PtOgSh1ivTZMfyEGj00X4I/story.html>

שלבי הפיתוח להתייחסות אתית:

1. הגדרת המוצר
א. הבנת הצורך העסקי – זיהוי הצורך או הבעיה אותה המערכת מנסה לפתור ובניית מערך הפיתוח.
יא. איסוף נתונים – זיהוי מקורות מידע, בתוך הארגון ומחוץ לו, אשר ישמשו לבניית המערכת והערכת הביצועים שלה.
2. אימון המוצר
א. ארגון הנתונים – עיבוד וסינון הנתונים הגולמיים לכדי צורה שתאפשר לאלגוריתמי הבינה המלאכותית לקבל את הנתונים ולערוך איתם חישובים.
ב. בניית מודל – הפעלת אלגוריתם של בינה מלאכותית בעיבוד המידע בניסיון למצוא תבניות המאפשרות הכללה.
3. אינטגרציה
א. הערכת המודל – בדיקת תוצאות שלב בניית המודל כדי להעריך אם המודל מדויק דיו כדי לאפשר שימוש בו.
ב. הפצה – חיבור רכיבי הבינה המלאכותית עם שאר חלקי המערכת והעברתם לשימוש נרחב.
4. ניהול שוק
א. מעקב אחר ביצועים – אחרי הפצת המערכת יש להמשיך לעקוב אחרי ביצועי המערכת כדי לוודא שהמערכת מבצעת את עבודתה בהתאם לציפיות.
ב. אקוסיסטם (חברתי, עסקי, רגולטורי) – במקביל לתהליך המתבצע בתוך הארגון, יש צורך להתייחס גם לשיקולים האתיים הנוצרים כתוצאה מקיום התהליך באקוסיסטם הישראלי. שילוב הבינה המלאכותית עלול להשפיע על המערכת החברתית, העסקית, הכלכלית, הצרכנית, הרגולטורית ועל מערכות נוספות, ויש לבחון האם וכיצד הבינה המלאכותית משפיעה על כך, גם לאחר שחרורה מהמפעל.

כיצד מייצרים את מפת השכיחות?

המבקשים לבנות את מפת השכיחות, עוברים על רשימת מקרי הבוחן, ובוחנים את ההשלכות ואת האירועים המדווחים בהתחשב ברשימת הערכים האתיים עליהם הצבענו בסעיף 2.1: הוגנות, אחריות, כיבוד זכויות אדם, הגנת סייבר, בטיחות וקיום שוק תחרותי. ממלאים את הטבלה בהתאם לאתגרים האתיים שהתגלו. המספרים בתוך התאים בטבלה מתייחסים למס' האירוע. בהמשך בודקים את כמות האירועים המצטברים מבין סך האירועים. תאים בטבלה בהם נמצאו מעט אתגרים וכשלים אתיים (שכיחות נמוכה של אירועים) נצבעו בצבע בז' וצהוב בעוד נקודות בהם נמצאו מספר רב יותר של אתגרים אתיים (שכיחות גבוהה של אירועים) נצבעו באדום. נציין כי לא התייחסנו הפעם לעוצמת הפגיעה האפשרית ולהיקפה בעת בחירת צבעים אלה, אלא רק למידת שכיחות הבעיות שאיתרנו. ייתכן כי כלי משוכלל יותר ייקח אף בחשבון את עוצמת הפגיעה ואת היקפה כאמור.

ההחלטה על מה מהווה שכיחות נמוכה, בינונית וגבוהה צריכה להתקבל בעת בחירת מס' האירועים עליהם מתבסס הארגון, ובהתחשב בתוצאות מילוי התאים במפה. לדוג' במקרה שלנו (ראו טבלה 2) יש לנו 10 אירועים ובמפה המקסימום של שכיחות אירועים לתא היא 3 ו-4 אירועים. על כן במקרה ויש שלושה או ארבעה אירועים מדווחים, צבענו את התא כשכיחות גבוהה של אירועים (נצבע באדום). בהתאם בחרנו 2 אירועים כשכיחות בינונית לאירועים בעייתיים (נצבע בצהוב), ואירוע אחד כשכיחות נמוכה (נצבע בבז'). הוועדה ממליצה למקבלי החלטות לדון בכובד ראש ולהציע פתרונות לאתגרים העולים לפי מפת השכיחות בהליך הפיתוח שלהן.

מכיוון שמפת השכיחות תלויה לחלוטין ברשימה איכותית של אירועי בוחן, ההנחה היא כי כל ארגון צריך לבחור קבוצה של אירועים שדומה למוצר אותו הם מפתחים, וההנחה היא כי קבוצת האירועים משתנה לאורך הזמן והניסיון האנושי.

טבלה 2 - מפת שכיחות אתגרים בשלבי פיתוח מערכת בינה מלאכותית

הבנת הצורך העסקי	איסוף נתונים	ארגון נתונים	בניית מודל	הערכת המודל	הפצה	מעקב אחרי ביצועים	אקוסיסטם	
	1,3,4		3,4	1,3,4	1,4,9	1,3,4,9	1,4,9	הוגנות
4			3,4				4	שקיפות
4			3,4				9	הסברתיות
1,2,3,4				1,3,4	3,4	3,4,5,6	2,5,6,9	אחריות
2,9	1	1,2	1	1			2,9	פרטיות
7,10						10	6,7,10	חופש בחירה
		2					2,9	אבטחת מידע
4,9			4		4	4	4,9	זכויות אדם
3,4	3,4		3,4	3,4	3,4,5,8	3,4,5	3,5,8	בטיחות
5					5,6	5		שוק תחרותי

מקרא:

1	מקרה בוחן 1 מערכת בינה מלאכותית לסינון מועמדים לעבודה
2	מקרה בוחן 2 שימוש בבינה מלאכותית להשפעה על תודעת האנשים
3	מקרה בוחן 3 חיזוי סיכון ממחלות
4	מקרה בוחן 4 מערכת להערכת מסוכנות של עצירים
5	מקרה בוחן 5 משחק בו שחקנים וירטואלים מונחי בינה מלאכותית צברו נשקים שוברי שיווין
6	מקרה בוחן 6 בוט שהפך לגזען
7	מקרה בוחן 7 בוט שמתחזה לאדם
8	מקרה בוחן 8 מכוניות אוטונומיות דרסה למוות הולכת רגל
9	מקרה בוחן 9 הטיות בזיהוי פנים
10	מקרה בוחן 10 מערכות המלצה של תוכן מציגות מידע שונה לאוכלוסיות שונות

שכיחות המקרים הבעייתיים הייתה נמוכה (מקרה בודד)
שכיחות המקרים הבעייתיים הייתה בינונית (2 מקרים)
שכיחות המקרים הבעייתיים הייתה גבוהה (3 מקרים ומעלה)

פרק 3 – בינה מלאכותית ואסדרה

הפרק הראשון בדוח זה הציג את ההיבטים הייחודיים של בינה מלאכותית ואשר רלוונטיים לדיון בנושא המשילות האתית והמשפטית בתחום הבינה המלאכותית. הפרק השני בדוח הציג את המלצות הצוות ביחס לעקרונות הערכיים והאתיים שיש להחיל על פעילות פיתוח ושימוש בטכנולוגית בינה מלאכותית. המלצות אלה גובשו בהתבסס על ההיבטים הייחודיים שאותרו בידי הצוות, בזיקה לדיונים דומים המתנהלים בגופים בינלאומיים ציבוריים ופרטיים. הפרק השלישי יציג את האפשרויות השונות בתחום האסדרה של בינה מלאכותית, כחלק מהדיון אודות התפקיד של המשפט ומוסדות משפטיים בקידום ודאות, מימוש עקרונות ערכיים ותמיכה בחדשנות של בינה מלאכותית.

הדיון נעשה ברמת הפשטה גבוהה, על מנת לאפשר "מבט על" על כיווני הפיתוח הנורמטיביים האפשריים והמומלצים. לצורך דיון זה, ובדומה לפרק הראשון והשני, במוקד פרק זה השאלה מהם המאפיינים הייחודיים של הטכנולוגיה ואופן הפיתוח והשימוש בה בהקשרי האסדרה, תוך התבססות על הנחת המוצא של ההתערבות המאוזנת – לא רק התערבות בדיעבד, אולם התערבות רק במידה הנדרשת ובאופן המינימלי הנדרש, על מנת לאפשר את החדשנות ואת המשך הובלת מדינת ישראל בתחום תוך שמירה על כללים אתיים ורגולטוריים ברורים.

3.1 אסדרה של בינה מלאכותית בפעילות פרטית

המטרה הכללית של אסדרה היא להביא להפנמה ולשינוי התנהגות של אנשים לפי שיקולים ערכיים של חברות (societies), כגון שמירה על בריאות ובריאות הציבור, מניעת ניצול צרכנים וכדומה. תחום הבינה המלאכותית מתייחס לטכנולוגיות שונות, עם שימושים מגוונים במשק, כאשר המגזר הפרטי מניע חלק ניכר מההתקדמות הטכנולוגית בתחום⁵⁷. כתוצאה מכך עולות שאלות לגבי האם צריך אסדרה, באיזו צורה, ובאילו מקרים.

פרק זה יתמקד באסדרה ביחס לפעילות פרטית בתחום הבינה המלאכותית⁵⁸, בהקשר לעקרונות והערכים שזוהו כיעדים לקידום אתי בסעיף 2.1.

יש לבחון את הסדרת הבינה המלאכותית במספר מעגלים – מעגל כללי – השפעה על החברה בכללותה; מעגל מפתחי הבינה המלאכותית והמשתמשים בה; ומעגל הלאומי של המדינה כיצרנית בינה מלאכותית וכמשתמשת בה.

במסגרת תפיסות האסדרה של טכנולוגיות מידע חדשניות, שנכונות גם לבינה מלאכותית, ניתן לציין מספר מאפיינים אותם יש לקחת בחשבון:

1. תחום הפיתוח של טכנולוגיות מבוססות בינה מלאכותית הן דינאמיות ביותר בכל הקשור להטמעה ואימוץ הטכנולוגיה על ידי החברה (society). ההתפתחות המהירה ואימוצה המהיר אינם מאפשרים לעיתים למוסדות העוסקים באסדרה לנתח את התופעה הטכנולוגית והשפעותיה ולהבין אותה, לצורך החלת מסגרת נורמטיבית. בנוסף, לעיתים תופעה טכנולוגית מביאה לשינוי תפיסות ערכיות וחברתיות באופן שמשפיע על הפרשנות הנורמטיבית ועל כן יש קושי בתגובה אסדרתית מהירה.

2. תחום מוצרי בינה מלאכותית הוא גלובלי ומושפע מפיתוחים טכנולוגיים חוצי-גבולות. לכך תורמים גם מאפייני הכלכלה הגלובאלית, שבעיקרה מאפשרת תנועה של מידע הון ושירותים מעבר לגבולות, באופן שמאתגר משטרי-אסדרה משפטיים.

3. עלייתן של חברות בעלות כוחות שוק חוצי-גבולות, פלטפורמות העברת מידע, מסחר ותיווך בשירותים ובמידע הכרוך בהם. חברות אלה הן בעלות עוצמה רבה בתחום המידע, ההון וכוח השוק, המשפיעות רבות על הסביבה הטכנולוגית, עסקית ואף פוליטית, העולמית והמקומית כאחד.

⁵⁷ ד"ר רועי שיינדורף, 2019, גיבוש אסטרטגיה ישראלית בתחום הבינה המלאכותית – היבטי המשפט הבין-לאומי, מזכר מטעם משרד המשפטים שהוכן עבור הוועדה.

⁵⁸ הפרק ממוקד באסדרה שעניינה הפעלת סמכות מנהלית ביחס לפעילות פרטית. בכל הקשור לפעילות שמבוצעת בידי מי שחל עליו המשפט המנהלי, (לדוג' שימוש ופיתוח בינה מלאכותית ע"י הממשלה), המשפט המנהלי עצמו מהווה הדין המסדיר. האופן שבו מקדמים הפנמה והטמעה של עקרונות מהותיים על גופים שחל עליהם המשפט המנהלי שונה מבחינה משפטית, מוסדית ותהליכית, ומצדיק דיון נפרד.

לענפי המשפט השונים יכולת להשפיע בצורות ישירות ועקיפות, בצורות משפטיות ולא-משפטיות (אתיות) על התחום. ישנם רגולטורים שקביעותיהם עשויות להשפיע על פיתוח בינה מלאכותית באופן ישיר, כמו קבלת החלטות לגבי הגנת פרטיות שאמורה להיות מוטמעת במערכות בינה מלאכותית. לעומת זאת, ישנם רגולטורים שקביעותיהם עשויות להשפיע על סביבת ההטמעה של הבינה המלאכותית, כמו החלטות בדבר מקומות שמותר יהיה למכונות אוטונומיות לנסוע ומקומות שאסור. לענפי המשפט השונים יש גם יכולת להשפיע על התמריצים הפועלים על השחקנים השונים, על מנת שיוצרו מנגנוני רגולציה עצמית, בדרך של קביעת אחריות או פטור מאחריות בחקיקה. כך ניתן אף לחשוב על תמריצים כלכליים להטמעת האתיקה בכל שלבי פיתוח המוצר ואחזקתו, בכל שרשרת הערך.

תובנות אלה על תפקיד המשפט ויכולתו לעצב סביבה דיגיטלית וסביבת שוק מבוססות על הניסיון המצטבר ביחסי הגומלין בין המשפט והאינטרנט. כך למשל, לסעיף 230 בחוק ה- Communication Decency Act בארה"ב, שקבע פטור למתווכים על מידע המועבר דרכם, הייתה השפעה מכרעת על עלייתן ושגשוגן של פלטפורמות חדשות. מנגד החקיקה הזו אפשרה לפלטפורמות אלה לקבוע בעצמן, באמצעים חוזיים בדרך כלל, את "כללי המשחק" הקשורים בהעברת מידע דרכן.

בהקשרי בינה מלאכותית, המגמה של האיחוד האירופי⁵⁹ היא להתאים כללי אחריות של המשפט הפרטי למציאות של מוצרים ושירותים מבוססי בינה מלאכותית, על בסיס הדגם של "אחריות מוחלטת". דגם ה"אחריות המוחלטת" משקף מודל התערבות משפטית אקטיבית, בו יש הטלת האחריות על היצרן או המשתמש לפי ההקשר, אלא אם הם יוכיחו את התקיימותם של תנאים מסוימים אשר יסירו מהם אחריות, או תנאים לקבלת פטור מאחריות.

בהמשך לכך, ניתן להעלות מודל ביניים של אסדרה משפטית המשלב גם היבטים אתיים, כלומר להטיל אחריות כל עוד אין לא עומדים בסטנדרטים מקובלים בתעשייה, אותם הרגולטור יצטרך לקבוע. מודל שכזה משלב בין הרצון להטיל אחריות למנוע נזק על הגורם היעיל, לבין החשש מפני אפקט מצנן על חדשנות. ניתן אף לחשוב על תמריץ בדרך של "מקל וגזר" כפי שנעשה בתחום השמירה על איכות הסביבה – הקלה בעונשים אם מאמצים קוד אתי בתחום והחמרה בעונשים אם לא אימצו קוד אתי כאמור⁶⁰.

לסיכום, הוועדה זיהתה כי קיימות מספר אפשרויות של אסדרה להתמודדות עם אתגרים של בינה מלאכותית (ראו טבלה 3). אפשרויות אלה יכולות להיות משפטיות או חוץ-משפטיות:

1. אסדרה באמצעות חקיקה באמצעות דין ייעודי – כלומר טיפול בחקיקה שקשורה להיבטי בינה מלאכותית, כגון עבירות כנגד "מחשב" בחוק המחשבים, שנועדו להגן עליו ועל המידע שבו, או חוק אחריות למוצרים פגומים וכד'.
2. אסדרה באמצעות פיתוח שיפוטי למשל של דיני הנזיקין או דיני חוזים.
3. אסדרה באמצעות תקינה מקצועית – אסדרה באמצעות מוסדות תקינה המשלבים נציגים מהממשלה, האקדמיה והחברה האזרחית.
4. אסדרה-עצמית לא משפטית – באמצעות כללים אתיים או סטנדרטים מקצועיים המפותחים בדרך כלל בידי הקהילה המקצועית הרלבנטית. למשל הקמת ועדות מייעצות ומפקחות (review boards) פנימיים, ציבוריים, משולבים עם נציגי אקדמיה, ממשל, תעשייה וחברה אזרחית שיכולים לתת החלטות מקדמיות בתהליך הפיתוח וההטמעה.

לכל אחת משיטות האסדרה יתרונות וחסרונות, באיזון שבין ניהול סיכונים, ודאות, תופעות לוואי, גמישות וחדשנות. איור 1 מתאר לדעת הוועדה את ההתאמה האופטימלית של אפשרויות אסדרה לפי אזורי הסיכון של המוצרים או השירותים מבוססי הבינה המלאכותית. כמו כן, טבלה 3, נועדה להוסיף מידע לגבי היתרונות והחסרונות של סוגי האסדרה השונים, ונועד לאפשר למקבלי החלטות לייצר דיון פורה לגבי סוגי האסדרה של בינה מלאכותית בגופי מטה הממשלה ובגופי הרגולציה.

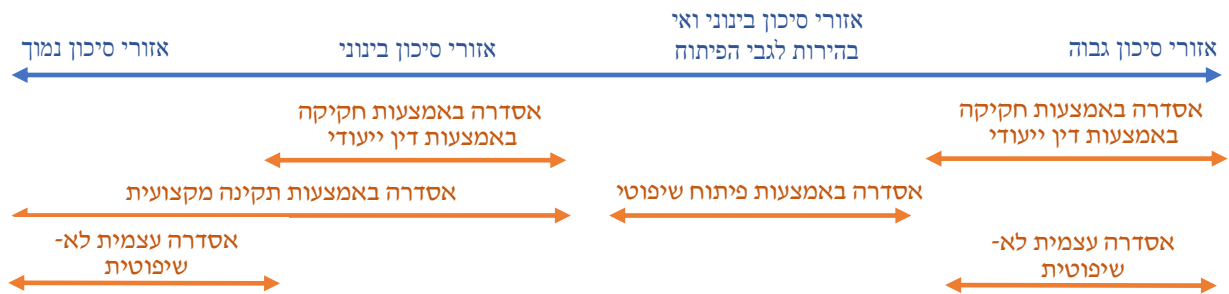
⁵⁹ European Commission, 2018, Liability for Emerging Digital Technologies, SWD (2018) 137 final - https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51633

⁶⁰ ראו למשל בהקשר זה את ה- US Federal Sentencing Guidelines for Organizations - <https://www.uscc.gov/guidelines/organizational-guidelines>

טבלה 3 מתמקדת בסוגי האסדרה לאור השאלה 'מה מסדירים'. אולם יש לקחת בחשבון גם את ההיבט המוסדי. אסדרה באמצעות חקיקה באמצעות דין ייעודי יכולה להתבצע על ידי (או בלעדיו) הקמת רשות ייעודית בחוק, שמומחית בתחום הבינה המלאכותית. יתרונות של רשות שכזו תהיה יכולת ההתמחות הגבוהה, נקודת מענה מרכזית ומהירה לפניות ולאתגרים השונים. מאידך, הקמת רשות ייעודית עלולה לגרום לרגולציה יתר ולפגיעה בחדשנות.

חשוב לציין כי ברוב המקרים אסדרה נעשית על ידי מגוון כלים ושיטות בו-זמנית. למשל, בנושא הפרטיות, קיימת רשות להגנת הפרטיות לצד חקיקה ראשית, חקיקת משנה, פיתוח שיפוטי ותקינה.

איור 1 – אפשרויות לאסדרה של בינה מלאכותית והתאמה לאזורי סיכון שונים



טבלה מס' 3 – אפשרויות אסדרה בתחום הבינה המלאכותית

סוג האסדרה	מאפיינים	יתרונות	חסרונות	המלצות הוועדה
אסדרה באמצעות חקיקה באמצעות דין ייעודי	חוק ייעודי או תיקון לחוקים אחרים שנאכף בידי רשות מדינתית קיימת או בידי גורמים פרטים	<ul style="list-style-type: none"> מגביר את הבהירות לגבי הערכים המוגנים מאפשר פיתוח שיפוטי קונקרטי על בסיסי קווים מנחים של המחוקק גמישות חלקית 	<ul style="list-style-type: none"> העדר התמחות מקצועית בארגון אחד אכיפה לאחר מעשה בלבד. עלול להגביר את חוסר ודאות לא מתערב במערכת יחסי הכוחות הקיימים שבו יתכן יתרון לגורמים מסוימים 	מתאים בעיקר לאזורי סיכון בינוני וגבוה
אסדרה באמצעות פיתוח שיפוטי	ללא חוק ספציפי	<ul style="list-style-type: none"> ללא חיכוך רגולטורי או משפטי ישיר גמישות מאפשר פיתוח שיפוטי 	<ul style="list-style-type: none"> בדרך כלל פיתוח שיפוטי יהיה במקרים מובהקים יותר של נזק – ולכן לא ברור שיאפשר מענה הולם לכל טווח הסיכונים לנזק העדר התמחות מקצועית בארגון אחד חוסר ודאות יתרון לשחקנים חזקים 	מתאים למצבים של סיכון בינוני ואי בהירות לגבי כיווני הפיתוח
אסדרה באמצעות תקינה מקצועית או של קהילה מקצועית/אתית	מאפשר ייבוא בידי המשפט בעתיד	<ul style="list-style-type: none"> גמישות לגיטימיות גבוהה בקהילה המקצועית תהליך שיתופי 	<ul style="list-style-type: none"> סיכון להדרה של המשפט וערכיו תלות במשפט לקבלת תוקף מחייב, פיקוח ואכיפה יתרון לשחקנים חזקים 	מתאים לאזורי סיכון בינוני ונמוך, ומתאים כמסגרת לפיתוח ובחינה של החלת העקרונות האתיים
אסדרה לא-משפטית	ללא נורמה משפטית (למשל החלת עקרונות אתיים)	<ul style="list-style-type: none"> גמישות לגיטימיות גבוהה בקהילה המקצועית 	<ul style="list-style-type: none"> סיכון להדרה של המשפט וערכי היסוד שלו (שוויון, הגינות, זכויות אדם) תלות בקהילה המקצועית לפיתוח חסר מנגנון אכיפה אמין יתרון לשחקנים חזקים 	מתאים לאזורי סיכון נמוך, שם מספיקה אסדרה לא-משפטית, ואזורי סיכון גבוה, שם ההתפתחות הטכנולוגית מהירה יחסית לאפיק המשפטי

בפרק 2 הציעה הוועדה כלי לקבלת החלטות בתחום האתי. במסגרת הכלי, הציעה הוועדה סדרה של שאלות למקבלי ההחלטות (ראה סעיף 2.2.1) ופרמטרים אותם זיהתה הוועדה כחשובים להילקח בתחום הבינה המלאכותית. פרמטרים אלה כוללים את: עוצמת הפגיעה הפוטנציאלית בפרט, עוצמת הפגיעה הפוטנציאלית בציבור, השפעה על הרצאת משאבים, השפעה על תודעה, עוצמת הפגיעה במקרה שיעשה שימוש לרעה או שהמוצר יצא משליטה, גיוון צוות המחקר והפיתוח, ועד כמה מהר אפשר לאתר תקלות. לפרמטרים אלה, מעבר להתייחסות האתית שעל מקבלי החלטות לתת את הדעת, יש השפעה על ההחלטה באיזו סוג אסדרה לנקוט.

לצד פרמטרים אלה, המצדיקים גישה מתערבת יותר של אסדרה, ציינו חברי הצוות את החשיבות של העלאת מודעות הציבור באופן כללי להיבטים החיוביים והמאתגרים, כאחד, של בינה מלאכותית וחינוך לאתיקה, שהם לא בסמכות ועדה זו.

3.2 עקרונות פעולה כללים המוצעים על ידי הוועדה

הוועדה מצאה לכונן להציע מספר עקרונות פעולה מנחים, נוכח מאפייניה הייחודיים של טכנולוגיית הבינה המלאכותית ויישומיה החברתיים אותם יש לכלול במסגרת פיתוח המודל האסדרתי.

א. התאמה לחקיקה ולתקינה בינלאומית וקידום מדיניות ישראלית בזירות הבינלאומיות

יש להדגיש כי כל מודל מוצע בתחום מחייב התאמה לחקיקה ולמדיניות המקובלת במדינות המפותחות, על מנת שישראל תוכל להישאר בחוד החנית של פיתוח הבינה המלאכותית וכדי לוודא שהרגולטור הישראלי לא מערים קשיים על מרכזי הפיתוח והתעשייה המקומיים, ביחס למדינות המפותחות בעולם. בתוך כך יש להתייחס לא רק למדינות האחרות ולמדיניות הנהוגה בהן, אלא אף להחלטות האזוריות ולהחלטות הגופים הבינלאומיים בתחום ובכלל זה בהחלטות ה-OECD והחלטות ה-The World Economic Forum (WEF).⁶¹ במסגרת מימוש עיקרון זה נדרשת הממשלה גם לקדם באופן פרואקטיבי את האינטרסים והערכים החשובים לקידום המדיניות הישראלית בזירות הבינלאומיות בהן מתנהל שיח זה, כדי להשפיע על ההסדרים טרם שהתגבשו.

יתר על כן, אם מדינת ישראל רוצה למצב את עצמה כמובילת דעה בזירה הבינלאומית, הוועדה ממליצה לייצר עמדה ממשלתית ברורה שתוצג בבמות הבינלאומיות החשובות. במזכר עבור הוועדה שנכתב על ידי ד"ר רועי שיינדרוף, המשנה ליועץ המשפטי לממשלה (משפט בין-לאומי) במשרד המשפטים, מופו הזירות הבינלאומיות החשובות העוסקות בתחום זה. נראה כי נכון יהיה לייצר עמדה ברורה ואחידה של מדינת ישראל, שתוצג בבמות אלה כאמור.⁶²

ב. מיפוי השחקנים השונים בתחום על מנת ליצור מסגרת מותאמת של אחריות ומערכת תמריצים

קידום ומימוש של מדיניות רגולטורית מבוסס גם על התאמה לתמריצים השונים שיש השחקנים הפועלים בשדה המוסדר. בהתאם לכך, הוועדה ממליצה למפות את השחקנים השונים המצויים בשרשרת הערך ובאקוסיסטם של המוצרים והשירותים בתחום זה, את התמריצים שלהם להפנים את העקרונות האתיים והמהותיים שתוארו בחלק הקודם. במיפוי יש לכלול גם את גורמי המחקר הבסיסי, המחקר היישומי, המחקר התעשייתי, הגורמים המיישמים וכדו'.

כפועל יוצא מכך ניתן יהיה לבחון את האחריות להטמעת ולמימוש העקרונות המנחים של כל שחקן. מוצע כי האחריות תקבע בהתאם לתפקיד היחסי בשרשרת ותוך הבנה של משמעות הטלת האחריות אל מול היכולת לעשות שימוש בטכנולוגיית הבינה המלאכותית באופן יעיל חברתית. בהקשר זה יודגש כי קידום תמריצים לשחקנים השונים יכול להיעשות בכל אחד מהמודלים האסדרתיים שתוארו בטבלה מס' 3, כלומר בדרך של הטלת אחריות, הבהרת אחריות, הטלת חובת ביטוח או אסדרה מראש. נראה

⁶¹ כך למשל, מכח החלטת ממשלה 4481, רשות החדשנות משתפת פעולה עם ה-WEF ומשתתפת ברשת בינלאומית לרגולציה של טכנולוגיות חדשניות (בין השאר, בינה מלאכותית).

⁶² ד"ר רועי שיינדרוף, 2019, גיבוש אסטרטגיה ישראלית בתחום הבינה המלאכותית – היבטי המשפט הבין-לאומי, מזכר מטעם משרד המשפטים שהוגש לעיון הוועדה.

כי יש מקום לבחון האם הדין האזרחי מספק מענה מעודכן לסוגי הסיכונים והפעילויות הקשורות בבינה מלאכותית.

ג. עיקרון האחראיות (accountability) והתאמה שלו לדינאמיות של תחום הבינה המלאכותית

רגולציה מודרנית של טכנולוגיה (למשל רגולציית הפרטיות) מביאה בחשבון את האתגר של בהתאמת מסגרת ערכית להקשר טכנולוגי וכלכלי מסוים. עקב כך עיקרון מנחה חשוב בדיני הפרטיות המודרניים הינו עיקרון האחראיות – אחראיות ההנהלה לבחון באופן קונקרטי את הסיכונים הנובעים לערך המוגן מהפעילות המוצעת, ולהתאים לה מסגרת ניהול סיכונים. יישום תפיסה זו בתחום הבינה המלאכותית מחייב להתמודד עם המאפיינים הייחודיים של הטכנולוגיה, ובהם הקושי לצפות במדויק את האופן שבו היא תפעל בנסיבות מסוימות, ולהסביר במדויק את האופן שבו היא פועלת.

לצורך יישום עיקרון האחראיות בהקשרי בינה מלאכותית, נראה כי יש להוסיף עיקרון רגולטורי המחייב את הארגון המיישם טכנולוגית בינה מלאכותית לקיים "מרחב ניסוי" ו- "מעגלי בקרה" בטרם מימוש בפועל של הטכנולוגיה. מובן כי רמת ועומק הניסוי הנדרש הינו בזיקה לעקרונות המהותיים שתוארו לעיל ולהיקף ההשפעה המשוערת של הטכנולוגיה⁶³.

ד. קידום בהירות נורמטיבית בשלבים קריטיים של שרשרת הערך של מוצרי בינה מלאכותית

השלבים הראשוניים של פיתוח מערכת הבינה מלאכותית - הבנת הצורך העסקי, איסוף הנתונים, ארגון הנתונים, בנייה והערכת המודל, הפצה ומעקב אחרי ביצועים - יכולים להיות משמעותיים לקידום הערכים והטמעת העקרונות. על כן, הוועדה סבורה כי קידום רגולציה המסייעת בשלבים אלה יכולה לסייע רבות בהטמעת העקרונות והערכים כאמור. ראינו כי קיים חסר במסדי נתונים מאוזנים לשימוש פתוח, על מנת לאפשר גם לחברות קטנות להצליח בתחום.

ה. הצורך במעגל בחינה קבוע של המדיניות הרגולטורית בידי הרגולטור

נוכח מאפייניו של תחום זה, ניתן להניח כי מדיניות רגולטורית ובחינה שנעשית במציאות טכנולוגית מסוימת תדרוש עדכון ובחינה בפרקי זמן קצרים יותר ממחזור הבחינה הרגיל של רגולציה. בנוסף, מוצע כי רשויות האסדרה ישלבו עיקרון תהליכי שלפיו יתאפשר ניסוי מבוקר של המדיניות הרגולטורית המיושמת בהקשר מסוים, בין היתר בשים לב למידת חדשנותו או הסיכונים הקשורים במימושו. בנוסף, עיקרון זה מאפשר להתמודד עם האתגר של איזון בין הרצון לחדשנות לבין חוסר הוודאות לגבי השפעתם על היבטים אתיים.

ו. בחינת שימוש בתפיסת סביבה רגולטורית מבוקרת (regulatory sandbox) –

הרעיון של יצירת מרחב ניסוי ליישומים יותר חדשניים ועם הסתברות לסיכון גבוהה, בתנאים מבוקרים צובר תאוצה. בשנים האחרונות במסגרת הניסיון לכייל את המדיניות הרגולטורית באופן שהיא אפקטיבית חברתית מבלי לפגוע בחדשנות, פותחה תפיסה של "ארגז חול" רגולטורי, המאפשרת לגורמים הפועלים בתוך "ארגז החול", לבחון בתנאים מבוקרים פעילויות חדשות, שעלולות להיות מפרות דין. לתפיסה זו ערך רב בתחום הבינה המלאכותית בשל הרצון לאפשר חדשנות מצד אחד והתמודדות עם סיכונים לא צפויים לאינטרסים חברתיים מצד שני. הוועדה ממליצה לייצר רגולציה מתאימה כדי לאפשר סביבה כזו.

על כן, בהתאם למדיניות הממשלתית הישראלית בנושא זה, ובכפוף לה, ובהמשך להמלצות דומות המתגבשות ב-OECD, הוועדה ממליצה לבחון את האפשרות לשימוש בתפיסה זו גם בתחום זה.

⁶³ ראו בעניין זה את הצו הנשיאותי האמריקאי מפברואר 2019, לגבי הצורך בקידום תקנים טכניים לבחינה ופריסה בטוחה של טכנולוגיות בינה מלאכותית <https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/>

ז. הממשק בין עקרונות הפעולה המוצעים לבין התשתית הרגולטורית הקיימת

במסגרת הפרק הראשון והשני נדונו ההיבטים הייחודיים של מערכות הבינה מלאכותית. מאחר שהטכנולוגיה היא טכנולוגיה כללית, יישומיה והסיכונים הנובעים ממנה הם תלויי הקשר. תחומים רבים מוסדרים כבר ברגולציה קיימת. על כן, ישנה חשיבות להבין את השיקולים, הערכים, האינטרסים והתועלות החברתיות האפשריות באותם התחומים, בטרם יוחלט על רגולציה חדשה או מותאמת. מאפיינים אלה אף משפיעים על התוצאה הערכית המבוקשת ועל הדרך לממש אותה.

נוכח מאפייני מערכות הבינה מלאכותית, ולאחר שהוצגו השלבים הקריטיים בפיתוח המערכת מבחינת הטמעת הערכים והעקרונות, הוועדה ממליצה להתחיל באיתור הרשויות המסדירות את נושא משאבי מידע המשמשים לעיבודי טכנולוגית בינה מלאכותית, וכן את אותן רשויות אסדרה האחראיות על שוק המוצר שייווצר. איתור רשויות אסדרה אלה יסייע בהסרת חסמים לשימוש בבינה מלאכותית ולהתאמה של המעטפת הרגולטורית לסיכונים הקיימים מהמערכת.

בפרט, כל רשות אמונה לבחון את השלכות מערכות הבינה מלאכותית על התחום אותו היא מסדירה. כך, למשל, משרד הבריאות אמון על בחינת תחום הבריאות ושילוב בינה מלאכותית בתוך המערך על כל ההשלכות הקשורות בכך. מן הראוי שמערך הסייבר יבחן את השלכות שילובה של מערכות בינה מלאכותית הן בהגנה על מדינת ישראל והן בהתמודדות אל מול שילוב מערכות בינה מלאכותית במדינות אחרות.

ח. היבטים מוסדיים רוחביים קיימים – הרשות להגנת הפרטיות

כפי שתואר בפרק הקודם, דיני הגנת הפרטיות במידע מסדירים את השימוש במידע אישי, ונועדו להגן על האוטונומיה של הפרט. עקב כך, לרשות להגנת הפרטיות תפקיד רוחבי כללי ותשתיתי בכל הקשור להסדרת השימושים במידע אישי בהקשרי בינה מלאכותית. כפועל יוצא מכך:

1. הוועדה ממליצה כי בכל הקשור לשימוש בינה מלאכותית הכרוך בעיבוד מידע אישי, או שיש לו השלכות על קבלת החלטות מבוססות מידע אישי המוסדרות באמצעות דיני הפרטיות, רשות האסדרה המובילה תהיה רשות הגנת הפרטיות, אשר תפעל בתיאום עם רשויות האסדרה הייעודיות, ככל שיהיו.

2. בהתאם לכך הוועדה ממליצה כי רשות הגנת הפרטיות תבחן את העקרונות המנחים שהוצגו לעיל ותגבש תכנית מימוש בתחומי המידע האישי בהקשר של בינה מלאכותית.

3. בפרט, נראה כי יש לוודא כי לרשות הגנת הפרטיות משאבים טכנולוגיים וחומריים לפתח מסגרת טכנולוגית ומשפטית מעודכנת לתחום אנונימיזציה של מידע. היכולת לבצע אנונימיזציה של מידע אישי, ברמת סמך סבירה, היא יכולת תשתיתית לפיתוח וקידום בינה מלאכותית⁶⁴.

ט. היבטים מוסדיים רוחביים קיימים – רשות התחרות (לשעבר הרשות להגבלים עסקיים)

טכנולוגית בינה מלאכותית מתוארת בהקשרים מסוימים כטכנולוגיה תשתיתית, המשפיעה או מכתיבה את אופן הפיתוח והשימוש בטכנולוגיות נוספות המבוססות עליה. במובנים אלה חלק מהפיתוח והיישום של טכנולוגית בינה מלאכותית במקום שבו משאבי המידע הנדרשים לפיתוח הטכנולוגיה או אופן המימוש של הטכנולוגיה מצויים בידי בעלי "כוח השוק", עלולים להביא לפגיעה בתחרות ההוגנת ובמקסום הערך של טכנולוגיה זו לחברה.

הדברים נכונים במיוחד לפלטפורמות מידע בעלות כוח שוק הנובע ממעמדן בכלכלת הרשת בשווקים "דו צדדיים". בהקשר זה נזכיר כי הוועדה מיפתה את נושא קיום השוק התחרותי כאחד מהערכים האתיים בדו"ח זה (ראה סעיף 2.1.6). בהתאם לכך, הוועדה ממליצה כי הרשות לתחרות תגבש דרכי התמודדות שמטרתן שמירה על תחרות הוגנת בתחום; ושמירה על צרכנים ועל נגישות הטכנולוגיה; ומניעת גלגול סיכוני הטכנולוגיה או עלויותיה באופן שאינו יעיל לשחקנים חלשים יותר במורד "שרשרת

⁶⁴ Polonetsky, Jules and Tene, Omer and Selinger, Evan, 2018, Consumer Privacy and the Future of Society, In: The Cambridge Handbook of Consumer Privacy, Eds. Evan Selinger, Jules Polonetsky and Omer Tene, <https://ssrn.com/abstract=3158885>

הערך". מן הראוי לציין, כי ככל הידוע לוועדה, רשות התחרות בוחנת בעת כתיבת הדוח את הצורך בגיבוש כלי התערבות רגולטורים בתחום המידע והטכנולוגיה⁶⁵.

י. הצורך במנגנון תיאום ממשלתי על-משרדי

מכיוון שטכנולוגיות הבינה המלאכותית תשפענה רבות על היבטים שונים של תחום האסדרה, הוועדה ממליצה להקים מנגנון תיאום פנים ממשלתי על משרדי. כל זאת, על מנת ליצור מדיניות אחידה, ברורה וקוהרנטית, בין כלל משרדי הממשלה. מכיוון שלכל החלטה בתחום הבינה המלאכותית השפעות רוחביות מיידיות, יש לבחון את כלל ההיבטים הרלוונטיים של כל החלטה על הרגולטורים השונים ועל החברה בכללותה.

בנוסף לתיאום הכולל והמנחה, אנו צופים שיעלה הצורך במקרים מסוימים בתיאום בפרויקטים רוחביים חוצי-משרדים, למשל בתחבורה חכמה, המשלבת את משרד התחבורה, משרד הפנים, ביטחון פנים וכדומה. בהתאם לכך מוצע להקים מנגנון תיאום ממשלתי בפורמט אפקטיבי כדי להעביר מידע מקצועי ורגולטורי בין הגורמים השונים ולאפשר להם קבלת החלטות מבוססת תמונה רחבה יותר.

בנוסף הוועדה ממליצה כי יוקם מנגנון תיאום ומוקד ידע ממשלתי כדי לסייע לממשלה ולרשויות האסדרה בקידום הנושא באופן קוהרנטי. בדיוני הוועדה עלה חשש מפני אי תיאום או סתירות שעלולות להיגרם כפועל יוצא של ההצעה להתבסס על רשויות האסדרה המגזריות. הוועדה לא דנה במאפייני המנגנון אלא ממליצה על הצורך בקיום פונקציה כזו במסגרת ההיערכות האסדרתית.

יא. רשויות האחריות על משאבי מידע

רשויות שיש להן אחריות בתחום משאבי מידע המשמשים לעיבודי טכנולוגיית בינה מלאכותית, מאסדרות את המשאב הבסיסי המניע את היכולת הטכנולוגית. לרשויות אלה תפקיד מפתח בבחינה האם המסגרת האסדרתית שהם מפעילים מתאימה להשגת התועלת החברתית המצרפית בשדה זה, תוך שמירה על שוק תחרות הוגן וחופשי ותוך שמירה על זכויות אדם שונות. במסגרת זו יש לתת את הדעת לא רק לסיכונים אלא גם למרחבי החדשנות ולאופן שבו התרת שימושים וקביעת אפשרויות שימוש אחריות במשאבי מידע יקדמו אינטרסים חברתיים.

הוועדה ממליצה כי רשויות האחריות על תחומי פעילויות המושפעים מהתוצרים של עיבוד משאבי מידע, יידרשו לבצע בחינה בהתאם לעקרונות העל. עליהן לבחון האם בעת פריסה של טכנולוגיות בינה מלאכותית או שימוש בהם בתחומי הפעילות המאוסדרים על ידם, יש צורך בהתאמה של המסגרת החלה כדי לקדם את ההגנה על האינטרסים המאוסדרים.

⁶⁵ ראו את קריאת הרשות לחדשנות <http://www.antitrust.gov.il/subject/166/item/35243.aspx>

נספח א' – הרחבה על המצב בארה"ב ובאירופה

בארצות הברית פורסם ב-11 בפברואר 2019 צו נשיאותי שנועד לשמר את ארצות הברית כמובילה בתחום הבינה מלאכותית⁶⁶. בצו מציינים את חשיבותה של טכנולוגיית הבינה מלאכותית בצמיחה הכלכלית של ארצות הברית, בביטחון הלאומי ובשיפור איכות החיים. בצו מציינים הנשיא כי ארצות הברית חייבת לקדם את פריצות הדרך הטכנולוגיות בתחום הבינה מלאכותית בקרב הממשל הפדרלי, התעשייה והאקדמיה. בנוסף צוין כי ארצות הברית מחויבת להוביל בפיתוח סטנדרטים טכנולוגיים מתאימים ולהפחית את החסמים לבדיקת ולהטמעת טכנולוגיית בינה מלאכותית באופן בטוח. כל זאת, על מנת לאפשר לתעשייה בתחום ליצור טכנולוגיה מבוססת בינה מלאכותית ולחברות קיימות לאמץ טכנולוגיה כזו. כך גם הובהר בצו כי ארצות הברית תקדם סביבה בינלאומית התומכת במחקר בינה מלאכותית אמריקאי וחדשנות ופותחת שווקים לתעשיות בינה מלאכותית אמריקאיות, תוך שמירה על היתרון הטכנולוגי שלה בבינה מלאכותית והגנה על טכנולוגיות הבינה מלאכותית הקריטיות שלה מפני רכישה על ידי מתחרים אסטרטגיים ואומות יריבות⁶⁷.

בנוסף, פורסם בארצות הברית מזכר בסוף חודש יולי 2018, בו מוצהר כי יש לדאוג לכך שארצות הברית תישאר מובילה בתחום הבינה מלאכותית. כל זאת משום שהדבר חשוב לביטחון הלאומי וכן לתחרות הכלכלית. על כן, על רשויות הממשל להשקיע בתחום הבינה מלאכותית על כל גווניו – לרבות למידת מכונה, מערכות אוטונומיות ויישומים בחזית של הטכנולוגיה האנושית. המזכר מציינים כי על רשויות הממשל לתעדף באופן כללי את ההשקעה בתחום על מנת לשמור את ארצות הברית בחזית הפיתוחים בתחום, ובאופן ייחודי על ארה"ב לתעדף מחקר ופיתוח בתחום המדע של הידע הקוונטי (Quantum Information Science)⁶⁸.

בצרפת הוגש, דוח בשנת 2018, בו הציעו ליצור מדיניות המותאמת לצרפת שאינה מועתקת ממדינות אחרות. הדוח המליץ להתמקד בארבעה תחומים (סקטורים) אסטרטגיים ולתת את הדעת תחילה למדיניות בינה מלאכותית בהם – בריאות, סביבה, תחבורה ותעשייה ביטחונית-הגנתית. בתוך כך, מציע הדוח ליצור פלטפורמות לכל סקטור כך שניתן יהיה לשתף את השחקנים השונים באותו תחום (תעשייה, ממשל, אקדמיה וכיו"ב) במידע ובהתפתחויות שונות בתחום. כמו כן ציינו בדוח כי יש לחשוב על פיתוח של טכנולוגיית בינה מלאכותית ירוקה וידידותית לסביבה. על מנת לאפשר דיון בסוגיות אתיות עתידיות, שלא חשבו עליהן עוד, הומלץ להקים ועדת אתיקה עצמאית שתוכל לבחון את כל ההשלכות הממשלתיות והפרטיות של טכנולוגיית הבינה מלאכותית וכן תוכל לדון בנושא מסוים או באופן כללי וליצור סטנדרטים בתחום לטובת פיתוח בינה מלאכותית בצורה אתית⁶⁹.

באיחוד האירופאי קבוצת המומחים HLEG, פרסמה את הדוח של כללים אתיים לבינה מלאכותית. התייחסנו לדוח זה במסמך, אולם היינו רוצים להוסיף. הדוח מציינים כי יש לשלב את הדרישות האתיות מבינה מלאכותית כבר בשלבים ההתחלתיים, לרבות בבחירת הצוות שיפתח את הטכנולוגיה, ובין היתר לדאוג גם לכך שהטכנולוגיה תסייע לכולם ללא אפליה, תוכל להסביר את פעילותה ולתת דין וחשבון עליה, לא תפלה לרעה, תכבד את האוטונומיה האנושית, תגן על הפרטיות, תהיה שקופה ובטוחה. יש להטמיע את הקוד האתי בתוך התרבות הארגונית של החברות בעת פיתוח המוצרים, הטמעתם והשימוש בהם וכן לדווח על כך לבעלי העניין. כך, יש לאפשר גיוון בצוותי הפיתוח, בהליך ההטמעה ובבחירת המוצר. לבסוף, יש לזכור כי ערכים אתיים עלולים לעיתים להתנגש ויש לתת לכך את הדעת גם כן. כך, למשל, שקיפות עלולה להביא לשימוש לרעה בזכות פרטיות. על כן, יש לתעד ולהסביר את ההחלטות המתקבלות תוך הסבר מהו הערך שייפגע על חשבון הערך האחר⁷⁰.

להרחבת הקריאה לגבי מקורות, דוחות ומאמרים בתחום ניתן לפנות לדוחות מוסד שמואל נאמן. בדוח הראשון למשל, בין היתר אוזכרו סינגפור, בריטניה, איחוד האמירויות הערביות, קנדה, יפן, הודו, האיחוד האירופאי ועוד. בדוח השני, למשל, אוזכרו האיחוד האירופאי, בריטניה, צרפת, סין, אינטל, IBM וגוגל כמו כן בפרויקט שיתוף פעולה בין התעשייה לאקדמיה, Future of Life⁷¹, ניתן למצוא מדיניות בתחום בינה

⁶⁶ <https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/>

⁶⁷ Ibid. Executive Order on Maintaining American Leadership in Artificial Intelligence, 2019.

⁶⁸ Mulvaney Mick and Kartsios Michael, 2018, Memorandum – FY2020 Administration Research and Development Budget Priorities, <https://www.whitehouse.gov/wp-content/uploads/2018/07/M-18-22.pdf>

⁶⁹ Ibid. Villani, 2018.

⁷⁰ Ibid. Ethical Guidelines for Trustworthy AI, 2019

⁷¹ <https://futureoflife.org/ai-policy>

מלאכותית מארצות שונות. ולבסוף, נציין כי בינואר 2019 נוסד כנס בינה מלאכותית, אתיקה וחברה בהוואי עם מחקרים מעודכנים בתחום⁷².

נספח ב' – מילון מושגים

בינה מלאכותית צרה (Narrow AI)

הכוונה לכלי בינה מלאכותית המסוגלים לפתור בעיות מסוימות או מגוון קטן של בעיות. כלים אלו יכולים להפגין יכולות דומות לאנושיות, או אפילו עולות עליהם, בעת ביצוע המשימות שנועדו לפתור אבל אין להן מגוון רחב של יכולות ובפרט אין להן מודעות עצמית, הכרה ומחשבה. לעומת בינה מלאכותית צרה, מערכות בינה מלאכותית רחבות הן מערכות המסוגלות להתמודד עם מגוון רחב של משימות, בדומה לבני אדם⁷³.

הטיה (Bias)

הטיה היא פגיעה או יתרון הניתן לקבוצת אוכלוסייה מסוימת בצורה לא-מידתית. מערכת מוטת מייצרת חסרון או יתרון לקבוצה אחת באוכלוסייה בצורה עקבית. בבחינת הטיות יש לפעמים דגש מיוחד על הגנה בפני הטיות כנגד קבוצות מוחלשות באוכלוסייה או קבוצות מוגנות.

הוגנות (Fairness)

הוגנות מייצגת מצב שבו יש חלוקה צודקת של משאבים, ללא הפליה כנגד (או בעד) אוכלוסיות מסוימות. הגדרת ההוגנות יכולה להשתנות בין חברה אחת למשניה וגם להשתנות במהלך הזמן עם התפתחות מושגי הצדק. לפיכך, הגדרת מושג ההוגנות נמצאת בדיון מתמיד.

למידת מכונה (Machine Learning)

למידת מכונה היא השיטה המובילה למימוש בינה מלאכותית. בשיטה זו, המכונה לומדת לבצע פעולה על-ידי התבוננות בדוגמאות של ביצוע המשימה. מתוך התבוננות זו, ועל-ידי ההערכה של מידת ההצלחה של כל ניסיון, המערכת מייצרת מדיניות על-פיה היא תתנהג בעת שהיא נדרשת לבצע פעולה. לצורך הדיון במסמך זה, למידה עמוקה (Deep Learning) היא דרך למימוש למידת מכונה.

⁷² <http://www.aies-conference.com/>

⁷³ להרחבה 9-10, Anastassia Lauterbach and Andrea Bonime-Blanc, 2018, The Artificial Intelligence Imperative, Praeger, p.