שרת החדשנות, המדע והטכנולוגיה

בס"ד

ירושלים, ה' בחשוון תשפ"ג
30 באוקטובר 2022

מדיניות רגולציה ואתיקה בתחום הבינה המלאכותית בישראל

1. אני מתכבדת לפרסם לשימוע פומבי ולהתייחסות הציבור עקרונות מדיניות בתחום הבינה המלאכותית (להלן גם: "מסמך המדיניות").

2. זוהי הפעם הראשונה בה מתפרסמים לשימוע ציבורי בישראל עקרונות מדיניות ממשלתיים, שינחו את דרך השימוש בטכנולוגית הבינה המלאכותית.

3. טכנולוגיה זו שינתה את חיינו ויצרה מפץ בעולם כולו, אולם לצד תרומתה המשמעותית, השימוש בבינה מלאכותית טומן בחובו אתגרים משמעותיים בהיבטים שונים, רבים מהם עבור המערכת הרגולטורית והמשפטית. הוא מעורר סיכונים וחששות שונים, החל מחשש לפגיעה בפרטיות ותקיפות סייבר וכלה באפליה פסולה בקבלת החלטות. כמו כן, הוא מעורר סוגיות משפטיות, רגולטוריות ואתיות מורכבות, כגון האחריות לתוצאות שנבעו מפעילות בינה מלאכותית, הצורך לגלות על שימוש בה או לתת הסבר להחלטותיה.

4. מסמך המדיניות מביא לידי ביטוי את התפיסה בה אני תומכת של "חדשנות אחראית": איזון בין הצורך להגן על הפרט בפני האתגרים הנובעים משימוש בטכנולוגיית הבינה המלאכותית, לבין האינטרס בקידום טכנולוגיה זו, תוך טיפוח סביבה עסקית מאפשרת ותומכת לתעשיית ההייטק והמחקר הישראליים.

5. עקרונות המדיניות המתפרסמים בזה לשימוע ציבורי גובשו, תוך התחשבות בעקרונות משפטיים ואתיים מקובלים בעולם והתאמתם לקונטקסט הישראלי. זאת, על מנת לעמוד בקצב ההתפתחות של טכנולוגיות הבינה המלאכותית ובמאפייניה הבינלאומיים, לצד המשך השמירה על מעמדה המוביל של ישראל בתחום הטכנולוגיה.

6. גיבוש מסמך מדיניות זה יסייע בקידום דירוגה הבינלאומי של ישראל בתחום הבינה המלאכותית, ויסמן מצפן לערכים שלאורם על טכנולוגיה זו להתפתח.

7. למסמך המדיניות מצורף נספח מקצועי מקיף, שהוכן על ידי גורמי המקצוע במשרד החדשנות, המדע והטכנולוגיה, ביחד עם מחלקת ייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים, אשר נרתמו למשימה חשובה זו, באופן הראוי לכל הערכה. הנספח המקצועי מניח סקירה נרחבת על תחום הבינה המלאכותית, מפרט את הממשק בין תחום זה לדין המצוי בישראל, סוקר את הרגולציה הבינלאומית ולבסוף מניח את המתודולוגיה לעקרונות המדיניות האסדרה והאתיקה המפורסמים בזה.

8. מסמך המדיניות המפורסם בזה, מהווה צעד משלים לצעדים נוספים שנעשו בממשלה בתחום הבינה המלאכותית בשנה האחרונה. קידום בניית תשתיות מיחשוביות, פרסום תכנית לאומית של משרד החדשנות בתחום הבינה המלאכותית ושיתופי פעולה שהותנעו לאחרונה בתחום הבינה המלאכותית עם ארה״ב וגופים נוספים.

9. בכוונתי להציג את מסמך המדיניות בכנס הקרוב של ארגון GPAI הבינלאומי בתחום הבינה המלאכותית, ארגון שרים אליו התקבלו מספר מצומצם של מדינות, שישראל הצטרפה אליו אך לאחרונה.

10. אבקש להודות למנכ״לית המשרד הילה חדד ובאופן מיוחד לעו״ד מאיר לוין, המשנה ליועץ המשפטי לממשלה, על כך שרתם את המחלקה הכלכלית במשרד המשפטים לעבודה משותפת ומאומצת עם צוות משרדי, מתוך הבנת חשיבותה. בפרט, תודות למשנה למנכ״לית משרד החדשנות, המדע והטכנולוגיה מר תום דן שריכז את העבודה על המסמך בצוותא עם עו״ד יוסף גדליהו מייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים, ליועמ״ש משרד החדשנות, המדע והטכנולוגיה (בפועל) עו״ד דני חורין, לעו״ד עמית אשכנזי על סיוע משמעותי בכתיבת המסמך, לעו״ד גאיה הררי הייט מלשכתי וכן לעו״ד ד״ר יובל רויטמן ראש אשכול רגולציה, ולמר גל קורן מהמחלקה הכלכלית בייעוץ וחקיקה על עבודתם המקצועית והמעמיקה. תודתי שלוחה גם לשותפים הנוספים שסייעו לעבודה בשני המשרדים.

11. מסמך המדיניות המצורף בזה הוא בסיס חשוב ראשון, והוא אף מניע תהליכי המשך חשובים, לרבות גיבוש כלי לניהול סיכונים והקמת מוקד ידע ממשלתי לתחום הבינה המלאכותית שיסייע לרגולטורים ולתעשייה.

12. אני מזמינה את הציבור, ובכלל זה אקדמיה, התעשייה וארגוני חברה אזרחית, להעיר הערותיו למסמך.

13. מצורף בזה מסמך המדיניות מטעמי ונספח מקצועי נלווה, לשימוע ציבורי .

אורית פרקש הכהן

שרת החדשנות, המדע והטכנולוגיה

ירושלים, ה׳ בחשוון תשפ״ג
30 באוקטובר 2022

עקרונות מדיניות רגולציה ואתיקה בבינה מלאכותית – שימוע ציבורי

בהמשך להחלטת ממשלה מספר 212 מיום 1.8.21 ״תכנית לקידום חדשנות, עידוד צמיחת ענף ההייטק וחיזוק המובילות הטכנולוגית והמדעית״ (להלן - ״החלטה 212״), שהטילה עליי להוביל את מדיניות הממשלה בתחום הבינה המלאכותית בנושאי רגולציה, מדיניות מידע ונתונים, אתיקה, שיתוף פעולה בין-לאומי אזרחי והטמעה במגזר הציבורי ולגבש תכנית לאומית, ובהמשך לעבודה מקצועית שנערכה בהתאם להנחייתי במשותף עם מחלקת ייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים, המצורפת למסמך עקרונות זה כנספח מקצועי, הריני מפרסמת להערות הציבור את עקרונות המדיניות, האתיקה והרגולציה הממשלתית שבכוונתי לקבוע ביחס לטכנולוגיות בינה מלאכותית.

ואלה עקרונות הרגולציה והאתיקה בתחום הבינה המלאכותית:

א. קביעת המדיניות בתחומי הבינה המלאכותית בשלב זה, תיעשה שלא באמצעות חקיקה רוחבית. בהתאם לתפיסה הנוהגת במדינות רבות, ובהתאמה לשלב הראשוני בה מצויה הטכנולוגיה, אין מקום בעת הזו לקידום רגולציה באמצעות חקיקה רוחבית שהיא ייחודית לתחום הבינה המלאכותית. יוער כי ייתכן ומדיניות זו תשתנה ככל שתחום הבינה המלאכותית יתפתח ויתבסס ומשמעויותיו תתבהרנה.

ב. אימוץ עקרונות אתיים לשימוש בבינה מלאכותית:

1. **בינה מלאכותית תשמש לקידום צמיחה, פיתוח בר קיימא והמובילות הישראלית בתחום החדשנות** - שימוש אחראי בבינה מלאכותית מהימנה הוא אמצעי לעודד צמיחה, פיתוח בר-קיימא, רווחה חברתית וקידום המובילות הישראלית בתחום החדשנות.
2. **האדם במרכז: כיבוד זכויות יסוד ואינטרסים ציבוריים** - פיתוח בינה מלאכותית, או שימוש בה, ייעשו תוך כיבוד שלטון החוק, זכויות יסוד ואינטרסים ציבורים ובפרט תוך שמירה על כבוד האדם ופרטיות.
3. **שוויון ומניעת אפליה פסולה** - בפיתוח ושימוש בבינה מלאכותית יובאו בחשבון הצורך בשוויון, גיוון, החשש להטיה במערכות בינה מלאכותית והסיכון לאפליה פסולה כנגד יחידים או קבוצות.
4. **שקיפות והסברתיות** - בפיתוח בינה מלאכותית ושימוש בה, יובאו בחשבון הצורך ליידע את מי שבא במגע עם בינה מלאכותית או מושפע מפעילותה והצורך במתן הסבר להחלטתה או לאופן שבו פעלה, בין היתר בשים לב למידת השפעתה, השלכותיה על מי שמושפע ממנה והאפשרויות הטכנולוגיות הזמינות.

5. **אמינות, עמידות אבטחה ובטיחות** - בפיתוח בינה מלאכותית ושימוש בה, יובא בחשבון הצורך בכך שמערכות בינה מלאכותית תהיינה אמינות, מאובטחות בטיחותיות לאורך כל מחזור חייהן, כך שבתנאי שימוש רגילים, שימוש צפוי או שימוש שגוי או תנאים מסוכנים אחרים - יפעלו כראוי ולא יהוו סיכון בטיחותי בלתי סביר. לשם כך יש לנקוט אמצעים סבירים בהתאם לתפיסות מקצועיות מקובלות לצמצם סיכוני בטיחות ואבטחת מידע בכל מחזור החיים של מערכות בינה מלאכותית.

6. **אחריותיות** - מפתחי בינה מלאכותית, מפעיליה או המשתמשים בה יגלו אחריות לתפקודה התקין, ולקיום העקרונות האחרים בפעילותם, בין היתר בשים לב לתפיסות ניהול סיכונים מקובלות ולאפשרויות הטכנולוגיות הזמינות.

ג. חלף מסגרת חקיקה רוחבית נוקשה ומחייבת, תפעל הממשלה בדרכים המוצעות להלן:

1. הרגולטורים הסקטוריאליים יבחנו את הצורך בקידום רגולציה קונקרטית בתחומם, תוך פעולה לאור מדיניות ממשלתית אחידה.

2. תקודם רגולציה המבוססת על ניהול סיכונים – כמפורט להלן בסעיף ד(1).

3. קידום רגולציה שתתאים במידת האפשר לנעשה במדינות מפותחות וארגונים בינלאומיים.

4. יועדפו שימוש בכלי רגולציה "רכים" ומתקדמים – כגון עקרונות אתיים לתחום הבינה המלאכותית (כמפורט בסעיף ב'), תקינה, המלצות רגולטור לאימוץ וולונטרי ורגולציה עצמית (מפוקחת או בלתי מפוקחת).

5. הרגולציה תתפתח באופן מודולרי ותוך עידוד נסיינות רגולטורית, ובכלל זה שימוש בפיילוטים וב"ארגזי חול" רגולטוריים, אשר יאפשרו הכנסה בטוחה של מערכות מבוססות בינה מלאכותית.

6. המשך עיצוב עקרונות הרגולציה ושאלת הצורך בהם, ייעשה תוך שיתוף הציבור ובכלל זה התעשייה, האקדמיה וארגוני מגזר שלישי.

ד. צעדים לביצוע:

1. **כלי לניהול סיכונים – תותנע עבודה לגיבוש כלי אחיד לניהול סיכונים ביחס לשימוש בבינה מלאכותית, שייצור שפה משותפת בין גורמי הממשלה והרגולטורים ובינם לבין גורמים פרטיים.** לשם כך, הריני מנחה את מנכ"לית משרד החדשנות, המדע והטכנולוגיה, לפעול להקמת צוות בין משרדי, בהובלת משרד החדשנות, המדע והטכנולוגיה, ובהשתתפות גורמים ממשרדי הממשלה הרלוונטיים, מאסדרים שונים ותעשייה, אשר יגבש המלצה לשרה לכלי זה. במסגרת גיבוש "כלי לניהול סיכונים", יתחשב הצוות הבין משרדי במכלול העקרונות האמורים מעלה, ויפעל במטרה לעודד חדשנות אחראית.

2. **הקמת מוקד ידע ותיאום ממשלתי שיעסוק בריכוז ותיאום סוגיית ההסדרה של בינה מלאכותית.** מוקד הידע יפעל במשרד החדשנות, המדע והטכנולוגיה, בהתאם לאחריות השרה והמשרד כמפורט בהחלטה 212, בסיוע ייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים ויעסוק, בין היתר, בנושאים הבאים:

א. יישום מדיניות רגולציה ואתיקה זו, ועדכונה לפי הצורך; ייעוץ למשרדי הממשלה ולרגולטורים בגיבוש מדיניות ורגולציה לגבי בינה מלאכותית; והנגשת מידע וכלים לשימוש אחראי בבינה מלאכותית עבור הממשלה והציבור.

ב. סיוע לגורמי הממשלה והרגולטורים הרלוונטיים במיפוי השימושים בבינה מלאכותית והאתגרים הנלווים להם בענפים שעליהם הם אמונים.

ג. הקמת פורום רגולטורים ופורום לשיתוף ציבור לתחום הבינה המלאכותית. יוקם פורום מקצועי פנים ממשלתי הכולל נציגי רגולטורים ומומחי טכנולוגיה מדיניות ומשפט, על מנת לקדם את התיאום ולדון בסוגיות משותפות; וכן פורום הכולל נציגי תעשייה, אקדמיה, ארגוני חברה אזרחית והציבור הרחב, כדי לדון בסוגיות של הסדרת בינה מלאכותית.

ה. **תימשך מעורבות אקטיבית של ישראל בפיתוח הרגולציה והתקינה הבינלאומית**. נוכח השפעתה וחשיבותה של הרגולציה והתקינה הבינלאומית על פעולתו של המשק הישראלי בתחומי הבינה המלאכותית, ובהתאם לאמור בהחלטה 212.

שרת החדשנות, המדע והטכנולוגיה

דברי הסבר

1. טכנולוגיות בינה מלאכותית נמצאות בשימוש מואץ במגזר הפרטי והציבורי ומחוללות מהפכה בתחומים רבים בחיינו. טכנולוגיות אלה משפיעות זה מכבר באופן דרמטי על תהליכי עבודה, ייצור, אספקה ותורמות להתייעלות, חדשנות, שיפור הבריאות, איכות החיים והעתיד.

2. לצד כל זאת, השימוש בבינה מלאכותית טומן בחובו אתגרים משמעותיים בהיבטים שונים, ובהם היבטים עבור המערכת הרגולטורית והמשפטית. הוא מעורר סיכונים וחששות שונים, החל מחשש לפגיעה בפרטיות ותקיפות סייבר וכלה באפליה פסולה בקבלת החלטות. כמו כן, הוא מעורר סוגיות משפטיות, רגולטוריות ואתיות מורכבות, כגון האחריות לתוצאות שנבעו מפעילות בינה מלאכותית, הצורך לגלות על שימוש בה או לתת הסבר להחלטותיה.

3. מדינת ישראל היא בעלת מוניטין עולמי ותעשייה מובילה בתחומי ההייטק המחקר והפיתוח. אולם למרות זאת, בכל הנוגע לתשתית הממשלתית ולהסדרה הממשלתית של בינה מלאכותית – ישראל מצויה בפער משמעותי. זאת, בין היתר, כפועל יוצא מכך שלא הייתה בישראל עד כה מדיניות ממשלתית סדורה בתחום הבינה המלאכותית.

4. במטרה להתמודד עם הפער האמור, מתפרסם מסמך מדיניות זה לשימוע, וזאת בהמשך להחלטת ממשלה מספר 212 מיום 1.8.21 "תכנית לקידום חדשנות, עידוד צמיחת ענף ההייטק וחיזוק המובילות הטכנולוגית והמדעית" (להלן- "החלטה 212"). במסגרת ההחלטה האמורה הטילה הממשלה על שרת החדשנות, המדע והטכנולוגיה להוביל את מדיניות הממשלה בתחום הבינה המלאכותית, וזאת כדלקמן:

> *"להטיל על שרת החדשנות, המדע והטכנולוגיה להוביל את מדיניות הממשלה בתחום הבינה המלאכותית בנושאי רגולציה, מדיניות מידע ונתונים, אתיקה, שיתוף פעולה בין-לאומי אזרחי והטמעה במגזר הציבורי האזרחי ונושאים נוספים... ולגבש תכנית לאומית בהתאם, תוך התחשבות בעקרונות משפטיים ואתיים מקובלים בעולם, כגון עקרונות ה-OECD על בינה מלאכותית שנקבעו בשנת 2019.*
>
> *במסגרת זאת, להטיל על שרת החדשנות, המדע והטכנולוגיה להוביל את שיתופי הפעולה הבינלאומיים האזרחיים בתחום, לייצג את ישראל בפורומים בינלאומיים אזרחיים, לרבות פורום השרים של ארגון ה OECD -ולהגן על האינטרסים של ישראל בפורומים האמורים".*

5. נקודת המוצא של עקרונות מסמך המדיניות היא מדיניות השרה לקידום "חדשנות אחראית". מדיניות שמטרתה היא לאזן באופן ראוי בין שמירה על הפרט בשימוש בטכנולוגיה זו, מול האתגרים שמעלה בינה מלאכותית, לבין החשיבות שבחיזוק ושמירת המובילות הטכנולוגית של ענף ההייטק הישראלי בכלל ובטכנולוגיה זו בפרט. זאת, במטרה להבטיח את האיתנות הטכנולוגית והביטחונית של מדינת ישראל לאורך זמן ולאור החשיבות האסטרטגית של קידום תחומי הבינה המלאכותית ומדעי הנתונים להמשך הצמיחה של המשק, להעלאת הפריון ולייעול המערכת הציבורית.

6. מסמך המדיניות מפורסם להערות הציבור, ומהווה נדבך נוסף וחשוב בעבודה הכוללת לקידום נושא הבינה המלאכותית בישראל.

7. יצוין כי ביום 4.9.22 הציגה שרת החדשנות, המדע והטכנולוגיה לוועדת השרים לחדשנות מדע וטכנולוגיה בממשלה את המלצות המועצה למחקר ופיתוח אזרחי (להלן – "המולמו"פ") לתחומי עדיפות לאומיים טכנולוגיים בחמש השנים הבאות, אשר גובשו בהמשך להחלטת ממשלה 212 ולהנחיית השרה. במסגרת זו, עוגנה פעם נוספת היות תחום הבינה המלאכותית (AI) ומדע הנתונים כתחום עדיפות לאומית טכנולוגי בישראל, כמו גם תחום הקוואנטום, לצד ובנוסף לנושאים נוספים כגון ביו קונברג'נס, פודטק, אנרגיות מתחדשות, חלל ובלו-טק.

8. כמו כן, ביום 18.7.22 השיק משרד החדשנות, המדע והטכנולוגיה את התוכנית הלאומית לבינה מלאכותית, בשיתוף עם רשות החדשנות, אגף התקציבים ושותפים נוספים. כן הותנע מהלך לבנית תשתית מיחשובית, ולהטמעת בינה מלאכותית במשרדי הממשלה, העצמת תעשיית החדשנות בתחום הבינה המלאכותית, וחיזוק מדיניות הדאטה הממשלתית בתחום, וזאת בהתאם להחלטת ממשלה 212.

9. בנוסף, נעשו מהלכים שמטרתם לחזק את ישראל כמובילה עולמית בתחום. כך, מדינת ישראל התקבלה לאחרונה, לראשונה, ותיוצג על ידי שרת החדשנות, בארגון GPAI – ארגון בינלאומי בתחום הבינה המלאכותית, בו חברות קבוצה מצומצמת של מדינות. כמו כן, הורחב שיתוף הפעולה מול ארה"ב ומול ארגון ה-OECD בתחום זה, באמצעות ביקורים מדיניים חשובים ותוכניות עבודה משותפות באמצעות צוותים משותפים ממשלתיים, ונחתמו הסכמים ייחודים בנושא בינה מלאכותית עם מספר מדינות, ובהן סינגפור.

10. למסמך מדיניות זה מצורף נספח מקצועי מעמיק שמפרט את בסיס המתודולוגיה למדיניות ומניח יסודות לגיבוש מדיניות רגולציה ואתיקה בנוגע לטכנולוגיות מבוססות בינה מלאכותית בישראל.

11. המסמך המצורף למסמך המדיניות כולל סקירה של האתגרים בבינה מלאכותית, הדין הקיים בישראל והממשק בינו לבין תחום הבינה המלאכותית, סקירה בינלאומית של ההסדרה וכן פירוט המתודולוגיה העומדת בבסיס עקרונות המדיניות המפורטות במסמך זה (להלן – "נספח לעקרונות המדיניות").

12. הנספח לעקרונות המדיניות מבוסס על עבודת מטה מקצועית ומעמיקה שנעשתה על ידי משרד החדשנות המדע והטכנולוגיה ומשרד המשפטים באמצעות מחלקת ייעוץ וחקיקה (משפט כלכלי), בהמשך לישיבות שהתקיימו במסגרת הצוות הבינמשרדי שהוקם מכוח סעיף 6(א) להחלטה 212.

13. בהתאם למסמך המדיניות המוצעת, בשלב זה אין מקום לקידום חקיקת מסגרת רוחבית על כל תחום הבינה המלאכותית. מוצע כי במקום זאת, יבחנו הרגולטורים השונים, כל אחד בתחומו, את הצורך בקידום רגולציה קונקרטית בתחומם, תוך שמירה על מדיניות ממשלתית אחידה ושיח ותיאום ממשלתיים שיעשו במסגרת פורום ידע ממשלתי המפורט במסמך המדיניות. מדיניות הרגולציה האחידה תהא מבוססת על ניהול סיכונים, תעקוב אחר הנעשה במדינות מפותחות ובארגונים בינלאומיים מקובלים ותהא גמישה ותאפשר גם שימוש באמצעים רכים (דוגמת תקינה וולונטרית ורגולציה עצמית במקרים המתאימים לכך) ובאמצעים רגולטורים מתקדמים.

14. כן מוצע אימוץ עקרונות אתיים משותפים עבור בינה מלאכותית מהימנה, על בסיס עקרונות שאושרו ב-OECD, תוך ביצוע התאמות נדרשות. אימוץ העקרונות האתיים יוכל לסייע לארגונים ולרגולטורים במסגרת פעילותם בתחום הבינה המלאכותית. מוצע כי העקרונות לא יחליפו את המסגרת המשפטית הקיימת, והם ישמשו בסיס לשיח משותף, הכוונה, ותיאום לפי ההקשר הנורמטיבי והרגולטורי.

15. לעקרונות ה-OECD חשיבות ייחודית, משום שגובשו בתיאום בין המדינות המפותחות מתוך שותפות ורצון לעשות שימוש מועיל בבינה מלאכותית באופן שמכבד את עיקרון החוקיות, ערכים דמוקרטיים ואינטרסים ציבוריים. מדינות אלה הסכימו על עקרונות אלו כקווים מנחים בתחום זה, תוך גישור על השוני בהיבטים המשפטיים, הרגולטוריים והכלכליים ביניהן. עקב כך העקרונות משקפים שיח מפותח שיש לו ערך לגיבוש מדיניות פנים מדינתית. בנוסף, יש לו חשיבות במרחב הבינלאומי כמכנה משותף מוסכם בין המדינות המפותחות.

16. מדיניות זו מהווה אבן יסוד ומצפן ראשון בתחום הבינה המלאכותית של ישראל ותקדם את ישראל כגורם משפיע בעיצוב הרגולציה והתקינה הבינלאומית.

17. מעבר לעקרונות המדיניות המפורסמים בזה מתניע מסמך המדיניות צעדים חשובים נוספים שיש לבצעם:

18. גיבוש כלי אחיד לניהול סיכונים ביחס לשימוש בבינה מלאכותית, שייצור שפה משותפת בין גורמי הממשלה והרגולטורים ובינם לבין גורמים פרטיים.

19. הקמת מוקד ידע ותיאום ממשלתי שיפתח מומחיות מקצועית וטכנולוגית, ויהיה בעל ראייה רוחבית ואחריות לקידום האינטרס הכלל-משקי בעידוד הטכנולוגיה והשגת היעדים הממשלתיים. המוקד יסייע לגורמי הממשלה והרגולטורים הרלוונטיים להבין ולמפות את השימושים הקונקרטיים במערכות מבוססות בינה מלאכותית הנעשים על ידי המפוקחים בענף המוסדר שעליו הם אמונים; האתגרים, החששות והסיכונים הכרוכים בכך; והמענים האפשריים.

20. עוד יהיה אמון מוקד הידע של המשרד על הקמת פורום רגולטורים ופורום לשיתוף ציבור לתחום הבינה המלאכותית. במסגרת כך יוקם פורום מקצועי פנים ממשלתי הכולל נציגי רגולטורים ומומחי טכנולוגיה מדיניות ומשפט, על מנת לקדם את התיאום ולדון בסוגיות משותפות; וכן פורום הכולל נציגי תעשייה, אקדמיה, ארגוני חברה אזרחית והציבור הרחב, כדי לדון בסוגיות של הסדרת בינה מלאכותית.

אורית פרקש הכהן

מסמך מקצועי

עקרונות מדיניות, רגולציה ואתיקה
בתחום הבינה המלאכותית

משרד המשפטים
MINISTRY OF JUSTICE | وزارة العدل

ייעוץ וחקיקה
OFFICE OF LEGAL COUNSEL AND LEGISLATIVE AFFAIRS
المشورة والتشريع

משרד החדשנות
המדע והטכנולוגיה

תוכן עניינים

תקציר מנהלים ותודות 13
1. חלק ראשון: רקע 21
1.1. מהי בינה מלאכותית? 21
1.2. חשיבות מדיניות רגולציה ואתיקה בתחום הבינה המלאכותית 23
2. חלק שני: המצב המשפטי והרגולטורי בהיבטי בינה מלאכותית בישראל 25
2.1. התפתחות המשפט והרגולציה להסדרת בינה מלאכותית 25
2.2. העיסוק הממשלתי בבינה מלאכותית 27
2.3. תחולה כללית של הדין והרגולציה בהיבטי בינה מלאכותית 29
2.3.1. דיני פרטיות 29
2.3.2. אפליה אסורה 30
2.3.3. דיני חוזים 30
2.3.4. הגנת הצרכן 31
2.3.5. דיני נזיקין 32
2.3.6. הסדרה ייעודית מגזרית 33
2.3.7. משפט מנהלי 34
3. חלק שלישי: פעילות ארגונים בינלאומיים ומדינות מפותחות 35
3.1. המלצות ה-OECD בתחום הבינה המלאכותית 36
3.2. טיוטת החקיקה של האיחוד האירופי 38
3.3. מועצת אירופה וקידום אמנה בנושא בינה מלאכותית 39
3.4. מדיניות הרגולציה בארה״ב 41
3.5. מדיניות הרגולציה בבריטניה 44
4. חלק רביעי: סוגיות ואתגרים המתעוררים בקשר לבינה המלאכותית 46
4.1. אפליה 47
4.1.1. סיכונים לאפליה אסורה במערכות מבוססות בינה מלאכותית 48
4.1.2. האתגרים בהתמודדות עם אפליה במערכות מבוססות בינה מלאכותית 50
4.2. מעורבות אנושית 52
4.2.1. מהי מעורבות אנושית? 53
4.2.2. יתרונות המעורבות האנושית 54
4.2.3. אתגרים ביחס למעורבות אנושית 55
4.3. הסברתיות 58
4.3.1. על הסברתיות ותופעת ״הקופסא שחורה״ 58
4.3.2. יתרונות וחסרונות הדרישה להסברתיות 60
4.4. גילוי 62
4.4.1. מה טיבה של דרישת הגילוי? 63
4.4.2. היתרונות והחסרונות להצבת דרישת גילוי 63

4.5. אמינות, עמידות, אבטחה ובטיחות 65
4.5.1. אמינות המערכת 66
4.5.2. עמידות המערכת ואבטחתה 68
4.5.3. בטיחות המערכת 70
4.6. אחריותיות 71
4.6.1. האתגרים בשימוש בהסדרים הקיימים להטלת אחריות 73
4.7. פרטיות 76
4.7.1. האתגרים לפרטיות בהם מתאפיין השימוש במערכות מבוססות בינה מלאכותית 77
5. חלק חמישי: דרכי התמודדות עם סוגיות ואתגרים אלה 80
5.1. אפליה 80
5.2. מעורבות אנושית 81
5.3. הסברתיות 83
5.4. גילוי 84
5.5. אמינות, עמידות ובטיחות 85
5.6. אחריותיות 87
5.7. פרטיות 88
6. חלק שישי: עקרונות מוצעים למדיניות רגולציה ואתיקה לתחום הבינה המלאכותית 90
6.1. אימוץ מדיניות רגולציה לתחום הבינה המלאכותית 92
6.1.1. החשיבות של מדיניות רגולציה ומעמדה 92
6.1.2. רכיבי המדיניות המוצעים לאימוץ 93
6.2. אימוץ עקרונות אתיים לתחום הבינה המלאכותית 102
6.2.1. החשיבות של אימוץ עקרונות אתיים ומעמדם 102
6.2.2. התבססות על עקרונות ה-OECD 103
6.2.3. העקרונות האתיים המוצעים לאימוץ 103
6.3. מיסוד מוקד ידע ותיאום ממשלתי להסדרת בינה מלאכותית 109
6.4. הקמת פורום רגולטורים ופורום לשיתוף ציבור לתחום הבינה המלאכותית 110
6.5. מיפוי השימושים בבינה מלאכותית והאתגרים הנלווים להם בענפים מוסדרים 111
6.6. מעורבות אקטיבית בפיתוח הרגולציה והתקינה בפורומים בינלאומיים 112
6.7. התוויית שפה אחידה באמצעות מסגרת מומלצת (וולונטרית) לניהול סיכונים 112

תקציר מנהלים ותודות

ביום 1.8.2021 קיבלה הממשלה את החלטה מס׳ 212 שעניינה ״תכנית לקידום חדשנות, עידוד צמיחת ענף ההייטק וחיזוק המובילות הטכנולוגית והמדעית״[1] (להלן: החלטת ממשלה 212). בין היתר, הטילה הממשלה במסגרת זו על שרת החדשנות, המדע והטכנולוגיה להוביל את גיבוש מדיניות הממשלה בתחום הבינה המלאכותית, לרבות בהיבטי מדיניות רגולציה ואתיקה.

בהתאם להחלטה זו, ועל מנת לקדם את המטרות שנקבעו במסגרתה, נערך במשותף מסמך זה על ידי גורמי המקצוע במשרד החדשנות, המדע והטכנולוגיה ומחלקת ייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים, וזאת לבקשת שרת החדשנות, המדע והטכנולוגיה הגב׳ אורית פרקש-הכהן ובהנחייתה, וכן הכווינו המשנה ליועצת המשפטית לממשלה (משפט כלכלי) עו״ד מאיר לוין, ובהכוונת מנכ״לית משרד החדשנות, המדע והטכנולוגיה הגב׳ הילה חדד-חמלניק.

ריכזו את העבודה על המסמך מר תום דן המשנה למנכ״לית משרד המדע, החדשנות והטכנולוגיה ועו״ד יוסף גדליהו ממחלקת ייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים. נטל חלק מרכזי בכתיבת המסמך עו״ד עמית אשכנזי, היועץ המשפטי למערך הסייבר הלאומי (בדימוס). ליוו מקרוב את העבודה על המסמך עו״ד דני חורין, היועץ המשפטי (בפועל) למשרד המדע, החדשנות והטכנולוגיה וד״ר עו״ד יובל רויטמן, ראש אשכול רגולציה בייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים. כן סייעה לעבודה ולהעברת דגשי שרת החדשנות, המדע והטכנולוגיה, עו״ד גאיה הררי הייט. תודות נוספות לכל מי שנטל חלק ובמיוחד למר גל קורן על עבודתו המסורה.

כמו כן, סייעו לעבודה על המסמך ולגיבוש ההמלצות שותפים מייעוץ וחקיקה ומהממשלה, וביניהם: עו״ד משה אוסטר, עו״ד שרית פלבר, עו״ד סדריק (יהודה) צבע, עו״ד טל וורנר-קלינג, עו״ד אביטל שטרנברג, עו״ד ד״ר עמרי בן-צבי, עו״ד שרה גולד, עו״ד רני נויבואר, עו״ד ד״ר תמר קלהורה, עו״ד ד״ר ליטל הלמן, עו״ד איל זנדברג, עו״ד מיכל אברהם, מר זיו קציר, עו״ד שרון שמש עזריה, עו״ד ראובן אידלמן, עו״ד ניר גרסון עו״ד שירן מימון, הגב׳ נועה אלקין, מר אלון פרבר והגב׳ נעם אמיר.

נקודת המוצא הממשלתית, כפי שהציבה הממשלה בהחלטה 212, היא כי יש לפעול על מנת לקדם את המובילות הטכנולוגית של ענף ההייטק הישראלי, ולהבטיח את האיתנות הטכנולוגית והביטחונית של מדינת ישראל לאורך זמן. זאת, לאור החשיבות האסטרטגית של קידום תחומי הבינה המלאכותית ומדעי הנתונים להמשך הצמיחה של המשק, להעלאת הפריון ולייעול המערכת הציבורית בפרט, והכל תוך התחשבות בעקרונות משפטיים ואתיים מקובלים בעולם.

בשנים האחרונות הפך המונח בינה מלאכותית למטבע לשון שגורה הבאה לתאר מכונות (מחשבים) הפועלות באופן שיכול להיתפש כחכם, מורכב או תבוני. למערכות מבוססות בינה מלאכותית שימושים רבים, כגון נהיגת כלי רכב עצמאיים (אוטונומיים), פענוח תצלומי רנטגן, הערכת סיכוני אשראי, מסחר בניירות ערך ובחינת מועמדים לתעסוקה. כיום מערכות אלה נמצאות בשימוש מואץ במגזר הפרטי והציבורי ומחוללות מהפכה בענפים שונים במשק. בעתיד הקרוב, על פי ההערכה,

[1] החלטה מספר 212 של הממשלה מיום 01.08.2021 **״תכנית לקידום חדשנות, עידוד צמיחת ענף ההייטק וחיזוק המובילות הטכנולוגית והמדעית״**, https://www.gov.il/he/departments/policies/dec212_2021.

מערכות כאלה צפויות להשפיע באופן משמעותי, ולעיתים אף מהפכני, על תהליכי עבודה, ייצור ואספקה ולהביא להתייעלות, חדשנות, ושיפור איכות החיים בתחומים רבים.

ברם, לצד היתרונות המשמעותיים, השימוש בבינה מלאכותית טומן בחובו אתגרים משמעותיים עבור המערכת המשפטית והרגולטורית ברחבי העולם ובישראל. כך, הוא מעורר סיכונים וחששות שונים וכן סוגיות משפטיות ורגולטוריות מורכבות, ובכלל זה ניתן למנות עניינים אלה:

אפליה – אף שהשימוש במערכות מבוססות בינה מלאכותית עשוי למתן אפליה הנובעת מהטיות אנושיות וממבנים חברתיים, מתעורר בעקבותיו חשש מאפליה בין היתר בשל אפשרות להטיה בנתונים המשמשים לאימון המערכת (כגון מאגרי מידע לא ייצוגיים / המשקפים אפליה קיימת), או שקילת משתנים המתואמים עם מאפיינים מפלים (כך למשל, מקום מגורים, הנלמד מכתובת או מיקוד, הוא נתון שלעיתים קרובות קורלטיבי (proxy) לשיוך הקבוצתי, למאפיינים אתניים או למעמד סוציו-אקונומי, ועלול להביא לתוצאות מפלות). ישנם אתגרים משפטיים וטכנולוגיים שונים העלולים להקשות על ההתמודדות עם אפליה במערכות מבוססות בינה מלאכותית.

מעורבות אנושית – מאפיין מרכזי שמייחד מערכות מבוססות בינה מלאכותית הוא יכולתן לפעול באופן אוטונומי. מעורבות אנושית בפעילות המערכת (כגון בפיקוח שוטף או במנגנון ערעור), עשויה לשפר את תהליך קבלת ההחלטה ולצמצם טעויות של המערכת; לחזק את מידת האחריותיות כלפי המערכות; להגביר את הלגיטימציה של ההחלטה ולהפחית את הפגיעה בכבודו של מושא ההחלטה. מנגד, קיים חשש כי המעורבות האנושית לא תהיה אפקטיבית ואף תסב נזק. על רקע זה, מתעוררות שאלות כגון מתי צריכה להיות דרישה למעורבות אנושית, וכיצד נכון לעצב את האינטראקציה בין הגורם האנושי לבין המערכת על מנת לנצל את היתרונות היחסיים של כל צד ולמנוע הטיות.

הסברתיות – הסברתיות היא היכולת להציג בצורה שניתנת להבנה על ידי בני אדם את אופן פעולת המערכת או ההחלטה שלה. בינה מלאכותית מבוססת על יכולות חישוביות היוצרות מודל תחזיות וקבלת החלטות שלעיתים לא ניתן "לחלץ" אותו מתוך המערכת (תופעה המתוארת כ-"קופסא שחורה" (black box)). קבלת החלטה שלא ניתן להבין או להסביר על ידי מכונה, עלולה להיות שגויה וכן לעורר חשש לפגיעה בכבוד האדם ובאוטונומיה שלו, בכך שהוא לכאורה נתון להחלטה שרירותית. לכן מתעוררות שאלות כלליות באילו מצבים ראוי לדרוש כי יינתן הסבר ביחס לאופן פעילות המערכת, ומה צריכה להיות רמת הפירוט של ההסבר שיינתן בכל תרחיש רלוונטי. שכן על אף יתרונותיה, בהקשרים שונים קיימת מורכבות בהחלת הדרישה להסברתיות באופן רחב.

גילוי – עם התקדמות הטכנולוגיה והשימוש ההולך והגובר במערכות מבוססות בינה מלאכותית לשם קבלת החלטות ובמערכות צ'אט-בוט (chat-bot), גוברת האפשרות שאנשים לא יהיו מודעים לכך שלמערכות אלה תפקיד משמעותי "מאחורי הקלעים" בקבלת ההחלטות בעניינם, או שהם משוחחים עם מערכות כאלה (ולא עם אנשים). במקרים מסוימים, לגופים המפעילים מערכות אלה יהיה תמריץ להסתיר זאת, לעיתים תוך ניסיון להטעות את משתמש הקצה. על רקע זה, מתעוררת השאלה מה מידת הגילוי הנדרשת לגבי מעורבות מערכות מבוססות בינה מלאכותית.

אמינות, עמידות, אבטחה ובטיחות – מערכות מבוססות בינה מלאכותית חשופות לטעויות ותקלות טכניות, או למניפולציות מכוונות. קיים חשש שמערכת תסבול מביצועים ירודים משום שאינה מסוגלת לבצע כהלכה את המשימה שיועדה לה, וכן חשש שגורם חיצוני ישבש את פעילותה על ידי ניצול של נקודת תורפה הטבועה בה. לרוב מפתחי המערכות והמשתמשים בהן יהיו בעלי האינטרס המרכזי להתמודד עם חששות אלה. עם זאת, במקרים מסוימים, עשויה להיות הצדקה לשקול התערבות רגולטורית, על מנת לוודא כי מערכות מבוססות בינה מלאכותית הן אמינות, עמידות, מאובטחות ובטוחות במידה מספקת. זאת, בפרט בנסיבות בהן מתקיימות הצדקות מקובלות להתערבות רגולטורית, כדוגמת החצנות שליליות הכרוכות בפעילותן של מערכות אלו.

אחריותיות – המאפיין האוטונומי של מערכות מבוססות בינה מלאכותית, והקושי לצפות את פעולתן, עלולים במקרים מסוימים לערער את המבנה המקובל המושתת על הימצאות אדם במוקד ההתרחשות שככלל נושא באחריות למעשיו. ככל שהמעורבות האנושית בהחלטה או בפעולה מצטמצמת ורחוקה יותר, מתעוררות שאלות בנוגע לנשיאה באחריות המוסרית, החברתית והמשפטית (פלילית או אזרחית) בגין טעויות שנגרמו אגב השימוש במערכות אלה. בכלל זה, מי נושא באחריות? מה סוג האחריות שניתן לייחס לו? ומה המשטר המתאים לבחינת שאלת האחריות (למשל רשלנות, אחריות קפידה ומוחלטת, או נמל מבטחים)? כמו כן, עולות שאלות בנוגע להגברת האחריותיות באמצעות קידום נורמות ארגוניות המבטאות נטילת אחריות, כגון ביצוע תהליך עיתי להערכת סיכונים או מינוי אחראי ארגוני להתמודדות עם סיכונים הקשורים לבינה מלאכותית.

פרטיות – פיתוח ושימוש במערכות מבוססות בינה מלאכותית ככלל מצריך שימוש בנתונים רבים, שחלקם עשויים לכלול מידע אישי, ולהיות מוסדרים באמצעות דיני הגנת הפרטיות. לפיכך, עשויים להתעורר אתגרים ייחודיים הנוגעים לצורך להבטיח את ההגנה הנדרשת לזכות לפרטיות. כך למשל, לעיתים מבוקש "לאמן" מערכת במידע שנאסף למטרות אחרות, באופן שעלול לאתגר את העמידה בעיקרון צמידות המטרה. כמו כן, קשיי ההסברתיות עלולים לאתגר את היכולת לקבל ההסכמה מדעת מנושא המידע, או לעמוד בדרישת השקיפות ובחובת היידוע. בנוסף, השימוש במערכות מבוססות בינה מלאכותית עלול ליצור מתח בין הצורך בנתוני עתק (Big Data), לבין מחיקת מידע עודף בהתאם לעיקרון צמצום המידע. כך גם, קיים חשש שייעשה שימוש במערכות אלה לשם זיהוי חוזר של מידע שעבר התממה (אנונימיזציה). על כן יש להידרש להיבטי פרטיות בפיתוח ושימוש מערכות מבוססות בינה מלאכותית, וגם במסגרת גיבוש מדיניות רגולטורית בתחום זה.

המצב הרגולטורי והמשפטי הקיים בישראל (בענפים שונים כגון דיני החוזים, הנזיקין, הגנת הצרכן והגנת הפרטיות), עשוי לחול באופן שמסדיר פיתוח ושימוש במערכות מבוססות בינה מלאכותית, ובהתאם גם יכול להיות רלוונטי לטיפול בחלק מהאתגרים והסוגיות המוזכרות לעיל ועניינים נוספים. עם זאת, הניסיון מלמד כי בין התפתחויות טכנולוגיות באשר הן לבין המערכת המשפטית והרגולטורית נוצרות פעמים רבות נקודות חיכוך, וכפועל יוצא מכך, לא אחת עולה צורך בהתאמה של המשפט והרגולציה הקיימים בעקבות הופעת טכנולוגיות חדשות. על רקע זה, יש להניח כי יהיה צורך בעיסוק של מקבלי ההחלטות בכלל, וגורמי הממשלה והרגולטורים בפרט, באתגרים והסוגיות שנסקרו לעיל ובעניינים נוספים הקשורים לפיתוח ושימוש במערכות מבוססות בינה מלאכותית.

מדיניות הרגולציה והאתיקה בתחום הבינה המלאכותית, המוצעת במסמך זה, נועדה להתוות את האופן שבו תבחן הממשלה את השאלה האם יש לקבוע כללי התנהגות לפיתוח ושימוש בבינה מלאכותית במגזר הפרטי ואת האופן לעשות כן. זאת, במטרה לקדם פיתוח ושימוש בטכנולוגיות בינה מלאכותית ואף להפחית חסמים משפטיים ורגולטוריים הניצבים בפניהן, ובד בבד לצמצם פגיעות אפשריות בזכויות יסוד ואינטרסים ציבוריים אגב פיתוח ושימוש בבינה מלאכותית.

<u>**עקרונות מוצעים למדיניות רגולציה ואתיקה לתחום הבינה המלאכותית**</u>

על אף המורכבות הרבה של תחום הבינה המלאכותית, על יתרונותיו והאתגרים הטמונים בו, לא גובשה עד עתה מדיניות ממשלתית אחידה בתחום זה. עיקר המסמך הוא עקרונות מוצעים למדיניות אתיקה ורגולציה בתחום הבינה המלאכותית בישראל, תוך התחשבות בעקרונות משפטים ואתיים מקובלים בעולם. לצורך גיבוש עקרונות אלה נבחנו המאפיינים הייחודים של הבינה המלאכותית, וכן האופן בו הארגונים הבינלאומיים ומדינות מפותחות עוסקות בתחום זה. המדיניות המוצעת מבקשת לאזן בין הצורך בוודאות ובהירות לבין האינטרסים הציבוריים והזכויות שעל הפרק והחשש מהתערבות שאינה מוצדקת העלולה לפגוע בחדשנות. כן כולל המסמך עקרונות נוספים המבקשים ליתן כלים טובים יותר ליישום מדיניות הרגולציה המוצעת.

להלן יפורטו בתמצית העקרונות המוצעים במסמך זה במישורים השונים:

1. <u>**אימוץ מדיניות רגולציה לתחום הבינה המלאכותית**</u>

נקודת המוצא להצעה זו, היא שישנה חשיבות רבה לכך שרגולציה המשפיעה על פיתוח בינה מלאכותית והשימוש בה, תקודם או תותאם לפי מדיניות רגולציה ממשלתית אחידה וקוהרנטית. זאת, על מנת להשיג את יעדי המדיניות של הממשלה; לקדם את תחום הבינה המלאכותית; להגן על זכויות יסוד ואינטרסים ציבוריים; ולצמצם את החשש לפגיעה בחדשנות הטכנולוגית.

בהתאם לכך, מוצע לאמץ מדיניות רגולציה בכלל, ולכוון את גורמי הממשלה ובפרט הרגולטורים הרלוונטיים להתחשב בה במסגרת פעילותם בקביעת רגולציה ביחס לבינה מלאכותית.

בפרט, מוצע לאמץ את רכיבי מדיניות הרגולציה המפורטים להלן, שגובשו בין היתר בשים לב ל"עקרונות מנחים לאסדרה מיטבית" הקבועים בחוק עקרונות האסדרה, התשפ"ב-2021; למסמכי מדיניות רגולציה שפרסמו המדינות המובילות בעולם בתחום ההסדרה של בינה מלאכותית; להמלצות שניתנו ולעבודות קודמות שנעשו במסגרת הפעילות הממשלתית בתחום; ולדיוני הצוות הבין-משרדי שעסק ברגולציה ואתיקה במהלך גיבוש התכנית הלאומית לבינה מלאכותית.

<u>רכיבי מדיניות הרגולציה המוצעים הם</u>:

ראשית, מוצע כי ככל שתקודם רגולציה שמטרתה להסדיר פיתוח ושימוש בבינה מלאכותית, הדבר ייעשה ברמת הענף שבו היא נדרשת, על בסיס צורך קונקרטי, ובהובלת הרגולטור האמון עליו. זאת, תוך שמירה על מדיניות ממשלתית אחידה באמצעות תיאום, אך לא באמצעות רגולציה רוחבית (למשל, לא באמצעות חקיקה ייעודית המשתרעת על כל תחום הבינה המלאכותית).

שנית, מוצע כי רגולציה ביחס לפיתוח ושימוש בבינה מלאכותית, ככל שתאומץ, תביא בחשבון את הרגולציה הנוהגת במדינות מובילות בתחום ושאומצה על ידי ארגונים בינלאומיים, ותותאם במידת האפשר לרגולציה רלוונטית שנקבעה במדינות ובארגונים כאמור. זאת, בין היתר, כדי להימנע מהצבת חסמים רגולטוריים ייחודיים בפני כניסת מערכות מבוססות בינה מלאכותית לישראל.

שלישית, מוצע כי רגולציה המסדירה פיתוח ושימוש בבינה מלאכותית, תהיה מותאמת ככלל לסיכונים הנשקפים מסוג הטכנולוגיה ומהשימוש הספציפי שאותו היא נועדה להסדיר, ותהיה תוצר של ניהול סיכונים שביצע הרגולטור ביחס אליו. כך, שככלל, הרגולציה לא תחול באופן אחיד על טכנולוגיות ושימושים שרמת הסיכונים והחששות לגביהם שונה באופן ניכר.

רביעית, על מנת לעודד את הפיתוח והשימוש בה ולהפיק את התועלות החברתיות והכלכליות הנובעות מכך, מוצע כי במקרים המתאימים ייעשה שימוש ברגולציה מאפשרת. בכלל זה, מוצע לבחון שימוש בכלי רגולציה "רכים" ומתקדמים, כגון עקרונות אתיים לתחום הבינה המלאכותית, תקינה, המלצות רגולטור לאימוץ וולונטרי ורגולציה עצמית (מפוקחת או בלתי מפוקחת).

חמישית, מוצע כי רגולציה שנועדה להסדיר פיתוח ושימוש בבינה מלאכותית תקודם ותתפתח בשלבים ובאופן גמיש בהתאם להתפתחויות הטכנולוגיות. במסגרת זו, אף מוצע כי ייעשה שימוש בנסיינות רגולטורית, ובכלל זה בפיילוטים וב"ארגזי חול" רגולטוריים, אשר יאפשרו הכנסה בטוחה של מערכות מבוססות בינה מלאכותית והפקה של התועלות החברתיות והכלכליות מכך.

שישית, מוצע כי רגולציה המסדירה פיתוח ושימוש בבינה מלאכותית, תפותח ותקודם תוך שיתוף בעלי הידע והמומחיות ובעלי העניין, ובכלל זה נציגי התעשייה (כולל "חברות הזנק"), האקדמיה וארגוני חברה אזרחית. זאת, במידה הנדרשת בנסיבות העניין, ועל מנת שהרגולציה תהיה מבוססת על תשתית מקצועית וטכנולוגית איכותית ותבטא איזון בין הזכויות והאינטרסים השונים.

במסגרת גיבוש ההמלצות בדבר מדיניות הרגולציה ניתן משקל של ממש, לכך שמדובר בתחום חדשני ומורכב, שההתפתחויות הטכנולוגיות במסגרתו הן מהירות ועשויות להקדים במקרים רבים את גיבוש הרגולציה. עוד נלקחה בחשבון, השונות הרבה שיש בין השימושים המגוונים בבינה מלאכותית, והעובדה שהמדובר בתחום שיש בו שימושים בעלי רגישות רבה או חשש לסיכון ולפגיעה בזכויות יסוד ואינטרסים ציבוריים, בצד שימושים שאינם מעוררים את אותה הרגישות או החשש. לבסוף, נלקח בחשבון הצורך ללכת "עקב בצד אגודל" בתחום חדשני זה, מתוך רצון להוסיף ולעודד את המשק הישראלי כגורם מוביל מבחינה בינלאומית בתחום הפיתוח והשימוש בבינה מלאכותית, כמו גם מתוך הרצון לפעול באופן התואם לנעשה במדינות מובילות בתחום זה בעולם.

2. <u>אימוץ עקרונות אתיים לתחום הבינה המלאכותית</u>

בהתאם להחלטת הממשלה 212, ובדומה למהלכים נוספים המקודמים במדינות ובארגונים שונים בעולם, מוצע לאמץ עקרונות אתיים משותפים עבור תחום הבינה המלאכותית בישראל, על בסיס עקרונות ה-OECD שנסקרו לעיל, כדי לסייע לרגולטורים ולארגונים בתחום זה.

העקרונות מאפשרים שפה משותפת, לגבי אופן ההתמודדות עם חששות וסיכונים מוכרים הכרוכים במערכות מבוססות בינה מלאכותית, לרבות הסוגיות והאתגרים שצוינו לעיל. כן הם מאפשרים לגשר בין ערכים, יישומים טכנולוגיים, והיבטים משפטיים, תוך ממשק לתפיסה הבינלאומית המתהווה של "בינה מלאכותית מהימנה". אימוץ ופרסום העקרונות במסגרת המדיניות מאפשר גם

שקיפות ושיח ציבורי לגבי העקרונות כנקודת מוצא למדיניות רגולציה לתחום הבינה המלאכותית. העקרונות יוכלו אף לסייע בקידום מדיניות רגולציה קוהרנטית בתחום הבינה המלאכותית, בכך שהם מכוונים את הרגולטורים להיבטים המרכזיים שיש להידרש אליהם, ומגבירים את הוודאות בקרב המגזר הציבורי והפרטי בעניין המדיניות הממשלתית לגבי בינה מלאכותית.

מוצע כי העקרונות לא יהיו בעלי מעמד משפטי, ובהתאם לכך הם לא מחייבים גורמים פרטיים או ציבוריים לפעול או שלא לפעול בדרך כלשהי, ואינם מחליפים את המסגרת המשפטית החלה או מהווים כלי לפרשנות משפטית. עם זאת, עקרונות אלה, משקפים היבטים בהם ראוי להתחשב במסגרת פיתוח ושימוש בבינה מלאכותית ובעת קביעת רגולציה בתחום זה.

העקרונות המוצעים מבוססים על עקרונות ה-OECD בהתאמות שונות, וזאת בשים לב לתפיסה הכוללת שלפיה יש להתחשב ולפעול בהתאם למדיניות המקובלת במדינות המפותחות בעולם.

העקרונות האתיים המוצעים הם:

א. **בינה מלאכותית לקידום צמיחה, פיתוח בר-קיימא ומובילות ישראלית בחדשנות** – שימוש אחראי בבינה מלאכותית מהימנה הוא אמצעי לעודד צמיחה, פיתוח בר-קיימא, רווחה חברתית וקידום המובילות הישראלית בתחום החדשנות.

ב. **האדם במרכז: כיבוד זכויות יסוד ואינטרסים ציבוריים** – פיתוח ושימוש בבינה מלאכותית יש לעשות תוך כיבוד שלטון החוק, זכויות יסוד ואינטרסים ציבוריים ובפרט תוך שמירה על כבוד האדם ופרטיות.

ג. **שוויון ומניעת אפליה פסולה** – בפיתוח ושימוש בבינה מלאכותית יש להביא בחשבון את הצורך בשוויון וגיוון, ואת החשש להטיה במערכות בינה מלאכותית והסיכון לאפליה פסולה כנגד יחידים או קבוצות.

ד. **שקיפות והסברתיות** – בפיתוח ושימוש בבינה מלאכותית, יש להביא בחשבון את הצורך ליידע מי שבא במגע עם בינה מלאכותית או מושפע מפעילותה על כך, וכן את הצורך במתן הסבר להחלטתה או לאופן שבו פעלה. זאת, בין היתר בשים לב למידת השפעתה, השלכותיה על מי שמושפע ממנה והאפשרויות הטכנולוגיות הזמינות.

ה. **אמינות, עמידות, אבטחה ובטיחות** – בפיתוח ושימוש בבינה מלאכותית יש להביא בחשבון את הצורך בכך שמערכות בינה מלאכותית תהיינה אמינות, מאובטחות ובטיחותיות לאורך כל מחזור חייהן, כך שבתנאי שימוש רגילים, שימוש צפוי או שימוש שגוי או תנאים מסוכנים אחרים, הן יפעלו כראוי ולא יהוו סיכון בטיחותי בלתי סביר. לשם כך יש לנקוט אמצעים סבירים, בהתאם לתפיסות מקצועיות מקובלות של ניהול סיכונים, על מנת לצמצם סיכוני בטיחות ואבטחת מידע בכל "מחזור החיים" של מערכות בינה מלאכותית.

ו. **אחריותיות** – מפתחי בינה מלאכותית, מפעיליה או המשתמשים בה, יגלו אחריות לתפקודה התקין, ולקיום העקרונות האתיים האחרים בפעילותם, בין היתר בשים לב לתפיסות ניהול סיכונים מקובלות ולאפשרויות הטכנולוגיות הזמינות.

3. **מיסוד מוקד ידע ותיאום ממשלתי להסדרת בינה מלאכותית**

מוצע להפקיד בידי גורם ממשלתי אחד את ריכוז ותיאום סוגיית ההסדרה של בינה מלאכותית. ישנה חשיבות לכך שגורם זה יפתח מומחיות מקצועית וטכנולוגית, ויהיה בעל ראייה רוחבית ואחריות לקידום האינטרס הכלל-משקי בעידוד הטכנולוגיה והשגת היעדים הממשלתיים.

גורם זה יעסוק ביישום מדיניות רגולציה ואתיקה זו, וגיבוש המלצות לעדכונה לפי הצורך; בייעוץ למשרדי הממשלה ולרגולטורים בגיבוש מדיניות ורגולציה לגבי בינה מלאכותית; ובהנגשת מידע וכלים לשימוש אחראי בבינה מלאכותית, ובכלל זה כלי ניהול סיכונים, עבור הממשלה והציבור.

4. **הקמת פורום רגולטורים ופורום לשיתוף ציבור לתחום הבינה המלאכותית**

מוצע למסד פורום מקצועי פנים-ממשלתי הכולל נציגי רגולטורים ומומחים לטכנולוגיה, מדיניות ומשפט, על מנת לאפשר למידה משותפת ולקדם את התיאום הממשלתי, כמו גם לדון בסוגיות משותפות. זאת, בדגש על תחומים עתירי "חיכוך" בין רגולטורים שונים, ובין היתר במטרה לבחון במשותף את הצורך בעדכון מדיניות הרגולציה לתחום הבינה המלאכותית.

כן מוצע להקים פורום הכולל נציגי תעשייה, אקדמיה, ארגוני חברה אזרחית והציבור הרחב, כדי לדון בסוגיות של הסדרת בינה מלאכותית. מטרה מרכזית של פורום זה היא לאתר את התחומים העיקריים שבהם יש צורך בהתאמה של הרגולציה או מדיניות הרגולציה בתחום הבינה המלאכותית, כדי להסיר חסמים לפעילות כלכלית או חברתית רציפה ויעילה או לגבש מענים מותאמים לצרכים. במסגרת פורום זה, ניתן יהיה לדון בהתפתחויות הטכנולוגיות, בשאלה אם כלי רגולציה מסוימים נותנים מענה לאתגרים השונים ולעקרונות שפורטו לעיל ובעניינים נוספים.

5. **מיפוי השימושים בבינה מלאכותית והאתגרים הנלווים להם בענפים מוסדרים**

בשנים האחרונות חל גידול משמעותי בהיקף השימושים במערכות מבוססות בינה מלאכותית והן מוטמעות כמעט בכל ענף במשק. בהתאם, ועל מנת שהחלטה על רגולציה נדרשת או נכונה תתקבל על בסיס מצע עובדתי מספק, מוצע כי גורמי הממשלה והרגולטורים הרלוונטיים יפעלו להבנה ומיפוי של השימושים הקונקרטיים במערכות מבוססות בינה מלאכותית הנעשים בענף המוסדר שעליו הם אמונים; האתגרים, החששות והסיכונים הכרוכים בכך; והמענים האפשריים.

6. **מעורבות אקטיבית בפיתוח הרגולציה והתקינה בפורומים בינלאומיים**

מוצע כי גורמי ממשלה הפועלים במסגרת פורומים וארגונים בין-לאומיים, או עומדים בקשרי עבודה עם מדינות מפותחות אחרות בתחום זה יפעלו יחד לקידום מדיניות רגולציה ואתיקה מאוזנת ותואמת למדיניות רגולציה זו וליעדים הממשלתיים. זאת, בשים לב להחלטת ממשלה 212, שהטילה על שרת החדשנות, המדע והטכנולוגיה "להוביל את שיתופי הפעולה הבין-לאומיים האזרחיים בתחום, לייצג את ישראל בפורומים בין-לאומיים אזרחיים, לרבות פורום השרים של ארגון ה-OECD".הצעה זו נובעת מהבנה כי לתקינה הבינלאומית שתגובש, כמו גם לתוצרי עבודה נוספים של מדינות וארגונים בינלאומיים העוסקים בסוגיות אלה, תהיה השפעה גלובלית משמעותית, כולל על הנעשה בישראל ועל יכולתו של המשק הישראלי להיות מוביל עולמי בתחום הבינה המלאכותית.

7. **התוויית שפה אחידה באמצעות מסגרת מומלצת (וולונטרית) לניהול סיכונים**

נוכח החשיבות של גיבוש מדיניות ממשלתית אחידה, וכן מתוך הרצון להגביר את הוודאות המשפטית בתחום זה, מוצע לפתח או לאמץ כלי אחיד לניהול סיכונים ביחס לשימוש בבינה מלאכותית, שייצור שפה משותפת בין גורמי הממשלה והרגולטורים ובינם לבין גורמים פרטיים.

השפה האחידה תסייע למגזר הפרטי להעריך את הסיכונים הנלווים לשימוש מסוים בבינה מלאכותית, וכן לרגולטורים לבחון את הסיכונים הכרוכים בשימוש בה והצורך בהתערבות.

1. חלק ראשון: רקע

1.1. מהי בינה מלאכותית?

בשנים האחרונות הפך המושג בינה מלאכותית למטבע לשון שגורה הבאה לתאר מכונות (מחשבים) הפועלות באופן שיכול להיתפש כחכם, מורכב או תבוני. ואולם, ריבוי הניסיונות ליצירת הגדרה פורמאלית עבור הבינה המלאכותית מרמז כי מדובר במשימה שאינה פשוטה כלל.

את הניסיונות השונים להגדרת הבינה המלאכותית ניתן לחלק לשתי קבוצות עיקריות – הגדרות טכנולוגיות והגדרות פונקציונאליות. ההגדרות הטכנולוגיות מתרכזות על פי רוב ביכולות הבסיסיות של כלי הבינה המלאכותית, ובשיטות ליישומן, וניתן לחלקן לשתי שיטות מרכזיות: האחת, מזוהה עם תחום מדעי הנתונים (data science)[2] ומתארת את הבינה המלאכותית בתור כלל הפעולות הטכנולוגיות למיצוי מידע והפקת תובנות מתוך מאגרי נתונים. השנייה, מזוהה עם תחום למידת מכונה (machine learning) ולפיה בינה מלאכותית היא היכולת של מכונה ללמוד לבצע פעולה ולטייב את ביצועה, בהסתמך על נתונים, דוגמאות וניסיון מצטבר. ראוי לציין כי שתי שיטות אלו גם יחד, אינן מתייחסות למושגים כגון "תבוניות" או "חשיבה", אלא הן מתייחסות למהות הטכנולוגית של איסוף ועיבוד נתונים במטרה להגשים משימה אלגוריתמית נתונה. לעומת זאת, הגדרות פונקציונאליות מתייחסות או מגדירות בינה מלאכותית באמצעות השוואתה לתהליכי חשיבה או הסקה אנושיים. הדגש בסוג זה של הגדרה הוא על יכולתן של מכונות לפעול באופן אשר יכול להיתפש כתבוני, או אשר מחקה את דפוסי הפעולה האנושיים. לצד זאת, במקרים רבים אוסף הפיתוחים הטכנולוגיים אשר חוסה תחת מטריית הבינה המלאכותית, משמש למספר גדול של משימות בהן לא יכולה הבינה האנושית להוות מדד או גורם מתאים להשוואה.

לצורך הדיון במסמך זה, תחום הבינה המלאכותית הוא שם כולל להתפתחות בתחום טכנולוגיות המידע והתקשורת ומדעי הנתונים[3] המאפשרת קבלת החלטות, ייצור תחזיות, או ביצוע פעולות על ידי מחשב ברמת עצמאות גבוהה, באופן המדמה או מסוגל להחליף בינה אנושית.

---

[2] בין בינה מלאכותית למדעי הנתונים קיים קשר משמעותי. ראו את מסקנות דוח ההיגוי המייעצת לות"ת לנושא מדעי הנתונים: המועצה להשכלה גבוהה, הוועדה לתכנון ולתקצוב, דוח ועדת ההיגוי המייעצת לות"ת לנושא מדעי הנתונים, דצמבר 2020, זמין כאן.
בעמוד 10: "תחום מדעי הנתונים (data science) מתייחס לאיסוף, ניהול, עיבוד, ניתוח וויזואליזציה של נתונים המשוייכים למנעד רחב של דיסציפלינות אקדמיות ואפליקציות מסחריות, כמו גם לחיזוי, ניתוח משמעות, הוצאת מסקנות ופיתוח כלים הנשענים על הנתונים הנאספים. **תחומי הליבה** של המחקר במדעי הנתונים (מסודרים להלן לפי "מחזור החיים" של נתונים) הינם: שיטות לאיסוף נתונים על ידי ניסויים, דגימה, כרייה ומנועי חיפוש; שיטות לאחסון והנגשת הנתונים לניתוח, כולל חישוב (מבוזר) על מערכות בסיסי נתונים/מידע, ושל נתוני עתק; פיתוח מתודולוגיות ניתוח נתונים כגון רשתות עצביות ולמידה עמוקה, למידה על ידי חיזוקים ובקרה מסתגלת, למידה סטטיסטית והסקה סטטיסטית; בניית אלגוריתמים ומודלים למיון וחיזוי, תיקופם, וניצולם ליצירת ידע והנגשתו (ויזואליזציה); התייחסות לנתונים מיוחדים כגון ניתוח תמונות וראייה ממוחשבת, עיבוד שפה טבעית, דיבור ושמיעה; ניתוח רשתות, נתונים אורכיים, נתונים עתיים ומרחביים ונתוני הישרדות. כל זאת תוך התייחסות להיבטים אתיים ביחס לפרט (כגון פרטיות וחיסיון), ולחברה (כגון הדירות ותקפות השימוש בתוצאות).
תחומי המעטפת של המחקר במדעי הנתונים הינם תחומים בעלי זיקה עמוקה ואפילו חפיפה חלקית עם תחומי הליבה, וכוללים, בין היתר: בינה מלאכותית (החלק שאינו מכוסה בתחומי הליבה), אופטימיזציה, חקר ביצועים, רובוטיקה, תורת המשחקים, תהליכים סטוכסטיים, אינפורמציה, בקרה, ניתוח אותות, ביו-אינפורמטיקה, ביולוגיה חישובית, רפואה דיגיטלית/מותאמת אישית, אפידמיולוגיה, אקונומטריקה, פסיכומטריקה, סייבר, חישוב קוונטי, וכן ניהול ובניית בסיסי נתונים/מידע".

[3] שם.

יצוין, כי ההגדרה שמציע ה-OECD לבינה מלאכותית מתמקדת בהיבט הפונקציונלי של "מערכת" העושה שימוש בבינה מלאכותית, והיא גובשה בוועדת מומחים בשנת 2018.[4] הגדרה זו נכללה בהמלצות ה-OECD בשנת 2019 (ראו להלן פרק "המלצות ה-OECD בתחום הבינה המלאכותית").[5] בהתאם להגדרה זו, מערכת בינה מלאכותית היא "מערכת ממוכנת שיכולה, בהתאם למטרות מוגדרות בידי אדם, להפיק תחזיות, המלצות, או החלטות המשפיעות על סביבות פיזיות או וירטואליות" כאשר "מערכות בינה מלאכותית מתוכננות לפעול ברמת עצמאות משתנה".[6] הגדרה דומה נכללה בטיוטת חקיקה באיחוד האירופי, הכוללת גם פירוט של סוגי הטכנולוגיות בנספח לחקיקה (ראו שם),[7] ובהגדרה המוצעת בטיוטת חקיקה קנדית.[8]

הבינה המלאכותית מבוססת כאמור במידה רבה על למידת מכונה, כלומר תהליכים שבהם התוכנה "לומדת" מהמידע שהיא אוספת או מקבלת ומפיקה תובנות הנדרשות לפעולתה. "לימוד" המכונה נעשה בין היתר דרך הצגה של דוגמאות רבות לתוצאות המבוקשות, וניסוי וטעיה.[9] למידת מכונה מאפשרת יצירת "מודל" המייצג תובנות מהמידע, ויכולה לשמש גם לפרש את תוצאותיו.[10]

הבינה המלאכותית מסוגלת לקבל החלטות או לגבש המלצות באופן עצמאי ומהיר, ובכך להחליף שיקול דעת אנושי בשורה של מצבים כאשר יש לה יתרונות בהיקף הפריסה ובמהירות התגובה. כמו כן, בינה מלאכותית מסוגלת לייצר תובנות ותחזיות על בסיס נתונים רבים באופן שפעמים רבות נחשב לבלתי אפשרי בידי אדם.

בינה מלאכותית מבוססת למידת מכונה נשענת בהקשרים רבים על היכולת של המכונה ללמד את עצמה, ובפועל "לכתוב לעצמה" את התוכנה באופן דינמי. כשבינה מלאכותית מבוססת על למידת

---

[4] OECD, AI in Society, 2019, Chapter 1, https://www.oecd-ilibrary.org

[5] OECD Council, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449 (Adopted on May 22, 2019; 2021), available at https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449 (להלן-המלצות ה-OECD).

[6] ההגדרה במסמך ה- OECD - *AI system*: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

[7] ההגדרה בטיוטת החקיקה שפורסמה בידי נציבות האיחוד האירופי,
'artificial intelligence system' (AI 'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;
EU Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Mending Certain Union Legislative Acts, 21.04.21, article 3(1) https://eur-lex.europa.eu/legal-content (EU Draft AI Regulation);
טיוטת החקיקה מצוייה בדיונים ועוברת שינוים.

[8] HOUSE OF COMMONS OF CANADA, BILL C-27, An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts
Part 3, article 2.
artificial intelligence system means a technological system that, autonomously or partly autonomously, processes data related to human activities through the use of a genetic algorithm, a neural network, machine learning or another technique in order to generate content or make decisions, recommendations or predictions. (*système d'intelligence artificielle*)
https://www.parl.ca/DocumentViewer.

[9] אופן הלימוד מושפע בין היתר מהיבטים אלה:
א. למידה מפוקחת או לא מפוקחת.
ב. למידה אקטיבית מול למידה פסיבית.
ג. מידת ה-"סיוע" המתקבל מהגורם המלמד.
ד. קיום שלב מקדים של למידה לפני הפעלת המערכת בתנאי אמת, לבין מערכת שנפרסת ללא שלב מקדים של למידה.

[10] השיטות בלמידת מכונה כוללות שיטות ששימשו חוקרים בעבר. שיטות אלה כוללות רגרסיה לינארית ולוגית, עצי החלטה, principle component analysis וגם רשתות נוירונים עמוקות (deep neural networks).

מכונה, לצד התועלות בכך, משמעותה המעשית היא שהתוכנה "מגדירה לעצמה" את אופן פעולתה. היבט זה עשוי להקשות על התחקות אחר הליך קבלת ההחלטות של המכונה שהוביל לתוצאה.

ועדת המומחים של ה-OECD ציינה כי אופן פיתוח בינה מלאכותית דומה לאופן פיתוח תוכנה, ותיארה את השלבים הבאים בפיתוח ושימוש בבינה מלאכותית: (1) עיצוב (design) – איסוף נתונים ובניית מודל קבלת ההחלטה; (2) בדיקה ואימות של המודל וכיול שלו, ובדיקת ביצועים במימדים ובשיקולים שונים; (3) ביצוע פיילוט בתנאי ההפעלה של המערכת; (4) הפעלה של המערכת ובקרה רציפה על עמידתה בהגדרות ובדרישות השונות בשימוש בפועל.

לבינה המלאכותית שימושים רבים, כגון נהיגת כלי רכב עצמאיים (אוטונומיים), פענוח תצלומי רנטגן, הערכת סיכוני אשראי, מסחר בניירות ערך ובחינת מועמדים לתעסוקה. בהתאם מערכות בינה מלאכותית מעורבות ביחסים בין צרכנים לעוסקים; בין עסקים לעסקים אחרים; בין בעלי מקצוע ללקוחות; ביחסי עבודה; בין גופים במגזר הציבורי; ובין המגזר הציבורי לכלל הציבור.

1.2. <u>חשיבות מדיניות רגולציה ואתיקה בתחום הבינה המלאכותית</u>

מדיניות הרגולציה והאתיקה בתחום הבינה המלאכותית, המוצעת במסמך זה, נועדה להתוות את האופן שבו תבחן הממשלה את השאלה האם יש לקבוע כללי התנהגות לפיתוח בינה מלאכותית או שימוש בה במגזר הפרטי ואת האופן לעשות כן. זאת, במטרה לקדם פיתוח ושימוש בטכנולוגיות בינה מלאכותית, ובד בבד לצמצם פגיעות אפשריות בזכויות יסוד ואינטרסים ציבוריים.

הצורך בגיבוש ואימוץ מדיניות רגולציה ואתיקה, ופעילות ליישומה, נובע מכמה גורמים מרכזיים:

<u>ראשית</u>, מאפייני הבינה המלאכותית עשויים לעורר שאלות בעניין התערבות רגולטורית, ומדיניות רגולציה ואתיקה תסייע בבחינת הצורך בהתערבות רגולטורית ואופי ההתערבות.

בהתאם לתפיסות מקובלות של רגולציה, יש לבחון התערבות באופן הפעילות של המגזר הפרטי, באמצעות כללים רגולטוריים (דהיינו, כללים מחייבים וניתנים לאכיפה שנועדו להסדיר פעילות כלכלית או חברתית המבוצעת על ידי גורמים פרטיים[11]), בעיקר במקום שבו קיימים כשלי שוק או כאשר יש חשש לפגיעה בזכויות יסוד או באינטרסים ציבוריים, שאינם מטופלים כראוי על ידי מנגנוני השוק.[12] זאת, כמובן, תוך מודעות לעלות הכרוכה בהתערבות הרגולטורית, ותוך השוואתה לעלות הכרוכה בכשל השוק או בפגיעה באינטרס המוגן כאמור.

בכבל תהליך של קביעת רגולציה, בעת קידום רגולציה לגבי טכנולוגיה חדשה, עולה החשש מפני אמצעים רגולטוריים שעלולים להוות נטל עודף, ובכך עלולים לפגוע, שלא לצורך, בקידום החדשנות הטכנולוגית. אתגר זה מלווה את המדינות השונות והרגולטורים השונים הפועלים בהן, בעת שהם בוחנים שאלות הנוגעות לקביעת רגולציה בתחום הבינה המלאכותית.

[11] ראו הגדרת "אסדרה" בסעיף 3 לחוק עקרונות האסדרה, התשפ"ב-2021.
[12] ראו סעיף 1 לחוק עקרונות האסדרה, התשפ"ב-2021.

על רקע זה, המדיניות המוצעת יכולה לסייע בבחינת הצורך בהתערבות רגולטורית ובבחינת אופייה. זאת, בין היתר בשים לב לאיזון הנדרש בין הרצון לקדם חדשנות טכנולוגית לבין הצורך להגן על זכויות יסוד ואינטרסים ציבוריים, ולפי תפיסת ניהול הסיכונים הבוחנת את היחס בין עוצמת הסיכון והסיכון להתרחשותו, לבין סוג האמצעים הנדרשים לניהולו.

<u>שנית</u>, כללים רגולטוריים עלולים לשמש כחסם בפני פיתוח או שימוש בבינה מלאכותית, ומדיניות רגולציה ואתיקה תועיל בהכוונה לצמצום מאוזן ונכון של חסמים אלה.

רגולציה מחילה מגבלות או איסורים על פעילות של גורמים פרטיים, ובהתאם לכך לעיתים היא עלולה למנוע או להגביל פיתוח בינה מלאכותית או שימוש בה. כך למשל, רגולציה עשויה לאפשר פעילות מסוימת על ידי אדם בלבד (כגון נהיגה עם "ידיים על ההגה" או זיהוי "פנים-אל-פנים" בפתיחת חשבון בנק) או על ידי איש מקצוע בלבד (כגון אבחון רפואי או ייעוץ משפטי), ובכך לחסום או להגביל מערכת אוטונומית מבוססת בינה מלאכותית המסוגלת לבצע פעילות דומה.

בהתאם לכך, המדיניות המוצעת עשויה להוביל ולתמוך בקידום מהלכים להפחתת חסמים, הנובעים מרגולציה קיימת או חדשה שאינה מותאמת לטכנולוגיה מבוססת בינה מלאכותית, וכפועל יוצא מונעת או מגבילה פיתוח ושימוש בבינה מלאכותית שלא לצורך. זאת, בין היתר, באמצעות עיצוב נכון של הכללים הרגולטוריים או על ידי שימוש בכלי רגולציה מתקדמים כגון סעיפי "שקיעה" (sunset clauses) ו"ארגזי חול" רגולטוריים (regulatory sandbox).

<u>שלישית</u>, מדיניות רגולציה ואתיקה צפויה להגביר את הוודאות העסקית, ולחזק את אמון הציבור בטכנולוגיות בינה מלאכותית, ובכך לעודד פיתוח ושימוש בבינה מלאכותית.

ככל טכנולוגיה חדשה, בחלק מהשימושים בבינה מלאכותית צפויה אי-ודאות לגבי אופן ההשפעה שלהם על זכויות יסוד ואינטרסים ציבורים, כמו גם על בעלי עניין בהקשרים שונים. לחששות אלה עלולה להיות השפעה כפולה: האחת, שגורמים עסקיים הפועלים לפיתוח והפצה של מערכות מבוססות בינה מלאכותית ימנעו מכך בשל החשש מרגולציה עתידית שתהיה חוסמת או מגבילה; השנייה, שהציבור הרחב שמיועד להשתמש במערכות אלה או להיות מושפע מהן, יירתע מכך.

לפיכך, מעבר להגנה על זכויות יסוד ואינטרסים ציבורים, למדיניות המוצעת חשיבות גם בהקניית ודאות לגבי אופן בחינת קידום כללים והסדרים שיחולו על בינה מלאכותית, במידת הצורך, וזאת, בהלימה לנעשה במדינות המפותחות שהם שווקי היעד של הטכנולוגיה. לקיומה של מדיניות בהירה, המעידה על בחינת האינטרסים השונים בידי הגורמים הציבוריים המופקדים עליהם, תפקיד חשוב בקידום הוודאות לעוסקים בתחום ולהגברת אמון הציבור בשימושים חדשניים בטכנולוגיה.

2. חלק שני: המצב המשפטי והרגולטורי בהיבטי בינה מלאכותית בישראל

2.1. התפתחות המשפט והרגולציה להסדרת בינה מלאכותית

לצד היתרונות למשק ולחברה הנובעים מההתפתחות הטכנולוגית בכלל, ומהתפתחות טכנולוגיות בינה מלאכותית בפרט, וכפי שבין היתר יתואר בחלקים הבאים בהרחבה, התפתחויות אלה מייצרות גם אתגרים משמעותיים עבור המערכת המשפטית והרגולטורית, ברחבי העולם ובישראל.

הניסיון מלמד כי בין התפתחויות טכנולוגיות באשר הן, לבין המערכת המשפטית והרגולטורית, נוצרות פעמים רבות נקודות חיכוך. זאת, בעיקר משום שלא אחת הטכנולוגיות החדשות משנות את מצב הדברים שהיה קיים טרם הופעתן או את אופן התנהגותם של יחידים והחברה, ואגב כך מעוררות סוגיות חברתיות, משפטיות ורגולטוריות (יש המתייחסים למצב זה כ"חדשנות משבשת" (disruptive innovation) ו"שיבוש רגולטורי" (regulatory disruption)).[13]

כתוצאה מכך, לא אחת עולה צורך בהתאמה של כללי המשפט והרגולציה הקיימים עקב הופעת טכנולוגיות חדשות, ובעשורים האחרונים מקבלי ההחלטות עוסקים במלאכה זו.[14] ההתאמה נעשית, ככלל, באמצעות פרשנות, תיקון הכללים, או קביעת כללים חדשים. מלאכה זו אינה קלה, וכרוכה בהתמודדות עם קשיים מהותיים, כגון הקצב המהיר של ההתפתחות הטכנולוגית, הצורך במידע ובמומחיות לשם הסדרתה, והמאפיינים הבינלאומיים הנלווים לה. עמד על הקושי הכרוך במלאכה זו, בין היתר, השופט סולברג בפרשת איגוד האינטרנט הישראלי:

> "בידוע, כי המשפט מדדה בעצלתיים אחר חידושי העולם, וכי החקיקה אינה מדביקה את קצב התקדמות המדע וישׂומיו. מפרי-החוק מסתגלים לקידמה מהר יותר מאוכפיו. זו אקסיומה. לראשונים אין עכבות; לאחרונים יש. שנים רבות חלפו מעת המצאת המחשב ועד לחקיקת חוק המחשבים, התשנ"ה-1995. לא חלפו אז דור או שניים במונחי מחשב, והחוק כבר נמצא מיושן, כי המחוקק לא צפה – ולא יכל לצפות – את חידושי הטכנולוגיה. אך לא רק עולם המשפט עומד נבוך. גם מדע הפסיכולוגיה נתקל בתופעות חדשות של התמכרות ופגיעות בנפש, ומנסה ללמוד דרכי התמודדות עדכניות 'תוך כדי תנועה'; כך גם הסוציולוגיה, ושאר ענפים מתחום מדעי החברה, הטבע והרוח. אין תימה אפוא שגם עולם המשפט איננו ערוך עדיין עם מלוא הכלים העומדים לרשותו".

---

[13] Nathan Cortez, *Regulating Disruptive Innovation* 29 Berkeley Technology Law 175 (2014).

[14] כך לדוגמה, בישראל הוקמו בשנים האחרונות מספר ועדות שעסקו בהתאמת המשפט והרגולציה להתפתחויות טכנולוגיות, עליהן נמנות, בין היתר: ועדת המשנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית (2019); צוותי העבודה הבין-משרדיים לגיבוש התכנית הלאומית לבינה מלאכותית, בהובלת משרד החדשנות, הטכנולוגיה והמדע בהתאם להחלטת ממשלה 212 (2021); הוועדה להתאמת המשפט לאתגרי החדשנות והאצת הטכנולוגיה (2021); הוועדה לבחינת אסדרת פעילות הרשתות החברתיות (2021); הוועדה לגיבוש אמצעים להגנה על הציבור ונושאי משרה בשירות הציבור מפני פעילות ופרסומים פוגעניים, כמו גם בריונות ברשת האינטרנט (2020); הוועדה הציבורית לבחינת חוק הבחירות (דרכי תעמולה), התשי"ט-1959; הוועדה לבחינת אסדרה של הנפקת מטבעות קריפטוגרפיים מבוזרים לציבור (2019); והצוות המקצועי לבחינת "כלכלת הפלטפורמה" – צורות העסקה ייחודיות שנוצרו במשק, בהובלת זרוע העבודה (2022).

> "[...] אין לכחד: הסדרה חוקית של הנעשה במרחב הוירטואלי היא מורכבת ומסובכת. קשה להגדיר את הדין הרצוי, וגם לא בנקל ניתן להחיל את הדין המצוי. לא בכדי יש שהגיעו למסקנה כי מדובר בתחום שראוי להסדרה חוקית; יש הסבורים כי הפסיקה היא המסגרת הראויה לעשיית ההתאמות הנדרשות לעידן האינטרנט; ואלה ואלה מתלבטים גם לגבי מידת השיתוף של קהילת משתמשי האינטרנט בעיצוב הכללים במרחב הוירטואלי וביישומם".

הנחיית היועץ המשפטי לממשלה מס' 1.2500,[15] עוסקת באופן קידום ופיתוח המשפט במסגרת היחסים שבין טכנולוגיה, ערכים ומשפט. ההנחיה עוסקת ב"ביצוע המעבר מן העולם הפיסי אל העולם הדיגיטלי, תוך ניהול סיכונים וקיום התכליות הנדרשות", הן בקשר "לשירותים ממשלתיים וציבוריים" והן במסגרת "הסדרים החלים על השוק הפרטי", המוגדרים כ-"הסדרים דיגיטליים". ההנחיה מתייחסת לגישה הפרשנית בעת עיצוב הסדר דיגיטלי, הכוללת בין היתר בדיקה לאיתור התכליות של ההסדר המשפטי שהופך לדיגיטלי, הסיכונים במעבר להליך דיגיטלי, ואופן ההתמודדות איתם באמצעים טכניים, נהלים ומשפטיים. על פניו, נראה כי לגישה זו רלוונטיות גם להחלת הדין הקיים על פיתוח ושימוש במערכות מבוססות בינה מלאכותית.

טכנולוגיות הבינה המלאכותית מצויות כבר כיום בשימוש בידי גופים פרטיים וציבוריים בישראל ובעולם. בדומה לאופן שבו התפתחו המשפט והרגולציה ביחס לטכנולוגיות אחרות, ניתן להצביע על מאפיינים מרכזיים הרלוונטיים גם לאופן קידום ופיתוח המשפט והרגולציה ביחס לטכנולוגיות בינה מלאכותית. המאפיין העיקרי והחשוב ביותר הוא שמערכת הדינים החלה על כל פעילות באופן כללי, הן בתחום המשפט הפרטי והן בתחום המשפט הציבורי, חלה גם על בינה מלאכותית.

כפי שיתואר להלן, השימוש במערכות מבוססות בינה מלאכותית מעורר סוגיות ואתגרים שונים, ובכלל זה אפליה; שקיפות; צורך במתן הסבר; אבטחת מידע והגנת סייבר; ופרטיות. ניתן להניח, כי חלק מסוגיות ואתגרים אלו יצריכו חשיבה מחודשת על הדין הקיים ואולי אף קביעת כללים חדשים במקרים המתאימים; אולם נראה כי חלק מרכזי מהסוגיות והאתגרים האלה צפויים לקבל מענה במסגרת הדין הקיים, בין אם במישרין ובין אם באמצעות פרשנות מתאימה.[16]

---

[15] הנחיית היועץ המשפטי לממשלה 1.2500, כללים מנחים לגיבוש הסדרים דיגיטליים, י"א תשרי התש"ף, 10 אוקטובר 2019.

[16] הסקירה בדבר אופן הפיתוח המשפטי של הדין באופן כללי לגבי בינה מלאכותית, ובתחומי משפט קונקרטיים מחייבת בחינה מפורטת נוספת. באופן כללי נראה כי הגישה העולה מהנחיית היועץ המשפטי לממשלה 1.2500 רלבנטית במובן זה שבחלק מהשימושים החברתיים של בינה מלאכותית ניתן יהיה להתאים את ההסדר הנורמטיבי באמצעות פרשנות הדין הקיים בכלים הקיימים, אולם לעיתים שיקולים הקשורים בשינוי חלוקת אחריות או שינוי נטלי ראיה יובילו לצורך בהסדר ספציפי. ראו הנחיית היועץ המשפטי לממשלה 1.2500, בעמוד 12. בספרות הועלו שאלות הקשורות באחריות, סיבתיות, ונזק הקשור בתוכנות ובבינה מלאכותית. המענה על שאלות אלה ייבחן בהמשך.

2.2. **העיסוק הממשלתי בבינה מלאכותית**

החלטת ממשלה 212, כמו גם העבודה שנעשתה לקראת עריכת מסמך זה, התבססו, בין היתר, על עבודות קודמות שנעשו במהלך השנים האחרונות בנוגע לתחום הבינה המלאכותית.

בשנת 2018 ריכז המטה לביטחון לאומי במשרד ראש הממשלה (להלן: המל״ל) את הבחינה האסטרטגית בכל הנוגע לטכנולוגיית הבינה המלאכותית. במסגרת זאת ריכז המל״ל, בין היתר, את הקשר עם ״המיזם הלאומי למערכת נבונות בטוחות״ בראשות פרופ׳ איציק בן ישראל ופרופ׳ אביתר מתניה,[17] כאשר תחום האתיקה והמשפט נדון בצוות משנה של המיזם, בראשות פרופ׳ קרין נהון ובהשתתפות מומחים מהאקדמיה, מהממשלה ומהתעשייה.[18] הוועדה קראה לבנות תכניות ידע והכשרה בתחום, לאמץ מערכת עקרונות אתיים בחברות וארגונים, לזהות עקרונות אתיים החשובים למקבלי החלטות ולהטיל אחריות על כלל העוסקים בבינה מלאכותית לפעול באופן חוקי ואתי. בנוסף, הוועדה פיתחה כלי עבור מקבלי ההחלטות, לבחינת האתגרים האתיים במערכות בינה מלאכותית. בתחום הרגולציה, הוועדה מיפתה את סוגי הרגולציה הקיימים והמליצה, בין היתר, להקים מנגנון תיאום פנים-ממשלתי על-משרדי במטרה ליצור מדיניות רגולטורית אחידה, ברורה וקוהרנטית; לבחון את התפיסה של סביבה רגולטורית מבוקרת; לסייע לרשויות המאסדרות את משאבי המידע בהסרת חסמים; ולהטיל על רשות התחרות לגבש דרכי התמודדות שמטרתן שמירה על תחרות הוגנת, על צרכנים ועל נגישות הטכנולוגיה.[19] בנוסף המל״ל סייע לרכז את עמדות ישראל לצורך ליווי הדיונים ב-OECD על עקרונות לתחום הבינה המלאכותית בשנת 2019.

בהמשך לכך, בשנת 2020 הפורום לתשתיות לאומיות למחקר ופיתוח (פורום תל״ם) הקים ועדת בדיקה לבחינת הצורך בהתערבות ממשלתית לשם האצת התפתחות תחום הבינה המלאכותית ומדע הנתונים, בראשות ד״ר ארנה ברי.[20] באותה שנה פרסמה הוועדה את המלצותיה, ובמסגרת זו, בין היתר, המליצה על יצירת סביבה רגולטורית מאפשרת ומעודדת לבינה מלאכותית.

בשנת 2021 הפיצו מחלקת ייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים והרשות הלאומית לחדשנות טכנולוגית (להלן: רשות החדשנות) ״קול קורא״ משותף, שהזמין פניות ציבור לגבי חסמים רגולטוריים ורגולציה אפשרית לתחום הבינה המלאכותית בישראל, בדגש על ניסוי והטמעה של טכנולוגיות בתחום זה,[21] וזאת על רקע עיסוקן בבינה מלאכותית ובנסיינות רגולטורית.

---

[17] בן ישראל, י, מתניה, א׳ ופרידמן ל׳, (עורכים), (ספטמבר 2020), המיזם הלאומי למערכות נבונות בטוחות להעצמת הביטחון הלאומי והחוסן המדעי-טכנולוגי: אסטרטגיה לאומית לישראל, דו״ח מיוחד לראש הממשלה, חלק א׳: תמצית והמלצות, ספטמבר 2020.
https://www.google.com/url
בהתאם לאמור בדין וחשבון של המיזם, ראשי המיזם התניעו את המיזם ביולי 2018, וגייסו מאות מומחים מהממשלה מערכת הביטחון האקדמיה והתעשייה, אשר פעלו בהתנדבות - 14 צוותים, וצוות מתכלל.

[18] פרופסור קרין נהון (יו״ר) ואח׳, **דין וחשבון של ועדת משנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית**, נובמבר 2019 http://ekarine.org.

[19] שם.

[20] פורום תל״ם בראשות ד״ר ארנה ברי, ועדת בינה מלאכותית ומדע הנתונים, דצמבר 2020, https://www.google.com/url.

[21] ראו: המחלקה הכלכלית במחלקת ייעוץ וחקיקה במשרד המשפטים הרשות הלאומית לחדשנות טכנולוגית, פנייה לציבור לקבלת מידע בנוגע לרגולציה וחסמים רגולטוריים ביחס לתחום הבינה המלאכותית, 15.06.21, https://innovationisrael.org.il.

בנוסף, בשנת 2021 הפיצה מחלקת ייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים "קול קורא" לקבלת סקירה בנושא יישומי בינה מלאכותית בתחום השירותים הפיננסיים. בהמשך לכך, בשנת 2022 הוגש למחלקה ופורסם לציבור דו"ח בנושא "בינה מלאכותית במגזר הפיננסי: שימושים נפוצים, אתגרים וסקירה השוואתית של התמודדות רגולטורית", המסכם מחקר שביצעו חוקרים בכירים מאוניברסיטת תל-אביב (להלן: דו"ח בינה מלאכותית בשירותים פיננסיים).[22] הדו"ח כולל סקירה מקיפה של השימושים הנפוצים בבינה מלאכותית במגזר הפיננסי (ובכלל זה שימוש בבינה מלאכותית לחיתום אשראי וביטוח, לפעולות מסחר בניירות ערך, לייעוץ השקעות וניהול כספים, לשירות לקוחות ולצ'אט בוטים), ובחינה של ההזדמנויות הנובעות מהשימוש בטכנולוגיה זו לצד האתגרים המשפטיים והרגולטוריים שהיא מייצרת. בנוסף, הדו"ח מתייחס לאתגרים הכלליים שמעוררת הבינה המלאכותית בתחומים נוספים, וכולל הצגה מפורטת של יוזמות ההסדרה בעולם. כמו כן, הדו"ח כולל את המלצות החוקרים באשר לאופן שבו יש להמשיך את העיסוק בתחום זה.

בנוסף, קודמו בשנים האחרונות פעולות סקטוריאליות הקשורות להסדרת היבטים שונים של שימוש בבינה מלאכותית בהקשרים ספציפיים (כגון כלי רכב עצמאיים (אוטונומיים) וסוגיות של פרטיות. להרחבה על אודות אלה ראו להלן בחלקים "דיני פרטיות ו"הסדרה ייעודית מגזרית").[23]

בהמשך לעבודות הנסקרות לעיל, במסגרת החלטת הממשלה 212 הוחלט על השקעה משמעותית בתחום זה, תוך ריכוז ההובלה לגיבוש התוכנית הלאומית בידי שרת החדשנות, המדע והטכנולוגיה, בשיתוף עם גופים נוספים, ובכלל זאת הוועדה לתכנון ולתקצוב הפועלת במועצה להשכלה גבוהה (ות"ת) ורשות החדשנות. ברקע הדברים יצוין כי במהלך השנים פעלו הות"ת ורשות החדשנות לקידום תחום הבינה המלאכותית. כך, בין היתר הות"ת פועלת לפיתוח והגדלת מספר החוקרים והסגל האקדמי בתחום מדעי הנתונים כחלק מפעילותה האסטרטגית בתחום זה. כמו כן, החל משנת 2018 ממקדת רשות החדשנות מאמצים בקידום מטרותיה בתחום הבינה המלאכותית.[24] מאמצים אלה כוללים פעילות תמיכה ישירה במסגרת תפקידיה (כגון תמיכה בפרויקטים טכנולוגיים ובהון אנושי) ופעילות לקידום מדיניות שמטרתה לעודד את החדשנות (כגון אפיון תשתיות טכנולוגיות וקידום רגולציה תומכת).[25] ברשות החדשנות פועלת גם תכנית תל"מ לבינה מלאכותית, שמטרתה תכלול וניהול התשתיות הלאומיות לבינה מלאכותית בפורום תל"מ.

---

[22] ראו: משרד המשפטים, מחלקת ייעוץ וחקיקה (משפט כלכלי) וכן ד"ר ארי אחיעז, פרופ' אסף חמדני, פרופ' דן עמירם, וד"ר קובי קסטיאל מאוניברסיטת תל אביב, בינה מלאכותית במגזר הפיננסי: שימושים נפוצים, אתגרים וסקירה השוואתית של התמודדות רגולטורית https://www.gov.il/he/departments/news/ai_report.

[23] בשנת 2020, פרסמה הרשות להגנת הפרטיות במשרד המשפטים סקירה בנושא "דירוג חברתי בראי הזכות לפרטיות", ובה מציגה הרשות את הסיכונים הכרוכים בשימוש במידע רב ובנתוני עתק (big data) ובאלגוריתמים, לצורך "דירוג חברתי" של אנשים להשגת מטרות שלטוניות. בשנת 2022 פרסמה הרשות להגנת הפרטיות מסמך בנושא "חובת יידוע במסגרת איסוף ושימוש במידע אישי", ובמסגרתו נאמר כי חובת הגילוי לפי סעיף 11 לחוק הגנת הפרטיות, התשמ"א-1981, החלה בעת איסוף מידע מאדם, חלה גם בהקשר של **"מערכות לאיסוף מידע וקבלת החלטות מבוססות אלגוריתם או בינה מלאכותית."** הרשות להגנת הפרטיות מציינת "הוראות כי סעיף 11 לחוק הגנת הפרטיות מטילות חובת יידוע בשלב איסוף מידע גם על גורמים האוספים מידע אישי באמצעות מערכות לקבלת החלטות מבוססות אלגוריתם (לרבות מערכות מבוססות בינה מלאכותית) וגם על גורמים האוספים מידע אישי לשם שימוש בו במסגרת מערכות שכאלו".

[24] רשות החדשנות, **מרוץ העוצמה הטכנולוגית ממשיך מה נדרש כדי שישראל תמשיך להוביל את תחום הבינה המלאכותית**, דו"ח החדשנות 2019-תמונת מצב (להלן – דו"ח החדשנות), https://innovationisrael.org.il .

[25] דוח החדשנות 2019.

2.3. תחולה כללית של הדין והרגולציה בהיבטי בינה מלאכותית

חלק זה ימשיך בסקירה של הדין הכללי בישראל, מתחומים שונים, אשר חל, ככלל, גם על מצבים המערבים בינה מלאכותית. נציין דוגמאות שונות לתחולה אפשרית של הדין הקיים, באופן שמסדיר פיתוח מערכות מבוססות בינה מלאכותית או שימוש בהן. לשם הגיוון, דוגמאות אלה לקוחות מתוך ענפי משפט שונים (כגון דיני חוזים, נזיקין, הגנת הצרכן ופרטיות), וכן יעסקו בדינים מיוחדים שמטרתם הגנה על אוכלוסייה מסוימת (כגון צרכנים או קטינים).

מטרת חלק זה היא להמחיש באמצעות דוגמאות כיצד הדין והרגולציה הקיימים יכולים להיות רלוונטיים לטיפול בהיבטי בינה מלאכותית, אולם הדברים נכתבים לשם ההמחשה בלבד, ואין בהם כדי להשליך על שאלות משפטיות או פרשניות שטרם הוכרעו על ידי הגורם המוסמך.

2.3.1. דיני פרטיות

חוק הגנת הפרטיות, התשמ"א-1981 קובע עוולות ועבירות הקשורות בפגיעה בפרטיותו של אדם, והוראות הקשורות באיסוף מידע אישי לצורך עיבוד ממוחשב במאגר מידע. עקב כך, בכל הקשור לפעולות עיבוד מידע הנעשית כחלק מתהליכי "נתוני עתק" או למידת מכונה, החוק צפוי לחול.

בהתאם להוראות חוק-יסוד: כבוד האדם וחירותו וכן חוק הגנת הפרטיות, פגיעה בפרטיות אפשרית אם ניתנה הסכמה מדעת, או לפי חוק. בנוסף, נדרש למסור מידע לנושא המידע אודות השימוש לשמו נאסף המידע. כמו כן, יש לעשות שימוש במידע שנאסף רק למטרה אשר לשמה הוא נמסר (כלומר, לא ניתן לאסוף מידע בהסכמה מדעת למטרה אחת ולאחר מכן לעשות בו שימוש למטרה אחרת בלא קבלת הסכמה נוספת או היתר חוקי אחר).[26]

הרשות להגנת הפרטיות במשרד המשפטים פרסמה גילוי דעת, שבו היא מתייחסת, בין היתר, לפרשנותה בעניין חובת היידוע במסגרת איסוף ושימוש במידע אישי כאשר נעשה שימוש בבינה מלאכותית.[27] במסגרת זו מציינת הרשות, כי כשאיסוף המידע נעשה באמצעות מערכות אוטומטיות, יש לוודא שיוצגו לנושא המידע כל הפרטים הנדרשים בהתאם להוראות החוק. עוד מציינת הרשות, כי כשעיבוד המידע נערך באמצעות מערכות אוטומטיות, על הליך היידוע לפרט על אופן פעולת המערכות, ככל שהדבר רלוונטי לגיבוש ההסכמה וככל שפירוט זה אפשרי מבחינה משפטית, טכנולוגית, ומסחרית. עוד ממליצה הרשות, כי יוסבר לנושא המידע על אודות פרטי המידע בהם עשויות המערכות להשתמש במסגרת השימוש במידע הנוגע אליו, והמקור של פרטי מידע אלו.

לצד ההוראות שצוינו לעיל, חוק הגנת הפרטיות ותקנות הגנת הפרטיות (אבטחת מידע), התשע"ז-2017 קובעים הוראות בעניין אבטחת מידע שבמאגר מידע. בנוסף, בהתאם להוראות החוק קיימת זכות לעיין במידע שבמאגר המידע וכן לתקן אותו. כאמור לעיל, על פני הדברים, ועל פי עמדת הרשות להגנת הפרטיות, הוראות אלה ואחרות צפויות לחול אף ביחס לשימוש בבינה מלאכותית.

---

[26] ראו: עת"מ 24867-02-11 **אי.די.איי חברה לביטוח בע"מ נ' רשם מאגרי המידע, הרשות למשפט טכנולוגיה ומידע במשרד המשפטים** (2012).

[27] הרשות להגנת הפרטיות, חובת יידוע במסגרת איסוף ושימוש במידע אישי, https://www.gov.il.

2.3.2. <u>אפליה אסורה</u>

על גופים שעליהם לא חל המשפט המנהלי, חל חוק איסור הפלייה במוצרים בשירותים ובכניסה למקומות בידור ולמקומות ציבוריים, התשס״א-2000. חוק זה מחיל חובות שוויון ואיסור אפליה על פעולות מסוימות במגזר הפרטי ועל גופים פרטיים, ובין היתר אוסר על פרסום הכולל אפליה כאמור, ואף מעביר את הנטל בתביעה אזרחית בין השאר בהתאם לתוצאה מפלה. בהיעדר הסדר חוקי ספציפי, במקרים שבהם שימוש בבינה מלאכותית יגרום לפעולה שתוצאתה אפליה אסורה לפי החוק – על פני הדברים, החוק והוראותיו האמורות עשויים לחול על מי שאחראי לפעילות הבינה המלאכותית. כך למשל, עשויה להיות להן רלוונטיות במקרה של קבלת החלטה מפלה על ידי מערכת מבוססת בינה מלאכותית, או מקום שמוצע פרסום בהתבסס על פילוח המבוצע על ידי בינה מלאכותית. עם זאת, ראוי להדגיש כי תחולת החוק היא רק על הפעילות המוגדרת בו.[28] הפרת חוק זה היא עוולה בנזיקין, כך שהוא מאפשר גם תביעה נזיקית. עם זאת, ייתכן שהעוולה המתאימה תהיה עוולת הרשלנות. כלומר, האם האופן שבו שימשה או פותחה המערכת היה עוולתי.

2.3.3. <u>דיני חוזים</u>

חוק החוזים (חלק כללי), התשל״ג-1973 (להלן: חוק החוזים), לא מונע ביצוע פעולות משפטיות באמצעים ממוחשבים, משום שאין בו ככלל דרישות של צורה וטכניקה. כיום נערכות עסקאות על סמך התקשרויות מול אתרי אינטרנט, בלי שעצם ביצוע הפעולה מעורר קושי משפטי. בהתאם לסעיף 23 לחוק החוזים ככלל אין דרישה צורנית לעריכת חוזה, ולכן ככלל אין מגבלה על שימוש במחשבים ובתקשורת אלקטרונית לעריכת חוזים.[29] בהקשר זה קבעה הפסיקה, בזיקה לפסיקה האמריקאית, הוראות לעניין היצע וקיבול ביחס לחוזים ממוחשבים.[30]

עמדה זו מבוססת גם על מסקנות ועדת המסחר האלקטרוני שפעלה במשרד המשפטים, על בסיס התפיסה בדבר הגמישות של הדין הקיים ויכולתו לחול גם על מצבים מורכבים מבחינה טכנולוגית שיתעוררו בעתיד.[31]

עם זאת, קיומו של אלגוריתם המעורב ביצירת חוזה שונה משימוש במדיה אלקטרונית בלבד, ומעלה שאלה שנדונה בוועדת המסחר האלקטרוני בדבר היכולת של ״סוכן אלקטרוני״, כלומר תוכנה הפועלת לפי הנחיות מי שערך אותה, לייצר התקשרויות באופן עצמאי. על אף שוועדת מסחר אלקטרוני דנה בנושא, בסופו של דבר הכלל שהציעה לא נכלל בהצעת חוק מסחר אלקטרוני.[32]

---

[28] ראו סעיף 3 לחוק: ״מי שעיסוקו בהספקת מוצר או שירות ציבורי או בהפעלת מקום ציבורי, לא יפלה בהספקת המוצר או השירות הציבורי, במתן הכניסה למקום הציבורי או במתן שירות במקום הציבורי, מחמת גזע, דת או קבוצה דתית, לאום, ארץ מוצא, מין, נטיה מינית, השקפה, השתייכות מפלגתית, גיל, מעמד אישי, הורות או לבישת מדי כוחות הביטחון וההצלה או ענידת סמליהם.״
״שירות ציבורי״ – שירותי תחבורה, תקשורת, אנרגיה, חינוך, תרבות, בידור, תיירות ושירותים פיננסיים, המיועדים לשימוש הציבור;
״שירותים פיננסים״ – שירותי בנקאות, מתן אשראי וביטוח;
״שירותי תחבורה״ – אוטובוסים, רכבות, תובלה אווירית, אניות, שירותי הסעה והשכרת רכב;
לביקורת על גישה זו ולהצעת גישה מרחיבה לתחולת החוק ראו: רונן אברהם, בנפרד ועדיין שווה? על דרכי ההתמודדות עם מקרי הפליה הנופלים בנפרד מתחולת חוק איסור הפליה, עתיד להתפרסם בספר אליעזר ריבלין.

[29] ראו: משרד המשפטים, הוועדה לבדיקת משפטיות הכרוכות במסחר אלקטרוני דו״ח חלקי, ירושלים, אייר תשס״ד, מאי 2004. (להלן – ועדת מסחר אלקטרוני) עמוד 16, וכן ראו: תמר קלהורה, הצעת חוק מסחר אלקטרוני: מטרות ועקרונות, חוקים ב 2010, 49-5.

[30] ראו למשל: תא (מרכז) 18763-04-15 **ויוה מדיה בע״מ נ' Google Ireland Ltd**, פסקאות 18-12.

[31] ראו ועדת מסחר אלקטרוני.

[32] ראה: קלהורה, בעמוד הערת שוליים 39.

על כן, ובהיעדר דין ספציפי העוסק בכך, לו תתעורר שאלה בדבר תחולת חוק החוזים על השימוש בבינה מלאכותית, יאפשר הדין הקיים מענים שונים. בין היתר, יהיה מקום לבחון שאלות בדבר ייחוס לצד המפעיל בינה מלאכותית להצעה להתקשר בחוזה, או למסירת הודעת קיבול להצעה להתקשר בחוזה, כוונה לבצע פעולה זו, כך שיישא בתוצאות הפעולה המשפטית.[33] עמדה זו נובעת מהאופן שבו מפותחת בינה מלאכותית ומשולבת בפעילות הארגון, המקדימה בדרך כלל לשימוש בבינה מלאכותית, והמעידה על רצונו וכוונתו של מפעיל הבינה המלאכותית לבצע פעולות משפטיות באמצעותה.[34]

שאלה נוספת נוגעת לקיומה של חובת גילוי בנוגע לשימוש בבינה מלאכותית על ידי מי שעושה בה שימוש בדבר עצם השימוש (ראו להלן פרק "גילוי"), וזאת מכוח החובה לנהוג בתום לב במשא ומתן לפי סעיף 12 לחוק החוזים,[35] או בשל ההוראה בעניין הטעיה לפי סעיף 15 לחוק. זאת, על רקע תכונות הבינה המלאכותית ויכולותיה, והאפשרות כי השימוש בבינה מלאכותית יקנה לצד המתקשר יתרון מהותי בניהול המו"מ לקראת החוזה.

תחום נוסף הוא החוזים האחידים. השימוש בבינה מלאכותית חושף את הפעילות לסיכונים שונים, ובין היתר סיכוני אבטחת מידע. במידה שמנסח חוזה אחיד יעביר את הסיכון הכרוך בשימוש בבינה מלאכותית לצד השני, יהיה מקום לבחון שאלות הנוגעות לקיומה של תניה מקפחת בחוזה אחיד.

2.3.4. <u>הגנת הצרכן</u>

חוק הגנת הצרכן, התשמ"א-1981 קובע הוראות שמטרתן הגנה על צרכנים. בהתאם לכך, מקום שבינה מלאכותית מעורבת בתהליך יצירת עסקה צרכנית, עשויות לחול על מי שמפעיל אותה החובות המוגברות הקשורות בגילוי בהתאם לחוק זה. בין היתר, יתכן שיהיה צורך לבחון אם אותם פערי הידע בין הצרכן לבין העוסק המפעיל את הבינה המלאכותית מתקיימים או אף מתגברים עקב כך שכנראה המידע שמצוי באופן בלעדי אצל העוסק כעת יהיה גדול עוד יותר.

בסעיף 7א לחוק הגנת הצרכן נקבע איסור על פרסום או שיווק פוגעניים לקטינים.[36] בהתאם לסעיף נקבעו גם תקנות הגנת הצרכן (פרסומת ודרכי שיווק המכוונים לקטינים), התשנ"א- 1991, שמטרתן מניעת ניצול חוסר ניסיונם ופערי הכוחות הקוגניטיביים שבין עוסקים לקטינים. הוראות אלה עשויות להיות רלוונטיות אף הן ככל שנעשה שימוש בבינה מלאכותית בתקשורת עם קטינים. יצוין, כי לאחרונה הודיעה הרשות להגנת הצרכן ולסחר הוגן כי היא רואה באתר אינטרנט המציע תכנים לקטינים כמי שאחראי לפרסומות המוצגות לצד התוכן, ללא קשר למידת השליטה או הידיעה

[33] יובהר כי מוקד הדיון בהקשר זה אינו דרישות הצורה לפי סעיף 23 לחוק החוזים, כי אם לעניין הביטוי "גמירת דעת" הנדרש כחלק מקיום "הצעה" לפי סעיף 2 לחוק ולעניין "קיבול" לפי סעיף 5 הדורש כי הקיבול יהיה "על דעתו" של צד לחוזה.
"2. הצעה
פנייתו של אדם לחברו היא בגדר הצעה, אם היא מעידה על ***גמירת דעתו*** של המציע להתקשר עם הניצע בחוזה והיא מסויימת כדי אפשרות לכרות את החוזה בקיבול ההצעה; הפניה יכול שתהיה לציבור."
5. קיבול
הקיבול יהיה בהודעת הניצע שנמסרה למציע והמעידה על ***גמירת דעתו*** של הניצע להתקשר עם המציע בחוזה לפי ההצעה.

[34] כמפורט לעיל, בעמ' 16-19.

[35] ראה: גבריאלה שלו ואפי צמח, דיני חוזים (מהדורה רביעית), 2019, עמוד 12-13.

[36] בהתאם לסעיף 7א.(א): "לא יפרסם אדם פרסומת ולא ינקוט דרך שיווק אחרת אם הפרסומת או דרך השיווק כאמור עלולות להטעות קטין, לנצל את גילו, תמימותו או חוסר ניסיונו או לעודד פעילות שיש בה כדי לפגוע בגופו או בבריאותו הגופנית או הנפשית."

בפועל של האתר על אודות פרסומות אלה, ובכלל זה לאפשרות שהפרסומות נובעות ממידע שנאסף ממכשיר הקצה שבו הן מוצגות.[37]

2.3.5. **דיני נזיקין**

שימוש במערכות מבוססות בינה מלאכותית הגורם לנזק עשוי לחשוף את המזיק לאחריות בנזיקין בגין הנזק, בתנאי שניתן יהיה להוכיח את יסודות האחריות בנזיקין.[38] במסגרת "קול קורא" שפרסמה מחלקת ייעוץ וחקיקה (משפט אזרחי) במשרד המשפטים,[39] הוצגו באופן כללי שאלות מרכזיות הקשורות בדיני הנזיקין והשימוש בכלי רכב עצמאיים, אשר יש להן רלוונטיות לבחינת קביעת חבות נזיקית.[40] דו"ח בינה מלאכותית בשירותים פיננסיים העלה שאלות דומות לגבי אופן יישום משטרי האחריות המקובלים בנזיקין לגבי שימוש במערכות מבוססות בינה מלאכותית.[41]

הדיון באחריות נזיקית בנסיבות קונקרטיות הוא תלוי נסיבות והקשר, וגם נגזר מהמסגרת הנזיקית החלה, כגון עוולת הרשלנות או הפרת חובה חקוקה, או העוולות הפרטניות. בלי לקבוע מסמרות, נראה כי תיתכן אפשרות לתחולת דיני הנזיקין על סיטואציות שבהן נגרם נזק או התנהגות עוולתית הקשורות בשימוש בבינה מלאכותית. דיני הנזיקין, במיוחד עוולת המסגרת של עוולת הרשלנות, גמישים דיים כדי להתמודד עם מצבים עובדתיים חדשים ועם שינויי הזמן, לרבות שינויים טכנולוגיים, משום שהם שואלים שאלות עקרוניות וכלליות שניתן לבחון אותן במצבים חדשים מגוונים. במסגרת יישום יסודות עוולת הרשלנות יעלו, בין היתר, שאלות אלה: האם יש חובת זהירות? האם חובת הזהירות הופרה? האם נגרם נזק עקב כך? האם הנזק הוא במתחם הסיכון שיצרה ההתנהגות? האם הוא לא רחוק מדי? שאלה מרכזית היא, האם מי שבונה מערכת של בינה מלאכותית מסוימת יכול וצריך לצפות שייגרם נזק מסוים כתוצאה מהפעלתה או השימוש בה. המסקנה הנורמטיבית תלויה הן בהיבטים עובדתיים טכנולוגים והן בשיקולי מדיניות.

[37] ראו עוד את עמדת הרשות להגנת הצרכן וסחר הוגן בעניין ת"צ 40801-09-20, המטילה אחריות על אתרי תוכן למניעת פרסום אסור לקטינים. הרשות קבעה עמדה זו למרות הטענה כי אופן הצגת הפרסומות אינו בשליטה ישירה ומודעות של מנהלי אתרי התוכן, וכי הם עשויים להיות מושפעים ממידע על אודות המשתמש המתקבל ממכשיר הקצה. https://www.law.co.il/news/2022/07/05/consumer-protection-minors

[38] "הטלת האחריות בנזיקין מותנית ככלל בהתקיימותם של שני יסודות אחריות הכרחיים: היסוד האחד הוא התנהגות עוולתית, דהיינו מעשה או מחדל או פעילות של המזיק או מצב שבו הוא נמצא, שהם תנאי והצדקה להטלת אחריות נזיקית עליו. היסוד השני של האחריות הוא קיומה של זיקה מספקת בין ההתנהגות העוולתית לבין הנזק, זיקה הנדרשת לצורך ייחוס הנזק למזיק. צויין כי שילוב שני יסודות אלה לרוב הכרחי אף לא תמיד מספיק, וכי יש שנדרש יסוד אחריות נוסף, בעיקר זה הובחן אם תוצאות הטלת האחריות הן תוצאות רצויות. עוולת הרשלנות אכן מושתת ת על שלושת יסודות אלה". ישראל גלעד, דיני נזיקין – גבולות האחריות, נבו 2012, פסקה 6.2, עמוד 418.

[39] ראו: משרד המשפטים, מחלקת ייעוץ וחקיקה, אסדרת השימוש ברכבים אוטונומיים - פנייה לקבלת עמדת הציבור בתחום הנזיקין והביטוח, 08.02.2021, https://www.gov.il.

[40] השאלות המרכזיות המועלות בקול הקורא בתחום אחריות נזיקית לרכב עצמאי (אוטונומי) כמצע לעיצוב ההסדר הראוי: "חסרונה של אישיות משפטית" משום שהרכב נשלט בידי מחשב, "הוספת גורם אחריות חדש" (משום שיש שילוב של גורמי בקרה), "העדר צפיות והעדר קשר סיבתי בין שלב התכנון והייצור לבין התנהגות הרכב האוטונומי" (עקב היות הרכב לומד), "מורכבות וחדשנות המידע הנדרש לצורך הוכחת רשלנות בתביעת נזקי רכוש", " too much information" (איסוף מידע רב המהווה יעד לתקיפה ומקשה על הבנת התמונה), "אי וודאות בנוגע לתוחלת הנזק הצפוי", "תמחור ביטוח בהעדר מידע", "אחריות תורמת של משתמשי הדרך והמשתמשים ברכב", "העדר מנגנוני תקינה ורישוי". על רקע האמור מועלים לדיון גם פתרונות ובהם: הצורך במשטר של אחריות מוחלטת והגדרת מאפייניה ומי חב לפיה, היחס בין הסדרי האחריות לבין ביטוח, היחס בין רכבים רגילים לרכבים אוטונומיים, סוגיית המידע הנאסף.

[41] ראו דו"ח בינה מלאכותית בשירותים פיננסיים, בעמוד 102-103, המעלה שאלות לגבי הטלת אחריות על תוכנה, חלוקת האחריות שבין המוסד הפיננסי, כותב התוכנה, והלקוח עצמו, היבטי הצפיות, וכן החשש שהטלת אחריות קפידה תביא לחסמי כניסה לשימוש בטכנולוגיה זו.

נראה עוד, כי דיני הנזיקין מאפשרים באופן עקרוני לבחון את חלוקת האחריות בין מפתח ומשתמש בבינה מלאכותית, בהתאם למנגנון חבות ביחד ולחוד. בהתאם למנגנון זה, האחריות המיוחסת לגורם אחד אינה מוציאה לרוב את האחריות של גורם אחר במישור היחסים שלהם מול הניזוק. חלוקת האחריות בין המזיקים מבוססת על "מבחן האשם המוסרי" שמקצה את חלק הארי של האחריות למי שאשמו המוסרי רב יותר.

<u>2.3.6. הסדרה ייעודית מגזרית</u>

לצד הוראות המשפט הפרטי, והוראות החלות מכוח אסדרת דיני פרטיות במידע, חלות גם הוראות בתחומי אסדרה ספציפים במשק. בהתאם לאופן שבו התפתחה האסדרה בתחום טכנולוגיית המידע, נראה כי מאסדרים יוכלו, במקרים המתאימים, בכפוף לסמכותם על פי הדין, לקבוע הוראות וכללי התנהגות פרטניים לגבי הבינה המלאכותית. לצד זאת, ייתכן כי כללים מסוימים יהיו טעונים עיגון בחקיקה ראשית, וזאת בהמשך לאופן הצגת הדברים בהנחיית היועץ 1.2500.[42]

כך, לדוגמה, בשנים האחרונות קידמה הממשלה את תיקון מספר 130 לפקודת התעבורה,[43] שמטרתו להוות תשתית חוקית לניסוי בשימוש ציבורי בכלי רכב עצמאיים במרחב הציבורי, לגבש תשתית ידע בנושא, ולבסס את אמון הציבור בכלי רכב עצמאיים.[44] תיקון זה לפקודת התעבורה קובע דרישות רגולטוריות ממי שמפעיל רכב עצמאי, ובכלל זה בתחום הבטיחות והגנת הסייבר, בטרם הפעלת הרכב, וכן בכל מחזור החיים של השימוש בו במרחב הציבורי.[45] התיקון מבוסס על קביעת העקרונות הנדרשים בחקיקה ראשית, ופירוט בתוספת לחוק או בתקנות מכוח הפקודה, והשלמת ההוראות במסגרת הוראות פרטניות לבעל ההיתר.[46]

כמו כן, ובכל הנוגע לסמכותם של רגולטורים בתחומים ספציפיים לקבוע הוראות המתייחסות לבינה מלאכותית, תיעשה עבודה לבחינת ההיבטים הקשורים ביישומי בינה מלאכותית בתחום הפיננסי, וזאת בהמשך לדו"ח בינה מלאכותית בשירותים פיננסיים. במסגרת הדו"ח, ממליצים החוקרים לבחון ולהסדיר את השימושים בבינה מלאכותית בתחום הפיננסי באופן ספציפי וייעודי, ולעסוק ברגולציה הפיננסית כמקשה אחת ולהסדירה באופן תואם, אך לא במשותף עם ענפי משק אחרים. דו"ח זה, לרבות ההמלצה האמורה, ייבחנו על ידי צוות מקצועי שיעסוק בנושא.

---

[42] הנחיית היועץ המשפטי לממשלה 1.2500, בעמ' 8-9.

[43] ראו: חוק לתיקון פקודת התעבורה (מס' 130), התשפ"ב-2022, ס"ח 2967, עמ' 772, י"ב באדר ב' התשפ"ב, 15.03.22.

[44] בהתאם לסעיף 16ג לתיקון לפקודת התעבורה: **"מטרתו של סימן זה לקבוע הסדרים שיאפשרו הפעלה של רכב עצמאי בלא נהג, בדרך, למטרת ביצוע ניסוי, תוך שמירה על בטיחות הנוסעים ברכב ועוברי הדרך ושימוש במגוון טכנולוגיות, כדי להביא לגיבוש תשתית ידע לגבי בטיחות הרכב העצמאי, יכולתו להשתלב באופן בטוח בין עוברי הדרך ולתת שירות לנוסעים והשפעתו על התנועה בדרך, לאפשר הנגשה של הידע האמור לציבור, ולבסס את אמון הציבור בו".**

[45] ראו סעיף 16ה (חובת קבלת היתר הפעלה) ו- 16ו (תנאים למתן היתר הפעלה) לתיקון לפקודה. יוער כי בהתאם לסעיף 16לב (שמירת דינים) לתיקון לפקודה נקבע כי: **"אין בהוראות לפי סימן זה כדי לגרוע מחובתו של בעל היתר לעמוד בהוראות כל דין, ובכלל זה חוק פיצויים לנפגעי תאונות דרכים, התשל"ה-1975, ובהוראות הדין לעניין חובת ביטוח, ובכלל זה הוראות פקודת הביטוח"**. בנושא זה ראו: מחלקת ייעוץ וחקיקה (אזרחי), **קול קורא לקבלת עמדת הציבור בתחום הנזיקין והביטוח בעניין אסדרת השימוש ברכבים אוטונומיים**, 08.02.2021, https://www.gov.il.

[46] ראו סעיף 16 ח וסעיף 16יג(ב) לפקודת התעבורה.

2.3.7. משפט מנהלי

כללי המשפט המינהלי מתווים מבחינה משפטית את דרך הפעולה של גופים מינהליים – משרדי ממשלה, יחידות סמך, חברות ממשלתיות וכד׳. שילוב הבינה המלאכותית בפעילותם של גופים ציבוריים אלה – אם ככלי תומך החלטה או כחלק מביצוע שלבי קבלת ההחלטה המנהלית בידי אלגוריתם – מחייב יישום של עקרונות המשפט המנהלי.

בתוך כך עשויות לעלות שאלות לגבי הסמכות המינהלית, ההליך המינהלי ודרך הפעלת שיקול הדעת המינהלי, כמו גם בנוגע להליך הביקורת השיפוטית עצמו ולאכסניה המוסדית הרלוונטית לבחינת החלטות אלגוריתמיות. בנושא זה שעניינו תחולת המשפט המנהלי על היבטי הבינה המלאכותית, תתקיים עבודת מטה של הגורמים המשפטיים הרלבנטיים בממשלה, בהובלת מחלקת ייעוץ וחקיקה במשרד המשפטים, שמטרתה לסייע ליועצים המשפטים במשרדים בקידום שימוש בבינה מלאכותית באופן העונה על דרישות המשפט המנהלי.

לסיכום, הדין הקיים מאפשר התמודדות מסוימת עם חלק מהאתגרים המתעוררים עקב שימוש במערכות מבוססות בינה מלאכותית, ובכלל זה אפליה, שקיפות, הצורך בהנמקה, אבטחת מידע, סייבר ופרטיות.

3. חלק שלישי: פעילות ארגונים בינלאומיים ומדינות מפותחות

בפרק זה נסקור את הגישה של ארגונים בינלאומיים ומדינות מובילות לקידום מדיניות רגולציה ואתיקה לבינה מלאכותית. מאפיין מרכזי בעולם הוא קידום מהיר של שיח בעל אופי משפטי, רגולטורי ואתי ביחס לבינה מלאכותית. במסגרת זאת, פורסמו מסמכים רבים שמטרתם הצהרה על עקרונות שיש לקחת בחשבון בעת פיתוח ושימוש בבינה מלאכותית.

מסגרות מדיניות אלה כוללות לרוב התייחסות לתועלות המשמעותיות הנובעות מהבינה המלאכותית, ולצדן התייחסות לאתגרים המתעוררים בקשר לבינה המלאכותית, וסקירה של אמצעים להתמודדות עם אתגרים אלה. קביעת מסגרות המדיניות נעשית ככלל, על מנת להגן על זכויות יסוד ואינטרסים ציבוריים, לחזק את הוודאות המסחרית והמשפטית, ולהגביר את אמון הציבור בבינה המלאכותית ולעודד שימוש בה. מכנה משותף מרכזי למסגרות המדיניות הוא התייחסות לעקרונות כגון אחריותיות, בטיחות, שקיפות, מניעת אפליה, וכיבוד זכויות אדם. בנוסף מסגרות המדיניות כוללות אמצעים ליישום העקרונות וכן שיטות לאיתור שימוש בבינה מלאכותית המייצרת סיכון גבוה לעקרונות וערכים אלה, ושיטות לניהול הסיכונים על מנת לאפשר פיתוח ושימוש בבינה מלאכותית באופן מהימן.

לצד מכנים משותפים אלה, ניתן לראות שוני במאפייני המדיניות, בשני היבטים מרכזיים. מאפיין מרכזי אחד הוא **רמת הפיתוח והפירוט של העקרונות והאמצעים לקידומם** – במדינות מסוימות העקרונות והאמצעים ליישומם מפורטים באופן כללי מאוד (כגון סינגפור). לעומת זאת, בטיוטת החקיקה של האיחוד האירופי על בינה מלאכותית לדוגמה, מוצעות דרישות מפורטות כבר בשלב זה של התפתחות הטכנולוגיה. מאפיין מרכזי שני הוא **המעמד הנורמטיבי המוצע לקידום המדיניות** – כלומר האם מוצע לקדם אותה בחקיקה ייעודית רוחבית, או שמא מוצע לקדם אותה על בסיס התשתית המשפטית הקיימת תוך קידום הסדרים נקודתיים בהתאם לצורך. כך לדוגמה, האיחוד האירופי מציע לקדם חקיקה רוחבית ומקיפה, הכוללת גם הוראות מפורטות לגבי הדרישות ממי שמפתח ומשתמש בבינה מלאכותית. מנגד, בבריטניה מוצעת מדיניות ממשלתית המבוססת על קביעת מספר עקרונות כלליים, אשר נועדו להנחות את הרשויות הרגולטוריות בלבד. אופן קידום המדיניות נובע, בין היתר, מהגישה הרגולטורית הכללית הנוהגת, וממאפייני השווקים הטכנולוגיים. כך לדוגמה, באופן כללי, האיחוד האירופי מוביל מגמה של הסדרה משפטית של שווקי טכנולוגיית המידע, בעוד שבארה"ב הגישה היא של צמצום התערבות מדינתית והסדרה ככלל באמצעות כוחות השוק. לצד זאת, לנוכח החשיבות הצפויה לטכנולוגיית הבינה המלאכותית, במימד הגיאופוליטי ובהיבטי הסחר הבינלאומיים קביעת הסדרים משפטיים ורגולטוריים בתחום הבינה המלאכותית משמשת גם ככלי לקידום הובלה מדינתית בתחום,[47] הן בגיבוש ההסדרים כמודל למדינות אחרות, והן כדי להשפיע על השווקים הבינלאומיים.[48]

---

[47] על הגישה הכללית של הובלה אירופאית באמצעות חקיקה, ראו:
Anu Bradford, The Brussels Effect, How The European Union Rules the World, Oxford 2022, Chapter 5 (Bradford).

[48] לנוכח התלות הגוברת של פעילויות פרטיות וציבוריות אירופאיות בשימוש בשירותי מיחשוב (כגון שירותי ענן) מחוץ לאירופה, הכוללים גם הוצאת מידע אישי ורגיש אחר מאירופה, יש מוביל האיחוד מגמה של "ריבונות דיגיטלית" באמצעות חקיקה. מגמה זו כוללת הסדרים משפטיים שמטרתם לחול על שירותים ומידע המוצעים באיחוד האירופי גם אם מקומם הפיזי הוא מחוץ לאיחוד. ראו:

עם זאת, גישות שונות אלה עלולות להביא ליצירת חסמי סחר ושיתוף פעולה. על רקע זה, מגמה מרכזית בתחום הגלובלי היא שיתוף פעולה גובר בין האיחוד האירופי לארה״ב לשם קידום תפיסה משותפת בנושא זה, על אף הגישות השונות. בפרט, האיחוד האירופי וארה״ב מקדמים שיתוף פעולה לגיבוש הבנות ומכנים משותפים בנושאי המדיניות המרכזיים בתחום הבינה המלאכותית. שיתוף פעולה זה מבוסס על הבסיס המשותף של עקרונות ה-OECD, והוא כולל שיתוף מידע על תקנים לניהול סיכוני בינה מלאכותית ושיטות להערכת מהימנות של בינה מלאכותית.[49]

3.1. המלצות ה-OECD בתחום הבינה המלאכותית

בשנת 2019 אישר ה-OECD (Organization for Economic Cooperation and Development) שורת המלצות בתחום הבינה המלאכותית (Recommendation of the Council on Artificial Intelligence).[50] ההמלצות כוללות שני חלקים: ״עקרונות לניהול ושימוש אחראי בבינה מלאכותית מהימנה״ (Principles for responsible stewardship of trustworthy AI), ו-״מדיניות לאומית ושיתוף פעולה בינלאומי לבינה מלאכותית מהימנה״ (National Policies and international cooperation for trustworthy AI). זהו המסמך הראשון שגובש על ידי ארגון בין-לאומי מוביל, אשר מעגן עקרונות כלליים לבינה מלאכותית. לצד פיתוח השיח הנורמטיבי בתחום זה, מטרת ההמלצות היא גם לעודד תאימות (interoperability) בין מדינות ה-OECD ואחרות במדיניות ובהסדרים המיישמים את העקרונות. זאת, כדי לקדם את הפעילות הכלכלית הבינלאומית בתחום הטכנולוגיה והמידע, כמנוף לקידום יעדי הפיתוח והצמיחה של מדינות ה-OECD.

ההמלצות אומצו על ידי כל מדינות ה-OECD ומדינות נוספות.[51] בשנת 2019 הן שימשו כמקור השראה למסמך העקרונות של ה-G20,[52] ארגון המאגד את 20 הכלכלות הגדולות בעולם שבו חברות 19 מדינות והאיחוד האירופי, ובהן מדינות נוספות שאינן חלק מה-OECD.[53] אף שהעקרונות אינם מחייבים, המלצות ה-OECD הן בדרך כלל בעלות השפעה משמעותית על מי שאימצו אותן,[54] וכן בקביעת סטנדרטים בינלאומיים ובהכוונת מדיניות.[55] בקרב מדינות ה-OECD יש גישות שונות לאופן קידום טכנולוגיה ולתפקיד החקיקה והרגולציה בתחום הטכנולוגיה. עקב כך להמלצות שהתקבלו בהסכמה בין המדינות החברות, ובפרט בין מדינות האיחוד האירופי לארה״ב ולמדינות כמו יפן שגישתן כאמור שונה, נודעת חשיבות רבה כרך בינלאומי המשקף מדיניות מוסכמת.

Frances G. Burwell and Kenneth Propp, The European Union and the Search for Digital Sovereignty, Building ״Fortress Europe״ or Preparing for a New World, Atlantic Council, Future Europe Initiative, 3-6 (Burwell and Propp) https://www.atlanticcouncil.org.

[49] US-EU Joint Statement of the Trade and Technology Council, 16.05.2022, Paris-Saclay, France, https://www.state.gov; במפגש שנערך במאי 2022 סיכמו נציגי המדינות על קידום שיתוף פעולה בתחום הבינה המלאכותית, הכולל שיתוף במסגרת קידום עקרונות ה-OECD, שיתוף מידע על **תקנים לניהול סיכוני בינה מלאכותית ושיטות להערכת מהימנות של בינה מלאכותית.**

[50] ראו המלצות ה-OECD, לעיל ה״ש 5.

[51] ראו בקישור לעיל: מדינות אלה כוללות את: אוסטרליה, אוסטריה, בלגיה, קנדה, צ׳ילה, צ׳כיה, דנמרק, אסטוניה, פינלנד, צרפת, גרמניה, יוון, הונגריה, איסלנד, אירלנד, ישראל, איטליה, יפן, קוריאה, לטביה, ,ליטא, לוקסמבורג, מקסיקו, הולנד, ניו זילנד, נורבגיה, פולין, פורטוגל, סלובקיה, סלובניה, ספרד, שוודיה, שוייץ, טורקיה, הממלכה המאוחדת, ארצות הברית. וכן מדינות נוספות: קולמוביה, קוסטה ריקה, ארגנטינה, ברזיל, מצרים, מלטה, פרו, רומניה, סינגפור, ואוקראינה.

[52] ראו את המידע המופיע לצד המלצות ה- OECD לעיל, וכן ראו את הצהרת הסיכום של המפגש ביוני 2019: https://www.oecd.org, וכן: https://www.g20-insights.org.

[53] בין מדינות אלה סין, רוסיה, ערב הסעודית, הודו, אינדונזיה, דרום אפריקה, ארגנטינה, ברזיל.

[54] על מעמדן של המלצות ב-OECD ראו OECD Directorate for Legal Affairs.

[55] OECD, 42 countries adopt new OECD Principles on Artificial Intelligence, https://www.oecd.org ההודעה מטעם ה-OECD מציינת למשל את ההשפעה של עקרונות ה- OECD לגבי הגנה על הפרטיות במידע, שהתקבלו ב- 1980, והיו בסיס לקידום מדיניות וחקיקה בתחום הפרטיות, ומשמשות מכנה משותף אחיד ברמה הבינלאומית.

במסגרת שיתוף הפעולה המתואר לעיל בין האיחוד האירופי לארה"ב, מוזכר נושא התיאום במדיניות בינה מלאכותית, והמלצות ה-OECD ופעילותו מצוינות כנקודת מוצא משותפת.[56]

ישראל היא חברה ב-OECD ולקחה חלק בדיונים על ניסוח ההמלצות. בנוסף, החלטת ממשלה 212 מנחה להתחשב בין היתר בעקרונות ה-OECD כחלק מקידום המדיניות הלאומית בתחום הבינה המלאכותית. להמלצות ה-OECD חשיבות במסגרת גיבוש ההמלצות שיפורטו בהמשך מסמך זה, הן בתחום ה"עקרונות אתיים לבינה מלאכותית", והן בהצבעה על כיוונים אפשריים לקידומן בפועל באמצעות הסדרים משפטיים ורגולטוריים. יצוין כי ה-OECD מקדם במקביל עבודה על מדיניות רגולציה בתחום הטכנולוגיות המתקדמות, ובהן בינה מלאכותית.

<u>הפרק הראשון להמלצות</u> כולל עקרונות לניהול אחראי של בינה מלאכותית מהימנה, שמיועדים לחול על ארגונים ציבוריים ופרטיים המפתחים בינה מלאכותית או משתמשים בה. העקרונות נועדו לקדם תפיסה של בינה מלאכותית שבה "האדם במרכז" (Human-Centered AI). זאת, לנוכח עליית חשיבותה של טכנולוגיה זו והשפעת התחזיות, ההמלצות או ההחלטות שלה על בני אדם.[57] על רקע מאפייני טכנולוגיית הבינה המלאכותית, ועדת המומחים של ה-OECD המליצה על נושאים שצריכים להיות בעדיפות לצורך קידום המטרה של "בינה מלאכותית שבה האדם במרכז".[58] נושאים אלה באו לידי ביטוי <u>בעקרונות שנכללו בהמלצות, והם</u>: (1) תרומה של הבינה המלאכותית לפיתוח בר קיימא ורווחה חברתית; (2) כיבוד ערכים השמים את האדם במרכז וערך ההגינות; (3) השימוש במערכות בינה מלאכותית ואופן פעולתן צריך להיות שקוף; (4) מערכות בינה מלאכותית צריכות להיות עמידות ובטוחות; (5) נדרשת אחריותיות לתוצאות של הבינה המלאכותית.[59] מרכיב משותף לעקרונות בפרק הראשון של מסמך ההמלצות הוא שיש להחיל אותם בהתאם להקשר ולרמת הסיכון הנובעת מהטכנולוגיה, ובשים לב לתועלות החברתיות מבינה מלאכותית.

<u>בפרק השני להמלצות</u> מוצעים כלי מדיניות שנועדו לקדם את העקרונות ובאופן כללי "בינה מלאכותית מהימנה" (Trustworthy AI) במסגרת מדינות ה-OECD, כגון קידום מדיניות לאומית לתחום הבינה המלאכותית וכן קידום הון אנושי בתחום והשקעה במחקר ופיתוח.

מאז גיבוש ההמלצות בשנת 2019, ה-OECD משקיע משאבים ותשומת לב בקידום יישום העקרונות באמצעות שיח רצוף על אופן מימוש העקרונות והכלים המעשיים הנדרשים לממשלות וארגונים. פעילות זו תומכת בקידום הבנה והתמודדות עם בינה מלאכותית, וכן שואפת לייצר מכנה משותף בינלאומי גם בהיבטים מעשיים של מימוש העקרונות. במסגרת זאת ה-OECD הקים פורטל הכולל מידע רב בנוגע למדיניות בתחום הבינה המלאכותית במגוון היבטים.[60] <u>תוצר מרכזי אחד</u>, של פעילות זו הוא הצעה לשיטה לסיווג ומיון מערכות מבוססות בינה מלאכותית.[61] שיטה זו חשובה לצורך איתור ההזדמנויות והסיכונים, והתאמה של המענה בתחום ניהול הסיכונים לסוג המערכת

---

[56] EU-US Joint Statement of the Trade and Technology Council, section 19. https://ec.europa.eu.
[57] דברי ההסבר והרקע לעקרונות מופיעים בפרסום שיצא במקביל מה-OECD, שנקרא "Artificial Intelligence in Society".
בהתאם לכך, האמור בטקסט מבוסס על ההסברים המופיעים בפרסום זה, שייקרא להלן בטקסט "המסמך המלווה" OECD, Artificial Intelligence in Society, 2019, https://www.oecd-ilibrary.org, Chapter 4, p.82.
[58] OECD, AI in Society, p. 83.
[59] שם.
[60] OECD AI Policy Observatory, https://oecd.ai/en/.
[61] OECD, OECD Framework for the Classification of AI Systems: a tool for effective AI policies, https://oecd.ai/en/classification.

ולסיכונים הכרוכים בה. פעילות מרכזית נוספת היא קידום מכנה משותף באשר לאופן ניהול סיכוני הבינה המלאכותית, בשים לב לפעילות רבה של ארגוני התקינה השונים בעולם בתחום זה.

3.2. טיוטת החקיקה של האיחוד האירופי[62]

האיחוד האירופי מבוסס על מספר אמנות רב צדדיות בין מדינות אירופה, שעל בסיסן ניתן לקדם בין היתר חקיקה במדינות החברות בתחומים הקבועים באמנות.[63] נציבות האיחוד האירופי, הרשות המבצעת של האיחוד, מקדמת במסגרת מדיניות האיחוד, בתחום הדיגיטלי בכלל ובתחום הבינה המלאכותית בפרט, חקיקה רוחבית שנועדה לחול על רוב השימושים האזרחיים בבינה מלאכותית. הפרלמנט האירופי ומועצת השרים קראו לנציבות לקדם חקיקה בתחום זה.[64]

בעקבות שימוע ציבורי ומחקר, זיהתה הנציבות שש בעיות מרכזיות העלולות לנבוע מפיתוח ושימוש בבינה מלאכותית, המצדיקות התערבות באמצעות חקיקה. בעיות אלה הן: (1) סיכון מוגבר לבטיחות ואבטחה; (2) סיכון מוגבר לפגיעה בזכויות אדם וערכי האיחוד; (3) היעדר כלים ומשאבים נדרשים בידי רשויות הרגולציה לפקח על פיתוח בינה מלאכותית ועל כך שהשימוש בה נעשה בהתאם לכללים; (4) אי-ודאות ומורכבות ביחס לכללים החלים על בינה מלאכותית שמרתיעים עסקים מפיתוח או שימוש בבינה מלאכותית; (5) החשש שחוסר אמון בבינה מלאכותית יפגע בפיתוח בינה מלאכותית באירופה ויפחית את התחרותיות של הכלכלה האירופית; (6) כלי מדיניות מבוזרים מייצרים חסמים לסחר ומסכנים את הריבונות הדיגיטלית של האיחוד.[65]

בהמשך לכך, בשנת 2021 וכחלק מהמדיניות האירופית הכוללת לקידום בינה מלאכותית,[66] הנציבות פרסמה טיוטת חקיקה בתחום הבינה המלאכותית. החקיקה המוצעת על ידי נציבות האיחוד האירופי ונדונה בפרלמנט האירופי קובעת כללי התנהגות והסדרה מפורטים שמטרתם מימוש העקרונות הכלליים של בינה מלאכותית מהימנה (Trustworthy AI), ובכך היא מהווה את מודל המדיניות החקיקתי המפורט ביותר מסוגו בעולם. זאת, משום שמודל מדיניות זה כולל **הסדרה רוחבית בחקיקה** של טכנולוגיות הבינה המלאכותית, באופן שהוא מגדיר את תפיסת ניהול הסיכונים הכוללת, את התבחינים לדרישות הרגולטריות, ואת הדרישות הרגולטוריות.

מטרות החקיקה הן: (1) להבטיח כי מערכות בינה מלאכותית שנעשה בהן שימוש בשוק האירופי המשותף בטוחות ומכבדות את החוק הקיים בנושא זכויות יסוד וערכי האיחוד; (2) להבטיח ודאות משפטית כדי לאפשר השקעות וחדשנות בבינה מלאכותית; (3) לקדם משילות ואכיפה אפקטיבית של החוק הקיים בנושא זכויות יסוד ודרישות בטיחות החלות על מערכות בינה מלאכותית; (4) לאפשר פיתוח של שוק אחיד לבינה מלאכותית בטוחה ומהימנה ולמנוע פיצול של השוק.[67] (5) לקדם תפקוד תקין של השוק המשותף, על ידי קביעת כללים אחידים לבינה מלאכותית, וצמצום החשש

---

[62] חלק זה בסקירה מבוסס על גם הסקירה בדו"ח בינה מלאכותית בשירותים פיננסיים, עמודים 28-36.

[63] European Union, European Union – Principles, countries, history, https://european-union.europa.

[64] European Parliament, European Parliamentary Research Service, Artificial intelligence act, January 2021, (EPRS), https://www.europarl.europa.eu , EPRS, p. 2-3.

[65] EPRS, p.3.

[66] מדיניות זו כוללת ארבעה יעדי מדיניות מרכזיים: ליצור תנאים לפיתוח ושימוש בבינה מלאכותית באיחוד האירופי, להפוך את האיחוד למקום הנכון לעשייה – ממעבדה לשוק, לוודא כי טכנולוגיות בינה מלאכותית משרתות את האוכלוסיה (חקיקת הבינה המלאכותית הינה חלק מיעד מדיניות זה), ובניית הובלה אסטרטגית בסקטורים. מקור: מצגת של נציבות האיחוד האירופי.

[67] EU Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, Brussels 21.04.21, COM(20221) 206 Final, https://eur-lex.europa.eu ("Draft AI Regulation").

להסדרים שונים שיפגעו במסחר בין המדינות. לצד המטרות הקשורות בכללי שוק אחידים בתוך השוק האירופי המשותף, מטרה מוצהרת נוספת של האיחוד האירופי, היא להוביל את השיח העולמי לגבי נורמות לפיתוח טכנולוגיות מבוססות בינה מלאכותית ולשימוש בהן.[68]

שיטת ההסדרה המוצעת בטיוטת החקיקה האמורה מבוססת על תפיסת ניהול סיכונים, שבה יש התאמה בין עומק הרגולציה לבין רמת הסיכון הנשקפת לזכויות יסוד ולבטיחות. טיוטת החקיקה מבחינה בין מערכות בינה מלאכותית אשר הסיכון בהן הוא "בלתי מקובל" ולכן הן אסורות; מערכות שלגביהן "הסיכון גבוה" ולכן תוקף בחינה מראש; מערכות שלגביהן "הסיכון מוגבל" ולכן הוא מחייב בשקיפות; או מערכות ב"סיכון נמוך או מינימלי" אשר לא חלות עליהן חובות.

מערכות שיוגדרו בסיכון גבוה יידרשו להירשם במרשם כללי אירופאי, ולעבור בדיקה מקדימה. במידה שמדובר בבינה מלאכותית המשולבת במוצר, הבדיקה תיעשה במסגרת בדיקת המוצר. יצרני מערכות בתחום שאינו דורש בחינה מוקדמת לפי חוק כיום, יידרשו לבצע בדיקה עצמית המראה שהם עומדים בדרישות. כמו כן, מערכות אלה יצטרכו לעמוד במגוון דרישות, ובכלל זה עריכת ניהול סיכונים; בדיקת הטכנולוגיה; עמידות; משילות מידע; שקיפות; מעורבות אנושית; והגנת סייבר.

3.3. <u>מועצת אירופה וקידום אמנה בנושא בינה מלאכותית</u>

מועצת אירופה (Council of Europe) היא ארגון של מדינות אירופה שהוקם לאחר מלחמת העולם השנייה כדי לקדם את התיאום ושיתוף הפעולה בעניין זכויות האדם באירופה.[69] כיום הארגון מונה 46 מדינות.[70] אמנה מרכזית שגיבש הארגון היא האמנה האירופית לזכויות אדם וחירויות יסוד (The European Convention on Human Rights and Fundamental Freedoms).

מועצת אירופה הובילה שתי אמנות מרכזיות שמטרתן קביעת כללים להגנה על זכויות אדם בשימוש בטכנולוגיית מידע. <u>אמנה אחת</u> עוסקת בהגנה על הפרטיות, משנת 1981;[71] <u>והאמנה השנייה</u> עוסקת בפשיעה ממוחשבת, משנת 2001.[72] פעילות נוספת של מועצת אירופה במרחב הדיגיטלי היא גיבוש כללים שמטרתם התמודדות עם ביטויי שנאה במרחב המקוון ובמרחב הפיזי, תוך כיבוד חופש

---

[68] EU Commission, Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions, Fostering a European approach to Artificial Intelligence, Brussels 21.04.2021, COM(2021) 205 final, p.5.
כן ראו Bradford ו- Burwell and Propp בהערת שוליים 194 לעיל.

[69] Council of Europe, About the Council of Europe – Overview, https://www.coe.int/en/web.

[70] הארגון כולל את המדינות המייסדות (ממערב אירופה) וכן מדינות שהצטרפו בשנות ה- 90 של המאה ה- 20 לאחר התפרקות הגוש הסובייטי. ראו שם.

[71] The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS (No. 108)).
האמנה קובעת עקרונות להגנה על הפרטיות בעת **איסוף ועיבוד מידע ממוחשב.** האמנה היתה המסמך הבינלאומי המחייב הראשון בתחום הגנת המידע האישי. היא נערכה במקביל לדיונים שנערכו בארגון ה- OECD על המלצות ה-OECD בנושא הגנה על הפרטיות, שאושרו בשנת 1980. האמנה תוקנה בשנים האחרונות כדי להתאים להתפתחויות הטכנולוגיות.
Council of Europe, Convention 108 and Protocols, https://www.coe.int.

[72] Council of Europe, The Budapest Convention (ETS No. 185) and its Protocols, https://www.coe.int/en;
מדינות משקיפות במועצת אירופה ובהן ארה"ב, קנדה, ויפן השתתפו במשא ומתן על ניסוח האמנה והן גם חתומות עליה, ועקב כך האמנה מהווה הסדר משפטי שיש לו חשיבות בינלאומית מעבר למדינות החברות במועצת אירופה בלבד. לאמנה הצטרפו 67 מדינות, וגם ישראל חברה באמנה כמדינה משקיפה במועצת אירופה. מאז עריכת האמנה נערכו לה פרוטוקולים נוספים כדי לעדכן חלק מהסדרים הנדרשים.

הביטוי.[73] באופן כללי, מועצת אירופה עוסקת בשימושים שלטוניים בטכנולוגיה (כלומר אספקת שירותים ציבורים ואכיפת חוק), למעט עניינים הקשורים בביטחון לאומי (national defense).[74]

בשנת 2019 מינתה מועצת השרים של מועצת אירופה ועדת אד-הוק בנושא בינה מלאכותית (CAHAI).[75] הוועדה פעלה עד שנת 2021, והגישה דו"ח מפורט ובו המלצות על נושאים באמנה חדשה של מועצת אירופה בתחום הבינה המלאכותית,[76] שהועברו להכנה לאמנה על ידי ועדת ההמשך (CAI).[77] זוהי היוזמה המובילה כיום לאמנה בינלאומית בתחום הבינה המלאכותית.

הוועדה מציינת את חשיבות ניסוח האמנה באופן שיאפשר הצטרפות של מדינות מחוץ לאירופה, וזאת בדומה לאמנה להגנה על הפרטיות ולאמנת פשעי הסייבר. זאת, על מנת להגביר את השפעת האמנה ולייצר כללי שוק שווים לגורמים הרלוונטיים ובהם התעשייה והאקדמיה.[78] בהמשך לכך, הוועדה ממליצה להתחשב בעיסוק בנושא בעקרונות ה-OECD, באיחוד האירופי, וב-UNESCO (שלא יידון במסמך זה). ישראל היא מדינה משקיפה בתהליך גיבוש האמנה המתקיים כיום.[79]

<u>בתמצית, ההמלצות המרכזיות שנכללו בדו"ח הן</u>:

1. **צורך במסגרת משפטית מחייבת** – לצד ההכרה בתועלות שיש בבינה מלאכותית, לצורך התמודדות עם הסיכונים האפשריים, הוצע כי יש צורך במסגרת משפטית המבוססת על הסטנדרטים של מועצת אירופה בתחום זכויות האדם, הדמוקרטיה[80] ושלטון החוק. המסגרת צריכה לכלול עקרונות כלליים ונורמות ספציפיות,[81] ולצדם עשויים להידרש הסדרים נוספים, מחייבים או וולונטריים, וברמה הסקטוריאלית.[82] הוועדה מציינת כי יש להימנע משכפול או פיצול של הסדרים קיימים בתחום המשפט הבינלאומי ודיני זכויות האדם, ולכן יש לבחון בזהירות את

---

[73] Counsel of Europe, ADI/MSI-DIS Committee of Experts on Combating Hate Speech, https://www.coe.int/en.

[74] בהתאם לסעיף 1(d) לאמנת מועצת אירופה.

[75] Council of Europe, CAHAI – Ad hoc Committee on Artificial Intelligence, https://www.coe.int/en.

[76] Council of Europe, Ad Hoc Committee on Artificial Intelligence (CAHAI), Possible elements of a legal framework on artificial intelligence based on the Council of Europe's standards on human rights, democracy and the rule of law, Strasbourg 03.12.21, https://rm.coe.int/cahai-2021-09, (hereinafter: CAHAI 2021).

[77] ההמלצות אינן חלות בתחום הביטחון הלאומי שאינו במנדט של מועצת אירופה, CAHAI, section 6.

[78] CAHAI, section 7.

[79] במסגרת עבודת הוועדה היא ביצעה סקירה של הנעשה בעולם. בסקירה זו נכלל מאמר מאת פרופסור בן ישראל מתניה וד"ר ליהי פרידמן, הסוקרים בהרחבה את המלצות ועדת המשנה בראשות פרופסור קרין נהון, לגבי הגישה שהוצעה בידי וועדת המשנה לעקרונות אתיים ולאסדרה.
Counsel of Europe, CAHAI, Towards Regulation of AI Systems, Global perspectives on the development of a legal framework on Artificial Intelligence (AI) systems based on the Council of Europe's standards on human rights, democracy and the rule of law, https://edoc.coe.int/fr/intelligence-artificielle.

[80] הדו"ח כולל התייחסות ספציפית לחשש לפגיעה בדמוקרטיה עקב שימוש בבינה מלאכותית, למשל באמצעות עיצוב דעת הקהל, או אמצעים קונקרטים יותר כגון מיקרו-טריגטינג, עריכת פרופילים של משתמשים, ומניפולציה על תוכן, CAHAI 2021, section 36.
בהמשך לכך, הדו"ח גם מציין כי בעוד שאיתור פגיעה בזכויות אדם היא יחסית ברורה, עלול להיות קושי לאתר פגיעה של בינה מלאכותית בדמוקרטיה או בשלטון החוק, ולכן מציע כי בחלק מהמקרים הפרה של זכויות אדם (כגון הזכות להתכנס), תהיה אינדיקציה לכך, CAHAI 2021, section 52.

[81] הוועדה מציעה לכלול באמנה הוראות בנושאים הבאים: 1. מניעת נזק לא חוקי, כולל הבהרת הקשר בין נזק לא חוקי לערכים המוגנים; 2. יחס שווה והעדר אפליה של יחידים בידי מערכות בינה מלאכותית כדי למנוע הטיה לא מוצדקת ושימושים שמביאים לתוצאות מפלות; 3. שמירה על שוויון בין המינים; 4. זכויות לאנשים בסיכון ובמצבי סיכון; 5. משילות מידע בהתאם לתפיסות משילות מידע באמנת הפרטיות (ראו לעיל), ובהתבסס על העקרונות בה. 6. עמידות (robustness); 7. בטיחות והגנת סייבר; 8. שקיפות (transparency); 9. הסברתיות (explainability); 10. יכולת קיום ביקורת (auditability); 11. אחריותיות (accountability); 12. התחשבות בקיימות (sustainability); 13. רמה נדרשת של פיקוח אנושי (human oversight).

[82] CAHAI 2021, section 3.

אופן עיצוב ההסדרים.[83] אופן עיצוב ההסדרים צריך לכלול זכויות לאנשים המושפעים מבינה מלאכותית, וכן חובות על מי שמפתח ומשתמש בבינה מלאכותית.[84]

2. **קביעת כללים מידתית** – הוצע כי הדרישות בקשר לפיתוח מערכות בינה מלאכותית והשימוש בהן, צריכות להיות מידתיות, ביחס לאופי הסיכון שכרוך במערכות אלה ביחס לערכים המוגנים. עקב כך הומלץ לקבוע קריטריונים להשפעה של מערכות בינה מלאכותית,[85] כאשר על הסיווג לכלול מספר רמות סיכון לערכים המוגנים. עקרונות בסיסיים המאפשרים את הקביעה לגבי הסיכון (כגון דרישות שקיפות) צריכים לחול לגבי כל המערכות המבוססות על בינה מלאכותית.[86] במסגרת זו אף הוצע לקבוע "שימושים אסורים" בבינה מלאכותית בתחומים שבהם נוצר סיכון "בלתי נסבל" (unacceptable) לערכים המוגנים, וכאשר לאחר בחינה אובייקטיבית והתחשבות בכל החלופות נמצא כי אין אמצעי אחר לצמצם את הסיכון.[87]

3.4. <u>מדיניות הרגולציה בארה"ב</u>

מדיניות ההסדרה המרכזית קבועה בחוזר מטעם משרד הניהול והתקציב (Office of Management and Budget) בבית הלבן,[88] שפורסם בשנת 2020 בעקבות צו נשיאותי 13859 בנושא זה (להלן: חוזר OMB). מטרת החוזר היא להסיר חסמים לפיתוח בינה מלאכותית ושימוש בה, תוך הגנה על טכנולוגיה אמריקאית, על הביטחון הלאומי והכלכלי האמריקאי, על פרטיות, זכויות אזרח, וערכים נוספים כגון חרות, שלטון החוק וכיבוד קניין רוחני. לצד זאת, משרד המדע שבבית הלבן החל בסדרת דיונים ציבוריים לבחינת הצורך ב"מגילת זכויות אדם דיגיטלית" לתחום הבינה המלאכותית.[89] בהצעת חוק פדרלית מקיפה להגנה על פרטיות שפורסמה לאחרונה,[90] מוצע לקבוע

---

[83] CAHAI 2021, section 15.
[84] CAHAI 2021, section 17.
[85] CAHAI 2021, section 18.
[86] CAHAI, section 5.
[87] הוועדה מציינת שימושים כגון מערכות העושות שימוש בביומטריה לזהות, למיין, לקטלג או להסיק תכונות או רגשות של יחידים, בפרט אם הן מאפשרות ניטור המוני. בנוסף יש לבחון מערכות "דירוג חברתי" שמשפיעות על גישה לשירותים חיוניים.
"דירוג חברתי" ("social scoring") עוסק בקביעת ציון לאדם על בסיס התנהגותו בשורה של פרמטרים ותחומים, תוך פיתוח הרעיון של "דירוג אשראי" למשל, הקובע את רמת סיכון או מהימנות האשראי של אדם, לדירוג רחב יותר.
הרשות להגנת הפרטיות מתארת זאת כדלקמן:
"דירוג חברתי הוא שם כולל למנגנון דירוג שיטתי ומתמשך של אזרחים או תושבים, או משתתפים בקהל משותף מסוים, על פי מידע הנוגע אליהם, אשר נאסף ממקורות שונים ומוצלב עם סוגי מידע אחרים, ואשר עשוי להשפיע על היבטים שונים הנוגעים לחייהם, לרבות דרך הענקת ציון, על פי פרמטרים שנקבעו, והצגת "מקומם" בדירוג ביחס למדורגים אחרים. מערכות של דירוג חברתי הן אפוא כאלו המאפשרות לגורמים (ציבוריים ופרטיים) להשתמש במידע על אודות פרט ספציפי לשם דירוגו על פי מדדים שונים באופן השוואתי לפרטים אחרים, כאשר דירוג זה עשוי - להיות שקוף בפני גורמים שונים, לרבות פרטים אחרים, ולהשפיע על היבטים שונים של חייו של אותו הפרט ממנו נאסף המידע."
ראו: הרשות להגנת הפרטיות, דירוג חברתי בראי הזכות לפרטיות:
סקירת רקע בעניין שימוש במערכות לדירוג חברתי, 21.04.21, https://www.gov.il/; CAHAI, section 21.
[88] Executive Office of the President of the United States of America, Office of Management and Budget, Memorandum for the Heads of Executive Departments and Agencies, Guidance for Regulation of Artificial Intelligence Applications, M-21-06, 17.11.2020.
[89] The White House, CYMI: WIRED (Opinion): Americans Need a Bill of Rights for an AI-Powered World, 22.10.2021, https://www.whitehouse.gov.
וראו עוד:
The White House, Readout of White House Listening Session on Tech Platform Accountability, 08.09.2022, " Principles for Enhancing Competition and Tech Platform Accountability With the event, the Biden-Harris Administration announced the following core principles for reform: 1. **Promote competition in the technology sector…5. Increase transparency about platform's algorithms and content moderation decisions. …6. Stop discriminatory algorithmic decision-making",** https://www.whitehouse.gov.
[90] United States House of Representatives, House Committee on Energy and Commerce, House and Senate Leaders Release Bipartisan Discussion Draft of Comprehensive Data Privacy Bill, 03.06.2022, https://energycommerce.house.gov.

עקרונות לשימוש באלגוריתמים המבוססים על מידע אישי (כגון איסור על שימוש מפלה),[91] וכן להטיל על רשות הסחר הפדרלית (ה-FTC) לפקח על שימוש בהם במקרים הקבועים בהצעת החקיקה. הדיון יתמקד בחוזר ה-OMB המציג את המדיניות הרגולטורית.

נקודת המוצא של חוזר OMB היא החשיבות של קידום התעשייה והחדשנות בארה"ב, והתחשבות בשיקול זה, ככל הניתן, ובמסגרת הדין, בפיתוח כללים רגולטוריים ובפעילות לא רגולטורית אחרת של הרשויות. בהתאם לכך, במסגרת החוזר נדרשים הרגולטורים לבחון אם הרגולציה מייצרת חסמים שעלולים לפגוע בתחרותיות התעשייה האמריקאית; וכן לבחון את התועלות הנובעות מבינה מלאכותית, לצד הסיכונים הכרוכים בה שעשויים להצדיק התערבות מצדם. על רקע זה, החוזר מגדיר את העקרונות הבאים לגבי מדיניות רגולציה בהקשרי בינה מלאכותית:

א. קידום אמון הציבור בבינה מלאכותית – על הרגולציה לשרת את אמון הציבור בטכנולוגיה באמצעות קידום אפליקציות מהימנות, עמידות וראויות לאמון.

ב. שיתוף הציבור.

ג. מהימנות מדעית ואיכות המידע – התפיסה הרגולטורית צריכה להיות מבוססת על מידע ותהליכים טכנולוגיים ומדעיים. בהתאם לכך על התהליך הרגולטורי להביא בחשבון ולשקף את החוזקות, החולשות, והשיפורים הצפויים עקב שימוש בבינה מלאכותית, התמודדות עם סיכון ועם הטיה, השפעה אפשרית על תחרות, פרטיות וקבלת החלטות בידי היחיד, היבטים הקשורים לביטחון הלאומי, ושימוש ראוי בתוצאות של הבינה המלאכותית.

ד. הערכת סיכונים וניהול סיכונים – הפעילות של הרשויות בתחום הבינה מלאכותית הן בהקשר הרגולטורי והן בהקשר לא רגולטורי צריכות להיות מבוססות על יישום עקבי של הערכת סיכונים וניהול סיכונים. תפיסת ניהול הסיכונים נועדה להגדיר את הסיכונים המקובלים וסיכונים שמשקפים אפשרות של נזק לא מקובל, או נזק שעלותו גבוהה מתועלתו. ניהול הסיכונים צריך להשוות את הסיכון הנובע מהשימוש בבינה המלאכותית ומהסיכון הקיים בלעדיה, ואם הסיכון הנובע מהשימוש בה נמוך יותר, הגם שהוא קיים, יש לבחון לאפשר אותו. ככל הניתן יש להפעיל תפיסת ניהול סיכונים זהה בין ענפי המשק השונים.

[91] 117 Congress 2nd session, A Bill to provide consumers with foundational data privacy rights, create strong oversight mechanisms, and establish meaningful enforcement, section 207.

ה. שקלול תועלות ועלויות – בקידום רגולציה יש לבחון את מכלול התועלות והעלויות הנובעות משילוב בינה מלאכותית בפעילות, תוך התחשבות בעלותה החברתית, בתועלת הגלומה בה ובהשפעותיה החלוקתיות. יש להשוות בין המצב הקיים לבין המצב הנובע משילוב הבינה המלאכותית, ובהיעדר מקור להשוואה, יש לשקול את הסיכונים והעלויות של אי-שימוש בה.

ו. גמישות – יש לשקול לקבוע כללים גמישים וניטרליים מבחינה טכנולוגית (כלומר, לא כללים מפורטים המתאימים רק לסוג ספציפי של מערכת), ולהימנע מכללים משפטיים נוקשים. יש לתת את הדעת לכללים החלים במדינות אחרות כדי שלא להקשות על התעשייה האמריקאית.

ז. הגינות ואי-אפליה – רשויות, במסגרת פעילותן הרגולטורית, צריכות להבהיר כיצד הן מתייחסות להשפעה של מערכות מבוססות בינה מלאכותית על הסיכון להתרחשות אפליה. בהקשר זה, יש לבחון את המצב לפני שילוב בינה מלאכותית ולאחריה.

ח. גילוי ושקיפות – גישת החוזר היא כי לגילוי ושקיפות בנוגע לשימוש בבינה מלאכותית תרומה לא רק להליך הרגולטורי אלא גם להגברת אמון הציבור. כשנעשה גילוי, עליו להיות בשפה ברורה המאפשרת למשתמשים לקבל החלטות מודעות. היקף הגילוי הנדרש מבוסס על הקשר, הערכת הנזקים האפשריים (כולל שימוש לרעה במידע הנמסר), היקף הנזקים, האפשרויות הטכנולוגיות ("technical state of the art"), והתועלות הפוטנציאליות משימוש.

ט. בטיחות ואבטחה – יש לעודד פיתוח מערכות בינה מלאכותית שהן בטיחותיות, בטוחות ופועלות כפי שיועדו לפעול (כלומר, משיגות כראוי את הפונקציה שלשמה נועדו), ושילוב היבטי בטיחות ואבטחה בשלבי העיצוב, הפיתוח, הפריסה והשימוש בבינה מלאכותית.

י. תיאום בין רשויות – יש לקדם תיאום בין הרשויות כדי לאפשר ולעודד שיתוף במידע ומומחיות, ולשפר את הוודאות והצפיות של מי שהמדיניות חלה עליו.

יא. שימוש בכלים לא רגולטוריים – רשויות יכולות להימנע מקביעת רגולציה ייעודית לבינה מלאכותית, ותחת זאת לעשות שימוש בהנחיות קונקרטיות להסרת אי-ודאות, שימוש בניסויים או קידום סטנדרטים וולונטריים.

יב. הסרת חסמים לשימוש בבינה מלאכותית – בין האמצעים המוזכרים בחוזר בהקשר זה ניתן למנות פעילות להנגשה של מידע הנדרש עבור בינה מלאכותית, מסירת מידע לציבור על הפעילות של הרשות בתחום כדי לעודד את אמון הציבור, השתתפות בפורומים הקובעים סטנדרטים וכללי התנהגות וולונטריים ופעילות לשם כך בפורומים בינלאומיים.

3.5. מדיניות הרגולציה בבריטניה

בספטמבר 2021 פרסמה בריטניה אסטרטגיה לאומית מעודכנת בתחום הבינה המלאכותית.[92] באסטרטגיה נכתב כי בהתאם לדו״ח של הפרלמנט הבריטי משנת 2018, אין הצדקה לקידום רגולציה רוחבית בתחום הבינה המלאכותית ויש יתרונות לטיפול במסגרת של רגולציה ענפית, משיקולים אלה: הצורך להעריך את הסיכון בתוך ההקשר הרחב של שימוש בטכנולוגיה; הצורך להתחשב במאפיינים הפרטניים ובאופן האסדרה של כל ענף במשק; הרצון לשמר גמישות בתגובה של הרגולטורים. לצד זאת, נכתב כי לנוכח המאפיינים המתבהרים של יישום סקטוריאלי של רגולציית בינה מלאכותית, יש שיקולים התומכים גם בקידום תפיסה רוחבית, ובפרט: עלולות להיות גישות סותרות בין מגזרים; יש חפיפה בין תחומי אסדרה; יש נטייה לבנות את האסדרה לבינה מלאכותית סביב אסדרה קיימת בלבד; הפעילות הבינלאומית הגוברת שמשפיעה על סקטורים רבים. כתוצאה מכך התפיסה הסקטוריאלית נבחנת שוב, תוך שיפור התיאום.

ביולי 2022 פורסמה לעיון הציבור טיוטת מדיניות הרגולציה הבריטית בנוגע לבינה מלאכותית (להלן: טיוטת מדיניות הרגולציה הבריטית).[93] נקודת המוצא לטיוטה היא קיומו של אקוסיסטם פעיל מאוד בתחום הבינה המלאכותית, הממקם את בריטניה במקום גבוה בעולם בדירוגים בתחום זה. המענה הרגולטורי הבריטי מבוסס על התמודדות עם השאלות והסיכונים הקשורים בבינה מלאכותית באמצעות שילוב בין הממשלה, רגולטורים, גופי תקינה טכניים והתעשייה.

טיוטת מדיניות הרגולציה הבריטית מציינת כי תחום הבינה המלאכותית במדינה משגשג, בין היתר בזכות המוניטין בדבר איכות הרגולציה הבריטית, המספקת בין השאר בהירות לגבי אופן פיתוח הרגולציה ונעשית תוך ביצוע הערכות השפעת רגולציה מקיפות. מאפיינים אלה, כך נטען, מגבירים את הוודאות ומאפשרים הגדלת ההשקעה בבינה מלאכותית בבריטניה וניתוב השקעות אליה.

בהמשך לכך, נכתב בטיוטת מדיניות הרגולציה הבריטית, כי על מנת לשמור על המעמד המוביל של בריטניה, יש לוודא שהכללים שיחולו על בינה מלאכותית עומדים בקצב ההתפתחות הטכנולוגית. ביחס לכך מציינת הטיוטה כי הבינה המלאכותית מוסדרת באופן חלקי בידי חוקים שונים שעוצבו לתכליות אחרות אבל חלים גם על בינה מלאכותית; הרגולטורים השונים נוקטים בפעולות כדי להיערך להשפעת הבינה המלאכותית; ובנוסף יש תפקיד חשוב לסטנדרטים בתחום זה ולכן יש לעודד פיתוח של תקנים שיהיו מובילים בעולם עבור פיתוח בינה מלאכותית ושימוש בה.

כמו כן, טיוטת מדיניות הרגולציה הבריטית עוסקת בחשש כי ריבוי הפעילות בתחום הסדרת הבינה המלאכותית – לרבות כללים וולונטריים, פיתוח תקנים, ועשייה רגולטורית – יוצרת אתגרים חדשים. בין היתר, הטיוטה מציינת את האתגרים הבאים: (1) אי-בהירות לגבי המסגרות החלות ואופן תחולתן, שעלולה להקשות על עסקים קטנים; (2) חפיפות בין מסגרות מסדירות; (3) אי-התאמה בין סמכויות הרגולטורים השונים; (4) פערים בחקיקה הנובעים מכך שהיא עוצבה ללא קשר לבינה מלאכותית. על רקע זה, מציינת הטיוטה את החשש שפערים אלה יביאו לפגיעה באמון צרכנים ועסקים, ויגבילו את הצמיחה והחדשנות בשימוש בבינה מלאכותית.

[92] United Kingdom, National AI Strategy, 22.09.2021, pillar 3, https://www.gov.uk.
[93] UK Government, Policy paper Establishing a pro-innovation approach to regulating AI, 20.07.22, https://www.gov.uk.

בהתאם לכך, על ידי נקיטת צעדים להגברת הבהירות והקוהרנטיות, המדיניות הבריטית מבקשת ליצור תפיסה רגולטורית תחרותית שמקדמת חדשנות ומחזקת את ההובלה הבריטית.

המדיניות הבריטית מציינת שייתכנו מספר גישות אסדרה. גישה אחת, שהיא גישת האיחוד האירופי, היא לקדם רגולציה רוחבית המבוססת על הסדרת בטיחות מוצרים. לעמדת המדיניות הבריטית, גישה כזו היא בעלת ערך כשמבקשים לתאם מדיניות בין מספר מדינות, אולם אין בה את היכולת להתאים את המדיניות באופן מפורט. גישה אחרת, היא ללא קביעת מגבלות רוחביות והשארת הסדרת הנושא לרגולטורים השונים בהתאם לשיקול דעתם. לצד התועלת בגמישות הרבה שבגישה כזו, לא יהיה בה תפיסה אחידה והיא עלולה להביא לבלבול. בהתאם לכך, הגישה הבריטית מציעה להגדיר את מאפייני הבסיס של בינה מלאכותית כדי למקד את השיח על מדיניות ההסדרה, אולם להשאיר את פיתוח הפרטים לרגולטורים המגזריים. רציונל מרכזי בגישה זו הוא שיש להסדיר את השימושים בטכנולוגיה, ולא את הטכנולוגיה ככזו.

על רקע זה, הגישה הכללית המוצעת לרגולציה היא תלוית הקשר; מכוונת לרגולציה רק במקרים של סיכון ממשי וברור (ולא במקרים של סיכון נמוך); קוהרנטית, מידתית וניתנת להתאמה (שימוש באמצעים רכים וולונטריים תחילה). הגישה הבריטית אינה מציעה תפיסה אחת הכוללת רשימה מרכזית של סיכונים ודרכי התמודדות, אלא מצפה מהרגולטורים השונים לפתח את אופן ניהול הסיכונים בהתאם לסוגי התרחישים. מטרת גישה זו גם לרתום את הידע והמומחיות הקיימים בידי הרגולטורים ביחס לסקטור הפעילות בו הם עוסקים.

עם זאת, על מנת שהגישה הסקטוריאלית תפותח באופן קוהרנטי, מוצעים במדיניות עקרונות-על חוצי מגזרים. העקרונות מבוססים על המלצות ה-OECD ומבהירים את המחויבות של בריטניה אליהם. העקרונות חוצי המגזרים הם: בטיחות, אבטחה ועמידות, שקיפות והסברתיות, שילוב שיקולי הגינות, הגדרת אחריות משפטית לבינה המלאכותית, הגדרת דרכי קבלת סעדים וערעור.

העקרונות נועדו לחול על הגורמים הפועלים בבינה מלאכותית, כאשר הרגולטורים יקבעו את תוכן העקרונות בפועל במגזר פעילות מסוים שעליו הם אמונים ואת אופן יישומו באותו מגזר. בהמשך לכך, ינקטו צעדים כדי לתאם את ההסדרים השונים בין הרגולטורים.

בהתאם לטיוטת המדיניות הבריטית, מוצע שלא לקדם את העקרונות בחקיקה, אולם לנקוט באמצעי מדיניות שונים כדי לתאם את הפעילות בין הרגולטורים השונים בתחום זה. המדיניות פורסמה להערות הציבור וצוין כי עד לסוף שנת 2022 יפורסם נייר מדיניות סופי.

4. חלק רביעי: סוגיות ואתגרים המתעוררים בקשר לבינה המלאכותית

לצד היתרונות המשמעותיים הגלומים בו, השימוש בבינה מלאכותית מעורר סיכונים וחששות שונים. סיכונים אלה תלויים בהקשר בו נעשה השימוש בבינה מלאכותית, ובין היתר באופן היישום הטכנולוגי וסוג השימוש.[94] ההתמודדות עם סיכונים וחששות אלה עשויה לעורר סוגיות משפטיות ורגולטוריות מורכבות, אשר יש בהן כדי ליצור אתגרים משמעותיים עבור המערכת המשפטית והרגולטורית ברחבי העולם ובישראל. על רקע זה, הפרק סוקר באופן כללי סוגיות, סיכונים ואתגרים נפוצים המתעוררים בעקבות פיתוח ושימוש בבינה מלאכותית, כפי שהם משתקפים מהספרות האקדמית ומדו״חות שונים, מישראל ומהעולם. הפרק הבא יעסוק בדרכי פעולה אפשריות להתמודדות עם הסוגיות, הסיכונים והאתגרים שיסקרו בפרק זה.

מטבע הדברים, הסקירה אינה ממצה את כל הסוגיות, הסיכונים והאתגרים שמתעוררים בקשר לבינה מלאכותית, ואופן הצגתם אינו ממצה את כל שידוע ונכתב לגביהם וכן משקף זווית מבט וניתוח שעשויה להיות חלקית. בנוסף, סדר הבאת הדברים אינו בהכרח משקף חשיבות של אתגר אחד למול משנהו או שכיחותו. עם זאת, מטרת הפרק היא להציף נושאים נפוצים העולים בשיח כחלק מהדיון בנוגע למדיניות הבינה מלאכותית בישראל. הסקירה נועדה לאפשר לגורמים השונים להכיר את הדברים, למפות אותם ולעקוב אחריהם.

יובהר כי פרק זה לא עוסק בפרשנות הדין הקיים בישראל, או בדין שנכון לאמץ באמצעות כללים או פרשנות בישראל. בנוסף, אין בהצגת הסוגיות, הסיכונים והאתגרים בפרק זה כדי לקרוא לפעולה ממשלתית, רגולטורית או משפטית מיידית, שאיננה היכרות ראשונית של הרגולטורים עם התחום. על אף קיומו של שיח אקדמי גובר בנושאים אלה, ופעילות רבה בתחום המדיניות, יש לזכור כי אנו מצויים בשלב ראשוני יחסית של ההתפתחות הטכנולוגית, ובהתאמה של המענים המשפטיים והרגולטוריים. ביחס לחלק ניכר מהסוגיות, הסיכונים והאתגרים שייסקרו להלן, הדיון מצוי בשלב מוקדם, וסביר להניח שככל שיחלוף הזמן הדברים יתבהרו ויתחדדו. לפיכך, בשלב זה מוצע לראות בסקירה זו מצע ראשוני לדיון ממשלתי, רגולטורי ומשפטי וכאמצעי לשיתוף הציבור בדיון זה.

בפרט, עבור גורמי הממשלה והרגולטורים הרלוונטיים, ולאור ההצעה להלן בעניין ״מיפוי השימושים בבינה מלאכותית והאתגרים הנלווים להם בענפים מוסדרים״, מוצע להסתייע בפרק זה עבור הבנה ומיפוי של הסוגיות, הסיכונים והאתגרים הכרוכים בשימושים קונקרטיים במערכות מבוססות בינה מלאכותית הנעשים בענף שעליו הם אמונים.

יודגש כי פרק זה מבקש להימנע מעיסוק מיוחד בשימושים של גופים מנהליים (משרדי ממשלה, יחידות סמך, חברות ממשלתיות וכד׳) במערכות מבוססות בינה מלאכותית. אמנם חלק ניכר מהסוגיות, הסיכונים והאתגרים הנסקרים בו, עשוי להתעורר גם ביחס לשימושים כאמור. אולם כפי שצוין לעיל, בכל הנוגע להיבטים המשפטיים של שימושי גופים מנהליים בבינה מלאכותית תתקיים עבודת מטה של הגורמים המשפטיים הרלוונטיים בממשלה, בהובלת מחלקת ייעוץ וחקיקה במשרד המשפטים, תוך עמידה על הדמיון והשוני בין שימושים כאמור לבין שימושים

[94] כך, שימוש בבינה מלאכותית בחקלאות עלול לעורר סיכונים בתחום הבטיחות, בעוד ששימוש בבינה מלאכותית בתחום החלטות אשראי עלול לעורר סיכונים בתחום הפרטיות ואיסור האפליה.

במערכות מבוססות בינה מלאכותית על ידי גופים פרטיים. בהתאם, דוגמאות משימושים של גופים מנהליים מהעולם, שיובאו להלן מפעם לפעם, יהיו למטרת המחשה של תופעות מסוימות או מורכבות טכנית, אך אינן מיועדות לעסוק באופן מיוחד באתגרים הנוגעים לגופים מנהליים.

4.1. __אפליה__

בבתי משפט במספר מדינות בארה"ב נעשה שימוש במערכות מבוססות בינה מלאכותית להערכת רמת הסיכון לרצידיביזם (פשיעה חוזרת) של עצורים ונאשמים, כדי לסייע בהחלטות שיפוטיות, כגון החלטה על הארכת מעצר או הפניה לחלופות מאסר. במחקר משנת 2016 בנוגע למערכת מסוג זה, COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), נמצא כי המערכת קבעה באופן סיסטמתי שהסיכוי של עצורים ונאשמים שחורים לרצידיביזם גבוה יותר מאשר זה של לבנים, וזאת גם כאשר לא היה בסיס לכך. למעשה נמצא כי כמות "השגיאות" של המערכת (שמשמעותן היא שהמערכת העריכה סיכוי גבוה לרצידיביזם שלא התממש בתוך תקופת הבקרה) הייתה כמעט כפולה ביחס לעצורים ונאשמים שחורים לעומת לבנים.[95]

אחד החששות הנפוצים בשיח בנוגע לשימוש במערכות מבוססות בינה מלאכותית, וכפועל יוצא מכך, אחד מההיבטים אשר יש מקום לבחון את אופן ההתמודדות עמם, הוא החשש שהפיתוח של מערכות אלו והשימוש בהן עלול להביא לאפליה אסורה. זאת, אף שהשימוש במערכות מבוססות בינה מלאכותית עשוי למתן את האפליה הנובעת מהטיות אנושיות ומבנים חברתיים, באמצעות קבלת ההחלטות על ידי אלגוריתם על בסיס מאפיינים אובייקטיבים ורלוונטיים.

הסיכון לאפליה אסורה בבינה מלאכותית נובע מהאפשרות להטיה בנתונים המשמשים לאימון המערכת או שימוש במאפיין רגיש באחד המשתנים הנאספים באופן שעלול להביא לתוצאות מפלות, או מהאפשרות להטיות מושרשות בתחום המומחיות המשמש לאימון הבינה המלאכותית.[96] סיכונים אלה נדונו בעבר בהקשרים של שימוש ביכולות מחשוב ונתונים מתקדמות, ובעיקר "נתוני עתק" ("big data"),[97] וקבלת החלטות אלגוריתמית,[98] אולם רמת המורכבות של מערכות מבוססות בינה מלאכותית, כמפורט להלן, מעלה חששות אלה לעיתים ביתר שאת.

להלן יוצגו סיכונים נפוצים העולים בספרות האקדמית ליצירת חשש לאפליה אסורה במערכות מבוססות בינה מלאכותית, והאתגרים המשפטיים והטכנולוגיים בהתמודדות עם חשש זה.

---

95 Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016).

96 מאפיין רגיש הינו או מאפיין כגון דת, גזע, מין, וכו, או נתון אחר שיש לו זיקה לאותו מאפיין (פרוקסי).

97 ראו למשל: Solon Barocas & Andrew D. Selbst, Big Data's Disparate Impact, 104 CALIF. L. REV. 671, 677-87 (2016); Tal Zarsky, Understanding Discrimination in the Scored Society, 89 Washington Law Review 1 (2015). Available at SSRN: https://ssrn.com/abstract=2550248.

98 ראו למשל: Danielle Keats Citron & Frank Pasquale, The Scored Society: Due Process for Automated Predictions, 89 WASH. L. REV. 1 (2014).

4.1.1. <u>סיכונים לאפליה אסורה במערכות מבוססות בינה מלאכותית</u>

כפי שיפורט להלן, וכפי שעולה מהכתיבה האקדמית בתחום, תוצאה מפלה המתקבלת עקב השימוש במערכות מבוססות בינה מלאכותית עלולה להיגרם מסיבות שונות, ובהתאמה ההתמודדות עם כל אחד מגורמי האפליה שונה.

כך לדוגמה, <u>אפשרות אחת</u> היא אפליה מכוונת (intentional discrimination). במקרה זה מפתחי המערכת או מפעיליה מעצבים אותה באופן מכוון כך שתיווצר אפליה. מערכות מבוססות בינה מלאכותית יכולות להקל על ביצוע אפליה מכוונת בהיותן כלי אפקטיבי לזיהוי ואפיון של השתייכות אדם לקבוצות אוכלוסייה שונות גם בנסיבות בהן שיקול זה לא אמור להיות גלוי. זאת, באמצעות יכולת מערכות אלה להסיק את ההשתייכות לקבוצה מסוימת על בסיס מידע ממגוון מקורות המלמדים על כך באופן עקיף. ניתן לעשות שימוש במידע זה על מנת להפלות במתכוון כנגד (או בעד) אותה קבוצה, למשל על ידי חסימת קבוצות אוכלוסייה ספציפיות מקבלת שירותים מסוימים.[99]

<u>אפשרות שניה</u> היא אפליה הנובעת מאופן בחירת הנתונים ועיצוב האלגוריתם המשמש לקבלת החלטות. החשש בהקשר זה הוא שעל אף שלמעצבי הבינה המלאכותית או הארגון המשתמש בה אין כוונה להפלות, אופן בחירת המשתנים (feature selection) המשמשים את המודל לקבלת ההחלטה, ועיצוב המודל, מייצרים אפליה הלכה למעשה.[100] לדוגמה, חוקרים גילו כי מערכת מבוססת בינה מלאכותית שנועדה לסייע לבתי חולים וחברות ביטוח בארה"ב לאתר מטופלים לתכנית טיפול ייעודית לחולים בסיכון גבוה, איתרה שלא במתכוון פחות מטופלים שחורים לתכנית. בדיעבד התברר כי אחד האינדיקטורים המרכזיים עליהם הסתמכה המערכת כדי לקבוע את רמת הסיכון נגע להיקף ההוצאות הרפואיות של המטופל עד אז. במחקר הסתבר כי מאחר שככלל היו למטופלים שחורים הוצאות רפואיות נמוכות יותר, הם קיבלו באופן שיטתי ציון נמוך מזה שניתן למטופלים לבנים, מה שפגע בזכאותם לתכנית הטיפול. כך, למרות שמשתנה ההוצאות הרפואיות של המטופל אינו מפלה כשלעצמו, השימוש בו יצר אפליה דה-פקטו באופן לא מכוון.[101]

---

[99] FREDRIK ZUIDERVEEN BORGESIUS, DISCRIMINATION, ARTIFICIAL INTELLIGENCE, AND ALGORITHMIC DECISION-MAKING, Study for the Counsel of Europe, 22 (2018) https://rm.coe.int/ (להלן: "המועצה האירופית")
במחקרו של פרופסור בורגיוס עבור מועצת אירופה הוא מציין את הסיבות הנפוצות הבאות, שחלקן יובהרו בטקסט, "To sum up, AI decision-making can lead to discrimination in at least six ways, which relate to (i) the definition of the target variables and the class labels; (ii) the labelling and (iii) collecting of the training data; (iv) the selection of the features; (v) proxies. And (vi) **organisations could use AI systems to discriminate** on purpose. AI can also lead to other types of unfair differentiation, or to errors";
לאחרונה פורסם מחקר מקיף מטעם מרכז המחקר של הפרלמנט האירופי, הסוקר באופן מקיף סוגים שונים של הטיות שעלולות להביא לאפליה אסורה. ראו:
European Parliament, European Parliamentary Research Services, Auditing the quality of datasets used in algorithmic decision-making systems, July 2022, https://www.europarl.europa.eu, (EPRS Study 2022) p. 5-16.

[100] שם, בעמ' 12.

[101] Ziad Obermeyer, Brian Powers, Christine Vogeli & Sendhil Mullainathan, *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 SCIENCE 447 (2019).

אפשרות שלישית קשורה לסיכון הנובע מהטיות המשקפות אפליה קיימת בחברה, המקבלת ביטוי בנתונים המשמשים לאימון מערכת הבינה המלאכותית._אפליה מסוג זה מתרחשת כאשר הנתונים עליהם המערכת "מתאמנת" או המדגם בו היא משתמשת משקפים אפליה קיימת בחברה.[102] אלגוריתם בינה מלאכותית "מתאמן" ו"לומד" על סמך הנתונים המוזנים לו. כאשר נתונים אלה משקפים אפליה קיימת, המערכת תשכפל ותנציח אפליה זו בהחלטותיה. דוגמה מוכרת נוגעת למערכת מבוססת בינה מלאכותית שהופעלה בחברה מסחרית לצורך גיוס עובדים. המערכת אומנה על בסיס דפוסי גיוס העובדים של החברה ב-10 השנים הקודמות, ובהמלצותיה היא שיקפה את הבולטות של גברים במשרות טכניות בתעשייה בתקופה זו בכך שהעדיפה באופן שיטתי גיוס של גברים למשרות אלו.[103] באופן דומה, נמצא בבריטניה שתוכנת מחשב שנועדה לסייע בבחינת מועמדותם של סטודנטים לרפואה שיקפה את העובדה שבעבר התקבלו ללימודי רפואה בעיקר גברים מקומיים, והפלתה לרעה נשים ומהגרים.[104]

השימוש במערכת עלול לא רק להנציח את האפליה הקיימת אלא אף להחריף אותה כתוצאה מהיזון חוזר של המערכת (feedback loop). כך למשל, תואר החשש שמערכת המשמשת את המשטרה לחיזוי פשע תעשה שימוש בנתונים היסטוריים מפלים על מנת להמליץ על אופן פריסת ניידות משטרה. מכיוון שפריסה רחבה יותר תוביל בסבירות גבוהה לעלייה בהיקף המעורבות המשטרתית באזור, לא מן הנמנע כי המערכת תלמד שצריך להגדיל אף יותר את הנוכחות המשטרתית באותם אזורים. כך למעשה תיווצר לולאה המזינה את עצמה, שתעצים את האפליה הקיימת בנתונים.[105]

אפשרות רביעית קשורה לכך שהמערכת "מתאמנת" על מאגרי מידע שאינם ייצוגיים מספיק (sample bias).[106] במצבים אלו, המערכת תתקשה לנתח בצורה מדויקת את ההשפעה של אלמנטים המאפיינים קבוצות אוכלוסייה שונות על ההחלטה. דוגמה בולטת לסוג זה של אפליה מצויה בתחום המחקר הרפואי. באופן היסטורי מחקרים רפואיים בוצעו בעיקר על גברים, ולכן היקף המידע הנוגע לרפואת גברים גדול משמעותית מזה הנוגע לרפואת נשים. כתוצאה מכך, כלי אבחון רפואיים המבוססים על בינה מלאכותית נחשפו למאגר מידע המכיל בעיקר מידע הנוגע לרפואת גברים, והם עלולים לעיתים ליצור אפקט מפלה בכך שיפעלו בצורה פחות יעילה ביחס לנשים.[107]

אפשרות חמישית היא שעשויה להיווצר אפליה כתוצאה משינוי הייעוד של המערכת, הקשרה או קהל היעד שלה. במקרים אלו, המערכת פותחה מתוך הנחה כי היא תשמש לקבלת החלטות בהקשר מסוים או בנוגע לאוכלוסייה מסוימת, ולכן פעילותה תהיה מותאמת לקבלת החלטות בהקשרים אלו או ביחס לפלחי אוכלוסייה זו. לכן, ניסיון לעשות שימוש במערכת זהה ביחס לאוכלוסיות אחרות עלול לגרום לתוצאות מוטות, שנובעות מכך שהיא לא תדע לנתח בצורה מדויקת את

---

[102] Borgeuis, p. 17.

[103] המועצה האירופית, לעיל ה"ש 99, בעמ' 15.

[104] Frederik Zuiderveen Borgesius, *Strengthening Legal Protection Against Discrimination by Algorithms and Artificial Intelligence*, 24 INT'L J. HUM. RTS. 1572, 1575 (2020).

[105] Lindsey Barrett, *Reasonably Suspicious Algorithms: Predictive Policing at the United States Border*, 41 N.Y.U. REV. L. & SOC. CHANGE 327, 337 (2017).

[106] Kirsten Lloyd, *Bias Amplification in Artificial Intelligence Systems*, 2 (2018). Available at: https://doi.org/10.48550.

[107] Isabel Straw, *The Automation of Bias in Medical Artificial Intelligence (AI): Decoding the Past to Create a better Future*, 110 ARTIFICIAL INTELLIGENCE MED. 101965, 101966 (2019).

המאפיינים השונים של ההקשר החדש או קבוצת האוכלוסייה החדשה.[108] ייתכנו אף מצבים בהם המערכת "תתאמן" על אוכלוסייה ממדינה מסוימת, אולם בפועל היא תוטמע במדינה אחרת, ואף לכך עלולה להיות השפעה מפלה על תוצאותיה של המערכת.

כאמור לעיל, התיאור של סיכונים אלה לאפליה אסורה אינו ממצה, אולם הוא נועד להציג את הצורך במודעות לאפליה אסורה עקב מאפייני הבינה המלאכותית.

4.1.2. <u>האתגרים בהתמודדות עם אפליה במערכות מבוססות בינה מלאכותית</u>

כמתואר לעיל, על אף ההבטחה הגלומה במערכות מבוססות בינה מלאכותית לצמצם אפליה שמקורה בהטיות המוכרות בשיקול דעת אנושי, קיים חשש כי השימוש במערכות עלול להביא דווקא להנצחת תופעות של אפליה קיימת ואולי אף להחרפתן או להתגבשות תופעות של אפליה מסוג חדש. חלק זה עוסק בקשיים המבניים, הן בצד הטכנולוגי והן בצד המשפטי, בהתמודדות עם אפקט מפלה שנוצר על ידי מערכות מבוססות בינה מלאכותית.

א. אתגרים טכנולוגיים:

על פניו, ניתן לחשוב על פתרון לחשש מאפליה באמצעות חסימה טכנית של המערכת מלהתאמן, ללמוד או להתחשב במסגרת תהליך קבלת ההחלטה בנתונים "אסורים" המבטאים שיקולים מפלים כגון מין, גזע, מוצא, דת, השקפה וכיו"ב. אולם, כפי שיפורט להלן, וכפי שתואר לעיל בקצרה, היכולת למנוע את ההכללה של מאפיינים מפלים כרוכה באתגרים לא מבוטלים.

<u>ראשית</u>, אף אם תהיה אפשרות לחסום את המערכת מלהתחשב באופן ישיר במאפיינים מפלים, עדיין ישנו חשש משמעותי כי המערכת תגיע לשקול אותם באמצעות התחשבות בפרמטרים אחרים אשר מתואמים עם המאפיין המפלה (proxy). כך למשל, מקום המגורים של הפרט שבעניינו מתקבלת ההחלטה (לפי כתובת או מיקוד), הוא נתון שלעיתים קרובות קורלטיבי לשיוך הקבוצתי, למאפיינים אתניים או למעמד סוציו-אקונומי. ישנו קושי אינהרנטי לאפיין מראש את כל הגורמים המקיימים קורלציה עם משתנים מפלים ולמנוע מהמערכת לשקול אותם, שכן החשש הוא שבאמצעות יכולות הניתוח הגבוהות של המערכת היא תמצא את הדרך להתחשב בקריטריונים המפלים. מעבר לכך, במקרים רבים הוצאתם גם לא תאפשר לשמר את יתרונות הבינה המלאכותית בניתוח כמויות עצומות של מידע. המערכת מסתמכת על קורלציות אלו, שפעמים רבות הן אף בלתי ניתנות לזיהוי על ידי גורמים אנושיים, על מנת לבסס ניבויים מדויקים יותר.[109]

<u>שנית</u>, מכיוון שבמקרים רבים קיים קושי מבני להסביר מנגנון קבלת ההחלטות של מערכות מסוימות המבוססות על בינה מלאכותית או כיצד הן הגיעו להחלטה מסוימת, יהיה קשה לאתר מקרים בהם המערכת הביאה בחשבון שיקולים מפלים, בין אם באופן ישיר ובין אם באופן עקיף, ואת המשקל שיוחס להם במסגרת קבלת ההחלטה (להרחבה ראו פרק "הסברתיות" להלן).[110]

[108] David Leslie, *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*, The Alan Turing Institute, 32 (2020), Available at: https://papers.ssrn.com.

[109] Anya E. R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1263-1265 (2020).

[110] Bert Heinrichs, *Discrimination in the Age of Artificial Intelligence*, 37 AI & SOC'Y 143, 150-151 (2022).

שלישית, במקרים רבים מאגרי מידע מקיפים ומגוונים ביחס לכל קבוצות האוכלוסייה פשוט אינם בנמצא. הדבר מפחית כאמור מאפקטיביות המערכת ביחס לקבוצות אוכלוסייה בלתי מיוצגות. חשוב להבהיר כי אין מדובר בבעיה נקודתית, אלא בבעיית רוחב שנובעת מהעובדה שיש פער היסטורי באיסוף נתונים ביחס לקבוצות מסוימות באוכלוסייה. אמנם, לעיתים ניתן להתמודד עם הבעיה באמצעות איסוף נתונים משמעותי על קבוצות אלו, אולם אפשרות זו מוגבלת ולא תמיד קיים תמריץ כלכלי מספק לביצועה.[111]

ב. אתגרים משפטיים:

אף אם מבחינה טכנולוגית ניתן יהיה להתגבר על האתגרים שתוארו לעיל, ישנם מספר אתגרים משפטיים שעלולים להקשות על ההתמודדות עם אפליה במערכות מבוססות בינה מלאכותית.

ראשית, כיום דיני איסור האפליה אוסרים הן על אפליה ישירה, דהיינו שימוש בקריטריונים מפלים כשלעצמם, והן על אפליה עקיפה, דהיינו פרקטיקות שיוצרות אפקט מפלה גם ללא כוונת מכוון (דוגמת שירות צבאי כתנאי לקבלה לעבודה, כאמצעי לאפליה של אוכלוסיות שאינן משרתות בצבא).[112] מבחינה משפטית, ההסתמכות על מערכות מבוססות בינה מלאכותית לקבלת החלטות עלולה לאתגר את ההכרה באפליה עקיפה. כמתואר לעיל, במקרים רבים המערכת עשויה להתבסס על משתנים בעלי ערך חיזויי הרלוונטיים מבחינה טכנולוגית לקבלת ההחלטה שהוגדרה עבורה, המגלמים למעשה אינדיקציה (proxy) לשיוך קבוצתי שאסור להשתמש בו.

אמנם על פני הדברים ובלי להביע עמדה באשר למצב המשפטי המצוי או הרצוי, נראה כי במקרים בהם הזיקה בין משתנה הפרוקסי לקריטריון המפלה היא מובהקת יותר, יהיה מדובר בשימוש אסור לפי דיני איסור אפליה. אולם, ככל שהזיקה רחוקה יותר, כך יהיה קשה לקבוע כי השימוש במשתנה גורם לאפליה עקיפה בהתאם לדיני איסור האפליה, ולפיכך אסור. במובן זה, השימוש בבינה מלאכותית עשוי לאתגר את דיני איסור אפליה, שכן המערכות המתבססות על תבניות והקשרים לא נוגעות בהכרח למאפיין המפלה, ולעיתים רחוקות ממנו, ועדיין עשויות להפלות.

שנית, דינים שונים המגבילים שימוש במידע עלולים למנוע יצירת מאגרי נתונים מגוונים ומאוזנים, על אף שמאגרים אלה מהווים תנאי הכרחי למערכת שאינה מפלה. דינים מגבילים כאמור, עשויים להשתייך לשני סוגים. **סוג ראשון** נוגע למצבים בהם הנתונים שמבוקש לכלול במאגר יהיו מוגנים. כך למשל, דיני הגנת הפרטיות, זכויות יוצרים, חסיונות וסודות מסחר מגבילים שימוש בנתונים מסוימים במאגרים ולכן עלולים למנוע את השימוש במידע הנדרש. לדוגמה, בניסיון ליצור אלגוריתם שיסייע בתהליך קבלתם של עובדים למקום העבודה, התבססות על נתוני עבר השייכים לחברה הקולטת בלבד עלולים לגרום להנצחת הטיות קיימות. ואולם, בהיעדר חובה רגולטורית לשיתוף מידע, יתעורר קושי להרחיב את מאגר הנתונים הקיים מגורם חיצוני שכן מידע כזה עשוי להיות מוגן באמצעות דיני הפרטיות (ביחס לעובדים) ודיני הקניין הרוחני (ביחס למעסיקים). **הסוג השני** של ההגבלות עניינו הגבלת השימוש במאגרי הנתונים עצמם. מאגרי נתונים רבים המשמשים ללמידת מכונה יהיו מוגנים בזכויות קניין רוחני, כגון סודות מסחריים או זכויות יוצרים, גם אם

[111] להשוואה ראו: EUR. COMM'N, DATA COLLECTION IN THE FIELD OF ETHNICITY (2017).

[112] חוק איסור הפליה במוצרים, בשירותים ובכניסה למקומות בידור ולמקומות ציבוריים, התשס"א-2000; חוק שוויון ההזדמנויות בעבודה, התשמ"ח-1988. להתייחסות לשני סוגי האפליה במשפט הישראלי ראו למשל: דנג"ץ 4191/97 **רקנט נ' בית-הדין הארצי לעבודה**, נד(5) 330, פס' 21-27 לפסק הדין של הנשיא ברק (2000).

כל פריט מידע לא יהיה מוגן בפני עצמו. משכך, שימוש במאגר עצמו על ידי גורם אחר ללא אישור עלול לעלות כדי הפרת הזכות של בעל הקניין הרוחני.[113] הדבר עלול להוות בעיה מכיוון שיצירת מאגר נתונים שכבר קיים בשוק משמעותה בזבוז משאבים וחוסר יעילות ואולי גם פגיעה במוצר הסופי. על רקע זה, נדרש לבצע איזון בין דרישה להרחבה וגיוון של מאגרי המידע ל"אימון" המערכת כדי לצמצם את החשש לאפליה ולדייק את המערכת, לבין האתגרים שבגיבוש וניהול מאגרי מידע נרחבים שכאלה, ובכלל זה שמירה על פרטיות המידע והגנה על סודות מסחריים וזכויות קניין רוחני.

4.2. מעורבות אנושית

אחד הציטוטים המפורסמים ביותר בסדרה "הממלכה הקטנה" ששודרה בבריטניה בתחילת המילניום הוא "The computer says no". אישה הכורעת ללדת; ילדה המגיעה לניתוח; עובדת בשיחת משוב על ביצועיה; או לקוח בסוכנות נסיעות – כולם נתקלו בפקידה המקישה באיטיות על מקלדת המחשב ומגיבה באדישות "המחשב אומר לא". הציטוט הזה, שהפך לשגור בבריטניה, נועד לבקר אנשי שירות לקוחות ופקידים ממשלתיים שמסתמכים אך ורק על המידע המופיע במחשב על מנת לקבל החלטות, לעיתים באופן שמנוגד לגמרי להיגיון הבריא.[114] חשש זה מלווה את שילוב המחשבים בארגונים כחלק מתהליך איסוף ועיבוד מידע וייצור תחזיות או החלטות מורכבות. כעת, כ-20 שנים לאחר שהסדרה שודרה, הביטוי קם לתחייה ומשמש לתיאור החשש שהשימוש במערכות המבוססות על בינה מלאכותית בתהליכי קבלת ההחלטות עלול להוביל למציאות דומה; מציאות שבה יסתמכו כמעט באופן בלעדי על המערכות בלי מעורבות אנושית בתהליך קבלת ההחלטה.

סוגיה מרכזית שמתעוררת בקשר לשימוש במערכות המבוססות על בינה מלאכותית, נוגעת למידת המעורבות האנושית הנדרשת במסגרת תהליך קבלת החלטה הנסמך על מערכת מסוג זה.

מאפיין מרכזי שמייחד מערכות מבוססות בינה מלאכותית הוא יכולתן לפעול באופן אוטונומי המדמה בינה אנושית ובמקרים רבים מסוגל לבצע פעולות חישוב ביעילות רבה יותר מבינה אנושית. בין היתר, הן מאפשרות לספק תחזיות או למצוא קורלציות בין פרטי מידע ללא מעורבות אנושית ולעיתים רבות אף באיכות גבוהה יותר. מערכות אלה יכולות באופן עקרוני להתאים עצמן לסביבה משתנה ולאתגרים בלתי צפויים (הגם שלא תמיד באופן מלא, ראו פרק "אמינות, עמידות, אבטחה ובטיחות" להלן).[115]

העיסוק במעורבות אנושית בבינה מלאכותית נעשה בשני הקשרים מרכזיים. <u>הקשר אחד</u> נובע מעיקרון ערכי-מוסרי שלפיו בעניינים שיש להם השפעה (משמעותית) על אדם, ההחלטה תתקבל על ידי אדם, כגון החלטה על שלילת חירות או החלטות שיש בהן מימד ערכי או אתי.[116] <u>הקשר שני</u> הוא תועלתני, ומטרתו להתמודד עם החשש מפני כשלים בטכנולוגיה עקב בעיות אינהרנטיות (כגון תכנון או ביצוע לקוי, אירוע לא צפוי, שימוש לרעה), באמצעות שילוב אדם בתהליך קבלת ההחלטות. מובן

[113] Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579 (2018).

[114] Doa A. Elyounes, *"Computer Says No!": The Impact of Automation on the Discretionary Power of Public Officers*, 23 VAND. J. ENT. & TECH. L. 451, 453 (2021).

[115] *Establishing a pro-innovation approach to regulating AI*, GOV.UK (Jul. 18, 2022), https://www.gov.uk.

[116] ביטוי פוזיטיבי לעיקרון זה בדין הקיים הינו הוראת סעיף 22 ל-GDPR העוסק בקבלת החלטות אוטונומיות לגבי אדם, ראו: Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, 2016 O.J. (L 119) 1, 46.

כי בין שני עקרונות אלה קיימים קשרי גומלין. מן העבר השני, וכפי שיתואר להלן, המעורבות האנושית עלולה גם לפגוע באיכות וביעילות תהליך קבלת ההחלטה, ומכאן שאינה חפה מקשיים. על כן, הצורך במעורבות אנושית, ומידת המעורבות הרצויה, הן סוגיות מורכבות אשר צריכות להיות מוכרעות בהתאם לנסיבות השימוש ועל בסיס הקשרים אלה.

לצורך הצגה של היבטים שונים של סוגיה זו יסקור הפרק את המאפיינים השונים של מעורבות אנושית בקבלת החלטות של בינה מלאכותית, ולאחר מכן את היתרונות במעורבות אנושית, ואת האתגרים במעורבות אנושית.

4.2.1. **<u>מהי מעורבות אנושית?</u>**

מעורבות אנושית מתייחסת לשילוב אדם בתהליך קבלת ההחלטה או הפעולה של מערכת מבוססת בינה מלאכותית שיש לה מאפיינים אוטונומיים, בשלב שבו היא פועלת.

ניתן לחלק את מידת המעורבות בתהליך לשלוש קטגוריות: <u>בקטגוריה הראשונה</u> נמנים מצבים בהם המערכת מקבלת את ההחלטה בעצמה, וגורמים אנושיים לא לוקחים כל חלק בתהליך קבלתה ואף לא מפקחים עליו; נהוג לכנות אותם כ-human **off** the loop. <u>בקטגוריה השנייה</u> מצויים מצבים שבהם המערכת מקבלת את ההחלטה בעצמה, אולם לגורם אנושי יש סמכות לפקח על ההחלטה ולהתערב במידת הצורך, כאשר הפיקוח יכול להתרחש בזמן אמת (תוך כדי ההחלטה) או בדיעבד (כגון בערעור על ההחלטה); נהוג לכנות מצבים אלו כ-human **on** the loop. <u>הקטגוריה השלישית</u> עוסקת במצבים שבהם נדרשת השתתפות של הגורם האנושי בדרך של קבלת ההחלטה עצמה, כך שהגורם האנושי מפעיל שיקול דעת עצמאי, אך ההחלטה מתקבלת בסיוע המערכת; נהוג לכנותם כ-human **in** the loop.

חלוקה אפשרית אחרת למידת המעורבות האנושית בתהליך קבלת החלטה מבחינה בין שישה מצבים: **אנושי** (manual) – כשהגורם האנושי מקבל את ההחלטה בעצמו ופועל ללא כל סיוע מהמערכת; **עצה** (advice) – כשהגורם האנושי מקבל את ההחלטה בעצמו, אך הוא עושה זאת תוך הסתייעות בהמלצה של המערכת; **הסכמה** (consent) – כשההחלטה מתקבלת על ידי המערכת עצמה אולם היא יכולה לצאת לפועל רק לאחר אישור של גורם אנושי; **וטו** (veto) – כשהמערכת מחליטה ופועלת בעצמה באופן אוטומטי, אולם לגורם האנושי יש הזדמנות להטיל וטו על ההחלטה של המערכת; **ערעור** (appeal) – כשהמערכת מחליטה בעצמה ופועלת באופן אוטומטי, אולם יש למושא ההחלטה אפשרות לערער עליה בפני גורם אנושי; **אוטונומי** (autonomous) – כשהמערכת מחליטה ופועלת בעצמה בלי ליידע את הגורם האנושי ובלי שניתן לערער על ההחלטה.[117]

[117] Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, VAND. L. REV. 7, 10-11 (Forthcoming 2023); Mica R. Endsley, *The Out-of-the-Loop Performance Problem and Level of Control in Automation*, 37 HUM. FACTORS 381, 385 (1995).

4.2.2. יתרונות המעורבות האנושית

מעורבות אנושית עשויה לסייע, בהקשרים מסוימים, בהתמודדות עם חלק מהאתגרים הכרוכים בשימוש במערכות מבוססות בינה מלאכותית. כפי שיפורט להלן, בהתאם להקשר השימוש ולסוג הסיכונים, היא עשויה לשפר את איכות תהליך קבלת ההחלטה ולצמצם טעויות של המערכת; לשפר את מידת האחריותיות של מערכות; ולהגביר את הלגיטימציה של ההחלטות ולהפחית את הפגיעה בכבודו של מושא ההחלטה.[118]

ראשית, מעורבות אנושית (וכאמור למעורבות אנושית יכולות להיות דרגות שונות) עשויה לעיתים לשפר את האיכות והדיוק של תהליך קבלת ההחלטה ולמנוע או לצמצם טעויות בהווה ובעתיד. טיעון מקובל הוא שאנשים מבינים בצורה טובה יותר את המטרה של ההחלטה, ויש להם אינטואיציה שחסרה אצל מערכות המבוססות על בינה מלאכותית. בהתאם לכך, במקרים מסוימים, לאנשים יש יתרונות יחסיים על פני מערכות המבוססות על בינה מלאכותית ומעורבותם יכולה לשפר את איכות תהליך קבלת ההחלטה. כאשר כתוצאה מכשל בפיתוח המערכת או בשימוש בה היא מגיעה להחלטה המסתמנת כשגויה ואדם יכול היה לזהות שגיאה זו, התרומה האנושית ברורה, שכן אדם יוכל לסטות מהחלטת המערכת ולתקנה במקרה נתון ולמקרים הבאים; אולם גם כאשר המערכת פועלת כהלכה, מעורבות גורם אנושי בתהליך קבלת ההחלטה עשויה לשפרו בטווח הקצר והארוך, בין היתר, על בסיס תובנות ביחס לעתיד שאינן מצויות בנתוני העבר עליהם מסתמכת המערכת. על רקע החשש מטעויות והטיות הכרוכות בשימוש במערכות אלה (ראו פרק "אפליה" לעיל וכן פרק "אמינות, עמידות, אבטחה ובטיחות" להלן), ובמיוחד בשלבים הראשונים של הטמעת השימוש במערכות, המעורבות האנושית צפויה לסייע להתמודד עם אתגרים אלה.[119]

שנית, הדרישה למעורבות אנושית עשויה לשפר ולהרחיב, כמו גם להביא לידי ביטוי, את האחריותיות (accountability) (ראו בעניין זה פרק "אחריותיות"); כאשר גורם אנושי מעורב בהחלטה של המערכת, ויש לו את הזמן, את המידע הרלוונטי או המומחיות ואת הסמכות לקבל את ההחלטה הסופית בעצמו או לפקח עליה, ניתן להביא באופן זה לידי ביטוי את האחריותיות של מפתחי או מפעילי הבינה המלאכותית לאופן פעולתה כמו גם לתוצאות של אותה החלטה.

שלישית, הצדקה נוספת למעורבות אנושית בתהליך קבלת ההחלטה היא שהחלטות שבהן מעורב גורם אנושי עשויות להיתפס כלגיטימיות יותר ולסייע בהגברת אמון הציבור בשימוש במערכות. מחקרים אמפיריים העלו כי אנשים סבורים שהחלטות שמתקבלות באופן אוטונומי על ידי מערכות מבוססות בינה מלאכותית, נתפסות ככאלה שמידת הלגיטימציה שלהן פחותה, בהשוואה להחלטות שמתקבלות על ידי גורמים אנושיים.[120] הסבר אחד לממצאים אלו, הוא שתפיסת הלגיטימיות של החלטה מסוימת מושפעת, בין היתר, מאופי ההליך שהוביל לה, ומהזכויות שמוקנות למושא ההחלטה בהליך. החלטה מסוימת תיתפס כלגיטימית יותר כשהיא התקבלה לאחר הליך הוגן (Due

[118] Kiel Brennan-Marquez, Daniel Susser & Karen Levy, *Strange Loops: Apparent* versus Actual Human Involvement in Automated Decision-Making, 34 BERKELEY TECH. L. J. 745, 746 (2019).

[119] Kiel Brennan-Marquez & Stephen Henderson, *Artificial Intelligence and Role-Reversible Judgment*, 109 J. CRIM. L. CRIMINOLOGY 137, 146-147 (2019); כפי שיפורט להלן, ישנם גם מקרים בהם המעורבות האנושית תפגע באיכות ההחלטה שתתקבל ולכן לא ניתן לקבוע באופן גורף כי מעורבות אנושית משפרת את איכות תהליך קבלת ההחלטה.

[120] Ari Waldman & Kirsten Martin, *Governing algorithmic decisions: The role of decision importance and governance on perceived legitimacy of algorithmic decisions*, BIG DATA & SOC'Y 1 (2022).

process), תוך מתן אפשרות למושא ההחלטה להישמע ולהשפיע על ההליך ושמירה על כבודו.[121] השימוש במערכות מבוססות בינה מלאכותית לקבלת החלטות מציב אתגרים משמעותיים לאפשרות להשיג יעדים אלו. זאת, הן משום שבמקרים רבים תהליך קבלת ההחלטה או הנימוקים שבבסיסה לא יהיו שקופים למושא ההחלטה (ראו פרק "הסברתיות" להלן),[122] הן משום שיכולת מושא ההחלטה להשמיע את קולו ולהשפיע על הליך קבלת ההחלטה מוגבלת.[123] הסבר נוסף הוא שהערעור בתפיסת הלגיטימיות נובע מכך שאנשים רגילים שהחלטות בעניינם בהקשרים מסוימים מתקבלות על ידי גורמים אנושיים, ומעדיפים שהליך קבלתן יביא בחשבון גם תכונות אנושיות כמו רגישות וחמלה.[124] מעורבות אנושית, ובפרט קבלת ההחלטה הסופית על ידי גורם אנושי, מאפשרת להתמודד עם קשיים אלו. הגורם האנושי יוכל לשמוע את מושא ההחלטה, לבחון מחדש את ההחלטה ולנמק אותה (בכפוף לאילוצים הטכנולוגיים, כמפורט בפרק ה"הסברתיות").

רביעית, יש הטוענים כי הכפפת אנשים להחלטות שמתקבלות על ידי מערכת מבוססת בינה מלאכותית בלבד, ללא מעורבות אנושית, עלולה לגרום, בהקשרים מסוימים, לפגיעה בכבודם של הגורמים שבעניינם מתקבלת ההחלטה. הסיבה המרכזית לכך היא שכאשר החלטות מתקבלות אך ורק על ידי מערכת ישנו חשש לפגיעה במרכיב האינדיבידואלי של מושא ההחלטה. למעשה ברוב המקרים הצורה שבה המערכת מקבלת החלטה היא באמצעות מציאת מתאם סטטיסטי בין המאפיינים של האדם לבין נתוני העבר, תוך ניסיון לשבץ את המקרה והאדם שבפניה לתוך קטגוריה של מקרי עבר. מכיוון שהתהליך שמבצעת המערכת לא מאפשר לאינדיבידואל להפגין את הייחודיות שלו, נטען כי למעשה מדובר בראייתו של הפרט כסט תכונות בלי לבחון את מכלול הזהות האנושית שלו באופן פרטני.[125] מעורבות גורם אנושי יכולה לספק למושא ההחלטה הזדמנות להביע את עצמו, ולוודא כי ההחלטה תתקבל על בסיס בחינת האדם כמכלול ולא כסט תכונות.

4.2.3. אתגרים ביחס למעורבות אנושית

לצד היתרונות המשמעותיים שניתן להשיג באמצעות מעורבות אנושית, שעיקרם הוצגו לעיל, החלה של דרישה למעורבות אנושית אינה מובנת מאליה. זאת, שכן כפי שיפורט להלן, קיים חשש כי במקרים רבים המעורבות האנושית לא תהיה אפקטיבית ואף עלולה להסב נזק.

ראשית, הציפייה שאנשים יוכלו להתערב באופן אפקטיבי, לפקח על פעילות המערכת ולתקן את טעויותיה – בפרט כשמדובר בטעויות שאינן נורמטיביות, אלא פרקטיות – אינה בהכרח ריאלית בהקשרים מסוימים. עצם השימוש במערכות הבינה המלאכותית מתבסס על ההערכה שהן מסוגלות בהקשרים מסוימים לחזות בצורה טובה יותר מבני אדם מהי ההחלטה הנכונה. על כן, ההנחה שאנשים יוכלו באותם הקשרים לשפוט בצורה טובה יותר מהמערכת האם ההחלטה שהתקבלה היא נכונה, אינה מתבקשת. הדבר נכון במיוחד ביחס למקרים שבהם אין הסבר מלא כיצד המערכת הגיעה להחלטה (ראו פרק "הסברתיות" להלן), אך רלוונטי גם למקרים אחרים

---

[121] Ari Ezra Waldman, *Algorithmic Legitimacy*, *in* THE LAW OF ALGORITHM 107, 110 (2020); Aziz Z. Huq, A Right to Human Decision, 106 VA. L. REV. 611, 656-671 (2020).

[122] Brennan-Marquez & Henderson, לעיל ה"ש 119, בעמ' 148; Crootof, Kaminski & Price II, לעיל ה"ש 117, בעמ' 54-55.

[123] Ric Simmons, *Big Data and Procedural Justice: Legitimizing Algorithms in the Criminal Justice System*, 15 OHIO ST. J. CRIM. L. 573, 579-580 (2018).

[124] שם, בעמ' 573-574.

[125] Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic accountability,* 92 S. CAL. L. REV. 1529, 1541-1545 (2019).

שבהם ככלל המערכת מתבססת על מספר עצום של נתונים ועל האינטראקציות ביניהם כדי לקבל החלטה, באופן שהמוח האנושי מתקשה לבצע.[126] אם הגורם האנושי המעורב לא יודע את הנימוקים להחלטה, או לא מבין כראוי את מכלול הנתונים והאינטראקציה שהובילו לקבלת ההחלטה, הוא עלול להתקשות בתיקון החלטות שגויות של המערכת.

שנית, לא אחת השימוש במערכות מבוססות בינה מלאכותית נעשה במטרה להתמודד עם העובדה שתהליך קבלת ההחלטות האנושי אינו אופטימלי. בפרט, הגם שמתעוררים חששות לא מבוטלים לאפליה בשימוש במערכות מבוססות בינה מלאכותית (ראו פרק "אפליה" לעיל), יש לזכור שמנגד קיימים חששות מקבילים ביחס לאפליה על ידי גורמים אנושיים. בנוסף, מחקרים מתחום הכלכלה ההתנהגותית מלמדים כי ישנן הטיות רבות שמשפיעות על תהליך קבלת ההחלטות האנושי, וגורמות לכך שחלק ניכר מההחלטות שמתקבלות על ידי אדם בסופו של דבר אינן אופטימליות. על-כן, במקרים מסוימים, הותרת ההחלטה הסופית בידי הגורם האנושי עלולה לרוקן מתוכן את אחד היתרונות המשמעותיים שניתן להשיג באמצעות השימוש במערכות מבוססות בינה מלאכותית והוא היכולת לנטרל, למצער בצורה חלקית, אפליה והטיות אלה.[127]

שלישית, ישנה שאלה כבדת משקל בנוגע למשמעות המעשית של מעורבות אנושית בתהליך קבלת החלטה, כאשר ניתנת המלצה ברורה של מערכת מבוססת בינה מלאכותית. מחקרים מתחום הכלכלה ההתנהגותית הראו כי ישנן הטיות ייחודיות המאפיינות את האינטראקציה בין מקבלי החלטות אנושיים לבין מערכות המבוססות על בינה מלאכותית. אחת ההטיות המרכזיות לעניין זה היא הטיית האוטומציה (automation bias), במסגרתה אנשים נוטים לאמץ את המסקנה אליה הגיעה מערכת אוטומטית (ובכלל זה מערכת מבוססת בינה מלאכותית), בלי להפעיל את אותה מידה של שיקול דעת כמו ביחס להחלטות שלא התקבלו על ידי מערכת. כתוצאה מכך, אנשים נוטים לפעול בהתאם להמלצות האלגוריתם גם כשיש ראיות סותרות, בגלל תפיסה כללית ש"האלגוריתם צודק" (the machine knows), וכפועל יוצא מתקשים בתיקון טעויות הנעשות על ידי המערכת.[128] מצדו השני של המתרס, ישנם ממצאים שלפיהם אנשים נוטים להימנע באופן לא רציונלי מאימוץ ההחלטה שהתקבלה על ידי מערכת, ומעדיפים את שיקול דעתם גם בהיעדר הצדקה ברורה לכך. הטיה זו, ההופכית לקודמת שצוינה, מכונה בספרות סלידה אלגוריתמית (algorithm aversion).[129] על רקע זה, ישנו חשש כי אותם גורמים אנושיים המעורבים בתהליך יתקשו להעריך בצורה נכונה את טיבה של ההחלטה שהתקבלה על ידי המערכת, וכתוצאה מכך יטעו בהחלטה אם לאמץ את המסקנה אליה הגיעה המערכת או לסטות ממנה.[130] כך שבפועל ספק אם המעורבות האנושית תהיה אפקטיבית בתיקון טעויות, והיא אף עלולה לגרוע מאיכות תהליך קבלת ההחלטות.

---

[126] Adrien Bibal, Michael Lognoul, Alexandre de Streel & Benoît Frénay, *Legal Requirements on Explainability in Machine Learning*, 29 Artificial Intelligence & L. 149, 156 (2021).

[127] Jeremy Waldron, It's All for Your Own Good, The New Yorker Review (9.10.2014) https://www.nybooks.com; איל זמיר ודורון טייכמן "ניתוח התנהגותי של החלטות שיפוטיות: הישגים ואתגרים" **משפט ועסקים** יט 57 (2015).

[128] Marina Chugunova & Daniela Sele, *We and It: An Interdisciplinary Review of the Experimental Evidence on How Humans Interact with Machines*, 99 J. Behav. Experimental Econ. 1, 19-20 (2022).

[129] Hasan Mahmud, A.K.M. NajmulIslam, Syed Ishtiaque Ahmed & Kari Smolander, *What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion*, 175 Technological Forecasting and Soc. Change 121390, 121390-121391 (2022).

[130] Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, 45 Computer L. Security Rev. 1, 14-18 (2022).

רביעית, ובהמשך לדיון בנקודה הקודמת, קיים חשש שגם במקרים המתאימים וללא הטיות ייחודיות האדם שמעורב בתהליך קבלת ההחלטה לא יסטה מהמלצות המערכת, מפאת החשש לנשיאה באחריות במקרה של טעות בשיקול דעתו (ראו גם פרק "אחריותיות" להלן). כך למשל, רופא עלול לחשוש לסטות מהמלצה של המערכת בנוגע לטיפול רפואי מומלץ על ידי המערכת מתוך חשש כי המטופל לא יגיב טוב לטיפול החלופי ויטען שהרופא פעל ברשלנות. באופן דומה, פקיד בבנק עלול לחשוש במיוחד לאשר מתן אשראי למי שהמערכת המליצה שלא לאשר לו, למרות שהפקיד סבור שיש הצדקה לאשר לו, וזאת בגין החשש כי במקרה שיהיה קושי בפירעון הוא ייאלץ לשאת באחריות בגין התוצאה השלילית מכיוון שסטה מהמלצת המערכת. חשש זה מהווה תמריץ משמעותי עבור הגורם האנושי המעורב להיצמד להמלצות המערכת ולא לסטות מהן.

חמישית, קיים חשש שהצורך במעורבות אנושית יפחית באופן משמעותי את התועלת הנובעת משימוש במערכות מבוססות בינה מלאכותית. אחת המטרות המרכזיות של השימוש במערכות אלה, הוא האופן שבו ניתן לייעל את תהליך קבלת ההחלטות באמצעותן. ככלל, מערכות אלה מסוגלות לפעול לאורך כל שעות היממה, הן לא "מתעייפות" ועלותן בטווח הארוך עשויה להיות נמוכה מתשלום לעובדים.[131] בהתאם, הדרישה למעורבות אנושית – שמחייבת להמשיך ולהעמיד תשומות אנושיות על מגבלותיהן ועלויותיהן – עלולה לפגוע ביעילות המושגת באמצעות השימוש במערכות אלה, וככל שתהיה רחבה יותר, הפגיעה ביעילות עלולה להיות משמעותית יותר.[132]

שישית, חשש נוסף הוא שההנחה כי אנשים מסוגלים לפקח באופן אפקטיבי על המערכת תגרום לתחושה לא מוצדקת של ביטחון. השימוש במערכות מבוססות בינה מלאכותית עלול להוביל לתוצאות לא רצויות במקרים שבהם המערכת אינה מושלמת, סובלת מהטיות וכדומה. ההנחה כי אנשים מסוגלים לפקח באופן אפקטיבי על המערכות באמצעות מנגנון המעורבות האנושית, יכולה לשמש כתמריץ לא מוצדק וגדול מדי לאימוצן של מערכות מבוססות בינה מלאכותית בשלב מוקדם מדי או במקרים שבהם הנזק שצפוי להיגרם כתוצאה מטעות בהחלטה של המערכת גבוה.[133] זאת בעוד שאלמלא מעורבות אנושית, שכאמור אינה יכולה בהכרח לתת מענה לחסרונותיה של המערכת, ייתכן שהשימוש היה נמנע, נדחה, מוגבל או מפוקח יותר.

לבסוף, יש חשש כי המעורבות האנושית תוביל לכך שהאחריות במקרה של טעות תעבור מכתפי מקבלי החלטות שהחליטו להשתמש במערכת או מפתחי המערכת, לגורם האנושי המעורב בתהליך קבלת ההחלטה, גם כשהדבר אינו מוצדק. למשל, במקרה שבו מקום עבודה יימנע מהעסקת עובדים מקבוצת אוכלוסייה מסוימת, יתכן כי הנטייה תהיה להאשים את הגורם האנושי המעורב גם אם המערכת באופן שיטתי דירגה מועמדים מאותה קבוצה ככאלה שמתאימים למקום העבודה במידה פחותה; זאת, גם אם לא בוצעו בדיקות נאותות מקובלות לצמצום החשש לאפליה או שלא הקפידו שהמערכת תוכל להסביר את החלטותיה וכפועל יוצא מידת השליטה של הגורם האנושי המעורב בתהליך הייתה חלקית.[134] יודגש כי לאו דווקא מדובר באחריות משפטית (פלילית או נזיקית), ובין

[131] Dominique Hogan-Doran, *Computer says "no": automation, algorithms and artificial intelligence in Government decision-making*, 13 JUD. L. REV. 1, 3-14 (2018).
[132] Green, לעיל ה"ש 130 בעמ' 23-21.
[133] שם, בעמ' 21-18.
[134] להרחבה ראו, Madeleine Clare Elish, *Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction*, 5 ENGAGING SC., TECH. & SOC'Y 40 (2019). כך לדוגמה, יש הטוענים כי הנטייה להסיט את האחריות לגורם האנושי שהיה מעורב בתהליך על חשבון גורמים אחרים באה כבר לידי ביטוי בתחום התעופה. שם, הנטייה היא להאשים את הטייס האנושי ולא את המפתחים של מערכת הטייס האוטומטי או המעצבים של הממשק שלה, גם

היתר למשל, ניתן לעשות שימוש בגורמים אנושיים כדי לשמור על תדמית המערכת באמצעות ניסיון להראות כי הכשל נגרם בגלל הגורם האנושי שהיה מעורב בתהליך, ולא בגלל תקלה במערכת.[135]

4.3. **הסברתיות**

במטרה להתמודד עם ניסיונות הונאה, משנת 2013 החלה רשות המסים ההולנדית לעשות שימוש במערכת מבוססת בינה מלאכותית שפותחה כדי לנסות לאתר הונאות בבקשות לקבלת קצבת ילדים. במשך שש שנים המערכת פעלה עד שהתברר שהשימוש בה הוביל לכך שכ-26,000 מבקשים הואשמו שלא בצדק בהונאה, וכפועל יוצא, נדרשו להשיב חלק מן הסכומים שקיבלו, סכומים שלעיתים הגיעו לעשרות אלפי אירו. החזרים אלו הביאו לכך שמשקי בית רבים, בעיקר משכבות חלשות באוכלוסייה, הידרדרו לעוני, לאובדן הבית ואף לחדלות פירעון. אחת הסיבות המרכזיות שהובילו להאשמה השגויה של המבקשים היא שפקידי רשות המסים ההולנדית, שהחליטו בסופו של דבר אם להאשים את המבקש בניסיון הונאה, קיבלו את התחזית של המערכת ללא הסבר ביחס לאופן בו היא פעלה. כתוצאה מכך, ברוב המקרים לא נותרה לפקידים ברירה אלא לאמץ את ההמלצות של המערכת, ובהתאם הם גם לא יכלו להסביר למבקשים מדוע נחשדו בהונאה.[136]

4.3.1. **על הסברתיות ותופעת "הקופסא שחורה"**

כמתואר לעיל, בינה מלאכותית מבוססת על יכולות חישוביות היוצרות מודל תחזיות וקבלת החלטות שלעיתים לא ניתן "לחלץ" מתוך המערכת. הדבר נובע מכך שבשונה מאלגוריתם רגיל, בבינה מלאכותית אופן "אימון" המערכת מוביל לכך שהתוכנה "כותבת את עצמה". במקרים אלו, גם מעצבי המערכת לא בהכרח יכולים להתחקות אחר החלטה פרטנית שלה. יכולות טכנולוגיות מתקדמות אלה מחריפות תופעה המתוארת כ-"קופסא שחורה" (black box) – מושג שנועד להמחיש מצב בו המידע המוזן למערכת (הקלט) ידוע, וכך גם התוצאה אליה הגיעה (הפלט); אולם דרך פעולת המערכת הפנימית אינה ידועה למי שמושפע ממנה, או אף אינה ניתנת להבנה פשוטה.

המשמעות של קבלת החלטה שלא ניתן להבין או להסביר על ידי מכונה, עלולה לעורר חשש לפגיעה בכבוד האדם ובאוטונומיה שלו, בכך שהוא לכאורה נתון להחלטה שרירותית. במקרים אלה מתעוררת שאלה כללית והיא באילו מצבים ראוי לדרוש כי יינתן הסבר ביחס לאופן פעילות המערכת ובעיקר מה צריכה להיות רמת הפירוט של ההסבר שיינתן בכל תרחיש רלבנטי. דרישה זו מכונה הסברתיות (explainability; לעיתים מתייחסים אליה גם כ-interpretability).

במקרים בהם מידת השליטה של הטייס הייתה חלקית. לעניין זה, ולאופן שבו הוא צפוי להשפיע גם על השימוש במכוניות אוטונומיות ראו: Tim Hwang & Madeleine Clare Elish, *Praise the Machine! Punish the Human! The Contradictory History of Accountability in Automated Aviation* (2017), Available at: https://papers.ssrn.com.

[135] Green, לעיל ה"ש 130, בעמ' 21-23.

[136] Amnesty Int'l, *Xenophobic Machines: Discrimination Through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal*, AI Index EUR 35/4686/2021, 11-26 (oct. 25, 2021), Available at: https://www.amnesty.org.

הסברתיות היא היכולת להציג בצורה שניתנת להבנה על ידי בני אדם את אופן פעולת המערכת או ההחלטה שלה.[137] הגם שתכונה זו מאפיינת חלק מהמערכות, היא לא תמיד מתקיימת. יש מערכות שמטבען ניתנות להסברה בקלות (explainable), למשל כאלה המשתמשות במודלים פשוטים יחסית על מנת לקבל החלטה או כאשר ניתן להציג באופן פשוט את התהליך שביצעה המערכת. עם זאת, כשהמערכת מבוססת על מודל מורכב יותר, מתבססת על מקורות מידע מרובים, או "לומדת" ומשנה באופן תדיר את אופן פעילותה, קיים קושי להסביר את התהליך שביצעה המערכת.[138] ביחס למערכות מסוג זה, לעיתים נדרש תוסף או אמצעי חיצוני למערכת, שפעמים רבות יפותח במקביל לה ויבוא על גביה, אשר יהא בכוחו להסביר את תהליך קבלת ההחלטה.[139]

אפשרות אחת לחלוקה של הדרישה להסברתיות, מבחינה בין התחומים השונים שלגביהם נדרש לספק הסבר. כך, במסגרת סקירה שחוברה על ידי מכון אלן טיורינג,[140] הוצע לחלק את דרישת ההסברתיות לשישה תחומים שהדרישה להסברתיות עשויה להיות רלוונטית אליהם. <u>ראשית</u>, ניתן לדרוש הסברתיות ביחס **לנימוקים** (rational explanation), משמע לתת הסבר לגבי הסיבות שהובילו לכך שהתקבלה ההחלטה הקונקרטית על ידי המערכת, בצורה שתהיה נגישה. <u>שנית</u>, ניתן לדרוש הסברתיות ביחס **לאחריותיות** (responsibility explanation), משמע להסביר מי לקח חלק בפיתוח המערכת, איך הטמיעו את המערכת, מי מנהל אותה ולמי ניתן לפנות כדי לערער על ההחלטה שהתקבלה. <u>שלישית</u>, ניתן לדרוש הסברתיות ביחס **למידע** (data explanation), קרי להסביר מה המידע עליו התבססה המערכת בהחלטה קונקרטית ואיך היא השתמשה בו, הן ביחס למידע שהפרט שבענייננו התקבלה החלטה סיפק והן ביחס למידע חיצוני שהמערכת נחשפה אליו. <u>רביעית</u>, ניתן לדרוש הסברתיות ביחס **להוגנות** (fairness explanation), כלומר להסביר מהם הצעדים שננקטו בתכנון המערכת ובהטמעה שלה על מנת לוודא כי ההחלטות שלה, ככלל, אינן מוטות, הוגנות ומובילות לתוצאות שוויוניות. <u>חמישית</u>, ניתן לדרוש הסברתיות ביחס **לבטיחות ולביצועים** (Safety and performance explanation), שמשמעותה הסבר לגבי הצעדים שננקטו על מנת למקסם את הדיוק, האמינות, הבטיחות והחוסן של המערכת. <u>שישית</u>, ניתן לדרוש הסברתיות ביחס **להשפעה** (impact explanation), משמע להסביר מה הצעדים שננקטו על מנת לפקח על ההשפעות של השימוש במערכת וההחלטות שלה על האינדיבידואל ועל החברה בכללותה.[141]

אפשרות נוספת לחלוקה של הדרישה להסברתיות, מבחינה בין מספר קטגוריות: (1) מתן הסבר כללי ביחס לפרמטרים המרכזיים עליהם בנוי המודל, אופן פעילות המערכת, ובאיזה שלב מעורב גורם אנושי (אם בכלל), בלי להתייחס להחלטה ספציפית; (2) מתן הסבר פרטני ביחס לפרמטרים הרלוונטיים שנלקחו בחשבון בגיבוש החלטה ספציפית; (3) מתן הסבר פרטני כיצד האינטראקציה בין הפרמטרים שנלקחו בחשבון השפיעה על ההחלטה הספציפית (ה"לוגיקה" שבבסיס ההחלטה).

---

[137] Gabriel Nicholas, *Explaining Algorithmic Decisions*, 4 GEO. L. REV. 711, 715 (2020).

[138] נמחיש את ההבדל באמצעות דוגמה פשוטה של מערכת שמטרתה להחליט אם לאשר הלוואה. הפיתוח של מערכת פשוטה יבוצע על בסיס כללים מוגדרים – לדוגמה ייקבע כי ההלוואה תאושר אם למבקש אין הלוואות נוספות, סכום ההלוואה אינו עולה על המשכורת החודשית ויש למבקש הון עצמי מינימלי של 10,000 ₪; המדובר בכללים שניתן להסביר. לעומת זאת, מערכות מורכבות המבוססות על למידה מכונה או רשת עצבית מלאכותית מסוגלות לבנות בעצמן את הכללים לאישור הלוואה מבלי להתבסס על מודל מבנה.

[139] Bibal, Lognoul, de Streel & Frénay, לעיל ה"ש 126, בעמ' 157-159.

[140] מכון טיורינג הוא מכון מחקרי בתחומי מדע הנתונים ובינה מלאכותית שהוקם על ידי מספר אוניברסיטאות בבריטניה ונתמך על ידי הממשלה, ראו אתר האינטרנט של המכון בכתובת: https://www.turing.ac.uk.

[141] David Leslie, *Explaining Decisions Made with AI*, The Alan Turing Institute, 20 (2020), Available at: https://ssrn.com.

4.3.2. יתרונות וחסרונות הדרישה להסברתיות

להסברתיות, בין שנעשית באופן וולונטרי ובין שנדרשת מכוח הדין, מספר יתרונות עיקריים:

ראשית, ההסברתיות מסוגלת להקל על זיהוי ותיקון טעויות של המערכת כשהן מתרחשות. כאשר המערכת ניתנת להסברה, מתאפשר להבין בצורה פשוטה ומהירה יותר מתי היא מגיעה לתוצאות שגויות ומה גורם לכך, ובהתאם לדרוש את תיקונה או את הפעלתה תחת פיקוח של גורם אנושי שיוכל לסטות במידת הצורך מהחלטה שקיבלה המערכת (ראו פרק "מעורבות אנושית" לעיל).

שנית, ההסברתיות עשויה לסייע לשפר את תהליך קבלת החלטות של הגורמים המשתמשים במערכת. למשל, רופא שנעזר במערכת בינה מלאכותית לשם אבחון יוכל לקבל ממנה מידע שיסייע לו בעתיד לקבל החלטות טובות יותר גם כאשר הוא לא נעזר במערכת.[142] בהקשר זה, הסברתיות רלוונטית גם ליחסים שבין מפתח מערכת המבוססת על בינה מלאכותית לבין הארגון המשתמש בה, שבהם עשוי להיות פער מידע אודות מאפיינים אלה. סגירת פער מידע זה מאפשר לארגון המשתמש להבין טוב יותר את אופן השימוש בטכנולוגיה ואת חלוקת הסיכונים הקשורה בה, ולעמוד מצדו בחובות החלות עליו כלפי מי שמושפע מהחלטותיו.

שלישית, במקרים רבים בלי הסברתיות לא תהיה אפשרות אפקטיבית להשתמש בכלים משפטיים על מנת לפקח אחר השימוש במערכות בינה מלאכותית. מבחינת מושא ההחלטה, בלי לקבל הסבר ביחס לסיבות שהובילו להחלטה, הוא יתקשה לדעת שהשימוש במערכת הביא להפרה של חובה בדין, להוכיח זאת ולנקוט בצעדים, ובכללם צעדים משפטיים, בהתאם. כך למשל, צרכן שהבנק סירב להעניק לו אשראי יתקשה להראות כי הסירוב היה בלתי-סביר בנסיבות העניין.[143] באופן דומה, במקרים של אפליה יהיה קשה לנקוט בהליכים לפי חוק איסור הפליה בשירותים ומוצרים.[144] מבחינת הרגולטורים, ללא הסברתיות יהיה קושי לוודא כי הגורמים המפוקחים עומדים בדרישות החוק,[145] למשל בהקשר של דיני איסור אפליה.[146] לבסוף, מבחינת בתי המשפט, ללא קבלת הסבר הולם על אופן פעולת המערכת, יתקשו השופטים לבחון את השימוש במערכות בינה מלאכותית ואת תוצאותיו בהיעדר מידע על טיב השימוש, הגורם לתוצאות והסיבה שהחליטה כפי שהחליטה.[147]

רביעית, ההסברתיות עשויה לשפר את היכולת של מושאי ההחלטה להתאים את ההתנהגות שלהם. כך, צרכן שמערכת המבוססת על בינה מלאכותית סירבה לתת לו אשראי יוכל להבין את הנימוקים שהובילו להחלטתה (למשל, פיגור מסוים בהחזרי הלוואה קודמת), ובהתאם להחליט אם לשנות את התנהגותו כך שיוכל לקבל בעתיד שירות, לפנות למתחרה או לוותר.[148]

---

[142] Nicholas, לעיל ה"ש 137, בעמ' 717.

[143] ס' 2 לחוק הבנקאות (שירות ללקוח), התשמ"א-1981; Bibal, Lognoul, de Streel & Frénay, לעיל ה"ש 126, בעמ' 156-153.

[144] חוק איסור הפליה במוצרים, בשירותים ובכניסה למקומות בידור ולמקומות ציבוריים, התשס"א-2000.

[145] William Magnuson, *Artificial Financial Intelligence*, 10 HARV. BUS. L. REV. 337, 353-354 (2020); Bibal, Lognoul, de Streel & Frénay, לעיל ה"ש 126, בעמ' 153-151.

[146] אחיעז, חמדני, עמירם וקסטיאל, לעיל ה"ש 22, בעמ' 20.

[147] Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829 (2019).

[148] Carlos Zednik, *Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence*, 34 PHIL. & TECH. 265, 274 (2021).

<u>לבסוף</u>, ההסברתיות צפויה להגביר את אמון הציבור ומוכנותו להשתמש במערכות בינה מלאכותית. הדבר יבוא לידי ביטוי הן ביחס לגורמים המשתמשים במערכת אשר ייטו לסמוך על ההחלטה שלה, ובהתאם לאמץ אותה כשזה נדרש,[149] והן ביחס למושאי ההחלטה של המערכת שככל הנראה יהיו נכונים יותר לאפשר למערכת להחליט בעניינם, ולקבל עליהם את ההחלטה.[150]

על אף יתרונותיה, בהקשרים שונים קיימת מורכבות בהחלת הדרישה להסברתיות באופן רחב. אמנם קיימות ומפותחות כיום טכנולוגיות מתקדמות שייעודן להסביר החלטות שהתקבלו על ידי מערכות מבוססות בינה מלאכותית, ולמעשה לפענח את אותה "קופסא שחורה".[151] ואולם, כפי שתואר לעיל, לא כל מערכת בינה מלאכותית ניתנת להסברה בקלות, אם בכלל. ככל שהמערכת מורכבת יותר, מסתמכת על נתונים רבים יותר ומבצעת קישורים רחבים יותר, כך גובר הקושי הטכנולוגי בהסברת התוצאה אליה היא הגיעה. על כן, אחד החששות המרכזיים שנוגעים להחלת דרישה להסברתיות הוא שקיום הדרישה באופן רחב כתנאי סף לשימוש בטכנולוגיה עלול להגביל את השימוש במערכות מתקדמות ומדויקות יותר. ככל שהמערכת מתחשבת ביותר סוגי פריטי מידע (ולעתים מדובר במאות ואלפי סוגי פריטי מידע) כך גדל הקושי בהסברתיות מלאה שלה. מכאן שעשוי להתקיים מתח (trade-off) בין איכות המערכת, לבין היכולת להסביר את החלטותיה ופעילותה. למשל, כיום יש סוגים של מערכות מבוססות בינה מלאכותית שמפאת מורכבותן והמודל המיושם באמצעותן, יש קושי במתן הסבר מהסוג שהוצג לעיל להחלטותיהן.[152]

זאת ועוד, הצבת דרישה להסברתיות באופן רחב וגורף עלולה לפגוע בהתקדמות הטכנולוגית. כשמדובר במערכות מורכבות, אפילו אם ניתן יהיה בסופו של דבר להסביר את האופן בו פועלת המערכת, הדבר יהיה כרוך בהשקעה של משאבים רבים. על כן, ובפרט כשמדובר בדרישה להסברתיות ברמה גבוהה, יתכן שמיזמים רבים ייפגעו בכדאיותם הכלכלית או שכדי להוציא אותם לפועל יהיה צורך להסיט משאבים מפיתוח מערכות מתקדמות יותר (בהקשרי יעילות ודיוק למשל), לצורך פיתוח האפשרות להסביר כיצד התקבלה ההחלטה. הדבר עלול בסופו של דבר להיות לא יעיל מבחינה חברתית. כך גם, דרישת ההסברתיות עלולה להוות חסם משמעותי עבור חברות קטנות המבקשות להיכנס לתחום הבינה המלאכותית, ואשר כתוצאה מהדרישה ייאלצו לפתח לצד המערכת גם את היכולת להסביר את האופן בו היא פועלת.[153]

לבסוף, מאחר שהסברתיות מצריכה שקיפות ומסירת מידע אודות הנתונים והמודל, עלולות להתעורר שאלות ביחס לזכויות מפתחי המערכת בתחום הקניין הרוחני, ובפרט בתחום הסודות

---

[149] Marina Chugunova & Daniela Sele, *We and It: An Interdisciplinary Review of the Experimental Evidence on How Humans Interact with Machines*, 99 J. BEHAV. EXPERIMENTAL ECON. 1, 32-33 (2022).

[150] Donghee Shin, *The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for explainable AI*, 146 INT'L. J. HUM-COMPUT. STUD. 102551 (2020); Andrea Ferrario, Michele Loi, *How Explainability Contributes to Trust in AI*, ACM Conference on Fairness, Accountability, and Transparency (2022), Available at: https://ssrn.com.
יש לציין, כי ביחס לשימוש בבינה מלאכותית על ידי גופים ציבוריים עולה לעיתים נימוק נוסף דאונטולוגי שמתייחס לכך שלתת הסבר לאדם אודות השימוש בכוח שמופעל עליו זה לכבד אותו בתור סובייקט, אולם נימוק זה אינו מובן מאליו כשמדובר בהסדרת פעילות פרטית. להרחבה ראו: Jerry Mashaw, Reasoned Administration: The European Union, the United States, and the Project of Democratic Governance, 76 GEORGE WASH. LAW REV.

[151] כך למשל, חלק מהמערכות הללו מסוגלות להצביע על החלקים הספציפיים בתמונה ששימשו את המערכת לניתוח תמונה לשם הגעה לזיהוי כלשהו, או יכולות לייצר "מפת חום" של השיקולים המאפשרת להבין את המשקל היחסי שניתן לשיקולים שונים בתהליך קבלת ההחלטה.

[152] Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J. L. & TECH. 889, 929-930 (2018).

[153] שם, בעמ' 893-894 ו-930; Deeks, לעיל ה"ש 147 בעמ' 1834.

המסחריים וזכויות היוצרים.[154] עם זאת, הסברתיות המבוססת על הצגת שיקולים מרכזיים, או הסברת ההחלטה ביחס לאדם מסוים, ללא חשיפת קוד המקור, אינה בהכרח מחייבת חשיפת סודות מסחריים.[155] במקרים אחרים, עשויות לעלות טענות ביחס לחשיפת סודות מסחריים של החברות המפתחות הקשורים לאופן שבו המערכת פועלת, שיצריכו התמודדות ואיזון בין האינטרסים.[156]

4.4. גילוי

בניסוי שנערך בשנת 2014 התכנס חבר שופטים במטרה לבחון האם "יוג'ין גוסטמן" (Eugene Goostman) הוא ילד אוקראיני בן 13 (שאינו דובר אנגלית כשפת אם), או שמא מדובר במערכת מבוססת בינה מלאכותית. לאחר כ-300 שיחות עם יוג'ין שבוצעו על ידי 30 שופטים, כל אחת במשך 5 דקות, לראשונה בהיסטוריה נטען כי מערכת מבוססת בינה מלאכותית הצליחה לעבור את מבחן טיורינג, לאחר שהצליחה להטעות 30% מהשופטים שסברו כי הם מדברים עם אדם.[157] ההתקדמות הטכנולוגית בתחום שנים האחרונות, מעוררת חשש כי היכולת של משתמש הקצה לדעת כי הוא בא באינטראקציה עם מערכת מבוססת בינה מלאכותית תהיה תלויה ברצון המפעיל לחשוף זאת.

עם התקדמות הטכנולוגיה והשימוש ההולך והגובר במערכות מבוססות בינה מלאכותית לשם קבלת החלטות, כמו גם השימוש במערכות צ'אט-בוט (chat-bot) המבוססות על בינה מלאכותית,[158] גוברת האפשרות שאנשים לא יהיו מודעים לכך שלמערכות אלה תפקיד משמעותי "מאחורי הקלעים" בקבלת ההחלטות בעניינם, או שהם משוחחים עם מערכות כאלה (ולא עם אנשים).[159]

כפי שיפורט להלן, השימוש במערכות מבוססות בינה מלאכותית נעשה על ידי מגוון רחב של גופים, ובין היתר על ידי חברות מסחריות וחברות טכנולוגיה גדולות, חברות הזנק (Start-up) וגופי ממשל (אם כי מסמך זה אינו נוגע להיבטים המיוחדים של השימוש במערכות על ידי גופים ציבוריים). במקרים מסוימים, לגופים אלו יהיה תמריץ לחשוף למשתמש הקצה כי הם עושים שימוש במערכות מבוססות בינה מלאכותית; ואולם, במקרים אחרים, ואלו המקרים בהם יתמקד חלק זה, יהיה לאותם גופים דווקא תמריץ להסתיר אותו, לעיתים תוך ניסיון להטעות את משתמש הקצה.

על רקע זה, מתעוררת שאלה, מה מידת הגילוי לגבי מעורבות מערכות מבוססות בינה מלאכותית (AI related transparency), הנדרשת מצד גורמים המשתמשים במערכות אלה.

יצוין כי יש העוסקים בהיבט זה כחלק משקיפות (Transparency) במשותף עם הסברתיות (Explainability). רב המשותף בין דרישות אלה, שכן גילוי בעניין השימוש בבינה מלאכותית ושקיפות לגבי אותו שימוש יכולים להיכלל כחלק מההסבר על תהליך קבלת ההחלטה, וגם

---

[154] Rita Matulionyte, *Reconciling Trade Secrets and Explainable AI: face recognition technology as a case study*, 44(1) EUR. INTELL. PROP. REV. 14 (2022).

[155] שם, בעמ' 10: "Overall, while the implementation of AI explainability principle will be compatible with trade secret protection in some cases, in other cases the application of more intrusive explainability techniques might interfere with trade secret protection"

[156] Matulionyte, לעיל ה"ש 154.

[157] *Turing Test success marks milestone in computing history*, UNIV. READING (June 08, 2014), https://archive.reading.ac.uk.

[158] תוכנות המתמחות בניהול שיחות טקסטואליות עם בני אדם, שנועדו להיחוות כטבעיות ככל האפשר.

[159] Nika Mozafari, Maik Hammerschmidt & Welf H. Weiger, *The Chatbot Disclosure Dilemma: Desirable and Undesirable Effects of Disclosing the Non-Human*, ICIS 2020 PROCEEDINGS. 6, 2 (2020), https://aisel.aisnet.org.

הסברתיות (בפרט כשמדובר בחשיפת המודל והנתונים ששימוש לאימון המודל) במהותה היא סוג של שקיפות מצד מפתחי ומפעילי המערכת. אולם במסמך זה נערכה הבחנה בין השניים, בין היתר בשים לב לכך שמסירת מידע מתמקדת בגילוי על האינטראקציה עם המערכת, בעוד שההסברתיות עוסקת בדרישה להסביר את דרך פעולתה של המערכת ושל החלטות ותחזיות שהתקבלו על ידה.

4.4.1. מה טיבה של דרישת הגילוי?

ניתן לחלק את האינטראקציה האנושית עם מערכות מבוססות בינה מלאכותית לישירה ועקיפה. אינטראקציה ישירה נוצרת כשהגורם האנושי בא בקשר בלתי אמצעי עם המערכת, כגון שהוא משוחח עם צ׳אט-בוט או כשהוא נחשף לתוכן שזויף באמצעות טכנולוגיית זיוף עמוק (deep fake). לעומת זאת, אינטראקציה עקיפה מתקיימת במצבים בהם במסגרת תהליך קבלת ההחלטה נעשה שימוש במערכת, אולם הקשר הישיר של מושא ההחלטה הוא עם גורם אנושי, שתפקידו לתווך את ההחלטה של המערכת, או להסתייע בה על מנת לקבל את ההחלטה בעצמו. כפי שיפורט להלן, הנימוקים להצבת דרישה לגילוי ביחס לאינטראקציה ישירה או עקיפה עשויים להיות שונים.

להמחשה, נתמקד באדם המעוניין בהלוואה מהבנק. כידוע, כיום גופים רבים עושים שימוש בצ׳אט-בוטים על מנת לנתב פניות ולספק מענה ראשוני של שירות הלקוחות, כך שלא מן הנמנע שבפתח תהליך בקשת ההלוואה יידרש המבקש לנהל אינטראקציה ישירה עם מערכת צ׳ט-בוט שתהיה מבוססת בינה מלאכותית. בשלב הבא, ולאחר שהגיש את בקשת ההלוואה, ייתכן שהבנק יעשה שימוש במערכת מבוססת בינה מלאכותית על מנת לבחון את הבקשה ולהחליט אם להיעתר לה ובאיזה תנאים. השימוש במערכת יכול להיעשות תוך החלפת שיקול הדעת של הפקיד, כשהמערכת קובעת באופן אוטונומי אם ובאיזה תנאים לאשר את הבקשה, או שהוא יכול להיעשות על מנת לסייע לפקיד להחליט בעצמו אם ובאיזה תנאים לאשר אותה. בעוד שבשלב הראשון האינטראקציה של המבקש עם המערכת היא ישירה, בשלב השני, האינטראקציה עם המערכת היא עקיפה.

בכל אינטראקציה עם מערכת מבוססת בינה מלאכותית, קיימת האפשרות כי אדם לא יהיה מודע לכך שהוא בא באינטראקציה עם מערכת כזו, ויטעה לחשוב שהוא מְתַקְשֵׁר עם אדם או שאדם (בלבד) מחליט בעניינו. אדרבה, לא אחת חברות מסחריות משקיעות משאבים רבים כדי לשוות לבוט מאפיינים אנושיים. כך לדוגמה, מערכות עזרה וירטואלית מבוססת קול, עלולות לגרום לטעות ולמחשבה שמדובר בגורם אנושי. זאת, באמצעות ניסיון לחקות את האופן בו אנשים מדברים (בין היתר, קיימות מערכות המחקות היסוס ועושות שימוש בביטויים כמו ״אמממ״).[160] מטרתה של הדרישה לגילוי, ככל שתתקיים, היא למעשה להבטיח כי לכל אדם אשר בא באינטראקציה עם מערכת המבוססות על בינה מלאכותית תהיה היכולת לדעת זאת.

4.4.2. היתרונות והחסרונות להצבת דרישת גילוי

ככלל, גילוי למשתמש הקצה הבא באינטראקציה עם מערכת מבוססת בינה מלאכותית יכול לסייע בהשגה של מספר מטרות, הן בשלב שלפני השימוש במערכת והן לאחר השימוש בה.

<u>ראשית</u>, הידיעה כי הגורם העומד בפניו משתמש במערכת מבוססת בינה מלאכותית, עשויה להוות אינדיקציה חיובית או שלילית עבור הפרט, שיכולה להשפיע על בחירתו כיצד לפעול, ובין היתר אם

[160] Fatimah Ishowo-Oloko, Jean-François Bonnefon, Zakariyah Soroye, Jacob Crandall, Iyad Rahwan & Talal Rahwan, *Behavioural Evidence for a Transparency–Efficiency Tradeoff in Human–Machine Cooperation*, 1 NATURE MACH. INTELLIGENCE 517, 517-520 (2019).

לרכוש מוצר מסוים או ממי לקבל שירות. בעוד שברור מדוע ידיעה זו עשויה להשפיע על המשתמש כאשר הוא בא באינטראקציה ישירה עם המערכת, למשל מכיוון שהוא מעדיף לשוחח עם גורם אנושי, הידיעה אם נעשה שימוש במערכת עשויה להשפיע גם כאשר המדובר באינטראקציה עקיפה. כפי שתואר גם בחלקים קודמים, השימוש במערכות מבוססות בינה מלאכותית יכול לשפר את דיוק ההחלטה; על כן, ייתכן למשל כי צרכנים יעדיפו לקבל שירות מחברה אשר מייצרת תחזיות עסקיות באמצעות שימוש בטכנולוגיית בינה מלאכותית על פני חברות אחרות. לעומת זאת, יהיו מצבים בהם השימוש במערכת דווקא ירתיע צרכנים מלבחור בקבלת שירות מסוים.

<u>שנית</u>, הידיעה כי מערכת מבוססת בינה מלאכותית הייתה מעורבת בקבלת החלטה בעניינו של הפרט, עשויה לאפשר לו לוודא שהמערכת עומדת בדרישות הרגולטוריות הרלוונטיות תוך שזכויותיו נשמרות. במובן זה, הידיעה על אודות האינטראקציה עם מערכת מבוססת בינה מלאכותית "פותחת את השער" עבור מי שמושפע מפעילותה לוודא את תקינות ההחלטה או הפעולה של המערכת, ואף של המערכת כולה, בשים לב לכל האתגרים הנסקרים במסמך זה ולסביבה הרגולטורית בה פועלת המערכת. ברי כי ככל שיתפתח שיח הזכויות והחובות סביב הפעלתן של המערכות, ובכפוף להסתייגויות שהוצגו לעיל, כך לידיעה כי נעשה שימוש במערכת יהיה תפקיד משמעותי יותר במיצוי הזכויות ועמידה על החובות של הגורמים הרלוונטיים.

<u>שלישית</u>, ניתן לראות את הגילוי כאקט של הגינות כלפי משתמש הקצה, כך שבמקרים מסוימים ייתכן כי ראוי להכיר במעין "זכות" שלו לדעת אם הוא בא באינטראקציה עם מערכת מבוססות בינה מלאכותית מטעמים מוסריים, וזאת אף שכנראה לא ינהג אחרת אם יגלה על המערכת.

נוסף לטעמים האמורים לעיל, שאינם נוגעים במישרין ליתרונות אפשריים מבחינת מפעיל המערכת, חשוב לציין כי לעיתים גם לו יהיה אינטרס לחשוף את השימוש שהוא עושה במערכת ולגלות על האינטראקציה. זאת משום שהחשיפה עשויה לשפר את התדמית של חברה בתחום הטכנולוגי, להצביע על יכולות מתקדמות יותר ביחס למתחרים, ולאפשר התנהלות בצורה שמותאמת לכך שנעשה שימוש במערכות מבוססות בינה מלאכותית (למשל, בתקשורת עם צ'אט-בוט ניתן לכתוב בשפה ישירה ופשוטה יותר). בפרט, בכל הנוגע לגילוי על אינטראקציה ישירה עם המערכת, לחברה עשוי להיות אינטרס לחשוף את השימוש בצ'אט-בוט במטרה לתת מענה ללקוחות המעדיפים להיות באינטראקציה עם המערכת ולא עם גורם אנושי, למנוע תסכול שיתעורר עקב ציפייה לא ריאלית ביחס ליכולות המערכת ומתוך תפיסה עקרונית כי עדיף להתנהל בשקיפות למול לקוחות.[161]

<u>מנגד</u>, לצד יתרונותיה, דרישה לגילוי עלולה לפעול ביעילות השימוש במערכת מבוססת בינה מלאכותית ולעיתים לפגוע בנכונות לעשות שימוש במערכות מבוססות בינה מלאכותית. במיוחד בנוגע לאינטראקציה ישירה, חשיפת השימוש במערכת עלולה לפגוע ביעילות. מחקרים מצאו כי משתמשים נוטים לסמוך פחות על מערכות מבוססות בינה מלאכותית ביחס לגורמים אנושיים אפילו אם רמת השירות זהה,[162] ובכלל זה בין היתר, לנהל אינטראקציות קצרות יותר עם צ'אט-

---

[161] Roberta De Cicco, Susana Cristina Lima da Costa e Silva & Riccardo Palumbo, *Should a Chatbot Disclose Itself? Implications for an Online Conversational Retailer, in* CHATBOTS RESEARCH AND DESIGN 3, 5 (2020).

[162] Conference Reprot, HAW. INT'L CONF. ON SYS. SCI., *Resolving the Chatbot Disclosure Dilemma: Leveraging Selective Self-Presentation to Mitigate the Negative Effect of Chatbot Disclosure,* 2916 (2021), https://scholarspace.manoa.hawaii.

בוטים; לרכוש פחות מוצרים; ובאופן כללי לא לשתף פעולה עם המערכת.[163] ייתכן כי חשיפת הגילוי תפגע ביעילות השימוש במערכת גם במקרים של אינטראקציה עקיפה, למשל בכך שתקל על ניסיונות זדוניים להשפיע על התוצאה שתתקבל על ידי המערכת, בין אם על ידי משתמש הקצה ובין אם על ידי מתחרים (להרחבה ראו פרק "אמינות, עמידות, אבטחה ובטיחות" להלן).

4.5. __אמינות, עמידות, אבטחה ובטיחות__

במרץ 2016 הוצג לעולם "טאיי" (Tay), בוט המתבסס על בינה מלאכותית, שנועד לחקות נערה אמריקאית בת 19 ולנהל שיחות אקראיות עם משתמשים בטוויטר. החברה המובילה שפיתחה אותו קיוותה שבאמצעות האינטראקציה עם משתמשים אחרים טאי יתפתח ויוכל לנהל שיחות מתוחכמות יותר ויותר, ואגב כך יסייע לה בפיתוח מערכות מתקדמות בעתיד. אלא שטאי שרד "באוויר" רק 16 שעות לפני שהחברה מיהרה להשבית אותו. התברר שבפרק הזמן הקצר שבו פעל טאי, הוא "למד" מהגולשים האחרים, ובהתאם התחיל להגיב בצורה מגונה, אנטישמית וגזענית. לטענת החברה המפתחת, מי שהיו אחראיים להידרדרות טאי הם הגולשים אשר לקחו חלק ב"מאמץ מתואם לנצל לרעה את כישורי הדיבור שלו כדי לגרום לו להגיב באופן לא הולם".[164] מקרה זה ממחיש את החשש שחלק מן המערכות המבוססות על בינה מלאכותית אינן מספיק אמינות, עמידות או בטוחות על מנת שניתן יהיה להפעילן בשגרה בלי לבצע התאמות מיוחדות.

מערכות מבוססות בינה מלאכותית נשענות על טכנולוגיות מידע ותקשורת, החשופות לטעויות ותקלות טכניות, או למניפולציות מכוונות. עקב כך, על מנת ליהנות מהתועלות הקשורות בבינה מלאכותית, יש חשיבות רבה לוודא את תפקודה התקין, הבטוח והמהימן של טכנולוגיה זו.

היקף השימוש הצפוי בבינה מלאכותית, במוצרים שכיום אין להם יכולת חישובית (כגון מכוניות, ומוצרים מסוגים שונים), מרחיב את היקף הרלוונטיות של סוגיות בטיחות הנובעות מסיכונים ממוחשבים. כך, כל עוד מדובר במוצרים שיש בהם רכיבים פיזיים בלבד, הסיכון הבטיחותי מהם נובע מאופן פעולתם הפיזי. שילוב רכיבים ממוחשבים מייצר סיכון מסוג חדש.

השימוש במערכות מבוססות בינה מלאכותית עשוי לייצר חשיפה לתוצאות שליליות, שעלולות להיגרם עקב כשלים בפעילות מערכות אלה. כפי שיפורט להלן, תוצאות אלה עלולות להיגרם הן בעקבות התממשות __כשלים פנימיים__ שמשפיעים על **אמינות** המערכת (Reliability) – מצבים שבהם המערכת סובלת מביצועים ירודים (poor performance) משום שהיא אינה מסוגלת לבצע כהלכה את המשימה שיועדה לה; הן בעקבות התממשותם של __איומים חיצוניים__ שמשפיעים על **עמידות** המערכת (robustness) – מצבים שבהם גורם חיצוני מצליח לשבש את פעילות המערכת על ידי ניצול של נקודת תורפה הטבועה בה (vulnerability).[165] לעיתים התממשות הכשלים הפנימיים או

---

[163] שם, בעמ' 2917; Xueming Luo, Siliang Tong, Zheng Fang & Zhe Qu, *Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases*, 38 MARKETING SCI. 913 (2019).

[164] Davey Alba, *It's Your Fault Microsoft's Teen AI Turned into Such a Jerk*, WIRED (mar. 25, 2016), https://www.wired.com.

[165] Ronan Hamon, Henrik Junklewitz & Ignacio Sanchez, *Robustness and Explainability of Artificial Intelligence – from technical to policy solutions*, JRC Technical Report, 14 (2020).

האיומים החיצוניים אף עלולה לגרום ליצירת סיכונים בטיחותיים וכלכליים משמעותיים. על רקע זה, יש חשיבות לאמינות ודיוק,[166] עמידות[167] ובטיחות[168] מערכות מבוססות בינה מלאכותית.

הכשלים האמורים הם ככלל אתגרים טכנולוגיים, נהליים או תהליכיים, ולרוב מפתחי המערכות והמשתמשים בהן יהיו בעלי האינטרס המרכזי להתמודד עמם, שכן הם אמורים להיות מעוניינים ליצור או להשתמש במערכות איכותיות ומוגנות. עם זאת, במקרים מסוימים, עשויה להיות הצדקה לשקול התערבות רגולטורית, על מנת לוודא כי מערכות מבוססות בינה מלאכותית הן אמינות, עמידות, מאובטחות ובטוחות במידה מספקת. זאת, בפרט בנסיבות בהן מתקיימות הצדקות מקובלות להתערבות רגולטורית, כדוגמת החצנות שליליות הכרוכות בפעילות מערכות אלה.

4.5.1. <u>אמינות המערכת</u>

לא אחת, למרות שמערכת המבוססת על בינה מלאכותית פועלת בתנאים שגרתיים, ואפילו אופטימאליים, התוצאות שאליהן היא תגיע לא יהיו מספיק טובות, עקב כשלים פנימיים. כלומר, מכיוון שאופן פיתוח המערכת או השימוש בה גורם לכך שלא תוכל לבצע כיאות את הפונקציה שיועדה לה. כשלים אלה, המשליכים על אמינות המערכת, עלולים להתרחש בכל שלב ב"מחזור החיים" שלה – בעיצוב המערכת; באיסוף וניתוח המידע והנתונים שעליהם מתבססת המערכת; ביצירת המודל שלפיו המערכת מקבלת החלטות; ובממשק הסופי בין המערכת לבין המשתמשים.

כך לדוגמה, <u>בנוגע לשלב אימון המערכת (AI training)</u>, החשש המרכזי הוא שהאימון לא ייעשה בצורה מספיק טובה, מה שיוביל לביצועים ירודים של המערכת ויגרע מאמינותה. אחד המקרים המרכזיים שבהם המערכת עלולה לסבול מביצועים ירודים הוא כאשר הנתונים ששימשו לאימון המערכת אינם מהימנים או מקיפים (תופעה המכונה הטיית ספקטרום (Spectrum bias). בהקשר זה, החשש הוא שכאשר המערכת תידרש להתמודד עם מצבים או קבוצות אוכלוסייה שמאפייניהם חורגים מהמנעד שעליו היא התאמנה התוצאות שלה יהיו פחות מדויקות, וייתכן שאף שגויות לחלוטין. כך למשל, אם האימון של מכונית עצמאית (אוטונומית) נעשה בישראל, ייתכן שבעת שימוש בה במדינה אחרת, היא לא תהיה ערוכה להתמודד עם נסיעות בתנאי שטח שאינם שכיחים בישראל כמו שלג או סופות חול, או שלא תדע כיצד להתמודד עם חיות על הכביש שאינן נפוצות בישראל כמו הקנגורו השכיח באוסטרליה. בהקשר זה, וכפי שתואר לעיל בפרק "אפליה", קיים גם חשש כי כאשר האימון של המערכת יתמקד בקבוצות מסוימות באוכלוסייה, היא תפעל באופן שיטתי בצורה פחות מדויקת ביחס לקבוצות אחרות באוכלוסייה. לדוגמה, במחקר שבחן מערכות לזיהוי פנים נמצא כי מספר מערכות, שפותחו על ידי חברות מובילות שונות, נטו לזהות ביתר קלות

---

[166] בטיוטת מסגרת התקינה של NIST (מכון התקנים הטכנולוגי האמריקאי) לבינה מלאכותית מוצע להגדיר דיוק, בהתבסס על תקן ISO/IEC TS 5723:2022 כ- "הקרבה של תוצאות, תצפיות, חישובים או הערכות לערכים אמיתיים או לערכים הנתפסים כאמיתיים". מדידת הדיוק צריכה להיות קשורה ליעדי בדיקה ברורים ופרטים לגבי שיטת הבדיקה, וכן מתועדים. במסגרת המוצעת של NIST מוצע להגדיר אמינות בהתאם לתקן ISO/IEC TS 5723:2022 כ- "יכולת של חפת לתפקד כמתוכנן, ללא כשל, לפרק זמן נתון בתנאים נתונים".
.NIST, AI RMF 2nd Draft for Public Comment, https://www.nist.gov

[167] במסגרת המוצעת של NIST מוצע להגדיר עמידות בהתאם לתקן ISO/IEC TS 5723:2022 כ-"יכולת של מערכת בינה מלאכותית לשמור על רמה של ביצוע בתנאים שונים". המשמעות היא לא רק ציפיה שמערכת הבינה המלאכותית תפעל כפי שתוכננה, אלא שגם תפעל כפי שתוכננה באופן שמצמצם סיכונים לאנשים גם בתנאים לא צפויים.

[168] בהתאם למסגרת המוצעת של NIST מוצע להגדיר בטיחות על פי תקן ISO/IECD TS 5723:2022,באופן ש- "מערכות בינה מלאכותית, לא יגרמו, בתנאים מוגדרים, לנזק פיזי או פסיכולוגי או יובילו למבצע שבו יש סיכון לחיי אדם, בריאות, רכוש או הסביבה". הפעלה בטוחה של בינה מלאכותית מצריכה תהליכי תכנון ופיתוח אחראים, מידע למיישמי הבינה המלאכותית לגבי אופן השימוש במערכת, וקבלת החלטות אחראית בידי המיישמים ומשתמשי הקצה.

פנים של גברים לבנים לעומת פנים של נשים שחורות, ככל הנראה בגלל ייצוג נמוך יותר לנשים שחורות במאגרי המידע והנתונים שעל בסיסן התאמנו המערכות.[169]

מאפיין נוסף של מערכות מבוססות בינה מלאכותית שעלול לגרום לתוצאה דומה לזו של הטיית הספקטרום, הוא היותן של חלק מן המערכות שבריריות (Brittle). כלומר, מערכות שעוצבו באופן כזה שאינן מסוגלות לבצע הכללות או להתאים עצמן לנסיבות משתנות. תכונות אלה הכרחיות למערכת מבוססת בינה מלאכותית, כיוון שלא תמיד יש יכולת ממשית לחשוף מערכת לנתונים המשקפים כל סיטואציה אפשרית או לחזות את כלל המצבים שעמם היא תידרש להתמודד בעתיד. כשמערכות שבריריות יידרשו להתמודד עם מצבים לא מוכרים, הדיוק שלהן ייפגע והסיכוי שיגיעו לתוצאה הנכונה יפחת באופן משמעותי.[170] בעיה זו חמורה במיוחד, כאשר יש כוונה לעשות שימוש במערכות מבוססות בינה מלאכותית בתחומים שבהם זמינות המידע והנתונים לוקה בחסר, וניתן לחשוף את המערכת מראש רק לחלק קטן מסוגי המקרים שעמם היא תידרש להתמודד.

<u>בשלב הבא, לאחר כניסת המערכת לשימוש (AI uses)</u>, אחד החששות המרכזיים הוא מהִשְׁתַּנּוּת במאפייני משתנה היעד של המערכת (כלומר, שינוי במאפיינים של מה שהמערכת נועדה לחזות). תופעה זו, המכונה "Concept Drift", מתייחסת לחשש כי ככל שעובר זמן רב יותר משלב האימון של המערכת, כך הנתונים ההיסטוריים שעליהם היא התאמנה ישקפו בצורה פחות טובה את המציאות שבה היא פועלת, ולכן ישנו חשש כי המערכת תהיה פחות מדויקת וחשופה לשגיאות בלתי צפויות. לשם ההמחשה, ביחס למערכת שמטרתה לבצע מסחר בניירות ערך (algo-trading), אחד החששות הוא ששינויים בשוק ניירות הערך עלולים לפגוע במידת הדיוק של המערכת; כך למשל, שהשימוש ההולך וגובר בשנים האחרונות בחברות רכישה למטרות מיוחדות (חברות SPAC),[171] עלול להוביל לעיוותים ולתוצאות שגויות של המערכת, מכיוון שייתכן שהמערכת לא נחשפה בתהליך האימון לחברות מסוג זה, שהשימוש בהן לא היה שכיח בעבר.[172]

חשש נוסף שקיים בשלב השימוש במערכת הוא מ"שכחה קטסטרופלית" (Catastrophic Forgetting). חשש זה נוגע לכך שכאשר מנסים ללמד מערכת לבצע מספר משימות, לעיתים הלימוד של משימות נוספות גורם לה "לשכוח" איך לבצע את המשימה הראשית.[173] כך לדוגמה, חוקרים מדרום קוריאה הקימו אתר אינטרנט שנועד לסייע, באמצעות מערכת מבוססת בינה מלאכותית,

---

[169] Joy Adowaa, Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifier (2017) (Unpublished B.A. Thesis, Massachusetts Institute of Technology), https://dspace.mit.edu.

[170] דוגמה מפורסמת לקושי של מערכת לבצע הכללות באה לידי ביטוי בתחום הנהיגה האוטונומית; התברר שמערכות מתקדמות שהצליחו לזהות בצורה מדויקת אוטובוס הסעות, טעו בסיווג ב-97% מהמקרים כאשר בתמונה האוטובוס נטה על צדו. ראו: Michel A. Alcorn, Qi Li, Zhitao Gong, Chengfei Wang, Long Mai, Wei-Shinn Ku & Anh Nguyen, *Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects* (2019), Available at: https://arvix.org/abs/1811.11553; Mary L. Cummings, Rethinking the Maturity of Artificial Intelligence in Safety-Critical Settings, 42 AI MAG. 6, 7-8 (2021).

[171] תאגיד אשר נרשם לבורסה כשלד בורסאי ללא פעילות במטרה לרכוש חברה פרטית ולהפוך אותה לציבורית בלי שהחברה הנרכשת תידרש לנהל הנפקה ראשונה לציבור. Johannes Kolb & Tereza Tykvova, *Going Public Via Special Purpose Acquisition Companies: Frogs do Not turn Into Princes*, 40 J. CORPORATE FIN. 80, 83-84 (2016).

[172] David Leslie, *Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector*, The Alan Turing Institute, 37 (2019).

[173] Ian J. Goodfellow, Mehdi Mirza, Da Xiao, Aaron Courville & Yoshua Bengio, *An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks*, 1 (2015), Available at: https://arxiv.org; James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran & Raia Hadsell, *Overcoming catastrophic forgetting inneural networks* (2017), Available at: https://arxiv.org.

להבחין בין תמונות אמתיות לבין זיופים שנעשו באמצעות טכנולוגיית זיוף עמוק (deep-fake). בהתחלה האתר עבד בצורה טובה, אולם לאחר מספר חודשים נעשה שימוש בטכנולוגיות חדשות ליצירה של זיופים עמוקים. כאשר החוקרים אימנו את המערכת להתמודד עם הסוגים החדשים של הזיוף, התברר להם כי המערכת "שכחה" איך לזהות את הסוג המקורי של הזיוף העמוק.[174]

חשוב לציין כי אין מדובר ברשימה ממצה של כשלים פנימיים ושל סוגי המקרים שמעוררים חשש לפגיעה במידת האמינות של מערכות מבוססות בינה מלאכותית. הסקירה לעיל אך נועדה להציג כמה כשלים מרכזיים, ולהמחיש כיצד עלולה להיגרם פגיעה באמינות מערכות אלה.

4.5.2. <u>עמידות המערכת ואבטחתה</u>

לצד החשש מפני כשלים פנימיים, קיים חשש כי מערכות מבוססות בינה מלאכותית יגיעו לתוצאות שגויות או שפעולתן תגרום לנזק עקב התערבות גורם חיצוני (adversary), לרוב בעל כוונות זדוניות, שינצל חולשה של המערכת על מנת לשבש את פעילותה או להשפיע על אופן פעילותה. חשש זה נוגע בעיקר לתחום המחקר של למידת מכונה אדוורסרית (Adversarial Machine Learning) העוסק בטכניקות המאפשרות לנצל, לשבש או לשטות במודלים חכמים, עם או בלי גישה למודל המערכת עצמו ולמידע שעליו המודל מתבסס, ובדרכים שבאמצעותן ניתן להתגונן מפני מתקפות כאלה.[175]

הגם שהחשש מתקיפה ושיבוש קיים כמעט ביחס לכל סוג של תוכנה, ובהתאם התפתחו ענפים באקדמיה, בתעשייה וברגולציה העוסקים בהגנה על מערכות תוכנה, בכל הנוגע למערכות מבוססות בינה מלאכותית חשש זה מחריף ואף עלול לבוא לידי ביטוי בסוגי תקיפה ושיבוש ייחודיים. זאת, משום שמגוון הדרכים שבהן ניתן לפגוע בפעילות של מערכות אלה רחב יותר.[176]

ניתן לחלק את הדרכים השונות לתקוף מערכת המבוססת על בינה מלאכותית לשלוש קטגוריות שונות: מתקפות שמטרתן לגרום למערכת **ללמוד** את הדבר הלא-נכון (Learn the wrong thing); מתקפות שמטרתן לגרום למערכת **לעשות** את הדבר הלא-נכון (Do the wrong thing); ומתקפות שמטרתן לגרום למערכת **לגלות** לתוקף את הדבר הלא-נכון (Reveal the wrong thing).

---

[174] Shahroz Tariq, Sangyup Lee & Simon S. Woo, *One Detector to Rule Them All: Towards a General Deepfake Attack Detection Framework* (2021), Available at: https://arxiv.org/abs/2105.00187; Charles Q. Choi, *7 Revealing Ways AIs Fail*, IEEE SPECTRUM (Sept. 25, 2021), https://spectrum.ieee.org.

[175] המכון הישראלי למדיניות טכנולוגיה **למידת מכונה אדברסרית: התפתחויות במחקר, סכנות והשלכות** 2 (2021) (להלן: "**למידת מכונה אדברסרית**").

[176] שם.

בנוגע למתקפות שמטרתן לגרום למערכת ללמוד את הדבר הלא-נכון, שיטה נפוצה לדוגמה, מכונה "תקיפת הרעלת מידע" (Data Poisoning Attacks). תקיפה זו מתייחסת למצבים שבהם התוקף משפיע על התנהגות המערכת על ידי מניפולציה של הנתונים המשמשים לאימון המערכת. ככל שהמערכת משתמשת במידע רב יותר לאימון, כפי שלמשל נדרש במערכת המבוססת על לימוד עמוק (deep learning), וככל שהמידע שבו משתמשת המערכת מגיע ממקורות לא מאובטחים או לא ידועים, כך גדלה החשיפה למתקפות מסוג הרעלת מידע.[177] מתקפות הרעלת מידע יכולות לשמש תוקף פוטנציאלי על מנת לפגוע בביצועים של המערכת (להפחית את הדיוק), לנטרל את המערכת, או ליצור דלת אחורית (back-door) שתאפשר לנצל את הפרצה בעתיד.[178] חלק משמעותי מהאתגר בהתמודדות עם תקיפות הרעלת מידע הוא שבעוד שדרך ההגנה המסורתית מפני מתקפות היא באמצעות יצירת חַיִץ בין פעילות המערכת לבין מידע חיצוני, עבור מערכות מבוססות בינה מלאכותית השימוש במידע חיצוני הוא הכרחי על מנת לאמן את המערכת לפעול כהלכה.[179]

באשר למתקפות שמטרתן לגרום למערכת לעשות את הדבר הלא-נכון, שיטה נפוצה למשל, מכונה "תקיפת התחמקות" (Evasion Attacks). תקיפה זו מתייחסת למצבים שבהם התוקף מבצע מניפולציה על הקלט המוזן למערכת במטרה לגרום לה לסווג את הקלט באופן שגוי. במקרים רבים המטרה של התקיפה היא לגרום לכך שהמערכת לא תצליח לסווג קלט מסוים (למשל, למנוע מתיבת הדואר האלקטרוני לסווג הודעה מסוימת כ"ספאם").[180] אחת השיטות לבצע תקיפת התחמקות היא באמצעות דוגמה אדוורסרית (Adversarial Example), קלט שעוצב באופן ייעודי כך שמצד אחד הוא יגרום למערכת לזהות אותו באופן שגוי, אבל שמצד שני יהיה בלתי ניתן לגילוי על ידי גורם אנושי שיבצע בקרה על המידע.[181] כך למשל, חוקרים הצליחו למצוא כי אפילו שינוי של פיקסל אחד בלבד בתמונה עלול לגרום לטעות בסיווג התמונה על ידי מערכות לזיהוי פנים.[182]

לבסוף, ובהתייחס למתקפות שמטרתן לגרום למערכת לגלות לתוקף את הדבר הלא-נכון, הכוונה היא למקרים שבהם התקיפה נעשית על מנת לחלץ מהמערכת מידע שהיא לא הייתה אמורה לגלות. להמחשה, שיטה מרכזית לגרום למערכת לגלות את הדבר הלא-נכון מכונה "היפוך מודל" (Model Inversion). במסגרתה התוקף, שאינו חשוף לכלל המידע המצוי בפני המערכת, יכול להסיק באמצעות ניתוח הפלטים של המערכת, תכונות הנוגעות למידע שמשמש את המערכת. החשש המרכזי לגבי סוג זה של מתקפות נוגע לפגיעה בפרטיות, שעלולה להיגרם עקב האפשרות שיתגלו פרטי מידע ממאגרי מידע שבמקרים רבים הם חסויים (להרחבה ראו פרק "פרטיות" להלן).[183]

---

[177] Avi Schwarzschild, Micah Goldblum, Arjun Gupta, John P Dickerson & Tom Goldstein, *Just How Toxic is Data Poisoning? A Unified Benchmark for Backdoor and Data Poisoning Attacks*, 1 (2021), Available at: https://arxiv.org.

[178] Hamon, Junklewitz & Sanchez, לעיל ה"ש 165, בעמ' 17.

[179] Jacob Steinhardt, Pang Wei Koh & Percy Liang, *Certified Defenses for Data Poisoning Attacks*, 1 (2017), Available at: https://arxiv.org.

[180] Zeinab Khorshidpour, Sattar Hashem & Ali Hamze, Attack, IEEE *16TH INT'L CONF. ON DATA MINING WORKSHOPS*, *Learning a Secure Classifier against Evasion Attack*, 295 (2016).

[181] **למידת מכונה אדברסרית**, לעיל ה"ש 175, בעמ' 3.

[182] Jiawei Su, Danilo Vasconcellos Vargas & Sakurai Kouichi, *One Pixel Attack for Fooling Deep Neural Networks* (2019), Available at: https://arxiv.org.

[183] Michael Veale, Reuben Binns & Lilian Edwards, *Algorithms That Remember: Model Inversion Attacks and Data Protection Law*, 376 PHIL. TRANSACTION ROYAL SOC'Y A 1, 4-6 (2018).

חשוב לציין, כי גם כאן אין מדובר ברשימה ממצה של איומים חיצוניים ושל החששות המתעוררים בקשר לעמידות ואבטחת מערכות בינה מלאכותית בפניהם. כמו כן, חשוב להבהיר, שחלק זה אינו עוסק באמצעים הטכנולוגיים הקיימים והמפותחים על מנת להתמודד עם איומים אלה. הסקירה לעיל אך נועדה להמחיש איומים חיצוניים מרכזיים, את החשיבות לעמידות המערכות בפניהם ואת הצורך לבחון התערבות רגולטורית על מנת להתמודד עם איומים אלו.

4.5.3. <u>בטיחות המערכת</u>

השימוש במערכות מבוססות בינה מלאכותית לא בהכרח יניב תמיד את התוצאה הרצויה. מדובר בחשש לא מבוטל, הנובע מהאפשרות שייעשה שימוש במערכת שככלל אינה מדויקת מספיק, וגם מהאפשרות שהמערכת לוקה בכשלים פנימיים או חשופה לאיומים חיצוניים. במקרים מסוימים החשש משגיאות של מערכות אלה לא גדול, אך במקרים אחרים הן עלולות לגרום פגיעה משמעותית בזכויות הפרט או באינטרסים ציבוריים, ובפרט בבטיחות המשתמשים בהן והמושפעים מהן.

ביחס לפגיעה בזכויות הפרט, חשוב להדגיש כי החשש לפגיעה כזו לא מתעורר רק במקרים של שימוש במערכות מבוססות בינה מלאכותית במגזר הציבורי.[184] גם שימוש במערכות אלו במגזר הפרטי, כגון על מנת להחליט אם לקבל עובד מסוים לעבודה או אילו עובדים לפטר, או על מנת לקבוע אם פרט מסוים זכאי לקבל אשראי ובאילו תנאים, עלול לגרום, במקרה של התממשות כשל פנימי או איום חיצוני, לפגיעה משמעותית בזכויות ובאינטרסים של הפרט.

באשר לפגיעה באינטרסים ציבוריים, למשל, קיים חשש כי השימוש במערכות מבוססות בינה מלאכותית בתחום השירותים הפיננסיים עלול לגרום לפגיעה ביציבות השווקים ובאינטרסים כלכליים רחבים. כך לדוגמה, בשנת 2013 הופץ פרסום כוזב כי התרחש פיצוץ בבית הלבן שהביא לצניחה של המדדים המרכזיים בבורסות בארה"ב. אחת הסיבות המרכזיות לצניחה המהירה הייתה כי חברות השקעה רבות עשו שימוש במערכות מבוססות בינה מלאכותית אוטונומיות למסחר בשוק ההון (Algo-trading). המערכות, שנחשפו לידיעה החדשותית הכוזבת, הגיבו באופן אוטומטי ומיידי לידיעה וגרמו לצניחה ניכרת של המדדים. אמנם השוק חזר לעצמו במהרה לאחר שהתברר שמדובר בדיווח כוזב,[185] אך דוגמה זו ממחישה כיצד שגיאות של המערכת, במקרה זה על רקע התממשות איום חיצוני, עלולות להוביל להשלכות כלכליות משמעותיות.[186]

לבסוף, התממשות כשלים פנימיים במערכת או איומים חיצוניים עלולה לגרום לסיכוני בטיחות משמעותיים. כך למשל, הגם שבאופן עקרוני שימוש במכוניות עצמאיות צפוי להפחית את מספר התאונות בכבישים, כשלים של המערכת ואיומים עליה עלולים לגרום לתאונות ולפגיעה בבטיחות משתמשי הדרך.[187] לשם ההמחשה, נמצא כי הדבקה של מדבקה קטנה על תמרור "עצור" עלולה לגרום לכך שהמערכת לא תצליח לזהות את התמרור, מה שעלול לשבש את מודל הנהיגה האוטונומי ולגרום למפגע בטיחותי באקראי או בתקיפה מכוונת (ראו דיון בתקיפת התחמקות (Evasion

---

[184] להרחבה על השימוש במערכות המבוססות על בינה מלאכותית בתחום הרווחה ראו גם: סיון תמיר **בינה מלאכותית בשירותי ממשל: הטמעת מערכות לקבלת החלטות מבוססות אלגוריתם בשירותי הרווחה** (2020).

[185] Patti Domm, *False Rumor of Explosion at White House Causes Stocks to Briefly Plunge; AP Confirms Its Twitter Feed Was Hacked*, *CNBC* (Apr. 23, 2013), https://www.cnbc.com.

[186] להתייחסות רחבה יותר להשלכות של השימוש במערכות המבוססות על בינה מלאכותית במגזר הפיננסי ראו אחיעז, חמדני, עמירם וקסטיאל, לעיל ה"ש 22.

[187] Zachary Arnold & Helen Toner, AI ACCIDENTS: AN EMERGING THREAT - WHAT COULD HAPPEN AND WHAT TO DO, CSET Policy Brief, 16 (2021), https://cset.georgetown.edu.

Attack) לעיל).[188] באופן דומה, קיימים סיכונים בשימוש במערכות מבוססות בינה מלאכותית בתחום הבריאות. כבר כיום נעשה בעולם שימוש במערכות אלה על מנת לסייע באבחון של מצבים רפואיים ועל מנת להמליץ לרופאים על אפשרויות טיפול. לנוכח רגישות הטיפול הרפואי, במקרים מסוימים כשלים של המערכת ואיומים עליה עלולים להביא לפגיעה ממשית בבריאות מטופל.

4.6. __אחריותיות__

ב-18 למרץ 2018 התנגשה מכונית באיליין הרצברג, כשחצתה את הכביש שלא במעבר חציה, וכתוצאה מהתאונה היא נפטרה מפצעיה. זו הייתה הפעם הראשונה שבה נהרג הולך רגל בתאונה שבה מעורבת מכונית עצמאית (אוטונומית). התאונה התרחשה בארה״ב במסגרת ניסוי, כאשר המכונית פעלה באופן עצמאי, כלומר לא בשליטתו של נהג הבטיחות. בעוד שרשויות האכיפה בארה״ב מיהרו להודיע כי לא ינקטו הליכים פליליים כנגד החברה שהפעילה את הניסוי אף שהמערכת לא הצליחה לסווג ולחזות כנדרש את מסלול ההליכה של הולכת הרגל, כשנתיים לאחר התקרית הוגש כתב אישום כנגד נהגת הבטיחות, הנהגת האנושית שתפקידה להתערב במקרה שהרכב אינו פועל כנדרש. לפי הנטען בכתב האישום, הנהגת גרמה ברשלנות למותה של הולכת הרגל מכיוון שהייתה עסוקה בטלפון הסלולארי שלה בזמן התאונה.[189] המשפט הפלילי בעניינה, שעדיין מתנהל נכון לכתיבת שורות אלה, מעורר עניין רב ברחבי העולם. שכן באופן עקרוני, וכפי שיפורט להלן, מקרים מסוג זה מעלים שאלות בנוגע לאפשרות להטיל אחריות בכלל, ואחריות משפטית (פלילית או אזרחית) בפרט, במקרה של שימוש במערכות מבוססות בינה מלאכותית.

האחריותיות (Accountability), ובכלל זה האחריות המשפטית (Legal liability), נמנית עם האתגרים המרכזיים הקשורים בשימוש במערכות מבוססות בינה מלאכותית. המאפיין האוטונומי של מערכות אלה, והעובדה כי הן מסוגלות לחקות שלבי ״מחשבה״ מסוימים המזוהים עם פעילות אנושית ולפעול בהתאם, עלולים במקרים מסוימים לערער את המבנה של האחריותיות, המושתת על הימצאות אדם במוקד ההתרחשות שככלל נושא באחריות למעשיו. ככל שהמעורבות האנושית בהחלטה או בפעולה מצטמצמת ורחוקה יותר, אתגר זה מתחדד, ומתעוררות שאלות בנוגע לנשיאה באחריות בגין טעויות שנגרמו אגב השימוש במערכות אלה, ובכלל זה, מי נושא באחריות? מה סוג האחריות שניתן לייחס לו? ומה משטר האחריות המתאים לבחינת שאלת האחריות?

המונח אחריותיות מתייחס ככלל לצורך בכך שיהיו גורמים מתאימים שיוודאו כי המערכת מפותחת ופועלת בהתאם לתפקיד שהוגדר לה ולסביבה הרגולטורית הרלוונטית. ניתן להגביר את מידת האחריותיות באמצעות קידום נורמות ארגוניות המבטאות נטילת אחריות, וזאת לצד הטלת אחריות משפטית, מוסרית או חברתית על הגורמים המתאימים כאשר הם לא עומדים בחובתם.

---

[188] **למידת מכונה אדברסרית**, לעיל ה״ש 175, בעמ׳ 3; Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, Dawn Song, IEEE/CVF CONF. COMPUT. VISION AND PATTERN RECOGNITION, *Robust Physical-World Attacks on Deep Learning Visual Classification* (2018), https://ieeexplore.ieee.org/document/8578273.

[189] Matt McFarland, *Uber self-driving car operator charged in pedestrian death*, CNN (Sept. 18.2020), https://edition.cnn.com.

<u>בנוגע לקידום נורמות ארגוניות</u>, המטרה היא להביא לכך שמפתחי המערכות והמשתמשים בהן יבינו כיצד הן משליכות על זכויות יסוד ואינטרסים ציבוריים, ויוכלו לנקוט בצעדים הנדרשים על מנת למנוע או לצמצם את החשש שהשימוש במערכות אלה יגרום לתוצאות שליליות.

במסגרת זו, לדוגמה, מפתחי מערכות מבוססות בינה מלאכותית או המשתמשים בהן יכולים לבחון את השלכות השימוש במערכת באופן עיתי במסגרת **תהליך של הערכת השפעה** (impact assessment) או הערכת סיכון (risk assessment). זאת, בין היתר, ביחס לסיכונים והאתגרים שנדונו לעיל. ביצוע תהליכי הערכת השפעה וסיכון יאפשר להקטין את החשש להתממשות סיכונים לא צפויים או לשימוש במערכת כאשר הנזק הצפוי עולה על התועלת. ככל שהערכות אלו יפורסמו לעיון הציבור או יועברו לגורמי פיקוח, תגדל גם האפשרות לעשות שימוש בכלים משפטיים (לדוגמה, סנקציות רגולטוריות) או חברתיים (למשל חרם צרכני) כדי להטיל אחריות על רקע תוצאות שליליות שהתרחשו עקב השימוש במערכת.[190]

באופן דומה, מפתחי מערכות מבוססות בינה מלאכותית והמשתמשים בהן יכולים **למנות אחראיים ארגוניים** שתפקידם יהיה להבטיח כי הארגון נוקט באמצעים המתאימים כדי להתמודד עם האתגרים והסיכונים שמעורר השימוש במערכות אלה. למשל, הם יבדקו ויוודאו כי הארגון עומד בסטנדרטים מקצועיים מקובלים, כי תהליכי הערכת ההשפעה נעשים כנדרש וכי מקבלי ההחלטות ערים לחששות הנוגעים לשימוש במערכת ובוחנים אותם כחלק מתהליך קבלת ההחלטות בארגון. זאת לדוגמה, בדומה להמלצת הרשות להגנת הפרטיות למנות ממונה על הגנת הפרטיות.[191]

<u>בנוגע להטלת אחריות</u>, היא יכולה להיעשות בשני מישורים עיקריים – במישור המשפטי ובמישור המוסרי-חברתי. אחריות משפטית (legal liability), מתייחסת לשימוש בכלים משפטיים או רגולטוריים שנועדו להתמודד עם מצב שבו יחיד או תאגיד אינו עומד בסטנדרטים המחייבים הנדרשים או מפר חובה שמוטלת עליו. לצד האחריות המשפטית, עלולה להתקיים גם אחריות מוסרית או חברתית (responsibility), כשהמערכת לא פותחה או פועלת כראוי.[192] למשל, אחריות חברתית יכולה להתבטא במוניטין תקשורתי שלילי או בכך שהציבור יפסיק להשתמש במערכת, מכיוון שהוא רואה בה את הסיבה להתרחשותה של תוצאה שלילית ולא מעוניין בהישנותה. ברם, פרק זה יתמקד בסוגיות הקשורות לאפשרות להטיל אחריות משפטית, ובעיקר אחריות נזיקית או פלילית, על רקע השימוש במערכות מבוססות בינה מלאכותית. זאת, לנוכח האתגרים הניצבים בפני המערכת הרגולטורית והמשפטית שמתעוררים בפרט בהקשרים אלה, ומשום חשיבותם של ההסדרים הללו ליצירת מבנה תמריצים הולם ולהגנה על זכויות ואינטרסים.

בהקשר זה, ראוי להקדים לדיון מספר דגשים: <u>ראשית</u>, יודגש כי מטרתו של פרק זה היא אך להציג ברמה כללית של הפשטה את האתגרים שהשימוש במערכות מבוססות בינה מלאכותית עלול לעורר ביחס לאחריות המשפטית, וכן כיווני חשיבה מקובלים שהובעו בעניין זה. בהתאם, אין בפרק זה משום הבעת עמדה ביחס לתחולת הדין הקיים או להסדרים הרצויים.

[190] Bernd Carsten Stahl, *Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence*, 62 INT'L J. INFO. MGMT. 102441, 102448 (2022).

[191] הרשות להגנת הפרטיות **מינוי ממונה הגנה על פרטיות בארגון ותפקידיו** (2022).

[192] *AI Principles Accountability (Principle 1.5)*, OECD (Sept. 19, 2022), https://oecd.ai/en.

שנית, יודגש כי ייתכן שחלק מההיבטים הנסקרים להלן זוכים למענה עקרוני במסגרת הדין הקיים בישראל, או באמצעות פרשנות מתבקשת להסדרים קיימים, אולם למיטב הידיעה אין קביעה פוזיטיבית עקרונית בעניינים אלה, הנדונים בספרות, דו"חות ומסמכים מישראל ומהעולם.

שלישית, יודגש כי כמובן ישנו פער יסודי בתכליות ובהסדרים השונים בנוגע לאחריות נזיקית ולאחריות פלילית. אולם לשם הניתוח ולנוחות הקריאה, ומאחר שהסקירה להלן נועדה לבחון את הסוגיות הרוחביות הקשורות בהטלת אחריות משפטית בגין תוצאות שליליות שנגרמו אגב שימוש במערכות מבוססות בינה מלאכותית, תיאור הסוגיות הדומות ייעשה ביחס לאחריות נזיקית ופלילית במשותף, תוך התייחסות להיבטים פרטניים של הסוגיות, הדוגמאות וההבדלים בנפרד. מובן כי ניתוח מלא של הסוגיה מחייב לבחון כל אחד מתחומים אלו – התחום הפלילי והתחום הנזיקי – באופן נפרד, ואופן הצגת הדברים להלן הוא אך לשם הניתוח ונוחות הקריאה כאמור.

4.6.1. האתגרים בשימוש בהסדרים הקיימים להטלת אחריות

כפי שיפורט להלן, השימוש במערכות מבוססות בינה מלאכותית עלול לאתגר את היכולת לעשות שימוש בהסדרים הקיימים להטלת אחריות, נזיקית או פלילית. במקרים מסוימים המאפיינים הייחודיים של מערכות אלה לא עולים בקנה אחד עם העקרונות העומדים בבסיס ההסדרים המשפטיים בתחומים הללו, וכפועל יוצא, נוצרים קשיים מהותיים ביחס לאפשרות להטיל אחריות כאשר שימוש במערכת הביא לתוצאה שלילית. נוסף על כך, גם במקרים שבהם ראוי להטיל אחריות, לא אחת השימוש במערכות עלול לעורר קשיים פרוצדוראליים בהוכחת האחריות.

קשיים אלו עשויים להתעורר ככלל על רקע המורכבות של המערכות והיכולת הייחודית שלהן "ללמוד" ולשנות את מאפייני הפעולה בהתאם למידע חדש שהן נחשפות אליו, לעיתים באופן אוטונומי. כתוצאה ממאפיינים אלו, וכפי שיפורט להלן, לא תמיד ברור כיצד לייחס את התוצאות השליליות הנובעות מהשימוש במערכות מבוססות בינה מלאכותית לגורם אנושי, ולמי.

א. קשיים מהותיים בהטלת אחריות

ראשית, קושי שעשוי להתעורר בקשר לשימוש במערכות מבוססות בינה מלאכותית, נוגע לסוגיית ייחוס האחריות לגורם ספציפי, ובפרט לשאלה על מי ניתן להטיל אחריות בגין התרחשות תוצאה שלילית בעקבות השימוש. לא אחת, המעורבות של מערכת מבוססת בינה מלאכותית בהתרחשות, גורמת לכך שלא ניתן בנקל לבחור את הגורם שראוי לייחס לו את האחריות. כשבאופן עקרוני, ניתן לחשוב על מספר גורמים מרכזיים שניתן לייחס להם אחריות, לבדם או ביחד עם אחרים.

בהקשר הנזיקי, שאלה זו רלוונטית הן במסגרת מערכת היחסים שבין המזיק לניזוק, הן (וככל הנראה ביתר שאת) במסגרת מערכת היחסים שבין גורמים שונים העשויים לחוב ביחד. *בהקשר הפלילי*, השאלה רלוונטית בקשר לחבות של חשוד או נאשם בביצוע עבירה לבד או עם אחר.

אפשרות אחת, היא להטיל אחריות על מפתח המערכת. על פניו, ובמיוחד כשמדובר במערכות מורכבות, המפתח הוא בעל הידע כיצד המערכת אמורה לפעול, איך נכון להשתמש בה ומהם הסיכונים שעלולים להתרחש בעקבות השימוש. אולם אפשרות זו אינה חפה מקשיים. שכן במקרים רבים מערכות מבוססות בינה מלאכותית ממשיכות להתפתח ולשנות את מתווה הפעילות שלהן גם לאחר שיצאו מידיהם של המפתחים. קושי נוסף הוא שאחד הגורמים המרכזיים לכך שמערכת לא פועלת כהלכה הוא מאגרי מידע לא מספיק איכותיים ששימשו לאימון שלה. הקושי מתעורר

כשמפתח המערכת הוא לא הגורם האחראי על השגת המידע, ולא הייתה לו אפשרות לדעת כי מאגר המידע לא איכותי מספיק (להרחבה בעניין זה ראו פרק "אמינות, עמידות ובטיחות" להלן). במקרים אלו, יהיה מקום לבחון אם להטיל את האחריות על ספק המידע או על המפתח.[193]

אפשרות שניה, היא לקבוע כי האחריות תוטל על מפעיל המערכת, קרי על הגורם, ככלל עסקי, שמשתמש במערכת לטובת פעילותו. היתרון בהטלת האחריות על המפעיל נובע מכך שבמרבית המקרים המפעיל הוא מי שמפיק את עיקר התועלת מהשימוש במערכת, ובהתאם ראוי גם להיות מי שיישא באחריות במקרה שנגרמה תוצאה שלילית בעקבות השימוש.[194] מעבר לכך, הטלת האחריות על המפעיל תתמרץ אותו להבין כיצד המערכת פועלת, לבצע ניתוח עלות-תועלת ביחס לשימוש בה, ולהפעיל אותה תחת מעטפת הולמת למניעת תוצאות שליליות (ראו למשל פרק "מעורבות אנושית" לעיל, שיכולה לשמש ככלי לצמצום החשיפה). החיסרון המרכזי הוא, שלא תמיד יש למפעיל אפשרות אמתית להבין כיצד פועלת המערכת, וחשוב מכך, פעמים רבות לא יהיו בידיו כלים איכותיים כדי לנקוט באמצעי הזהירות הנדרשים.

אפשרות שלישית, במקרים שבהם מעורב גורם אנושי, וכתלות במידת המעורבות שלו (להרחבה ראו פרק "מעורבות אנושית" לעיל), עשויה להיות נטייה לראות אותו כאחראי לתוצאות השליליות שנגרמו במסגרת השימוש במערכת, שכן מתפקידו לכאורה למנוע התרחשות תוצאות שליליות כאלה. ואולם, כפי שכבר תואר בהרחבה בפרק הנוגע למעורבות אנושית, נטייה זו לא בהכרח תהיה מוצדקת בגלל מגבלות הכרוכות במעורבות האנושית בפעילות מערכות מבוססות בינה מלאכותית. זאת, בפרט, במקרים שבהם אותו גורם אנושי נעדר יכולת אפקטיבית להתערב בפעילותה של המערכת, למשל כשהיא לא ניתנת להסברה או כשהממשק שלה אינו מאפשר בקרה יעילה.

אפשרות רביעית, ונעדרת ביסוס בעת הזו, היא שתתפתח גישה לפיה בהקשרים מסוימים ניתן יהיה להכיר באישיות משפטית נפרדת למערכת מבוססת בינה מלאכותית, בדומה לזו שמכירים בה לגבי תאגידים, ולהטיל אחריות על המערכת עצמה. כמובן, אפשרות זו מצריכה לעשות שימוש בקונסטרוקציה משפטית שעלולה לעורר קשיים לוגיים ופרקטיים (למשל, אם למערכת מבוססת בינה מלאכותית עשויות להיות זכויות ככל שהיא עלולה לשאת באחריות משפטית; מה היחס בין המערכת לבין מפתחיה ומפעיליה; ואם וכיצד ניתן לייחס למערכת יסוד נפשי). נוסף על כך, אפשרות זו מעוררת קושי לגבי השתת סנקציה על המערכת.[195] מכל מקום, הדבר אינו על הפרק לעת הזו.

אפשרות חמישית, היא, במקרים המתאימים, לבחון את האחריות של גורמים נוספים אשר היו מעורבים בתהליך כניסת המערכת לשימוש או בשימוש שגרם לתוצאה השלילית. כך למשל, במקרים מסוימים ייתכן כי יהיה ראוי להטיל אחריות על היבואן של המערכת, ככל שלא ביצע את הבדיקות הנדרשות לפני שהכניס אותה לשימוש בשוק המקומי; על המפיץ של המערכת, ככל שלא

---

[193] Herbert Zech, *Liability for AI: public policy considerations*, 22 ERA F. 147, 148-149 (2021).

[194] European Commission*, Liability for Artificial Intelligence and Other Emerging Digital Technologies*, PUBLICATIONS OFFICE, 39-42 (2019), Available at: https://data.europa.eu.

[195] יש שהציעו כי את הסנקציות הנהוגות כיום ניתן להחיל, בשינויים המחויבים, גם על המערכות. כך למשל, הוצע כי במשך תקופת "המאסר" המערכת תושבת ולא ניתן יהיה לעשות בה שימוש; או שכחלופה לעבודות שירות, המערכת תפסיק לפעול לטובת המטרות המסחריות שלה ותפעל לטובת הציבור; אפילו עלתה ההצעה שבמקום עונש מוות הקוד של המערכת ימחק לצמיתות. אולם ספק אם הצעות אלו מהוות כלי ענישה אפקטיבי אשר יספק מענה הולם להתמודדות עם מקרים בהם השימוש במערכת גרם לתוצאה שלילית. להצעות בכיוון זה ראו: Gabriel Hallevy, I, Robot - I, Criminal: When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses, 2010 SYRACUSE SCI. & TECH. L. REP. 1, 29-35 (2010).

עשה את הבדיקות הנדרשות לפני תחילת ההפצה; או על הבעלים של המערכת, קרי הגורם שלטובתו פותחה המערכת. עם זאת, חשוב לציין כי לעיתים רבות תהיה זהות בין חלק מהגורמים הנ"ל.[196]

<u>שנית</u>, עקרון יסוד הוא שככלל אין מטילים אחריות על גורם שאינו נושא באשמה ביחס לתוצאה השלילית שהתרחשה. וכפי שיפורט להלן, השימוש במערכות מבוססות בינה מלאכותית עלול להקשות על האפשרות לייחס אשמה בגין החלטותיהן ופעילותן לגורמים אנושיים. לפיכך הטלת אחריות עלולה לעורר שאלות שונות, והדבר מלמד על האתגר המשמעותי הגלום בהיבט האחריותיות. כך למשל, עלולות להתעורר שאלות בנוגע לבחינת עוולת הרשלנות *בדיני הנזיקין* ולאפשרות להוכיח התרשלות או הפרה של חובת הזהירות במסגרתה. כך גם *במשפט הפלילי*, עלולות להתעורר שאלות הנוגעות, בין היתר, לקיומו של היסוד הנפשי, בהיבטי מודעות וכוונה.

<u>שלישית</u>, קושי דומה לזה שמתעורר בקשר לייחוס אשמה, עשוי לעלות כאשר נדרש לבחון אם קיים קשר סיבתי משפטי בין המעשה לבין התוצאה השלילית. מכיוון שמבחן מרכזי המשמש לבחינת הקשר הסיבתי המשפטי בדיני הנזיקין ובמשפט הפלילי הוא מבחן הצפיות,[197] ומאחר שמאפיין בולט של מערכות מבוססות בינה מלאכותית הוא יכולתן לנתח מידע ולמצוא קשרים מעבר ליכולות אנושיות ולפעול על בסיס קשרים אלה, עלול להתעורר קושי לייחס לגורם אנושי יכולת לצפות את התרחשות התוצאה השלילית, ובהתאם להיווצר אתגר ייחודי בזיהוי קשר סיבתי משפטי.

ב. קשיים פרוצדוראליים בהטלת אחריות

לצד הקשיים המהותיים שנסקרו לעיל ביחס ליישום ההסדרים הקיימים בדיני הנזיקין ובמשפט הפלילי במקרים של שימוש במערכות מבוססות בינה מלאכותית, עלולים להתעורר גם קשיים פרוצדוראליים שבכוחם לפגום ביכולת לעשות שימוש בהסדרים הקיימים בתחומים הללו.

<u>ראשית</u>, השימוש במערכות מבוססות בינה מלאכותית עלול לעורר קשיי הוכחה משמעותיים, בפרט כאשר נעשה שימוש במערכות שאינן ניתנות להסברה (להרחבה ראו פרק "הסברתיות" לעיל). מערכות אלה פועלות במקרים רבים באופן בלתי צפוי שעלול להביא לתוצאות שליליות, כאשר הסיבות לכך שמערכת מסוימת לא פועלת כנדרש הן רבות, ולא אחת קשה לדעת מה הוביל לכשל (להרחבה ראו גם פרק "אמינות, עמידות ובטיחות" לעיל). מכיוון *שבמשפט האזרחי*, ככלל, נטל ההוכחה מוטל על הניזוק, *ובמשפט הפלילי* נטל ההוכחה מוטל על רשויות האכיפה, הרי שהשימוש במערכות מבוססות בינה מלאכותית צפוי להקשות על היכולת לאכוף במקרים הרלוונטיים. לצד זאת, יצוין כי גם *במשפט הפלילי* וגם *בדיני הנזיקין*, לעיתים קרובות נעשה שימוש בחזקות משפטיות, בין היתר לעניין הקשר הסיבתי, מה שעשוי להקל על אכיפת הדינים הרלוונטיים.

---

[196] Proposal for a Regulation of the European Parliment and of the Council Laying Down Harmonised Rules on Artificial intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM (2021) 206 final (Apr. 21. 2021).

[197] לדרישת הקשר הסיבתי המשפטי בנזיקין ראו למשל: ע"א 576/81 **בן שמעון נ' ברדה**, פ"ד לח(3) 001 (1984); לדרישת הקשר הסיבתי במשפט הפלילי ראו למשל: ע"פ 8827/01 **שטרייזנט נ' מדינת ישראל**, פ"ד נז(5) 506, פס' 27 לפסק הדין של השופט חשין (2003).

שנית, אף כאשר ניתן להוכיח את יסודות התביעה, השימוש במערכות מבוססות בינה מלאכותית עלול להיות כרוך בעלויות ניכרות. במקרים רבים, על מנת לפענח את הקוד של המערכת, להבין איך היא פועלת באופן כללי וכיצד פעלה במקרה הנתון, ולהסיק מכך כי היה כשל בפעילות המערכת שמצדיק הטלת אחריות, נדרשים משאבים רבים והסתמכות על חוות דעת של מומחים טכנולוגיים. גם זאת אפשר, רק במקרים שבהם הקוד גלוי או שהוא נחשף במסגרת ההליך המשפטי, דבר שאינו מובן מאליו בשים לב לכך שמדובר לא אחת בסוד מסחרי. כך גם, ייתכן שיתעורר קושי לאתר מומחה מתאים, שכן חלק גדול מהמומחים מועסקים על ידי חברות מסחריות שעוסקות בתחום.[198] כתוצאה מכך, הליך שינוהל בעקבות שימוש במערכת צפוי להיות במקרים רבים ארוך ויקר יותר באופן משמעותי לעומת הליך שגרתי. ייקור ההליך והימשכותו, עלולים ליצור "אפקט מצנן" מפני הגשת תביעה או נקיטה בהליכי אכיפה, גם במקרים שבהם הדבר רצוי מבחינה חברתית.[199]

יודגש כי קשיי ההוכחה האמורים אינם נוגעים רק לביסוס האשמה או לקשר הסיבתי המשפטי, אלא נוגעים גם לחלוקת האחריות בין הגורמים השונים שהיו מעורבים בהתרחשות התוצאה השלילית. *בדיני הנזיקין* הדבר עלול להקשות על חלוקת האחריות בין המזיקים (במצב של ריבוי מעוולים)[200] וכן על חלוקת האחריות בין המזיק לבין הניזוק (במצב של אשם תורם)[201]; *ובמשפט הפלילי* על קביעת סוג האחריות של הצדדים לעבירה,[202] וכן על קביעת מתחם העונש ההולם בהתאם לאשם היחסי של החשודים או הנאשמים.[203] האחריות בגין התוצאה השלילית שנגרמה עקב השימוש במערכות מבוססת בינה מלאכותית יכולה להתחלק בין מספר גורמים שחלקם מנויים לעיל. במקרים מסוימים חלוקת האחריות ביניהם תהיה מורכבת להוכחה, ובין היתר, תלויה גם בשיקולי מדיניות. חשוב לציין כי אמנם הקושי בחלוקת האחריות קיים גם כשלא נעשה שימוש במערכות מבוססות בינה מלאכותית, אך כאשר נעשה שימוש במערכות אלה האתגר מתעצם.

4.7. פרטיות

במקרים רבים פיתוח ושימוש במערכות מבוססות בינה מלאכותית מצריך שימוש בנתונים רבים, לעתים בהיקפים עצומים. בפרט, נעשה שימוש בנתונים רבים לצורך פיתוח מודלים (קרי, לשם למידה), לצורך בדיקת המודלים לאחר שפותחו לשם תיקונם או דיוקם, או לצורך עיבוד הנתונים הפרטניים לשם הפקת המלצה, תחזית או החלטה קונקרטית. כאשר הנתונים המהווים בסיס לפעילות זו מהווים מידע אישי, המוגן על ידי חוק הגנת הפרטיות, התשמ"א-1981 (להלן: "חוק הגנת הפרטיות"), הוראות החוק חלות על הפעילות. כך, במקרים רבים המידע המוזן לתוך המערכת לצרכים אלו (הקלט) וגם התוצר, קרי, ההמלצה, התחזית או ההחלטה המתגבשות כתוצאה משימוש בבינה מלאכותית (הפלט) עשויים לכלול מידע אישי. מכאן, שבכל אותם מקרים בהם נעשה שימוש במידע אישי בכל אחד מהשלבים שתוארו לעיל, יחולו הוראות חוק הגנת הפרטיות.[204]

---

[198] משרד המשפטים **אסדרת השימוש ברכבים אוטונומיים - פנייה לקבלת עמדת הציבור בתחום הנזיקין והביטוח** 9 (2021).
[199] European Commission, לעיל ה"ש 194, בעמ' 20-21.
[200] ס' 84 לפקודת הנזיקין.
[201] ס' 68 לפקודת הנזיקין.
[202] סי' ב' לחוק העונשין.
[203] ס' 40ט לחוק העונשין.
[204] השימוש במידע אישי יכול שיעשה במגוון של אופנים וסוגי קבצים, לרבות שמירה, עיבוד והעברה של קבצי טקסט, תמונות, קבצי וידאו, הקלטת קול, מידע ביומטרי ועוד.

בעת קביעת אופן תחולת החוק והחובות החלות על מי שמבצע את הפעילות הכרוכה באיסוף או עיבוד מידע אישי, יש לתת את הדעת לאתגרים הייחודים לפרטיות. אתגרים אלה נובעים מהמאפיינים הטכנולוגיים של "נתוני עתק" (Big data) ובינה מלאכותית.

4.7.1. <u>האתגרים לפרטיות בהם מתאפיין השימוש במערכות מבוססות בינה מלאכותית</u>

כפי שיפורט להלן, פיתוח ושימוש במערכות מבוססות בינה מלאכותית מעוררים מספר אתגרים ייחודיים הנוגעים לצורך להבטיח את ההגנה הנדרשת לזכות לפרטיות.

<u>ראשית</u>, השימוש במערכות אלה עלול לאתגר את היכולת לעמוד **בעיקרון צמידות המטרה**. אחד מעמודי התווך של דיני הגנת הפרטיות, הקבוע במפורש בחוק הגנת הפרטיות, הוא עקרון צמידות המטרה. בהתאם לעקרון זה, ניתן לאסוף מידע אישי ולהשתמש בו אך ורק למטרה אשר לשמה נמסר המידע, אלא אם נושא המידע עצמו הסכים לשימושים נוספים במידע אודותיו.[205] עקרון זה מאותגר בשלב פיתוח הבינה המלאכותית שכן במקרים רבים נדרש לעשות שימוש במידע שנאסף למטרות אחרות. כך למשל, דוגמא מובהקת לאי-התאמה של המטרה, היא מקרה של שימוש במידע שמתקבל לטובת הפעלה קולית לצורך אימון אלגוריתם לאיתור בעיות במצב רפואי של בני אדם. היכולת לקבל את הסכמתם של נושאי המידע לצורך השימוש הנוסף המתבקש, מעוררת קשיים מעשיים שונים, כגון הקושי לשוב ולפנות אל נושאי המידע הרלוונטיים ולקבל הסכמה מדעת לכך, בפרט כשמדובר במידע שנאסף בעבר הרחוק ובנושאי מידע רבים.

<u>שנית</u>, השימוש במערכות מבוססות בינה מלאכותית עלול לאתגר את היכולת לעמוד **בדרישת השקיפות ובחובת היידוע**. דיני הגנת הפרטיות נועדו בין היתר להבטיח את היכולת של נושאי המידע לשלוט במידע המתייחס אליהם, ולקבוע כיצד ייעשה בו שימוש. אחת הזכויות המרכזיות שחוק הגנת הפרטיות מקנה לאדם היא זכות יידוע שמאפשרת לקבל פירוט בנוגע למטרות ולשימושים שעבורם האדם מתבקש למסור את המידע אודותיו, למי יימסר המידע, והאם חלה עליו חובה חוקית למסור את המידע, או שמא מסירת המידע תלויה ברצונו או בהסכמתו.[206] חובה חוקית זו חלה בכל מקרה של פניה לאדם לקבלת מידע אודותיו, גם כשאיסוף המידע נעשה מכוח הסמכה בדין וגם כאשר הפניה לאדם נעשתה בעקבות בקשתו לקבלת שירות מסוים.[207] שקיפות משפרת גם את האפשרות של נושא המידע לממש את זכות העיון במידע, המוקנית לו בחוק הגנת הפרטיות,[208] באופן משמעותי ואפקטיבי, וכן את אחריותיות הארגון כפי שיוסבר בהמשך. אולם, במקרים רבים בהם נעשה שימוש במערכות מבוססות בינה מלאכותית שאינן ניתנות להסברה (ראו פרק "הסברתיות" לעיל), עלול להיווצר קושי בעמידה בדרישה זו. זאת מכיוון שחלק מההיבטים הנוגעים לשימוש במערכת לא יהיו ידועים למפעיל, ובהתאם לא ניתן יהיה למסור אותם לנושא המידע.

---

[205] ראו ס' 2(9) ו-8(ב) לחוק הגנת הפרטיות.
[206] ראו ס' 11 לחוק הגנת הפרטיות; ראו גם "חובת יידוע במסגרת איסוף ושימוש במידע אישי", לעיל ה"ש 27.
[207] שם, בפס' 4-5.
[208] ס' 13 לחוק הגנת הפרטיות.

שלישית, השימוש במערכות מבוססות בינה מלאכותית עלול להוביל לכרסום **בעיקרון ההסכמה מדעת**. חוק הגנת הפרטיות קובע כי הסכמת אדם לאיסוף ולשימוש במידע אישי אודותיו, נדרשת להיות "הסכמה מדעת".[209] הדרישה להסכמה מדעת מחייבת את מבקש המידע להציג לאדם נתונים בנוגע לאיסוף המידע עליו והשימוש בו, כך שבפני נושא המידע תעמוד תמונה מלאה בטרם יחליט האם להסכים למסירת המידע אודותיו.[210] הקושי בעמידה בדרישת השקיפות שתואר לעיל, והקושי לקיים את חובת היידוע בשל אופיין של חלק ממערכות בינה מלאכותית שאינן ניתנות להסברה, עלולים, במקרים מסוימים, לאתגר את היכולת לקבל ההסכמה מדעת מנושא המידע.[211]

רביעית, השימוש במערכות מבוססות בינה מלאכותית עלול ליצור מתח מובנה בין הצורך בנתוני עתק לבין מחיקת מידע עודף בהתאם **לעיקרון צמצום המידע**. האפשרות לקבל גישה ל"נתוני עתק" (Big Data), והיכולת לנתח מידע בהיקפים גדולים, טומנת בחובה פוטנציאל רב לקידום ופיתוח השימוש במערכות מבוססות בינה מלאכותית. שכן, כאמור לעיל, בשלב פיתוח המודל המערכת "מתאמנת" על המידע, כאשר בשלב איסוף המידע לא בהכרח ניתן לדעת מהן התובנות והתועלות שהמערכת תדע להפיק מהמידע הנאסף. המשמעות היא שאיסוף וניהול כמויות גדולות של מידע יכולים לסייע בפיתוח המערכות, אף שחלק מהמידע הנאסף כלל אינו חיוני למתן השירות עצמו. מנגד, איסוף ושמירה של מידע אישי עודף מגדילים את הסיכון לפגיעה בפרטיות, שכן המידע עשוי לדלוף ולהיחשף לגורמים לא מורשים. משכך, מחייבות תקנות הגנת הפרטיות (אבטחת מידע), התשע"ז-2017, כל בעל מאגר מידע אישי לבחון אחת לשנה אם הוא מחזיק במאגר מידע עודף, שאינו דרוש לו עוד.[212] מכאן, שכאשר המידע בו נעשה שימוש לצורך המערכת הוא מידע אישי, מתקיים מתח מובנה בין עיקרון צמצום המידע, הקובע כי יש לאסוף ולהשתמש אך ורק במידע הרלוונטי וההכרחי לשם הגשמת המטרה אשר לשמה המידע נאסף, לבין הרצון לספק למערכת מידע בהיקפים גדולים ככל הניתן.

חמישית, ניתן לעשות שימוש במערכות מבוססות בינה מלאכותית לטובת **הסקת מידע רגיש**. יכולת הסקת המסקנות וההיקשים של שימוש בנתוני עתק ולמידת מכונה מאפשרת לשלב ולנתח מרכיבי מידע שונים, וכך להסיק מסקנות הכוללות מידע רגיש ממידע לא רגיש (למשל מצב בריאותי מהרגלי רכישה). לדוגמה, מחקרים מראים, כי בנסיבות מסוימות בינה מלאכותית מסוגלת לזהות ממידע ביומטרי, כגון מבנה עצמותיו או אופן התנהגותו של אדם, דפוסים החוזים את נטיותיו והתנהגותו של אותו אדם ואשר משליכים על מצבו הבריאותי, נטייתו המינית ועוד.

---

[209] בעקבות תיקון שנערך לחוק הגנת הפרטיות בשנת 2007, "הסכמה" בסעיף 3 לחוק מוגדרת כ-"הסכמה מדעת, במפורש או מכללא".

[210] על היותה של חובת היידוע חלק מעקרון ההסכמה מדעת, ראו פסקאות 65-66 ו-78 לעמדת היועץ המשפטי לממשלה עס"ק (ארצי) 7541-04-14 **הסתדרות העובדים הכללית החדשה מרחב המשולש הדרומי - עיריית קלנסווה**, (פס' 144 לפסק הדין של השופט איטח (נבו 15.3.17).

[211] ראו למשל את הסקירה המקיפה בתחום הכלכלה ההתנהגותית:
Acquisti, Alessandro, Brandimarte, Laura, and Loewenstein, George, 'Secrets and Likes: The Drive for Privacy and the Difficulty of Achieving It in the Digital Age'. Journal of Consumer Psychology. 30(4), 736-758 (Oct. 2020), Available at SSRN: https://ssrn.com/abstract=3688497;
Solove, Daniel J., The Limitations of Privacy Rights (February 1, 2022). 98 Notre Dame Law Review - (Forthcoming 2023), GWU Law School Public Law Research Paper No. 2022-30, Available at SSRN: https://ssrn.com or http://dx.doi.org.

[212] תקנה 2(ג) לתקנות הגנת הפרטיות (אבטחת מידע), תשע"ז- 2017. ראו גם טיוטת מסמך מדיניות של הרשות להגנת הפרטיות בנושא צמצום מידע. המסמך זמין כאן.

שישית, יש חשש שייעשה שימוש במערכות מבוססות בינה מלאכותית לשם **זיהוי חוזר של מידע שעבר התממה**. בשל היכולת של מערכות מבוססות בינה מלאכותית לעבד מגוון רחב של נתונים ממגוון מקורות, השימוש בהן עלול להגביר את הסיכון לזיהוי נושאי המידע במערכי נתונים, לרבות כאלו המשמשים לאימון מערכות, אשר עברו תהליכי התממה (אנונימיזציה). החשש בהקשר זה גובר ככל שמדובר בשילוב של מידע ממקורות שונים ובהיקפים גדולים.

שביעית, השימוש במערכות מבוססות בינה מלאכותית עלול לעורר קושי **בהגדרת תפקידים ותחומי אחריות ביחס למידע אישי**. כאמור, מידע אישי עשוי לשמש בשלב פיתוח מערכת הבינה המלאכותית, בשלב בדיקת המערכת ובשלב השימוש בה. לעיתים בשלב פיתוח המערכת המפתחים מסתייעים במאגרים של צד שלישי, הכוללים מידע אישי, לצורך אימון המערכת. במקרים אלו הגדרת תחומי האחריות של כל אחד מהצדדים עלולה להיות סבוכה ומורכבת.

מכלל האמור לעיל, עולה כי בשים לב למאפייני טכנולוגיית הבינה המלאכותית, אשר עושה שימוש במידע בהיקף עצום, הן לשם פיתוחה ולמידתה והן במהלך השימוש בה, הפרטיות מהווה אתגר משמעותי, שיש להידרש אליו במסגרת הפיתוח והשימוש במערכות מבוססות בינה מלאכותית, וגם במסגרת גיבוש מדיניות רגולטורית בתחום זה.

5. חלק חמישי: דרכי התמודדות עם סוגיות ואתגרים אלה

פרק זה סוקר באופן כללי דרכי פעולה אפשריות להתמודדות עם הסוגיות, הסיכונים והאתגרים שנסקרו בפרק הקודם, כפי שהן משתקפות מהספרות ומדו״חות שונים, מישראל ומהעולם. כפי שצוין לעיל, גם בנוגע לדרכי ההתמודדות הסקירה להלן אינה מתיימרת למצות את כל דרכי הפעולה האפשריות, או לקבוע קביעות לגבי הדין החל או זה שראוי שיחול, אלא לעסוק בדרכי פעולה עיקריות ולבחון את יתרונותיהן וחסרונותיהן המרכזיים בהתאם למתאפשר, וזאת כפתח לבחינה עתידית. בדומה לאמור לעיל, גם פרק זה מיועד לשמש כבסיס להמשך דיון.

בפרט, עבור גורמי הממשלה והרגולטורים הרלוונטיים, ולאור ההצעה להלן בעניין ״מיפוי השימושים בבינה מלאכותית והאתגרים הנלווים להם בענפים מוסדרים״, פרק זה יכול לסייע בהבנה ומיפוי של דרכי הפעולה האפשריות להתמודדות עם הסוגיות, הסיכונים והאתגרים הכרוכים בשימושים קונקרטיים במערכות מבוססות בינה מלאכותית בענף שעליו הם אמונים.

5.1. <u>אפליה</u>

כאמור לעיל, ישנו קושי ממשי להתמודד עם החשש שהשימוש במערכות מבוססות בינה מלאכותית יביא להנצחת ואף להחרפת בעיית האפליה. אולם, ישנן מספר דרכים שמאפשרות להתמודד, ולו באופן חלקי, הן עם האתגרים הטכנולוגיים והן עם האתגרים המשפטיים, ובכלל זה:

א. מאגרי מידע מגוונים ואיכותיים

דרך מרכזית אחת להתמודדות עם אפליה, היא לדרוש כי מערכת מבוססת בינה מלאכותית ״תתאמן״ על בסיס מאגרי מידע רחבים ומגוונים, שיאפשרו לה להתחשב במאפיינים של כל קבוצות האוכלוסייה הרלוונטיות, תוך שמירה כי הנתונים לא משקפים אפליה היסטורית. ניתן לקבוע סטנדרטים שונים לגיוון הנתונים ולאיכותם, בין היתר, באמצעות אתיקה, רגולציה או תקינה. הדבר מחייב במקביל פתרון לבעיה משפטית של מערכות הדינים המגבילות שימוש במידע. פתרון חלקי לבעיה זו נמצא בחוות דעת מקיפה שנכתבה במשרד המשפטים בנוגע לשימוש בתכנים מוגנים בזכויות יוצרים, שקובעת כי ברוב המקרים שימוש כאמור יהווה שימוש הוגן בתכנים.[213] עם הזמן, יהיה צורך לייצר פתרונות משפטיים שיאפשרו שימוש בפרטי מידע גם על רקע הגבלות בתחומי משפט אחרים (דיני הגנת פרטיות, חסיונות וכדומה). מעבר לכך, ניתן לבחון לעודד ואף לחייב במקרים מתאימים שיתוף נתונים בין גופים על מנת לייצר מאגרי נתונים מגוונים ואיכותיים.

ב. ביצוע בדיקות מקדימות ועיתיות

דרך נוספת היא לדרוש כי לפני תחילת השימוש במערכת המבוססת על בינה מלאכותית ייבחנו החלטותיה ופעולותיה, ובפרט אם היא נוהגת בצורה דומה ביחס לאנשים בעלי מאפיינים שונים. במילים אחרות, ניתן לדרוש כי בשלב של התנסות המערכת, היא תיבחן על קבוצות אוכלוסייה שונות באופן שיוודא שהיא אינה מפלה לרעה. ככל שמדובר במערכת שניתנת להסברה, ניתן אף לבחון את האינטראקציה בין הפרמטרים שהמערכת משקללת לבין מאפייני שיוך קבוצתי חשודים בתחום בו היא פועלת, בניסיון לאתר משתנים קורלטיביים למאפיינים מפלים (משתני proxy). גם לאחר שהמערכת נכנסת לפעולה, דרישת בדיקה עתית של הנתונים והחלטות המערכת עשויה לאתר

[213] משרד המשפטים **גילוי דעת: שימושים בתכנים מוגנים בזכויות יוצרים לצורך למידת מכונה** (2022).

בשלב מוקדם נטייה לאפליה. במסגרת זו, ניתן אף לדרוש ביצוע של תהליך הערכת סיכונים שמטרתו לבחון מה רמת החשש שהשימוש במערכת יוביל לאפליה ומה מידת הפגיעה המשוערת במקרה שהחשש יתממש. ייתכן שבמקרים המתאימים ראוי שהבדיקה והפיקוח ייעשו על ידי גורמים חיצוניים (פרטיים או ציבוריים) שיוכלו להעריך את מידת החשש מאפליה.

מבחינה משפטית ורגולטורית, ישנן מספר אפשרויות ליישם את הדרישה האמורה לגיוון. דרך אחת היא לחייב חברות להצהיר כי בדיקות כאמור בוצעו או לחייבן לפרסם את קבוצות האוכלוסייה לגביהן נבחנה המערכת. כשמדובר בחברות שפעילותן כפופה לחובת רישוי העושות שימוש במערכות מבוססות בינה מלאכותית שיוצרות סיכון לאפליה, ניתן אף להתנות את הרישיון הרגולטורי בקיומן של בחינות גיוון כאמור. לחלופין, ניתן ליצור "נמל מבטחים" (safe harbor) משפטי, עבור חברות שייתבעו בגין החלטה מפלה של החברה – במקרה שהן יכולות להוכיח שהמערכת אומנה ונוסתה על מאגר נתונים מגוונים – וזאת במטרה לתמרץ אותן לנקוט אמצעי הזהירות הנדרשים.

ג. קביעת דרישה להסברתיות

אמצעי נוסף שבאמצעותו ניתן להתמודד עם החשש מפני אפליה הוא קביעה של דרישה להסברתיות (ראו פרק "הסברתיות" להלן). קבלת הסבר בנוגע למערכת ולתהליך קבלת ההחלטה על ידי המערכת במקרה הקונקרטי, תסייע לדעת האם הביאה בחשבון מאפיינים מפלים או משתנים קורלטיביים אליהם (ככל שניתן לאתרם), ולנקוט באמצעים המתאימים – טכנולוגיים או משפטיים – כדי להתמודד עם האפליה. כך לדוגמה, ניתן לבטל טכנית אפשרות להתחשב במשתנה proxy מסוים; לאכוף את דיני האפליה על ידי הרגולטורים הרלוונטיים; או לאפשר אכיפה פרטית בגין אפליה על ידי מושאי ההחלטה, ככל שייחשפו לנימוקים שעמדו בבסיס ההחלטה. עם זאת, יש לקחת בחשבון גם את הקשיים הנלווים לדרישה להסברתיות, כפי שהוצגו לעיל.

ד. עריכת מבחנים תוצאתיים

לבסוף, אמצעי שיכול לסייע בהתמודדות עם החשש מפני אפליה במערכות מבוססות בינה מלאכותית, הוא שימוש מוגבר במבחנים תוצאתיים על מנת לבחון אם התרחשה אפליה במקרה נתון, וזאת כמנגנון משלים או כחלופה אפשרית לצורך בהסברתיות.

כך למשל, ניתן לקבוע חזקה כי השימוש במערכת גורם לאפליה, ולכן פסול, כשיש שוני העולה על סף מסוים בין אנשים מקבוצות שונות באוכלוסייה. אפשרות נוספת, היא להתבסס על בחינה תוצאתית לצד יצירת "נמל מבטחים" שיספק הגנה במקרה של שימוש במערכות שנבדקו או במערכות ניתנות להסברה. חזקה או נמל מבטחים כאמור עשויים לתמרץ חברות לבכר שימוש במערכות כאלה כאמצעי להגנה מפני תביעות. לחלופין, ניתן לעשות שימוש בכלים משפטיים על מנת להבטיח יחס שווה לגורמים שונים, למשל על ידי הרחבת חובות הייצוג ההולם.[214]

5.2. <u>מעורבות אנושית</u>

לאור האמור לעיל, ברי כי לצד היתרונות המשמעותיים שניתן להשיג באמצעות מעורבות אנושית, ישנו צורך להתאים את המעורבות לנסיבות שבהן נעשה השימוש במערכת, על מנת לוודא כי התועלת שבמעורבות האנושית משמעותית. לעניין זה חשוב להדגיש כי אין מדובר רק על השאלה

[214] ראו למשל בעניין ייצוג הולם לנשים: מרכז המידע והמחקר של הכנסת **חקיקה בישראל הקובעת ייצוג הולם לנשים** (2005).

מתי צריכה להיות מעורבות אנושית, אלא נדרש גם לחשוב כיצד נכון לעצב את האינטראקציה בין הגורם האנושי לבין המערכת על מנת לנצל את היתרונות היחסיים של כל צד ולמנוע הטיות.

סוגיית הצורך במעורבות אנושית רלוונטית כמעט לכל השימושים בבינה מלאכותית. כך לדוגמה, בתחום הבריאות, עולה השאלה אם יש צורך שרופא יבקר דיאגנוזה או המלצות לטיפול שסיפקה מערכת מבוססת בינה מלאכותית; בתחום הפיננסי, עולה שאלה אם יש צורך בבקרה על שימוש במערכת בינה מלאכותית לייעוץ השקעות באופן שלא תתקבל החלטה על השקעה בהתבסס על מערכת מסחר אלגוריתמי בלבד, או אם יש צורך שפקיד יקבל החלטה על מתן או סירוב אשראי כדי שלא לחשוף את הגוף המלווה לסיכוני אשראי לא מקובלים או לשלול לשווא אשראי מלווה פוטנציאלי; בתחום התעסוקה, עולה שאלה אם יש צורך שהמעסיק יפקח על שימוש במערכת על מנת להעסיק או לפטר עובדים. עם זאת, יש להבחין בין סוגי המעורבות השונים, ולבחון איזה סוג של מעורבות אנושית נדרש (אם בכלל) ביחס לכל תחום ושימוש פרטני בבינה מלאכותית.

על מנת לקבוע את הדרישה למעורבות אנושית ביחס לכל תחום ושימוש בבינה מלאכותית, ואת היקפה במקרים שבהם היא נדרשת, יש לתת את הדעת למגוון הנסיבות והמאפיינים. בכלל זה, והגם שאין מדובר ברשימה ממצה, ניתן להתחשב במאפיינים אלה: ראשית, אופי ההחלטה והשפעתה – ככל שהסיכון שעלול להיווצר בעקבות ההחלטה גדול יותר, ובהתאם גובר החשש לפגיעה משמעותית בזכויות יסוד או באינטרסים ציבוריים ולהשפעה לרעה על מושא ההחלטה, כך מתעצמת, למצער בעת הזו, ההצדקה לשקול דרישה למעורבות אנושית באופן משמעותי יותר. כך למשל, ההצדקה למעורבות אנושית עשויה להיות גבוהה במיוחד כשמדובר בהחלטות בלתי הפיכות (כמו בתחום הרפואה).[215] שנית, אפקטיביות המעורבות האנושית – כך למשל, כאשר מדובר במערכות שאינן ניתנות להסברה, כפי שתואר לעיל, היכולת של הגורם האנושי להתערב באופן אפקטיבי בהחלטת המערכת נפגעת. שלישית, ובהנחה שיימצא כי נדרשת מעורבות אנושית, יש לבחון אם המעורבות האנושית צריכה להיות מראש (באמצעות קבלת ההחלטה הסופית על ידי אדם, או כשהוא יכול להטיל וטו בזמן אמת), או שניתן להסתפק במנגנון של ערעור בדיעבד.

מכל מקום, אין באמור לעיל כדי לקבוע אם ובאלו מקרים יש לדרוש מעורבות אנושית, אלא להצביע על אפשרויות וכיווני מחשבה בלבד. יחד עם זאת, נראה כי במקרים שבהם יימצא כי יש מקום לדרוש מעורבות אנושית כלשהי בתהליך קבלת החלטה, יש לבחון אם נדרש גם להסדיר בקרה על אותה מעורבות אנושית, כדי לוודא שתצליח להגשים את התכלית שלשמה נדרשה. כאשר מצד אחד, ניתן לבצע בקרה לגבי המקרים שבהם הגורם האנושי החליט להתערב ולסטות מההחלטה של המערכת, ולבחון לדוגמה האם הסטייה הייתה מוצדקת; מה הגורמים שהובילו לסטייה (כשל נקודתי או בעיה רוחבית במערכת); והאם הגורם האנושי הסתמך על המידע הרלוונטי במסגרת החלטתו. מצד שני, אפשר לבצע בקרה לגבי המקרים שבהם הגורם האנושי החליט שלא להתערב, תוך בחינה, בין היתר, האם הוא הפעיל שיקול דעת משמעותי; בחן את הנתונים הרלוונטיים; או הצליח להבין את האופן שבו המערכת הגיעה להחלטה ונימוקיה. כמו כן, יהיה מקום לשקול בחלוף זמן את מידת האפקטיביות של המעורבות האנושית ואת מידת תרומתה לתהליך.

[215] Brennan-Marquez & Henderson, לעיל ה"ש 119, בעמ' 146-147.

5.3. <u>הסברתיות</u>

כאמור לעיל, להחלת הדרישה להסברתיות ביחס לשימוש במערכות המבוססות על בינה מלאכותית יתרונות משמעותיים. ברם, לנוכח האתגרים הקשורים להחלת דרישה כזו, ייתכן שיש מקום לעשות זאת באופן יחסי ובהתאם לנסיבות השימוש במערכת. באופן דומה, ניתן להתמודד עם חלק מהקשיים שתוארו לעיל באמצעות צמצום היקף ההסבר שיינתן (ראו לעיל בפרק הקודם בחלק העוסק ב"הסברתיות"), או הגבלת הגורמים שיהיו חשופים להסבר.

<u>במסגרת בחינת האפשרות להחיל דרישה להסברתיות באופן יחסי ובהקשר מסוים</u>, יהיה מקום להתחשב בזהות הגוף המשתמש במערכת המבוססת על בינה מלאכותית והגוף המושפע מהשימוש, כמו גם באופי ההחלטה והשפעתה. לגבי **הגוף המבצע את השימוש במערכת**, הבחנה אפשרית היא בין שימושים שמבוצעים על ידי גופים פרטיים בעלי מאפיינים ציבוריים, לבין שימושים שנעשים על ידי גורמים פרטיים לחלוטין. ככל שהגוף נוטה יותר לכיוון הציבורי, עשויה להתחזק ההצדקה להחיל לגביו דרישה להסברתיות.[216] לעומת זאת, ביחס לגופים פרטיים, ייתכן שיהיה מקום להחיל את הדרישה להסברתיות באופן רוחבי או באופן סקטוריאלי רק בתחומים שבהם היא נדרשת. בפרט, דרישת ההסברתיות רלוונטית במיוחד כשמדובר בגופים פרטיים שההחלטות שלהם מפוקחות באופן הדוק על ידי רשויות השלטון, משום שלהחלטות אלה יש השפעה משמעותית על חיי הפרט, למשל בתחום הבריאות, הבנקאות או הביטוח.[217] בנוגע **לאופי ההחלטה והשפעתה**, ככל שהסיכון שעלול להיווצר בעקבות ההחלטה גדול יותר, ובהתאם יש חשש לפגיעה משמעותית יותר בזכויות ולהשפעה לרעה על מושא ההחלטה, כך תגדל חשיבותה של הדרישה להסברתיות. למשל, יש הצדקה חזקה יותר לדרוש הסבר לגבי מערכת שקובעת אם אדם יוכל לקבל הלוואה או להתקבל למקום עבודה, לעומת אלגוריתם שנועד להתאים תוכן וידאו או מוזיקה לצרכן.[218]

<u>בעניין הגורמים החשופים להסבר</u>, **מצד אחד**, ככל שהדרישה להסברתיות תהיה רחבה יותר, כך הפיקוח על פעילות המערכת צפוי להיות טוב יותר, והיתרונות שתוארו לעיל עשויים להתגשם. לדוגמה, החלת הדרישה להסברתיות כלפי כלל הציבור, כך שתשתרע אף מעבר למי שמושפעים מההחלטה, תאפשר לקיים שיח ציבורי רחב במשמעויות השימוש במערכת ולפקח עליה (כמובן שדרישה כזו תהיה אפקטיבית יותר אם ההסבר יהיה נגיש וברור לכל הציבור ולא רק ל"יודעי ח"ן"). באופן דומה, הצבת דרישה להסברתיות כלפי הרגולטור הרלוונטי תאפשר פיקוח אפקטיבי מצדו. על כן, יש מקום כי בגופים הרגולטוריים הרלוונטיים יהיו אנשי מקצוע בעלי המומחיות הטכנולוגית הנדרשת על מנת לבחון את ההסברים המתקבלים. **מצד שני**, דרישת הסברתיות צרה יותר תקל על מפתחי המערכת ומשתמשיה, לרבות במישור ההגנה על הקניין הרוחני שלהם,[219] ואולי אף תסייע לצרכנים שלרוב יתקשו להבין הסברים טכניים ומפורטים יתר על המידה.

---

216 ביחס לרשויות ציבוריות ראו ס' 2א לחוק לתיקון סדרי המינהל (החלטות והנמקות), התשי"ט-1958, וכן ברק-ארז **המשפט המנהלי** 423-438 (2010).

217 Bibal, Lognoul, de Streel & Frénay, לעיל ה"ש 126, בעמ' 151-153. לדוגמאות להחלה של חובת הנמקה במגזר הפרטי והמלצה להחיל חובת הנמקה רחבה על החלטות שמקבלים גופים פרטיים על בסיס ניתוח טכנולוגי של נתוני מידע ראו: מעין פרל "פרטיות, שליטה ופיקוח בעידן של נתוני עתק חובת הנמקה על החלטות אלגוריתמיות" **משפט, חברה ותרבות** ב 167, 188-192 (מיכאל בירנהק עורך 2019).

218 Miriam C Buiten, *Towards Intelligent Regulation of Artificial Intelligence*, 10 EUR. J. RISK REG. 44, 57-58 (2019).

219 Andrew D. Selbst, Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1122-1126 (2018).

בפועל, בקצה אחד של הסקאלה, יש מי שהציעו לחייב משתמשים במערכות בינה מלאכותית לפרסם דו"ח הערכת השפעה ביחס לכל שימוש באלגוריתם (algorithm impact statement – AIS), בדומה לאופן שבו מתפרסמים דו"חות ביחס להשפעה סביבתית או רגולטורית (Regulatory Impact Assessment – RIA) של פעולות מסוימות. בהתאם לגישה זו, יהיה מקום לפרסם לעיון הציבור מהי המערכת בה מבוצע השימוש; מה היה תהליך הפיתוח שלה; איך היא עובדת (בהתאם לרמות שתוארו לעיל); ומה צפויה להיות ההשפעה שלה. בקצה האחר של הסקאלה, ניתן לסייג את הזכות לקבל הסבר רק לפרט מושא ההחלטה של המערכת, ואפילו רק למקרים שבהם הוא זקוק לכך לשם הליך משפטי או כשנשלל ממנו שירות חיוני בעקבות החלטת המערכת.[220]

לבסוף, יש לתת את הדעת על כך שלא אחת מפתחי המערכת ומשתמשיה הם בעלי תמריץ משמעותי להשתמש במערכות הניתנות להסבר. כך למשל, לבנקים עשוי להיות תמריץ כלכלי להבין מדוע מערכת אשראי מאשרת או מסרבת לתת הלוואה ללקוח, כדי לוודא שההחלטה לא מגלמת סיכונים בלתי סבירים או מפסידה לשווא את הלקוח. באופן דומה, לחברות השקעות יכול להיות תמריץ כלכלי להבין מדוע מערכת ממליצה על השקעה מסוימת, כדי לוודא שההמלצה טובה או ליישם את התובנות ביחס להשקעות אחרות. במקרים אלה, לאור תמריציהם, מפתחי המערכת ומשתמשיה עשויים ליזום הסברתיות ולרצות בה, גם בהיעדר דרישה רגולטורית להסברתיות. ואולם, בחלק מהמקרים לא בהכרח יהיה לגורמים אלה תמריץ מספק להשקיע בקבלת הסבר, או תמריץ למסור את ההסבר שהתקבל לגורם חיצוני, ואז עשוי להתעורר צורך בהתערבות רגולטורית.

5.4. גילוי

על רקע האמור לעיל, ובפרט במקרים שבהם למפעיל מערכת מבוססת בינה מלאכותית אין תמריץ מספק לגלות למשתמש על אינטארקציה עם בינה מלאכותית, עשויה להיות הצדקה להתערבות רגולטורית בדמות דרישה כי בנסיבות מסוימות יהיה עליו לגלות על השימוש במערכת מסוג זה.

בבחינה השאלה אם ראוי להציב דרישה לגילוי עשויים להיות מספר שיקולים רלוונטיים. בכלל זה, ראשית, **ההשפעה על מצב משתמש הקצה**. ככל שהשימוש במערכת משפיע בצורה יותר משמעותית על זכויותיו, כך ייתכן שיהיה מקום להכיר באינטרס חזק יותר שלו לדעת כי הוא בא באינטראקציה עם מערכת מבוססת בינה מלאכותית. שנית, **ההקשר שבו מתקיימת האינטראקציה**. כשהשימוש במערכת נעשה במסגרת מערכת יחסים בה כבר הוכרה חשיבות השקיפות, כמו במסגרת מערכת יחסים צרכנית (מכוח דיני הגנת הצרכן)[221] או במסגרת דיני הבחירות,[222] כך ייתכן שתהיה הצדקה משמעותית יותר לדרוש גילוי. שלישית, יש לבחון כיצד הגילוי צפוי **להשפיע על התנהלות משתמש הקצה**. במקרים בהם הגילוי יכול לסייע למשתמש להתאים את התנהגותו, למצות את זכויותיו או לאכוף את החובות של המפעיל, כך ייתכן שתהיה הצדקה משמעותית יותר לדרישה זו. צריך בהקשר זה לקחת בחשבון את העדר היכולת לצפות באופן מלא את אופני השימוש על ידי הפרט במידע האמור. רביעית, יש לבחון את **ההשפעה של הגילוי על המפעיל**. ככל שהדרישה לגילוי תגרור פגיעה גדולה יותר באינטרסים העסקיים של המפעיל, ביעילות השימוש במערכת או בסודות מסחריים,

---

[220] שם, בעמ' 1133-1138.
[221] ראו למשל: ס' 4 לחוק הגנת הצרכן, התשמ"א-1981.
[222] ראו למשל: ס' 2א1 לחוק הבחירות (דרכי תעמולה), התשי"ט-1959.

יהיה צורך להביא היבט זה בחשבון גם כן. כך גם כאשר דרישת הגילוי מגדילה באופן משמעותי את עלויות העסקה או הנטל הרגולטורי באופן שאינו מוצדק מבחינת איזון האינטרסים.

5.5. אמינות, עמידות ובטיחות

שימושים רבים במערכות מבוססות בינה מלאכותית אינם מעוררים חשש משמעותי גם כשהמערכת לא מגיעה לתוצאות הרצויות. כך למשל, כשנעשה שימוש במערכות אלה על מנת לאפיין משתמשים ולהציע להם פרסומות שתואמות את העדפותיהם האישיות, שגיאת המערכת ככלל לא תעורר עניין חברתי או כלכלי ממשי. במקרים מעין אלו, כשהחשש משגיאה של המערכת לא משמעותי, ההצדקה להתערבות רגולטורית בנוגע למידת האמינות והעמידות של המערכת אינה גדולה, אם בכלל. לעומת זאת, כשהשימוש במערכת עלול, במקרה של התממשות כשל פנימי או איום חיצוני, לגרום לפגיעה בזכויות יסוד או באינטרסים ציבוריים – ובמיוחד כשקיים חשש לפגיעה בבטיחות המשתמשים במערכת או המושפעים ממנה – ישנה הצדקה משמעותית יותר להתערבות רגולטורית במטרה להגביר את מידת האמינות, העמידות, האבטחה והבטיחות של המערכת. זאת, בפרט בנסיבות בהן מתקיימות הצדקות מקובלות להתערבות רגולטורית, כדוגמת החצנות שליליות או פערי מידע.

מעבר לכך, ניתן להצביע על מספר גורמים שעשויים לחזק את ההצדקה להתערבות רגולטורית, בין היתר עקב הגדלת הסיכוי להתרחשותם של סיכונים בטיחותיים:

<u>ראשית</u>, ככל שהשימוש במערכת נעשה בקנה-מידה (Scale) רחב יותר, הן כשמשתמשים בה למגוון גדול יותר של משימות והן כשהיא משפיעה על מספר רב יותר של אנשים, כך שגיאה של המערכת עלולה לגרום לפגיעה משמעותית יותר. למשל, אם כל המכוניות העצמאיות של חברה מסוימת יעשו שימוש באותה מערכת, ויארע בה כשל, יגדל הסיכון לנזק רב שייגרם עקב תאונות.[223] במובן זה, קנה מידה רחב כשלעצמו עשוי לחזק את ההצדקה להתערבות רגולטורית על מנת לוודא שמערכת אמינה, עמידה ובטוחה, וזאת מאחר ששגיאות בהיקף רחב עלולות להיות בעלות משמעות חברתית או כלכלית ולהיתרגם לפגיעה בזכויות הפרט או באינטרסים ציבוריים, ובפרט בבטיחות.

<u>שנית</u>, כאשר נעשה שימוש במערכות שאינן ניתנות להסברה (המכונות ״קופסא שחורה״), גדל הסיכון להתרחשות אירועים בטיחותיים. במערכות אלו האפשרות לצפות מראש את כלל הסיכונים, לבחון את הסיכוי שיתממשו ולמנוע אותם מראש כבר בשלב פיתוח המערכת, מאתגרת יותר לעומת מערכות הניתנות להסברה. מעבר לכך, כאשר נעשה שימוש במערכות שאינן ניתנות להסברה, גם היכולת של גורם אנושי להתערב באופן אפקטיבי על מנת שיוכל לתקן שגיאות של המערכת מצטמצמת (להרחבה ראו פרק ״הסברתיות״ ופרק ״מעורבות אנושית״ לעיל).[224]

<u>שלישית</u>, כאשר נעשה שימוש במערכות שהצלחתן תלויה במהירות הפעולה שלהן, גובר החשש להתרחשות אירועים בטיחותיים, שכן ייתכן שמעורבות אנושית ב״זמן אמת״ איננה אפשרית. כך למשל, בתחום המסחר האלגוריתמי, המערכת נדרשת לפעול במהירות כדי שתהיה אפקטיבית. במקרים אלו, לגורם אנושי שיבקש להתערב בהחלטת המערכת לא תהיה אפשרות מעשית לנתח את ההחלטה מספק מהר ולתקן שגיאות במקרה של כשל בפעילות המערכת או איום עליה.[225]

---

[223] Arnold & Toner, לעיל ה״ש 187, בעמ׳ 17-18.
[224] שם, בעמ׳ 17.
[225] Thomas G. Dietterich, *Steps Toward Robust Artificial Intelligence*, 38 AI MAG. 3, 9 (2017).

יצוין כי בדומה לאמור בפרקים הקודמים, גם ביחס לסוגיית האמינות, העמידות, האבטחה והבטיחות של מערכות מבוססות בינה מלאכותית, היקף ההתערבות הרגולטורית (אם בכלל), ראוי להיבחן על פי מאפייני המערכת העומדת על הפרק ונסיבות השימוש בה. כאשר במסגרת הבחינה יש להביא בחשבון, בין היתר, את מידת החשיפה לכשלים פנימיים או איומים חיצוניים, ההסתברות להתממשותם וההשלכות הצפויות אם יתממשו; את קיומו ועוצמתו של תמריץ עצמאי, לרוב כלכלי, של מפתחי המערכת להתמודד עם כשלים ואיומים אלה; ואת העלויות הכרוכות ביישום הדרישות הרגולטוריות על מנת למתן את הסיכונים הנובעים מכשלים ואיומים אלה. בנוסף, יש לשקול מהו הכלי המתאים ביותר להתערבות רגולטורית ביחס לאותה מערכת ובנסיבות הנתונות.

לדוגמה, אמצעי מרכזי המצוי בידי רגולטורים, לשם התמודדות עם החשש בעניין כשלים פנימיים של מערכות מבוססות בינה מלאכותית ואיומים חיצוניים עליהן, הוא דרישה לביצוע הערכות תפקוד וסיכונים למערכות אלה (ראו להלן הצעה שעניינה "התוויית שפה אחידה באמצעות מסגרת מומלצת (וולונטרית) לניהול סיכונים"). מטבע הדברים, נודעת חשיבות לעיתוי ולתדירות ביצוע הערכות אלה, שיכולות להיעשות לאורך כל "מחזור החיים" של המערכת, דרך שלבי הפיתוח וההטמעה שלה ולאחר הטמעתה. כמו כן, ישנה חשיבות לאופן ביצוע הערכות אלה. כך למשל, ביחס לתהליך הערכת התפקוד של המערכת, בשלב פיתוח המערכת כדאי לבצע את ההערכה בסביבה שמדמה באופן הקרוב ביותר את הסביבה שבה תידרש המערכת לפעול לאחר כניסתה לשימוש. באופן דומה, הערכת התפקוד שתתבצע על מאגר נתונים חיצוני ושונה מזה שהמערכת התאמנה עליו, תאפשר לבחון כיצד היא מצליחה להתמודד עם מצבים לא מוכרים ובלתי צפויים.[226]

אמצעי נוסף העומד בפני הרגולטורים, הוא מעורבות אפשרית ביצירה של סטנדרטים טכנולוגיים ומקצועיים (best practices) למערכות מבוססות בינה מלאכותית, שיהיו מחייבים או מומלצים. סטנדרטים אלו יאפשרו מצד אחד להנגיש למפתחי המערכות את הכלים שבהם ניתן להשתמש כדי למתן את החשש משגיאה של המערכת, ומצד שני יאפשרו גם לקבוע סטנדרט מחייב שבכוחו לאזן בין החששות הכרוכים בשימוש במערכת לבין התועלות שניתן להשיג באמצעותה.[227]

אמצעי אחר, הוא קביעת דרישה רגולטורית, במקרים המתאימים, להסברתיות של המערכת או למעורבות אנושית בפעילות המערכת (להרחבה ראו פרקים "הסברתיות" ו-"מעורבות אנושית" לעיל). בין היתר, פיקוח מצד גורם אנושי על פעילות המערכת עשוי לאפשר תיקון שגיאות שלה ב"זמן אמת", בטרם ייפגעו זכויות הפרט או אינטרסים ציבוריים; והסברתיות של המערכת יכולה לסייע לזיהוי ותיקון כשלים פנימיים בפעילותה בשלב מוקדם יחסית, וכן לשפר את אפקטיביות המעורבות האנושית. כך שמעבר ליתרונות שצוינו לעיל, קביעת הדרישות הללו עשויה להועיל גם לחיזוק האמינות, העמידות, האבטחה והבטיחות של מערכות מבוססות בינה מלאכותית.

לבסוף, יצוין כי ניתן להפחית את מספר האירועים הבטיחותיים שייגרמו כתוצאה משימוש במערכות אלה, באמצעות מיסוד מנגנון שיתוף מידע על אודות אירועים בטיחותיים שהתרחשו או שנמנעו. במסגרת זו, התערבות רגולטורית יכולה להיות בקביעת דרישת שיתוף המידע ואופן

226 Hamon, Junklewitz & Sanchez, לעיל ה"ש 165, בעמ' 15.

227 Gary Marchant, *"Soft Law" Governance of Artificial Intelligence*, AI PULSE (2019), https://escholarship.org.

השיתוף (מתי, כיצד ועם מי ישותף), או במתן תמריצים לשיתוף פעולה מצד מפתחי המערכות והמשתמשים בהן (כגון על ידי הבטחת השמירה על מידע מסחרי רגיש של חברות המשתפות).[228]

5.6. <u>אחריותיות</u>

על רקע האתגרים בהחלת ההסדרים הקיימים על השימוש במערכות מבוססות בינה מלאכותית, עיון בספרות האקדמית מעלה מספר הצעות ביחס למשטר האחריות המשפטי שראוי להחיל, הן לגבי *דיני הנזיקין* והן ביחס *למשפט הפלילי*, בכל הנוגע למקרים שבהם השימוש במערכות אלה הביא לתוצאות שליליות. להלן יוצגו מספר חלופות מרכזיות שהוצעו למשטר האחריות ולהסדרים שדרכם ניתן לבחון את השימוש במערכות, תוך עמידה על היתרונות והחסרונות העיקריים של כל אחת מהאפשרויות האלה. כך גם, תתואר האפשרות להגביר את מידת האחריותיות באמצעות הטמעת נורמות ארגוניות חדשות המותאמות לשימוש במערכות מבוססות בינה מלאכותית.

א. משטרי האחריות והסדרים חלופיים אפשריים

<u>אפשרות אחת</u>, היא להמשיך לעשות שימוש במשטרי האחריות הקיימים, ובפרט בעוולות בדיני הנזיקין ובעבירות של רשלנות במשפט הפלילי, תוך ניסיון להתגבר על סוגיות ושאלות שמציב השימוש במערכות מבוססות בינה מלאכותית בפני יישום ההסדרים הקיימים.

<u>אפשרות נוספת</u>, היא קביעת הסדרים חדשים וייעודיים שיסדירו את האפשרות להטיל אחריות על רקע השימוש במערכות מבוססות בינה מלאכותית. בפרט, אחת האפשרויות המרכזיות ליצירת הסדר חדש היא להחיל במסגרתו משטר של אחריות מוחלטת או קפידה, שאינה תלויה באשם.

<u>אפשרות אחרת</u>, היא כי לצד ההחלה של משטר האחריות המועדף ביחס לשימוש במערכות מבוססות בינה מלאכותית, ייקבעו הסדרים חדשים הכוללים "נמלי מבטחים" (Safe Harbors) שיגנו באופן מוחלט מפני הטלת אחריות, או לחלופין, שיובילו לכך ששאלת האחריות תיבחן על בסיס משטר אחריות אחר (למשל תחת משטר אחריות של רשלנות במקום של אחריות חמורה).

ב. הטמעת נורמות ארגוניות להגברת האחריותיות

לצד עיצוב משטרי האחריות ועד לגיבוש הנורמות המשפטיות בצורה ודאית ויציבה, יש חשיבות לתהליכי הפנמה של הסיכון בידי הארגונים המפתחים והמשתמשים בבינה מלאכותית.

הטמעה של נורמות ארגוניות, בין אם באופן וולונטרי ובין אם כתוצאה מחובה רגולטורית (עוד בטרם קביעת היקף החובה המשפטית בדיעבד), תסייע לקידום ניהול הסיכונים וההגנה על זכויות יסוד ואינטרסים ציבוריים. זאת, במטרה להביא לכך שמפתחי המערכות והמשתמשים בהן יבינו כיצד מערכות אלה משליכות על זכויות יסוד ואינטרסים ציבוריים, ויוכלו לנקוט בצעדים הנדרשים על מנת למנוע או לצמצם את החשש שהשימוש בהן יגרום לתוצאות שליליות.

בכלל זה, לדוגמה, ניתן לעודד או לדרוש ממפתחי המערכות והמשתמשים בהן לבצע תהליכי הערכת השפעה או הערכת סיכון באופן עיתי, או למנות אחראיים ארגוניים שתפקידם להבטיח שהארגון נוקט באמצעים המתאימים כדי להתמודד עם האתגרים והסיכונים הנשקפים ממערכות אלה.

---

[228] Arnold & Toner, לעיל ה"ש 187, בעמ' 19.

5.7. פרטיות

בהתאם לדרישות הדין הישראלי ולעקרונות רגולציה מהעולם,[229] ועל רקע החשש לסיכונים לפרטיות, בחלק זה יתוארו אמצעים מקובלים להקטנת החשש לפגיעה בפרטיות. כמובן שאין מדובר ברשימה ממצה, וייתכנו אמצעים ומנגנונים נוספים להפחתת החשש לפגיעה בפרטיות.

ראשית, ובאשר לצורך בגיבוש הסכמה מדעת מקום בו נעשה שימוש במידע אישי, יצוין כי עמדת הרשות להגנת הפרטיות, בהתאם להמלצות שפרסמו גופים בינלאומיים,[230] היא כי כשאיסוף המידע נעשה באמצעות מערכות אוטומטיות (כגון "בוטים"), ובכללן מערכות מבוססות בינה מלאכותית, יש לוודא כי יוצגו לנושא המידע כל הפרטים הנדרשים לפי סעיף 11 לחוק הגנת הפרטיות. עוד הביעה הרשות עמדתה, כי כאשר עיבוד המידע נערך באמצעות מערכות אוטומטיות, במסגרת הליך היידוע יש לפרט על אופן פעולת המערכות האמורות, ככל שהדבר רלוונטי לגיבוש ההסכמה וככל שפירוט זה אפשרי מבחינה משפטית, טכנולוגית, ומסחרית. כן המליצה הרשות, כי יוסבר לנושא המידע על אודות פרטי המידע בהם עשויות המערכות להשתמש במסגרת השימוש במידע הנוגע אליו, והמקור של פרטי מידע אלו.[231] זאת ועוד, הרשות ציינה כי כשמערכת בינה מלאכותית מאפשרת להסיק על אדם מסקנות הנובעות מעיבוד מידע על אחרים בעלי מאפיינים דומים, יש ליידעו על כך.

המלצות נוספות שפורסמו על ידי גופים בינלאומיים, כוללים במסגרת חובת היידוע גם הודעה על כך שנושאי המידע מבצעים אינטראקציה עם בינה מלאכותית או מספקים מידע אישי אשר יעובד על ידי מערכת מסוג זה, ובמסגרת זו חלה החובה ליידע אותם מהן מטרות המערכת והשפעותיה האפשריות, מהו המידע המשמש את המערכת, ומה ההיגיון העומד בבסיס החלטותיה.[232] נדמה כי כדי ליתן מענה לאתגר הפרטיות יהיה מקום לקחת בחשבון אף היבטים אלה.

שנית, ובהתאם לעיקרון צמצום המידע, יש לפעול על מנת שהמטרות שלשמן יעובד המידע, יוגדרו בבירור, יתועדו ויתאימו לציפייה סבירה של נושאי המידע בנוגע לשימוש במידע הנוגע להם. בהקשר זה, יש מקום להביא בחשבון גם את הצורך בצמצום המידע האישי, המשמש את תהליכי פיתוח בינה מלאכותית, כך שהוא יהיה מוגבל למידע הרלוונטי והנדרש עבור המטרות שהוגדרו.

---

229 ראו למשל החלטת ה- Global Privacy Assembly, בו חברות רשויות להגנת המידע והפרטיות מרחבי העולם, בנושא בינה מלאכותית, משנת 2018. ההחלטה זמינה כאן.

230 ראו גם החלטת ה-Global Privacy Assembly בנושא בינה מלאכותית משנת 2020, ההחלטה זמינה כאן.

231 עמדת הרשות להגנת הפרטיות בעניין חובת יידוע במסגרת איסוף ושימוש במידע אישי, לעיל ה"ש 27, פס' 24.

232 ראו גם החלטת ה-Global Privacy Assembly בנושא בינה מלאכותית משנת 2020, לעיל ה"ש 230.

שלישית, יהיה מקום להביא בחשבון, במקרים מסוימים, גם את האפשרות לשימוש במנגנונים מגבירי פרטיות, אשר יסייעו בצמצום היקפי המידע האישי, כגון התממה (אנונימיזציה); שימוש במידע סינתטי (מידע אשר נוצר באופן מלאכותי וניתן להשתמש בו כתחליף למידע המקורי האמיתי); Federated learning (אימון המידע בכמה נקודות מקומיות, וחיבור התובנות למודל אחד, בלי לחלוק את המידע האישי);[233] הוספת "רעש";[234] Differential Privacy;[235] ועוד. שיטות נוספות להגברת פרטיות הרלוונטיות לשלב הסקת המסקנות לגבי אדם ספציפי הן עיבוד המידע בציוד הקצה של המשתמש ולא בשרת מרכזי, המרת המידע לקוד שאינו נקרא על ידי אדם ועוד. אף קביעת מדיניות לעניין מועדי מחיקת המידע המשמש לאימון המערכת, אשר תכלול קריטריונים לבחינת השאלה לשם מה נדרש המידע, האם שמירתו הכרחית, ואילו טכנולוגיות מגבירות פרטיות ניתן ליישם ביחס למידע אשר לא נמחק, תוכל לסייע בהתמודדות עם אתגר הפרטיות. יוזכר, כי בהתאם להוראת תקנה 2(ג) לתקנות הגנת הפרטיות (אבטחת מידע), התשע"ז-2017, כל בעל מאגר מחויב לבחון אחת לשנה האם הוא שומר במאגר מידע רב מן הנדרש למטרות המאגר.

לבסוף, ולצורך יישום ההיבטים שתוארו לעיל, יש מקום, במקרים המתאימים, לקבוע נהלים ומנגנוני יישום ופיקוח פנימיים על השימוש במידע אישי. במסגרת זאת, יוכלו הרגולטורים או הרשות להגנת הפרטיות לקבוע הוראות מתאימות בדבר מינוי ממונה הגנה על פרטיות בארגון,[236] כמו גם על ביצוע הכשרות מתאימות וכן על קביעת מנגנוני בקרה ופיקוח על שימוש במידע אישי. מעבר לכך, ייתכן כי במקרים המתאימים, יהיה ניתן לעשות שימוש בתסקיר השפעה על פרטיות בארגון לשם בחינת מידת הנחיצות של השימוש באמצעים שתוארו לעיל.[237] לא כל שימוש במערכת מבוססת בינה מלאכותית מעורר את אותה מידה של חשש בנוגע לשמירה על הפרטיות של נושאי המידע; בהתאם, הרגולטורים הרלוונטיים והרשות להגנת הפרטיות יוכלו לבחון באופן מעמיק ומפורט, ובשים לב לאופי המערכת ומידת הסיכון והרגישות שיש בה, מהם האמצעים המתאימים בנסיבות העניין, לרבות אלה שתוארו לעיל, לצורך הבטחת ההגנה הנדרשת לזכות לפרטיות.

---

[233] כגון טכניקת למידת מכונה המאמנת אלגוריתם על פני מספר התקני קצה מבוזרים, או שרתים המחזיקים בדגימות נתונים מקומיות מבלי להחליף אותן.

[234] קרי, שינוי חלק מהמידע האישי באופן רנדומלי, תוך שימור ערכים סטטיסטיים במערך הנתונים,

[235] שיתוף מידע פומבי על מערך נתונים על ידי תיאור דפוסי הקבוצות בתוך מערך הנתונים, מבלי לשתף מידע על פרטים במערך הנתונים עצמו.

[236] ראו המלצות הרשות להגנת הפרטיות בעניין "מינוי ממונה הגנה על הפרטיות בארגון ותפקידיו" (ינואר 2022). המסמך זמין כאן.

[237] ראו מדריך עזר מתודולוגי לעריכת תסקיר השפעה על הפרטיות שפרסמה הרשות להגנת הפרטיות (אוגוסט 2021). המדריך זמין כאן.

6. חלק שישי: עקרונות מוצעים למדיניות רגולציה ואתיקה לתחום הבינה המלאכותית

כאמור לעיל, במסגרת החלטת ממשלה 212 הטילה הממשלה על שרת החדשנות, המדע והטכנולוגיה להוביל את מדיניות הממשלה בתחום הבינה המלאכותית. בהתאם להחלטה זו מגבש משרד החדשנות, המדע והטכנולוגיה את מדיניות הממשלה בתחום הבינה המלאכותית שמטרתה קידום המובילות הטכנולוגית והמדעית של מדינת ישראל.

פרק זה מפרט עקרונות מוצעים למדיניות אתיקה ורגולציה בתחום הבינה המלאכותית בישראל, תוך התחשבות בעקרונות משפטים ואתיים מקובלים בעולם. לצורך גיבוש עקרונות אלה נסקרו המאפיינים הייחודים של טכנולוגיית הבינה המלאכותית, וכן האתגרים והסוגיות המתעוררים בעקבות פיתוח ושימוש במערכות מבוססות בינה מלאכותית, והאופן בו מדינות מפותחות עוסקות בתחום זה. המדיניות המוצעת מבקשת לאזן בין הצורך בוודאות ובהירות לבין האינטרסים הציבוריים והזכויות שעל הפרק, והחשש מהתערבות שאינה מוצדקת העלולה לפגוע בחדשנות.

<u>להלן עיקרי שבעת העקרונות המוצעים שיפורטו בפרק זה</u>:

(1) אימוץ מדיניות רגולציה לתחום הבינה המלאכותית

מוצע לאמץ מדיניות רגולציה שתכלול את הרכיבים שיפורטו להלן, ותסייע לממשלה בבחינת הצורך בהתערבות רגולטורית ביחס לפיתוח ושימוש בבינה מלאכותית ואופי ההתערבות:

א. אין בשלב זה מקום לקידום חקיקת מסגרת רוחבית על כל תחום הבינה המלאכותית; במקום זאת, במידת הצורך, על הרגולטורים הסקטוריאליים לבחון את הצורך בקידום רגולציה קונקרטית בתחומם, תוך פעולה לאור מדיניות ממשלתית אחידה.

ב. חשיבות לרגולציה התואמת לנעשה במדינות מפותחות וארגונים בינלאומיים.

ג. החלת רגולציה המבוססת על כלים ומסגרות לניהול סיכונים.

ד. בחינת השימוש ב"רגולציה רכה" וכלי רגולציה מתקדמים (דוגמת עקרונות אתיים, תקינה וולונטרית ורגולציה עצמית).

ה. פיתוח הרגולציה באופן מודולרי ועידוד נסיינות רגולטורית, תוך שמירה על ניטרליות טכנולוגית.

ו. עיצוב הרגולציה תוך שיתוף הציבור ובכלל זה התעשייה, האקדמיה וארגוני חברה אזרחית.

(2) אימוץ עקרונות אתיים לתחום הבינה המלאכותית

מוצע לאמץ עקרונות אתיים משותפים עבור תחום הבינה המלאכותית, על בסיס עקרונות ה-OECD שנסקרו לעיל, כדי לסייע לרגולטורים ולארגונים בתחום זה. מוצע כי העקרונות לא יהיו בעלי מעמד משפטי מחייב, אך ישמשו בסיס לשיח משותף ותיאום לפי ההקשר הנורמטיבי והרגולטורי.

<u>העקרונות המוצעים הם</u>:

א. בינה מלאכותית לקידום צמיחה, פיתוח בר-קיימא ומובילות ישראלית בחדשנות.

ב. האדם במרכז – כיבוד זכויות יסוד ואינטרסים ציבוריים.

ג. שוויון ומניעת אפליה פסולה.

ד. שקיפות והסברתיות.

ה. אמינות, עמידות, אבטחה ובטיחות.

ו. אחריותיות.

(3) מיסוד מוקד ידע ותיאום ממשלתי להסדרת בינה מלאכותית

מוצע להפקיד בידי גורם ממשלתי אחד את ריכוז ותיאום סוגיית ההסדרה של בינה מלאכותית. ישנה חשיבות לכך שגורם זה יפתח מומחיות מקצועית וטכנולוגית, ויהיה בעל ראייה רוחבית ואחריות לקידום האינטרס הכלל-משקי בעידוד הטכנולוגיה והשגת היעדים הממשלתיים.

גורם זה יעסוק ביישום מדיניות רגולציה ואתיקה זו, וגיבוש המלצות לעדכונה לפי הצורך; בייעוץ למשרדי הממשלה ולרגולטורים בגיבוש מדיניות ורגולציה לגבי בינה מלאכותית; ובהנגשת מידע וכלים לשימוש אחראי בבינה מלאכותית, ובכלל זה כלי ניהול סיכונים, עבור הממשלה והציבור.

(4) הקמת פורום רגולטורים ופורום לשיתוף ציבור לתחום הבינה המלאכותית

מוצע למסד פורום מקצועי פנים-ממשלתי הכולל נציגי רגולטורים ומומחי טכנולוגיה, מדיניות ומשפט, על מנת לקדם את התיאום ולדון בסוגיות משותפות; וכן פורום הכולל נציגי תעשייה, אקדמיה, ארגוני חברה אזרחית והציבור הרחב, כדי לדון בסוגיות של הסדרת בינה מלאכותית.

(5) מיפוי השימושים בבינה מלאכותית והאתגרים הנלווים להם בענפים מוסדרים

מוצע כי גורמי הממשלה והרגולטורים הרלוונטיים יפעלו להבנה ומיפוי של השימושים הקונקרטיים במערכות מבוססות בינה מלאכותית הנעשים על ידי המפוקחים בענף המוסדר שעליו הם אמונים; האתגרים, החששות והסיכונים הכרוכים בכך; והמענים האפשריים.

(6) מעורבות אקטיבית בפיתוח הרגולציה והתקינה בפורומים בינלאומיים

מוצע כי גורמי ממשלה הפועלים במסגרת פורומים וארגונים בין-לאומיים, או עומדים בקשרי עבודה עם מדינות מפותחות אחרות בתחום זה יפעלו יחד לקידום מדיניות רגולציה ואתיקה מאוזנת ותואמת למדיניות רגולציה זו וליעדים הממשלתיים. זאת, בשים לב להחלטת ממשלה 212, שהטילה על שרת החדשנות, המדע והטכנולוגיה "להוביל את שיתופי הפעולה הבין-לאומיים האזרחיים בתחום, לייצג את ישראל בפורומים בין-לאומיים אזרחיים, לרבות פורום השרים של ארגון ה-OECD".

(7) התוויית שפה אחידה באמצעות מסגרת מומלצת (וולונטרית) לניהול סיכונים

מוצע לפתח או לאמץ כלי אחיד לניהול סיכונים ביחס לשימוש בבינה מלאכותית, שייצור שפה משותפת בין גורמי הממשלה והרגולטורים ובינם לבין גורמים פרטיים. השפה האחידה תסייע למגזר הפרטי להעריך את הסיכונים הנלווים לשימוש מסוים בבינה מלאכותית, וכן לרגולטורים לבחון את הסיכונים הכרוכים בשימוש בה בתחום שעליו הם אמונים ואת הצורך בהתערבות.

6.1. אימוץ מדיניות רגולציה לתחום הבינה המלאכותית

מוצע לאמץ מדיניות רגולציה שתכלול את הרכיבים שיפורטו להלן, ותסייע לממשלה בבחינת הצורך בהתערבות רגולטורית ביחס לפיתוח ושימוש בבינה מלאכותית ואופי ההתערבות.

א. אין בשלב זה מקום לקידום חקיקת מסגרת רוחבית על כל תחום הבינה המלאכותית; במקום זאת, במידת הצורך, על הרגולטורים הסקטוריאליים לבחון את הצורך בקידום רגולציה קונקרטית בתחומם, תוך פעולה לאור מדיניות ממשלתית אחידה.

ב. חשיבות לרגולציה התואמת לנעשה במדינות מפותחות וארגונים בינלאומיים.

ג. החלת רגולציה המבוססת על כלים ומסגרות לניהול סיכונים.

ד. בחינת השימוש ב"רגולציה רכה" וכלי רגולציה מתקדמים (דוגמת עקרונות אתיים, תקינה וולונטרית ורגולציה עצמית).

ה. פיתוח הרגולציה באופן מודולרי ועידוד נסיינות רגולטורית, תוך שמירה על ניטרליות טכנולוגית.

ו. עיצוב הרגולציה תוך שיתוף הציבור ובכלל זה התעשייה, האקדמיה וארגוני חברה אזרחית.

6.1.1. החשיבות של מדיניות רגולציה ומעמדה

נקודת המוצא להצעה זו, היא שישנה חשיבות רבה לכך שרגולציה המשפיעה על פיתוח בינה מלאכותית והשימוש בה, תקודם או תותאם לפי מדיניות רגולציה ממשלתית אחידה וקוהרנטית. זאת, על מנת להשיג את יעדי המדיניות של הממשלה, לקדם את תחום הבינה המלאכותית, להגן על זכויות יסוד ואינטרסים ציבוריים, ולצמצם את החשש לפגיעה בחדשנות הטכנולוגית.

בהתאם לכך, מוצע לאמץ מדיניות רגולציה, ולכוון את גורמי הממשלה ובפרט הרגולטורים הרלוונטיים להתחשב בה במסגרת פעילותם בקביעת רגולציה ביחס לבינה מלאכותית. למשל, מוצע כי רגולטור ממשלתי השוקל להחיל רגולציה על שימוש מסוים בבינה מלאכותית בתחום שעליו הוא אמון, יביא בחשבון את החשיבות שבגיבוש אותה רגולציה בשים לב לנעשה במדינות המפותחות, יבצע הליך שיתוף ציבור משמעותי, ויעדיף, במקרים המתאימים, להשתמש בכלי רגולציה מתקדמים ובכלי נסיינות רגולטורית (דוגמת "ארגז החול" הרגולטורי (Regulatory Sandbox)).

בנוסף לכך, מוצע לאמץ את רכיבי מדיניות הרגולציה המפורטים להלן, שגובשו בין היתר בשים לב ל"עקרונות מנחים לאסדרה מיטבית" הקבועים בחוק עקרונות האסדרה, התשפ"ב-2021; למסמכי מדיניות רגולציה שפרסמו המדינות המובילות בעולם בתחום ההסדרה של בינה מלאכותית; להמלצות שניתנו ולעבודות קודמות שנעשו במסגרת הפעילות הממשלתית בתחום; ולדיוני הצוות הבין-משרדי שעסק ברגולציה ואתיקה במהלך גיבוש התכנית הלאומית לבינה מלאכותית.

רכיבי מדיניות הרגולציה המוצעים להלן עומדים כל אחד בפני עצמו, אך מתחברים יחד לכדי מדיניות שלמה. מטבע הדברים, לא כל הרכיבים ניתנים ליישום ביחס לכל רגולציה בתחום הבינה המלאכותית, וההתחשבות בהם ראויה להיות תמיד על פי ההקשר והנסיבות ובשים לב לשיטות העבודה המקובלות, אך מוצע כי ככלל יהיה מקום להתחשב בהם ולשאוף ליישמם.

6.1.2. רכיבי המדיניות המוצעים לאימוץ

א. אין בשלב זה מקום לקידום חקיקת מסגרת המשתרעת על כל תחום הבינה המלאכותית; במקום זאת, במידת הצורך, על הרגולטורים הסקטוריאליים לבחון את הצורך בקידום רגולציה קונקרטית בתחומם, תוך פעולה לאור מדיניות ממשלתית אחידה.

במסגרת רכיב זה במדיניות הרגולציה מוצע כי בשלב זה ככל שתקודם רגולציה שמטרתה להסדיר פיתוח בינה מלאכותית או שימוש בה, הדבר ייעשה ברמת הענף שבו היא נדרשת על בסיס צורך קונקרטי ובהובלת הרגולטור האמון עליו, תוך פעולה לאור מדיניות ממשלתית אחידה באמצעות תיאום, אך לא באמצעות חקיקה רוחבית ייעודית להסדרת תחום הבינה המלאכותית.

תחום הבינה המלאכותית רחב ומתייחס לטכנולוגיות שונות, להן שימושים מגוונים, המופיעות כמעט בכל ענף במשק, ובכלל זה בריאות, תחבורה, פיננסים, חקלאות וחינוך. נוכח הבסיס המשותף של הבינה המלאכותית מצד אחד, וגיוון השימושים מצד שני, שאלה מרכזית ביחס להסדרה של בינה מלאכותית היא האם יש להסדיר פיתוח ושימוש בבינה מלאכותית (ככל שיש צורך בכך) באמצעות רגולציה כלל-משקית אחת, או באמצעות רגולציות ספציפיות לכל ענף בנפרד.

אפשרות אחת היא רגולציה רוחבית וכללית שתשמש כמסגרת רגולטורית המסדירה פיתוח ושימוש בבינה מלאכותית בכלל הענפים, באופן הדומה לתחולת הרגולציה בתחום התחרות, הפרטיות והגנת הצרכן. אפשרות אחרת היא רגולציה ענפית, שתחול רק ביחס למגזר או לשוק מסוים שאותו היא נועדה להסדיר באופן ספציפי, כגון בריאות או בנקאות.[238] להמחשה, ההבדל יכול להיות בין קידום של חוק בינה מלאכותית שיכלול הוראות כלליות שיחולו על כל המשק, לבין קביעת כללים פרטניים על ידי רגולטור מסוים לגבי שימושים בבינה מלאכותית שבתחומו.

כמפורט בפרק "פעילות ארגונים בינלאומיים ומדינות מפותחות" לעיל, **האיחוד האירופי** מקדם טיוטה של חוק הבינה המלאכותית (Artificial intelligence act), המציעה מסגרת רגולטורית רוחבית ראשונה מסוגה לבינה מלאכותית. הטיוטה קובעת מסגרת משפטית שתחול על פיתוח ושימוש בכל המערכות המבוססות על בינה מלאכותית באיחוד האירופי (להרחבה ראו סקירת טיוטת החקיקה של האיחוד האירופי לעיל). לעומת זאת, מדינות מובילות אחרות, ובהן **ארה"ב**, **יפן**[239] **וסינגפור**[240], נוקטות בגישה שונה, שבבסיסה איתור הסיכונים וההתמודדות איתם ברמת הארגון, באמצעות שפה משותפת. שפה זו מאפשרת בירור וליבון מעמיקים לגבי הצורך והאופי של מעורבות רגולטורית, אך לא מקודמת לצידה רגולציה רוחבית לתחום.

מדינה מובילה נוספת שאימצה מדיניות של רגולציה ענפית לתחום הבינה המלאכותית היא **בריטניה**. כאמור בסקירה לעיל, בטיוטת מסמך מדיניות הרגולציה שפרסמה לאחרונה, העדיפה ממשלת בריטניה בצורה ברורה שימוש ברגולציה ענפית, וזאת לאור היתרונות שראתה בכך.

---

[238] יצוין שיש גם מי שהציעו רגולציה שתהיה רוחבית יותר, ותסדיר באופן כללי את התחום הטכנולוגי. כלומר, רגולציה שתסדיר חדשנות טכנולוגית, אולם על פניו נראה שמדובר ביריעה רחבה מידי.

[239] Ministry of Economy, Trade and Industry, Consumer Affairs Agency, Expert Group on how AI Principles Should be Implemented, AI Governance in Japan Ver. 1.1., https://www.meti.go.jp; Government of Japan, Ministry of Economy, Trade and Industry, Study Group on New Governance Models in Society, Governance Innovation ver.2 - A Guide to Designing and Implementing Agile Governance, 2021, https://www.meti.go.jp.

[240] Personal Data Protection Commission Singapore, Singapore's Approach to AI Governance, https://www.pdpc.gov.sg.

התפיסה הבריטית מבוססת על יצירת עקרונות משותפים לבינה מלאכותית, אולם יישום בהתאם לשימוש ותוך הסתמכות על הידע והמומחיות של גופי הרגולציה הסקטוריאליים.

אמנם, לקידום רגולציה רוחבית יתרונות בתחום האחידות והקוהרנטיות בהתמודדות עם טכנולוגיות בינה מלאכותית. במקרים רבים האתגרים הבסיסיים הכרוכים בבינה מלאכותית זהים או דומים, ללא קשר לענף שבו היא מוטמעת. הסדרת אתגרים אלה באמצעות רגולציות שונות ומקבילות עלולה ליצור אי-אחידות, פערים, סתירות וחוסר ודאות. זאת ועוד, סטנדרטים אחידים מאפשרים שפה משותפת בין גורמי הממשלה, גורמים פרטיים, ואף בין המערכות, בענפים שונים.

ואולם, השימושים בבינה מלאכותית מגוונים ונעשים ככלל בהקשר מסוים, כך שלא בנקל ניתן לקבוע רגולציה כוללנית שתתאים באופן מספק למאפיינים של כל שימוש ותטפל בכולם. בנוסף, נוכח ריבוי השימושים האפשריים בבינה מלאכותית, ממילא עשוי להתעורר צורך להתייחס לכל שימוש ותחום פעילות על בסיס מאפייניו ומכלול נסיבותיו במסגרת הרגולציה הרוחבית. כמו כן, גיבוש רגולציה רוחבית יהיה כרוך בהליך ממושך (בפרט כשיש צורך לעגן אותה בחקיקה), מה שעלול להביא לכך שבמועד סיומו התוצאה לא תענה לצרכים המקוריים, או תגרום לקיבוע תוצאות המשקפות את המציאות הטכנולוגית והחברתית הקיימת בשלב מסוים, ולהתיישן עם שינוי המציאות הטכנולוגית והחברתית בכל ענף. לבסוף, בעת הזו נראה כי מוטב להרחיב את התשתית העובדתית והמקצועית בעניין מלאכת הרגולציה של פיתוח ושימוש בבינה מלאכותית בישראל.

גם בדו"ח בינה מלאכותית בשירותים פיננסיים עסקו החוקרים בהבחנה בין רגולציה ענפית ורוחבית. הם הביעו עמדתם שלפיה "בשלב זה החסרונות של קביעת עקרונות אחידים שיחולו בתחומי משפט שונים עולים על היתרונות", ובהתאם המליצו כי "בשלב זה אין לקבוע עקרונות אחידים" ביחס לכל השימושים בבינה מלאכותית. לצד זאת, החוקרים הצביעו על "צורך בתיאום" כדי להתמודד באופן רגולטורי מתואם עם סוגיות בעלות מאפיינים רוחביים. בנוסף, הם המליצו לבחון את הרגולציה הפיננסית – שעמדה במוקד מחקרם – כמקשה אחת, עקב הקשר בין התחומים השונים הנכללים בה (בנקאות, ניירות ערך ושוק ההון), וכן הציעו להפקיד בידי גורם ממשלתי אחד את "תיאום סוגיית האסדרה של יישומי בינה מלאכותית בתחומים שונים" (להרחבה ראו הצעה בעניין "מיסוד מוקד ידע ותיאום ממשלתי להסדרת בינה מלאכותית" להלן).

ב. <u>חשיבות לרגולציה התואמת לנעשה במדינות מפותחות וארגונים בינלאומיים</u>

במסגרת רכיב זה במדיניות הרגולציה מוצע כי הרגולציה ביחס לפיתוח ושימוש בבינה מלאכותית, ככל שתאומץ, תביא בחשבון את הרגולציה הנוהגת במדינות מובילות בתחום ושאומצה על ידי ארגונים בינלאומיים, ותתאים עצמה במידת האפשר לרגולציה רלוונטית שנקבעה במדינות ובארגונים כאמור. זאת, הן על מנת למנוע פגיעה בזכויות יסוד ואינטרסים ציבוריים בעקבות רגולציה שאינה מעודכנת, הן כדי למנוע חסמים רגולטוריים ייחודיים בפני כניסת טכנולוגיה לישראל, והן כדי לתמוך ביכולתה של התעשיה הישראלית לייצא למדינות היעד העיקריות.

ארגונים בינלאומיים ומדינות רבות בעולם עוסקים בקידום עקרונות משפטיים ורגולציה לתחום הבינה המלאכותית. בין הארגונים נמנים לדוגמה ה-OECD, UNESCO, CAI (ועדה של מועצת אירופה), GPAI והאיחוד האירופי. עם המדינות המובילות נמנות למשל ארה״ב, בריטניה, קנדה, יפן וסינגפור. בנוסף, נעשית עבודה נמרצת בארגוני תקינה בינלאומית כמו ISO ו-NIST לפיתוח תקנים ל״בינה מלאכותית אחראית״. אלה, ורבים אחרים, עוסקים בהסדרת בינה מלאכותית, ויש להניח כי לתוצרי העבודה שלהם תהיה השפעה גלובלית, כולל על הנעשה בישראל. כבר היום, תוצרים של ארגונים ומדינות אלה מסייעים רבות לגיבוש המדיניות שמופיעה במסמך זה.

בשים לב לגודלם היחסי של שווקי ישראל, ולכך שטכנולוגיות שמפותחות בישראל מכוונות פעמים רבות בעיקר לשווקים הגלובליים, יש חשש כי יצירה של רגולציה ייחודית לישראל עבור תחום הבינה המלאכותית, עלולה, בין היתר, להיתרגם לחסם רגולטורי ייחודי בפני כניסת מערכות מבוססות בינה מלאכותית לישראל, ולהפסד התועלות הכלכליות והחברתיות מכך.

לפיכך, למעט במקרים חריגים, מוצע להימנע מרגולציה ייחודית בקשר לבינה מלאכותית, ולהתאים את הרגולציה בישראל לזו שאימצו ויאמצו המדינות המובילות בתחום והארגונים הבינלאומיים ככל האפשר. ניתן לצפות כי עשויות להתפתח גישות שונות, כפי שניכר כבר היום ביחס לשאלת הצורך בחקיקת מסגרת. במקרים כאלו, יהיה צורך לבחור בין גישות מקובלות שונות מהעולם, בהתחשב במכלול היתרונות והחסרונות של כל אחת. בדרך זו ניתן לוודא שהרגולציה בישראל בתחום הבינה המלאכותית תהיה תואמת, ככל הניתן, לתוכן הרגולציה המקובלת בשווקים הגלובליים (כגון התקנים המתפתחים), אף אם ההסדרים המשפטיים ואופן מימושם יהיה שונה.

הדבר עולה בקנה אחד גם עם המגמה ההולכת ומתגברת בשנים האחרונות, להתאמת הרגולציה בישראל לרגולציה המקובלת בעולם. ביטוי מרכזי למגמה זו מופיע בחוק עקרונות האסדרה הקובע כעקרון מנחה לאסדרה מיטבית כי רצוי שרגולציה תיקבע ככלל על בסיס כללים ואמות מידה מקצועיים שגובשו בארגונים בינלאומיים, או שחלים במדינות מפותחות עם שווקים משמעותיים. ביטוי נוסף למגמה זו מצוי ב״רפורמת היבוא״ שעברה במסגרת חוק ההסדרים בשנת 2021 ונועדה להתאים את הרגולציה על היבוא בישראל לרגולציה זרה ובכך להגביר תחרות ולהוזיל מחירים.

גם ועדת המשנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית בראשות פרופ׳ קרין נהון (ראו פרק ״העיסוק הממשלתי בבינה מלאכותית״ לעיל) המליצה על התאמת הרגולציה בתחום הבינה המלאכותית לנעשה במדינות מפותחות, ובלשון דו״ח הוועדה: ״הוועדה רואה חשיבות להתאים את האסדרה הנבחרת לחקיקה, למדיניות ולתקינה מקובלות במדינות המפותחות, על מנת שמדינת ישראל תוכל להישאר בחוד החנית של התחום״.

ג. <u>החלת רגולציה המבוססת על כלים ומסגרות לניהול סיכונים</u>

רכיב זה במדיניות הרגולציה מציע כי רגולציה המסדירה פיתוח ושימוש בבינה מלאכותית, תהיה מותאמת ככלל לסיכונים הנשקפים מסוג הטכנולוגיה ומהשימוש הספציפי שאותו היא נועדה להסדיר, ותהיה תוצר של ניהול סיכונים שביצע הרגולטור ביחס אליו. כך שככלל לא תחול באופן אחיד על טכנולוגיות ושימושים שרמת הסיכונים והחששות לגביהם שונה באופן ניכר.

לבינה המלאכותית מגוון רחב של שימושים, שאינם שווים זה לזה במידת הסיכון והרגישות הקשורה בהם (למשל, על פניו נראה כי יש הבדל בין שימוש בבינה מלאכותית כדי להציע לצרכן סרט או שיר לטעמו, לבין שימוש בה לאבחון רפואי או החלטת אשראי). על כן, ייתכן בהחלט שגם בתוך ענף רגולטורי מסוים, יהיה שוני במידת ההצדקה להתערבות רגולטורית בין שימוש אחד לאחר, הנובע מכך ששימוש אחד מעורר חששות וסיכונים גדולים יותר. בהתאם לכך, מוצע להבחין בין רמת הסיכון המשתנה של השימושים השונים במסגרת הרגולציה, ולגבשה באופן שיביא לידי ביטוי ביצוע הליך כלשהו לניהול סיכונים על ידי הרגולטור (ראו גם עקרון מוצע מס׳ 7 להלן).

יצוין כי באופן כללי תפיסות ניהול סיכונים בתחום הבינה המלאכותית באות לידי ביטוי בשני היבטים. <u>היבט ראשון</u> הוא ניהול הסיכונים על ידי הממשלה והרגולטורים כחלק ממדיניות הרגולציה. כך, לפי תפיסת ניהול סיכונים גוברת ההצדקה להתערבות רגולטורית בפיתוח ובשימוש בבינה מלאכותית, מקום שבו הסיכון לפגיעה בזכויות יסוד או אינטרסים ציבוריים גבוה. בהתאמה, ככל שהסיכון נמוך יותר, ההצדקה להתערבות מצטמצמת, וכללי ההתנהגות עשויים להיות בגדר המלצה בלבד. <u>היבט שני</u> הוא ניהול הסיכונים ברמת הארגון המפתח או משתמש בבינה מלאכותית. תפיסת ניהול סיכונים זו מבוססת על תפיסות מקובלות לניהול סיכונים, בין היתר בהתאם לעקרונות האתיים (ראו הצעה להלן), כדי לאתר את הסיכונים הנובעים מבינה מלאכותית לזכויות יסוד ואינטרסים ציבורים, ולנקוט באמצעים טכנולוגיים ונהליים לצמצם אותם לרמה מקובלת.

הממשק בין שני היבטים אלה, הוא שתפיסת ניהול הסיכונים של מדיניות הרגולציה מסייעת להפעלת סמכות הרגולטור בין היתר כדי לבחון האם ההתמודדות הארגונית הוולונטרית מספקת מענה הולם לסיכון לפגיעה בזכויות יסוד או באינטרסים ציבורים (ראו גם עקרון מוצע מס׳ 7 להלן).

טיוטת הרגולציה של האיחוד האירופי מציעה לקבוע סיווג למערכות בינה מלאכותית עם דרישות וחובות שונות המותאמות על פי ״גישה מבוססת סיכונים״. לפי גישה זו, יישומי בינה מלאכותית יוסדרו רק בהתאם לנדרש כדי לטפל ברמות סיכון קונקרטיות. הטיוטה מבחינה בין מערכות ב״סיכון בלתי מקובל״, ״סיכון גבוה״, ״סיכון מוגבל״, ו״סיכון מינימלי״, ומגדירה את התהליך לסיווג המערכות לתוך קטגוריות אלה. לשם ההמחשה, ב״סיכון בלתי מקובל״ נכללות מערכות המשמשות רשויות ציבוריות ל״דירוג חברתי״ (social scoring)[241] ומערכות שמשתמשות בטכניקות תת-הכרתיות מניפולטיביות; וב״סיכון גבוה״ נכללות מערכות לזיהוי ביומטרי, חינוך, תעסוקה ומערכות המשמשות לגישה לשירותים פרטיים חיוניים. בהתאם לסיווג המערכות, לפי רמות

[241] ״דירוג חברתי״ (״social scoring״) עוסק בקביעת ציון לאדם על בסיס התנהגותו בשורה של פרמטרים ותחומים, תוך פיתוח הרעיון של ״דירוג אשראי״ למשל, הקובע את רמת סיכון או מהימנות האשראי של אדם, לדירוג רחב יותר. ראו גם הערת שוליים 265 לעיל המפנה לגילוי דעת מטעם הרשות להגנת הפרטיות. הרשות להגנת הפרטיות, דירוג חברתי בראי הזכות לפרטיות:
סקירת רקע בעניין שימוש במערכות לדירוג חברתי, 21.04.21, https://www.gov.il/he/departments/publications/reports/social_ranking.

הסיכון, מוטלות חובות שונות על מפתחי המערכות והמשתמשים בהן. מאיסור על שימוש במערכת, דרך החלה של דרישות רגולטוריות לשם שימוש במערכת, ועד שימוש (כמעט) חופשי במערכת.

גם ועדת המשנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית הציעה חלוקה בין אזורי סיכון בינוני וגבוה (בהם נדרשת לשיטתה חקיקה); אזורי סיכון בינוני ונמוך (בהם תידרש תקינה); ואזורי סיכון נמוך (שלא דורשים מגבלה משפטית). בנוסף, הוועדה הציעה כלי שפיתחה, הכולל בין היתר קבוצה של שאלות ראשוניות לבחינת ההשפעה של מערכת מבוססת בינה מלאכותית, ובכלל זה מהי עוצמת הפגיעה הפוטנציאלית בפרט או בציבור, ומהי עוצמת הפגיעה הצפויה אם יעשו שימוש לרעה במוצר או שהוא יצא משליטה.

לבסוף יצוין, כי גם חוק עקרונות האסדרה מורה על התאמת הרגולציה לסיכונים. בהקשר זה, החוק קובע כי רגולציה מיטבית היא כזו הנקבעת במטרה להביא למירב התועלת לחברה ולמשק, תוך שקילת האינטרס המוגן וההשפעות הנובעות מקביעת הרגולציה או אי-קביעתה, עלות הציות לה, ו"ככלל, על בסיס ניהול סיכונים". בדברי ההסבר לחוק, הצורך בניהול הסיכונים הובהר כך: "מוצע כי אסדרה מיטבית תיקבע, ככלל, לפי עקרונות של ניהול סיכונים. זאת, מתוך הבנה שככלל, אין זה רצוי מבחינה כלכלית וחברתית לפעול באמצעות אסדרה כדי לאיין סיכונים באופן מוחלט".

ד. <u>בחינת השימוש ברגולציה "רכה" וכלי רגולציה מתקדמים</u>

לאור השלב הראשוני שבו מצויות הטכנולוגיה והרגולציה בתחום הבינה המלאכותית, ועל מנת לעודד את הפיתוח והשימוש בה ולהפיק את התועלות החברתיות והכלכליות הנובעות מכך, מוצע כי במקרים המתאימים ייעשה שימוש ברגולציה המאפשרת השגת מטרות אלה. בכלל זה, מוצע לבחון שימוש בכלי רגולציה "רכים" ומתקדמים, כגון עקרונות אתיים לתחום הבינה המלאכותית, תקינה, המלצות רגולטור לאימוץ וולונטרי ורגולציה עצמית (מפוקחת או בלתי מפוקחת).

ככלל, כאשר רגולטור ניצב בפני טכנולוגיה חדשה המשנה מהמציאות הרגולטורית שנהגה בשוק לפני כניסתה, ובענייננו בינה מלאכותית, עומדות בפניו מספר גישות אפשריות מרכזיות לתגובה.[242] <u>אפשרות אחת</u> היא חסימה (Block) של הטכנולוגיה החדשה, באמצעות קביעת כללים משפטיים האוסרים על כניסת הטכנולוגיה החדשה, או באמצעות מתן פרשנות לכללים משפטיים קיימים האוסרת את השימוש בה. <u>אפשרות שנייה</u> והפוכה, היא אי-התערבות (Free-Pass) בכניסת הטכנולוגיה החדשה, לאחר שנמצא שאין קושי או אינטרס פיקוחי המצדיק התערבות. <u>אפשרות שלישית</u>, היא יישום רגולציה קיימת (Old-Regs), דהיינו לאפשר את כניסת הטכנולוגיה החדשה, וליישם לגביה את הכללים המשפטיים הקיימים כך שיסדירו אותה בדומה למצב שלפני כניסתה. <u>אפשרות רביעית</u>, היא קביעת רגולציה חדשה (New-Regs), כלומר פיתוח מסגרת רגולטורית חדשה ומעטפת משפטית ייעודית שתעגן אותה, כדי להסדיר את כניסת הטכנולוגיה החדשה.

בעשורים האחרונים, ועם התפתחות עולם מדיניות הרגולציה באופן כללי וביחס לרגולציה של טכנולוגיה בפרט, התפתחו שיטות רגולציה המשלבות בין אפשרויות אלה, ובדרך זו מאפשרות ומעודדות חדשנות טכנולוגית. בכלל זה ניתן למנות שימוש בעקרונות אתיים המכוונים בין היתר למגזר הפרטי כגון אלה שיוצעו להלן; בתקינה; בהמלצות של הרגולטור הניתנות לאימוץ וולונטרי

[242] Eric Biber, Sarah E.Light, J.B. Ruhl & James Salzman, *Regulating Business Innovation as Policy Disruption: From the Model T to Airbnb*, 70 VANDERBLIT L. REV. 1561, 1605 (2019).

ואינן מחייבות; ורגולציה עצמית. הרעיון הבסיסי בכל אלה, הוא לעודד חדשנות תחת נכונות וולנטרית לפעילות אחראית ושימת עין מצד הרגולטור והממשלה. עניין זה משמעותי במיוחד לבינה מלאכותית, עקב הצורך לנהל את הסיכונים הכרוכים באימון האלגוריתמי וביכולת המשתנה לצפות את תוצאות פעולתו. המגמה המסתמנת בקרב מדינות העולם היא של קידום מהיר של תקינה וקביעת כללים במקביל לפיתוח הטכנולוגי, ומגמה זו מתורגמת לדרישה מהרגולטורים לקיים תהליך בחינה, מעקב ותגובה הנדרש לצורך אימוץ בינה מלאכותית בצורה מהירה יחסית. כתוצאה מכך, יש לבחון העשרת סל הכלים הרגולטורי בכלים המאפשרים למידה, בחינה והתנסות בשימוש בטכנולוגיה, בניהול הסיכונים הקשור בה, ובהתאמה של סל הכלים הרגולטורי אליה.

לשם ההמחשה, דרך אפשרית במקרים המתאימים לכך היא רגולציה עצמית (Self-Regulation), שמשמעותה העברת האחריות לפיתוח כללי ההתנהגות והוכחה על עמידה בהם למפתחי המערכת או לגורמים אחרים בתעשייה, לעיתים תוך שמירה על פיקוח שלטוני ברמות משתנות (Enforced Self-Regulation). רגולציה עצמית יכולה להיות מפוקחת (כשאחת החלופות היא רגולטור שאוכף על התעשייה את הכללים שנקבעו על ידה עבור עצמה) ואז היא דומה יותר לרגולציה המסורתית; והיא יכולה להיות בלתי מפוקחת ולמעשה "להיאכף" על ידי הגוף שאימץ אותה ביחס לעצמו באופן וולונטרי. ככלל, נראה שיהיה מתאים יותר לעשות שימוש ברגולציה עצמית כשיש אמון בין הרגולטור למפוקחים הפוטנציאליים; כשהרגולציה מותאמת באופן פרטני למפוקח מסוים (tailor made); ובעיקר כשהסיכונים הגלומים בטכנולוגיה בהקשר הקונקרטי הם מצומצמים.

<u>יתרון מרכזי</u> לרגולציה עצמית, הוא שחלק מפערי המידע והמומחיות בין הרגולטור למפוקחים נפתרים, עקב העובדה שהתעשייה נושאת בחלק מהאחריות להסדרת הפעילות בתחום. <u>בנוסף</u>, עצם הידיעה כי ככל שהרגולציה העצמית לא תיתן מענה הולם לסיכונים הרגולטור יצטרך לקבוע רגולציה שלטונית, מהווה תמריץ ליישום הולם של כלי הרגולציה העצמית. <u>כמו כן</u>, יש לזכור כי לגבי חלק ניכר מהסיכונים המתעוררים בהקשרי בינה מלאכותית, יש לתעשייה תמריץ משמעותי להתמודד עמם, באופן שעשוי לייתר במקרים המתאימים רגולציה שלטונית כופה (למשל, ניתן להניח שגם יצרני כלי רכב עצמאיים (אוטונומיים) דואגים לבטיחות השימוש בהם, וגם מפתחי מערכות לקבלת החלטות אשראי מעוניינים לוודא שהן לא יפלו כדי שאכן ישתמשו בהן).

מצד שני, <u>החיסרון העיקרי</u> לרגולציה עצמית קשור בחשש שהכללים שייקבעו במסגרת רגולציה עצמית לא יספקו את ההגנה הנדרשת על זכויות יסוד ואינטרסים ציבוריים, כפי שהיה ראוי וניתן להגן עליהם באמצעות רגולציה שלטונית. <u>בנוסף</u>, ככלל ישנו רצון ליצור כללים שיהיו ניתנים לאכיפה, כשבמצב של רגולציה עצמית שאינה מפוקחת לא תהיה אכיפה שלטונית, וגם במצב של רגולציה עצמית מפוקחת רמת האכיפה שונה מחמת ההבדל במעמד הכללים ובסנקציה שלצדם. <u>כמו כן</u>, קיים חשש מהקניית יתרון לגופים פרטיים גדולים שיגבשו את הרגולציה העצמית, ויקבעו את ה"סטנדרט", באופן שעשוי לשמר את כוחם בשוק ולהקשות על מתחרים קטנים וחדשים.

עקב כך יש חשיבות לעודד פיתוח ושימוש בכלי זה כדי להשיג את יעדי ההסדרה, אולם יש לעשות בו שימוש זהיר במקרים המתאימים בשים לב לחסרונותיו, וכן לעקוב אחר יישומו כדי לבחון את הצורך בהתערבות במידה שלא ניתנת הגנה נאותה לזכויות יסוד ואינטרסים ציבוריים.[243] יצוין כי

[243] Practical Guidance on Agile Regulatory Governance to Harness Innovation (hereafter the "Practical Guidance") C(2021)99/ADD1, p.7-8.

ישנו כמובן קשר בין שימוש בכלים רכים לבין גישת ניהול הסיכונים שמוצעת לעיל. בפרט, ככל שהסיכונים הכרוכים בפיתוח ושימוש בבינה מלאכותית גדולים יותר, יהיה מקום לנקוט במשנה זהירות בשימוש בכלי רגולציה רכים בשל החשש להתממשות סיכונים אלה, ולהיפך.

לבסוף יצוין, כי ועדת הבדיקה לבחינת הצורך בהתערבות ממשלתית לשם האצת התפתחות תחום הבינה המלאכותית ומדע הנתונים שהקים פורום תל״ם, בראשות ד״ר ארנה ברי, המליצה על פיתוח תקינה לבינה המלאכותית ״הן מבחינת אלגוריתמיקה ומודלים והן מבחינת הנתונים״. זאת, על מנת להבטיח פיתוח ושימוש אתי ובטוח בבינה מלאכותית ו״יכולת בקרה ומדידה של התוצרים ועמידתם בתקנים השונים. עבודה נמרצת בכיוון זה נעשית כיום בעולם, בין היתר על ידי ארגוני תקינה בינלאומיים כגון IEEE, ISO ו-NIST, ויש להניח כי תשמש כלי משמעותי בהסדרת התחום.

ה. <u>פיתוח הרגולציה באופן מודולרי ועידוד נסיינות רגולטורית, תוך שמירה על ניטרליות טכנולוגית</u>

ברכיב זה במדיניות הרגולציה מוצע כי רגולציה שנועדה להסדיר פיתוח ושימוש בבינה מלאכותית תקודם ותתפתח בשלבים ובאופן גמיש בהתאם להתפתחויות הטכנולוגיות, וכן מוצע כי ייעשה שימוש בנסיינות רגולטורית, ובכלל זה בפיילוטים וב״ארגזי חול״ רגולטוריים שיאפשרו הכנסה בטוחה של מערכות בינה מלאכותית. הכל, תוך שימוש בכלים ניטרליים מבחינה טכנולוגית (כלומר, לא כללים מפורטים ונוקשים המתאימים רק לסוג ספציפי של מערכת).

דומה כי האתגר המשמעותי ביותר הניצב בפני הרגולטורים בבואם להסדיר טכנולוגיה בכלל, ובינה מלאכותית בפרט, הוא הקצב המהיר והדינמיות המאפיינים את ההתפתחות הטכנולוגית (The Pacing Problem). המתח המובנה שבין נוקשות המערכת הרגולטורית לבין דינמיות ההתפתחות הטכנולוגית מייצר אתגר דו-כיווני. <u>מצד אחד</u>, הרגולציה לעיתים קרובות מתקשה להדביק את הקצב המהיר של ההתפתחויות הטכנולוגיות, ולהציע הגנה מספקת מפני הסיכונים הכרוכים בהן. שינוי ועדכון של הרגולציה כרוך לעיתים קרובות בהליכים שהשלמתם נמשכת זמן לא מבוטל, בין אם המדובר בנסיבות בהן יש צורך לתקן חקיקה ראשית, ובין אם נדרש תיקון חקיקת משנה או נהלים של הרגולטור. במצבים אלה הרגולציה ״נשארת מאחור״ ומותירה ואקום רגולטורי שמאפשר התפתחות טכנולוגית ללא פיקוח מספק. <u>מצד שני</u>, הקושי בעדכון רגולציה קיימת מביא לכך שהרגולציה מהווה לעיתים חסם בלתי מוצדק הבולם התפתחות טכנולוגית רצויה. במצבים אלו, הרגולציה – שפעמים רבות נקבעה בתקופה בה כלל לא היה ניתן להעלות על הדעת את ההתפתחות הטכנולוגית שאירעה – מונעת את הפיתוח או השימוש בטכנולוגיה ללא הצדקה.

אתגר נוסף הוא שלעיתים עולה צורך ברגולציה בשלב שבו עדיין אין ודאות בנוגע לשימושים וההשלכות של הטכנולוגיה החדשה, בפרט כשמדובר בטכנולוגיית בינה מלאכותית, מה שמשפיע על היכולת להסדיר אותה. יש לציין, כי קיים מתח מסוים (trade-off) בין הבנת השימושים וההשלכות של טכנולוגיה חדשה לבין מידת היכולת להשפיע על ההתפתחות שלה. שכן ככלל ניתן להסדיר יותר בקלות טכנולוגיה נתונה כשהיא עדיין ״צעירה״ ולא פופולרית, כאשר במצב זה לעיתים קרובות ההשלכות הבלתי צפויות והלא רצויות שלה עדיין אינן ניכרות בבירור; או שניתן להמתין ולזהות השלכות אלה, אבל אז ייתכן שיהיה קשה להסדיר אותה לאחר התקבעות השוק.

על רקע זה, ומסיבות נוספות לרבות פערי מידע והמומחיות, נודעת חשיבות לקיומם של מחזורי רגולציה לומדים וגמישים, הכוללים בחינה של הצורך בשינוי מדיניות לאחר שהתקבלה ויכולת התאמה מהירה שלה.[244] כלומר, יש יתרון לקביעת כלי המדיניות ועקרונות התוכן בהסדרים המאפשרים שינוי והתאמה, למשל עדיפות לעיגון הרגולציה בכלים גמישים יחסית כדוגמת תקנות ונהלים על פני חקיקה ראשית, מקום בו ניתן לקבוע רגולציה במתכונת זו. בהמשך לכך, יש חשיבות לכך שהסדרים אלו יבחנו ויעודכנו לעיתים תכופות בהתאם להתפתחות הטכנולוגית.

זאת ועוד, על רקע האמור לעיל, מתחדדת החשיבות של הטמעת כלים ויכולות שיאפשרו למערכת הרגולטורית להיות יותר לומדת, דינמית וגמישה. בהתאם, יש צורך בהטמעת יכולות של נסיינות רגולטורית ונכונות לקדם נסיינות ביחס לבינה מלאכותית. יכולות אלה קריטיות בעידן הנוכחי המתאפיין בהתפתחות טכנולוגית מואצת, ורלוונטיות במיוחד לגבי בינה מלאכותית. שכן נדרשת גישה זריזה ויעילה לרגולציה שתאפשר למצות את פוטנציאל התפתחויות בתחום זה, לאפשר ואף לעודד אותן, אך לעצב אותן באופן שלא פוגע בזכויות יסוד ובאינטרסים ציבוריים.

אחד הכלים המרכזיים המשמשים לנסיינות רגולטורית הוא "ארגז החול" הרגולטורי (Regulatory Sandbox) (להלן: הסנדבוקס), כלי מדיניות המאפשר לרגולטורים לייצר "סביבת ניסוי" רגולטורית. במסגרת זו, הרגולטור מבצע התאמות ברגולציה הקיימת כדי לאפשר למפוקחים להכניס טכנולוגיה חדשה באופן מבוקר, מצומצם ולזמן מוגבל.[245] בנוסף, הסנדבוקס מסייע לרגולטורים להתמודד עם פערי המידע והמומחיות בינם לבין השוק ועם קצב ההתפתחות המהיר של הטכנולוגיה, שכן רגולציה זמנית, מדודה וגמישה מאפשרת תהליך למידה "תוך כדי תנועה" ומקנה כלים להסרה מבוקרת וזהירה של חסמים ממוקדים.[246] המפוקחים מצדם מרוויחים מההזדמנות לפעול בסביבה "ידידותית", ובלי לשאת בנטל רגולטורי שלא חיוני לפעילותם.[247]

יצוין כי הגם שהשימוש הרווח בסנדבוקס הוא כאשר נדרשת הקלה או הגמשה ביחס לרגולציה הקיימת בשוק מוסדר, ניתן לעשות בו שימוש לעיתים גם במצב ההפוך שבו על פניו נראה כי נדרשת תוספת או הקשחה ביחס לרגולציה הקיימת בשוק שאינו מוסדר. מצב זה, המכונה "סנדבוקס הפוך", הוא שימוש בסנדבוקס לשם הוספת תנאי המגביל פיתוח או שימוש בטכנולוגיה, כתחליף לחסימתה או להגבלתה באופן נוקשה וקבוע יותר. למשל, לעיתים מבוקש להכניס לשוק טכנולוגיה חדשה שיוצרת סיכונים שטרם הוסדרו במישרין, אך לא מן הנמנע שיוסדרו בעתיד. במצב זה, הרגולטור יכול (בכפוף לסמכות ולהתנהלות המתאימה על פי כללי המשפט המנהלי) להציע למשתתפים להשתמש בסנדבוקס לצורך למידת הטכנולוגיה והשלכותיה, ואולי אף לקבוע תנאים ייעודיים שימנעו בטווח הקצר פגיעה באינטרסים פיקוחיים שבאחריותו. בתמורה לכך, הוא יוכל להימנע בשלב זה מהסדרה נוקשה ומהירה שתגביל את הפיתוח והשימוש בטכנולוגיה. הסנדבוקס ההפוך עשוי להתגלות כשימושי גם ביחס לבינה מלאכותית, שכן במקרים רבים הסיכונים לזכויות יסוד ואינטרסים ציבוריים הכרוכים בה אינם מוסדרים באמצעות רגולציה קודמת.

---

[244] OECD, Practical Guidance on Agile Regulatory Governance to Harness Innovation (hereafter the "Practical Guidance") C(2021)99/ADD1.

[245] דו"ח הצוות הבין-משרדי לרגולציה חכמה, בעמ' 129. המסמך זמין כאן.

[246] שם, בעמ' 130.

[247] אחיעז, חמדני, עמירם וקסטיאל, לעיל ה"ש 22, בעמ' 21; Lawrence G. Baxter, *Adaptive Financial Regulation and RegTech: A Concept Article on Realistic Protection for Victims of Bank Failures*, 66 DUKE L.J. 567, 600-01 (2016).

גם ועדת המשנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית הכירה באתגר הקשור לדינמיות ההתפתחות הטכנולוגית בתחום הבינה המלאכותית. בין היתר על רקע זה, המליצה הוועדה על בחינה קבועה ובפרקי זמן קצרים (בהשוואה לרגולציה בתחומים מסורתיים) של המדיניות הרגולטורית על ידי הרגולטור, וכן המליצה לשלב תהליכים של ניסוי מבוקר למדיניות, בפרט באמצעות סנדבוקס. עוד יצוין, כי גם בטיוטת הרגולציה האירופית, תחת פרק שעוסק ב"אמצעים לתמיכה בחדשנות", מוצע להשתמש בסנדבוקס כדי לבחון עמידה של מערכות מבוססות בינה מלאכותית בדרישות שקובעת הטיוטה.[248]

בנוסף, ישנה חשיבות לשימוש בכלים רגולטוריים וברגולציה שיהיו ניטרליים מבחינה טכנולוגית. כלומר, לא כללים משפטיים ורגולטוריים מפורטים המתאימים לסוג טכנולוגיה מסויים שעלול להשתנות במהרה, ולהפוך ללא אקטואלי, או אף להוביל להתקבעות השוק על פתרון טכנולוגי מסויים. זאת, בדומה לעקרון בעניין "גמישות" המוכר במדיניות הרגולציה בארה"ב כפי שהיא מופיעה בחוזר ה-OMB (ראו לעיל פרק "מדיניות הרגולציה בארה"ב").

ו. <u>עיצוב הרגולציה תוך שיתוף הציבור ובכלל זה התעשייה, האקדמיה וארגוני חברה אזרחית</u>

רכיב זה במדיניות הרגולציה מציע כי רגולציה המסדירה פיתוח ושימוש בבינה מלאכותית, תפותח ותקודם תוך שיתוף בעלי הידע והמומחיות בתחום, וכן בעלי העניין שצפויים להיות מושפעים ממנה. זאת, במידה הנדרשת בנסיבות העניין, ועל מנת שהרגולציה תהיה מבוססת על תשתית מקצועית וטכנולוגית איכותית ותבטא איזון בין הזכויות והאינטרסים השונים.

הטכנולוגיה משתנה תכופות, על ידי חברות בעלות אמצעים ומומחיות, המעצבות את הטכנולוגיה ובקיאות בה. ההתקדמות הטכנולוגית נעשית ככלל ללא מעורבות הרגולטורים, וחברות הטכנולוגיה נהנות מהגנה על הסודות המסחריים שלהן כלפי מתחרים. כך שהמידע בנוגע לטכנולוגיות בינה מלאכותית פעמים רבות לא מצוי באופן מלא בידי הרגולטורים, ולמעשה תחום הטכנולוגיה בכלל וטכנולוגיות הבינה המלאכותית בפרט מתאפיין במידה רבה באי-סימטריה של מידע. אף במקרים שבהם מסוגלים הרגולטורים להשיג מידע מחברות הטכנולוגיה, עלול להתעורר קושי להבין את המידע הטכני, ולחזות את השפעות הטכנולוגיות. המורכבות הטכנולוגית מחייבת בקיאות ומומחיות טכנית מצד הרגולטורים, והקדשת משאבים רבים לשם כך, שאינם תמיד בנמצא.

ככלל, יש צורך לעצב רגולציה על בסיס תשתית עובדתית איכותית, בשיתוף בעלי העניין, המידע והמומחיות ותוך שמיעת הציבור הרחב. מעבר לחובות הבסיסיות בעניין זה מתחום המשפט המנהלי, חוק עקרונות האסדרה קובע במפורש כי "אסדרה מיטבית" היא כזו שתהליך גיבושה וקביעתה מתבססים על נתונים הנוגעים לעניין ונעשים בהתאם לעקרון השקיפות ותוך שיתוף הציבור במידה הנדרשת בנסיבות העניין. צורך זה בא לידי ביטוי במיוחד כשמדובר ברגולציה של טכנולוגיה בכלל ובינה מלאכותית בפרט, על רקע פערי המידע והמומחיות הנסקרים לעיל.

בהתאם לכך, רכיב זה מדגיש את הצורך בשיתוף ציבור כחלק מהתמודדות עם פערי המידע והמומחיות, וכחלק מקיום השיח הדינמי על התפיסות הערכיות ביחס לשימושים בטכנולוגיה חדשה. הכוונה היא לשיתוף במידת האפשר ובשים לב למכלול הנסיבות של הציבור הרלוונטי ובעלי

[248] סעיף 53 לטיוטת הרגולציה האירופית. מעניין לציין כי הטיוטה נותנת עדיפות לספקים קטנים וחברות הזנק בגישה לסנדבוקס.

העניין, לרבות חברות קטנות בדגש על חברות ההזנק (סטארט-אפ), כדי להגביר את השקיפות, אמון הציבור, והמומחיות בתחום. לצד זאת, יש לפעול לחיזוק היכולות הטכנולוגיות והמומחיות אצל גורמי הממשלה והרגולטורים, על מנת שיוכלו לגבש עמדה מקצועית עצמאית.

6.2. <u>אימוץ עקרונות אתיים לתחום הבינה המלאכותית</u>

מוצע לאמץ עקרונות אתיים משותפים עבור תחום הבינה המלאכותית, על בסיס עקרונות ה-OECD שנסקרו לעיל, כדי לסייע לרגולטורים ולארגונים בתחום זה. מוצע כי העקרונות לא יהיו בעלי מעמד משפטי מחייב, אך ישמשו בסיס לשיח משותף ותיאום לפי ההקשר הנורמטיבי והרגולטורי.

<u>העקרונות המוצעים הם</u>:

א. בינה מלאכותית לקידום צמיחה, פיתוח בר-קיימא ומובילות ישראלית בחדשנות.

ב. האדם במרכז – כיבוד זכויות יסוד ואינטרסים ציבוריים.

ג. שוויון ומניעת אפליה פסולה.

ד. שקיפות והסברתיות.

ה. אמינות, עמידות, אבטחה ובטיחות.

ו. אחריותיות.

6.2.1. <u>החשיבות של אימוץ עקרונות אתיים ומעמדם</u>

<u>ראשית</u>, העקרונות מאפשרים שפה משותפת, הן בין מגזרי המשק השונים, הן מול הנעשה במדינות העולם המפותחות לגבי אופן ההתמודדות האחראי עם חששות וסיכונים מוכרים הכרוכים במערכות מבוססות בינה מלאכותית, לרבות האתגרים שנסקרו לעיל במסמך זה.

<u>שנית</u>, העקרונות מאפשרים לגשר בין ערכים, יישומים טכנולוגיים, והיבטים משפטיים, תוך ממשק לתפיסה הבינלאומית המתהווה של "בינה מלאכותית מהימנה". אימוץ ופרסום העקרונות במסגרת המדיניות מאפשר שקיפות ושיח ציבורי לגבי העקרונות כנקודת מוצא למדיניות רגולציה לתחום הבינה המלאכותית. בנוסף, לעקרונות חשיבות כמכנה משותף לשיח בינלאומי בנושא זה.

<u>שלישית</u>, העקרונות מסייעים לקידום מדיניות רגולציה קוהרנטית בתחום הבינה המלאכותית, בכך שהם מכוונים את הרגולטורים להיבטים המרכזיים שיש להידרש אליהם, ומגבירים את הוודאות בקרב המגזר הציבורי והפרטי בעניין המדיניות הממשלתית לגבי בינה מלאכותית.

אין כוונה להקנות לעקרונות אלה מעמד משפטי, ובהתאם לכך הם לא מחייבים גורמים פרטיים או ציבוריים לפעול או שלא לפעול בדרך כלשהי, ואינם מחליפים את המסגרת המשפטית החלה או מהווים כלי לפרשנות משפטית. עם זאת, עקרונות אלה משקפים היבטים בהם ראוי להתחשב במסגרת פיתוח ושימוש בבינה מלאכותית ובעת קביעת רגולציה בתחום זה.

6.2.2. **התבססות על עקרונות ה-OECD**

בהמשך להחלטת ממשלה 212, העקרונות שמוצע לאמץ מבוססים בעיקר על העקרונות המוצעים בהמלצות ה-OECD לקידום בינה מלאכותית מהימנה,[249] תוך התאמה שלהם ליעדי המדיניות הישראלית ובהתבסס על הנעשה במדינות מפותחות, לרבות התאמות שבוצעו במדינות אלה.

עד כה נכתבו בעולם מסמכים שונים הכוללים עקרונות אתיים או עקרונות משפטיים שהם רלוונטיים לתחומי העיסוק בבינה מלאכותית. מחקרים מונים עשרות מסמכים ובהם עקרונות לבינה מלאכותית מהימנה או אתית שנערכו על ידי ארגוני חברה אזרחית, חברות טכנולוגיות, ארגונים פרטיים, ממשלות, וארגונים בינלאומיים.[250] ועדת המשנה של המיזם למערכות נבונות גיבשה אף היא רשימה של עקרונות אתיים שיש להתחשב בהם.[251]

לצד זאת, לעקרונות ה-OECD שנסקרו לעיל (ראו פרק "המלצות ה-OECD בתחום הבינה המלאכותית") חשיבות ייחודית, משום שגובשו בתיאום בין המדינות המפותחות מתוך שותפות ורצון לעשות שימוש מועיל בבינה מלאכותית באופן שמכבד את עיקרון שלטון החוק, ערכים דמוקרטיים ואינטרסים ציבוריים. מדינות אלה הסכימו על עקרונות אלו כקווים מנחים בתחום זה, תוך גישור על השוני בהיבטים המשפטיים, הרגולטוריים והכלכליים ביניהן. עקב כך העקרונות משקפים שיח מפותח שיש לו ערך לגיבוש מדיניות פנים מדינתית. בנוסף, יש לו חשיבות במרחב הבינלאומי כמכנה משותף מוסכם בין המדינות המפותחות, לרבות במסגרת שיתוף הפעולה בין האיחוד האירופי וארה"ב כאמור לעיל.

6.2.3. **העקרונות האתיים המוצעים לאימוץ**

מוצע לאמץ את ששת העקרונות האתיים המשותפים לתחום הבינה המלאכותית המפורטים להלן. לאור השונות באופני השימוש בבינה מלאכותית, מרכיב מרכזי באופן התחולה של העקרונות הוא יישום לפי הקשר ובהתאם לתפיסות מקובלות של ניהול סיכונים.

א. **בינה מלאכותית לקידום צמיחה, פיתוח בר-קיימא ומובילות ישראלית בחדשנות**

שימוש אחראי בבינה מלאכותית מהימנה הוא אמצעי לעודד צמיחה, פיתוח בר-קיימא, רווחה חברתית וקידום המובילות הישראלית בתחום החדשנות.

עיקרון זה מצביע על החשיבות של שימוש אחראי בבינה המלאכותית להשגת יעדי המדיניות שמוגדרים בהחלטת ממשלה 212 ויעדי מדיניות מרכזיים נוספים.

גישה זו גם מכירה בכך שחדשנות טכנולוגית כגון טכנולוגיית הבינה המלאכותית יכולה לתרום משמעותית לרווחת הפרט, ולתחומים חשובים כגון בריאות הציבור, או צמיחה בת-קיימא. עקב

---

[249] ראו פרק "המלצות ה-OECD בתחום הבינה המלאכותית" לעיל.

[250] מחקר של אוניברסיטת הרווארד משנת 2020 סקר מסמכים אתיים רבים שהופקו בידי ארגונים שונים, את העקרונות הכלולים בהם, ואת אופן פרשנותם. ראו:
Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1, Available at SSRN: https://ssrn.com/abstract or http://dx.doi.org;
לסקירה משנת 2022 ראו: Nicholas Kluge Corrêa, Camila Galvão, James William Santos, Carolina Del Pino, Edson Pontes Pinto, Camila Barbosa, Diogo Massmann, Rodrigo Mambrini, Luiza Galvão, Edmund Terem, Worldwide AI Ethics: a review of 200 guidelines and recommendations for AI governance, *Computers and Society*, June 2022, https://arxiv.org/abs.

[251] דוח ועדת המשנה.

כך, יש צורך במבט מאוזן המביא בחשבון הן את הסיכונים והן את התועלות, ולבחון את אלה למול נקודת המוצא החברתית-טכנולוגית הקיימת.

בעולם של שווקים טכנולוגים גלובליים, ולנוכח השלב בו נמצאת הטכנולוגיה, יש לתת את הדעת לחשיבות של שימור עוצמה טכנולוגית ויכולת טכנולוגית. זאת, תוך פעולה במסגרת ערכית ואתית. על רקע המאפיינים של הפיתוח והשימוש בטכנולוגיה בשווקים גלובליים, ועל רקע מקומה של ישראל בשווקים אלה, יש לקחת בחשבון את הקשר בין עיקרון זה לבין העקרונות הנוספים.

ב. <u>האדם במרכז – כיבוד זכויות יסוד ואינטרסים ציבוריים</u>

פיתוח בינה מלאכותית, או שימוש בה, יש לעשות תוך כיבוד שלטון החוק, זכויות יסוד ואינטרסים ציבורים ובפרט תוך שמירה על כבוד האדם ופרטיות.

עיקרון זה מציין את נקודות המוצא הנורמטיביות בישראל ובמדינות המפותחות הכוללות את שלטון החוק והמשטר הדמוקרטי, וכן הגנה על זכויות יסוד ואינטרסים ציבורים.[252] הביטוי "האדם במרכז" ("Human-centered AI") מהווה מחולל מרכזי לצורך פירוש ופיתוח העקרונות המשפטיים והרגולטוריים בתחום הבינה המלאכותית.[253] תפיסה זו נועדה להזכיר כי בעת פיתוח ושימוש בטכנולוגיות חדשניות, האדם, ויחסי הגומלין שלו (במגוון הקשרים) עם החברה בה הוא פועל, מהווים נקודת מכוון נורמטיבית.

תפיסת "האדם במרכז" נועדה להבטיח שבעת פיתוח ושימוש בבינה מלאכותית, ובפרט הישענות על חישובים אלגוריתמיים מורכבים המייצרים החלטות או תחזיות, תינתן תשומת הלב הנאותה לערכים הרלוונטיים להקשר שבו נעשה השימוש.

עקרון זה נועד להעצים, לפי ההקשר והסיכון, את כיבוד הערכים הקיימים אגב פיתוח בינה מלאכותית ושימוש בה.

לצד הערכים הכלליים, מוזכרים בסעיף "כבוד האדם" ו"פרטיות". לכבוד האדם מעמד חשוב ויסודי כעוגן לזכויות אדם,[254] והוא מופיע בעקרון זה כביטוי לתפיסת "האדם במרכז" ועל מנת לבטא את האוטונומיה והרצון של הפרט.[255] הזכות לפרטיות מוזכרת לאור היותה זכות משמעותית בפעולות הקשורות ב"נתוני עתק" (איסוף ועיבוד מידע) ובעת קבלת החלטות על סמך מידע אישי. עקב כך, להגנה על הפרטיות רלוונטיות ניכרת לעקרונות האתיים המקובלים ביחס לבינה מלאכותית.

---

[252] ראו גם סעיף 1 לחוק עקרונות האסדרה, התשפ"ב-2021.

[253] ראו את האמור במסמך הלוואי להמלצות ה-OECD, הכולל התייחסות לשיקולים השונים שהוצגו בדיונים המכינים ובניסוח ההמלצות, OECD, AI in Society, 2019, Chapter 4, https://www.oecd-ilibrary.org.

[254] ראו: חוק יסוד: כבוד האדם וחירותו.

[255] **בדוח צוות המשנה ראו:**

כיבוד זכויות אדם והגנה עליהן - כיבוד כלל זכויות האדם ומודעות לצורך הגובר בהגנה על זכויות מסויימות שעלולות להיפגע יותר בעידן הבינה המלאכותית וביניהן:

א. <u>שמירה על שלמות הגוף -</u> הגנה מפני פגיעה בחייו או בגופו של אדם.
ב. <u>פרטיות</u> – לרבות מניעת פגיעה בפרטיות באיסוף המידע, ניתוח ועיבוד המידע, שיתוף המידע ובשימושים חדשים בו.
ג. <u>השמירה על האוטונומיה</u> - שמירה על חירות האדם. בין השאר, מניעת השפעה שאינה הוגנת על התנהגות היחיד, המרחב הפרטי שלו וזכותו "להיעזב לנפשו".
ד. <u>זכויות אזרחיות ופוליטיות</u> - לרבות חופש הביטוי וחופש הדת והמצפון.

עקרונות אלה מבקשים לצמצם איסוף מידע מזוהה, להבחין בין הגנה על מידע רגיל ומידע רגיש ולוודא כי מידע נאסף בהסכמה ולמטרה מוגדרת.[256]

ג. שוויון ומניעת אפליה פסולה

בפיתוח ושימוש בבינה מלאכותית יש להביא בחשבון את הצורך בשוויון וגיוון, ואת החשש להטיה במערכות בינה מלאכותית והסיכון לאפליה פסולה כנגד יחידים או קבוצות.

עיקרון זה משקף את התפיסה הבסיסית של הוגנות והתובנה המצטברת בדבר החששות להתרחשות אפליה פסולה בעקבות השימוש בבינה מלאכותית, שלעיתים אף תהיה בלתי חוקית. בכל הקשור לבינה מלאכותית, נעשה שימוש בהוגנות (fairness) כעיקרון הנוגע למניעת אפליה, ואשר נועד לבטא את הצורך בבחינה של מכלול הנסיבות הצריכות לעניין, על מנת לחשוף חשש להטיה אסורה, וזאת לנוכח מגוון התרחישים הטכנולוגים והחברתיים הרלוונטיים.[257] בין היתר, כמפורט באתגר האפליה לעיל, חששות אלה כוללים את האפשרות להטיה בנתונים המשמשים לאימון המערכת או שימוש במאפיין רגיש באחד המשתנים הנאספים באופן שעלול להביא לתוצאות מפלות, וכן את האפשרות להטיות מושרשות בתחום המומחיות המשמש לאימון הבינה המלאכותית.[258] הזכות לשוויון מוזכרת בין היתר עקב החשש כי חישובים אלגוריתמיים עלולים לשקף הטיות בחברה.[259]

מדובר בעיקרון שנועד לבטא את הצורך להביא בחשבון את ההיבטים הנוגעים לשוויון ולמניעת אפליה במסגרת פיתוח או שימוש בבינה מלאכותית. בהקשר הנוכחי, הכוונה אינה להגדיר במדויק את אופן תחולת הזכות לשוויון והאיסור להפלות, כי אם להצביע על הסיכון המוגבר לכך עקב מאפייני הטכנולוגיה, והחשיבות של מודעות בעלי העניין לכך.[260] בהמשך לכך, היקף השתרעות הזכות לשוויון או הגדרת אפליה פסולה, נובע, בין היתר, מאופי השימוש בבינה המלאכותית ומידת רגישותו.[261] במסגרת האמור, במקרים המתאימים, יהיה צורך לנקוט אמצעים, בהתאם להקשר ולרמת הסיכון, להתמודדות עם חששות אלה. כך למשל, נקיטת אמצעים על מנת לוודא שאין הטיה גלויה בנתונים, ונקיטת צעדים לצמצום הסיכון להטיה שאינה גלויה באמצעות בקרות ובדיקות.

---

256 ראו הסקירה לעיל בדבר היבטי פרטיות של בינה מלאכותית. ביחס לעקרונות ה- OECD ראו: OECD, AI in Society, p. 87-89.

257 ראו: OECD, AI in Society, p.90;
ראו גם בטיוטת ההנחיה הבריטית:

Embed considerations of fairness into AI

In many contexts, the outcomes of the use of AI can have a significant impact on people's lives - such as insurance, credit scoring or job applications. Such high-impact outcomes - and the data points used to reach them - should be justifiable and not arbitrary.

In order to ensure proportionate and pro-innovation regulation, it will be important to let regulators continue to define fairness. However, in any sector or domain we would expect regulators to:

- interpret and articulate 'fairness' as relevant to their sector or domain, decide in which contexts and specific instances fairness is important and relevant (which it may not always be), and design, implement and enforce appropriate governance requirements for 'fairness' as applicable to the entities that they regulate

258 מאפיין רגיש הינו או מאפיין כגון דת, גזע, מין, וכו' או נתון אחר שיש לו זיקה לאותו מאפיין (פרוקסי).

259 להרחבה ראו אתגר האפליה לעיל וכן: EPRS Study 2022, p.1.

260 ראו: EPRS Study p.1.

261 הדיון על אודות עיקרון השוויון ואיסור האפליה במשפט הישראלי הינו רחב ומורכב. ראו למשל במשפט החוקתי והמנהלי: דפנה ברק ארז, משפט מנהלי, נבו, עמוד 673- 721; לסקירה על המשפט הפרטי וביקורת עליה: רונן אברהם, בנפרד ועדיין שווה? על דרכי ההתמודדות עם מקרי הפליה הנופלים בנפרד מתחולת חוק איסור הפליה, עתיד להתפרסם בספר אליעזר ריבלין; בדיני העבודה: שרון רבין מרגליות, שלושה דורות של אפליה תעסוקתית: הישגים ומגבלות של המאבק לקיום שוויון תעסוקתי, עבודה חברה ומשפט טו, עמ' 7.

התמודדות עם הסיכון כראוי, כמו גם המודעות לו, יכולה להביא לשיפור קבלת ההחלטות הכוללת, תוך מיקסום התועלות מבינה מלאכותית והפחתת הסיכון להטיה.

ברמה המעשית, לצורך התמודדות עם החשש להטיה, יש חשיבות לניהול סיכונים לאורך כל "מחזור החיים" של מערכת בינה מלאכותית. זאת, בין היתר, באמצעות התמודדות עם הטיות באיסוף ועיבוד מקדים (pre-processing) של הנתונים; בשלב בניית המודל, האימון והערכה; ובשלב היישום ובדיקת ההשפעה על יחידים וקבוצות בעת ההפעלה.[262]

ד. <u>שקיפות והסברתיות</u>

בפיתוח בינה מלאכותית ושימוש בה, יש להביא בחשבון את הצורך ליידע מי שבא במגע עם בינה מלאכותית או מושפע מפעילותה ואת הצורך במתן הסבר להחלטתה או לאופן שבו פעלה, בין היתר בשים לב למידת השפעתה, השלכותיה על מי שמושפע ממנה והאפשרויות הטכנולוגיות הזמינות.

עיקרון השקיפות וההסברתיות נועד להתמודד עם האפשרות שאדם יקיים אינטראקציה עם בינה מלאכותית ולא ידע על כך, וכן עם הסיכונים האינהרנטיים הנובעים מהמפגש בין יכולת ההבנה האנושית וקבלת ההחלטות האלגוריתמית.[263] אמנם בני האדם הם אלה שמייצרים את הבינה המלאכותית, אולם היא אינה מובנת לרוב הציבור, וגם למומחים לבינה מלאכותית ייתכן שיהיה קושי להסביר אופן קבלת החלטה קונקרטית או תוצאותיה. היבט זה מקבל משנה חשיבות כאשר בינה מלאכותית מעורבת בהחלטה שיש בה פגיעה בזכויות של אדם. במקרים אלה גוברת החשיבות לקיומו של מידע על הפרמטרים והלוגיקה ששימשו בסיס לפעולת הבינה המלאכותית.

על כן, בהתאם לעיקרון זה ראוי ככלל כי מפתח בינה מלאכותית או משתמש בה יגלה למי שמקיים עמה אינטראקציה או מושפע מפעילותה את דבר השימוש בבינה מלאכותית; וכן יגיש, במידת האפשר והנדרש בנסיבות העניין, מידע אודות מאפייני ההחלטה הממוחשבת. כן ראוי כי יינקטו אמצעים סבירים בהתאם למקובל בתחום כדי לאפשר להתחקות אחר טעמי ההחלטה.

כמפורט לעיל, בעולמות הבינה המלאכותית נעשה שימוש תדיר בדימוי של קבלת ההחלטה האלגוריתמית כ"קופסא שחורה", שבה לא ניתן להבין את אופן קבלת ההחלטה או "לנהל שיח" עם מקבל ההחלטה על מנת לעמוד על התשתית העובדתית ושיקול הדעת המקצועי שהפעיל.

הסיכון האמור גובר בנסיבות שבהן עקב ההתייעלות הכרוכה במיחשוב ושימוש בבינה מלאכותית, יש המרה של תהליך שבו מעורב אדם לתהליך טכנולוגי בלבד ללא מעורבות אנושית (כגון רכישת מוצר או שירות, קבלת ייעוץ או המלצות).

---

[262] ראו: EPRS, p.1, 6, 13.

[263] דוח ועדת המשנה הפריד בין שני עקרונות אלה, כמפורט להלן:

א. <u>שקיפות</u>–הנגשה, בהתאם להקשר ולנסיבות, של מידע על התהליך עצמו ועל דרך קבלת ההחלטה. לכל הפחות יש לאפשר קבלת מידע על התהליך, על מנת לקבוע אם הוא עונה על העקרונות האתיים.

ב. <u>הסברתיות</u> – 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);

בדו"ח תלם גם הופרדו שני עקרונות אלה, עמוד 62.

בהתאם למסמך המלווה את המלצות ה-OECD, עיקרון ההסברתיות מיועד להתמודד עם האתגר האמור ועוסק בהבנה של תוצאה ספציפית של מערכת בינה מלאכותית.[264] בהתאם למסמך המלווה את המלצות ה-OECD, מוצעות שלוש אפשרויות (חלופיות) להסברים לאופן פעולת מערכת בינה מלאכותית, וזאת בזיקה להסברים הנדרשים לפי דיני הגנת הפרטיות האירופאים לקבלת החלטות אוטומטית. אפשרויות אלה הן: (1) הצגת הפרמטרים המרכזיים בהחלטה בהתאם לסדר חשיבותם; (2) הגורמים המכריעים בהחלטה; (3) הסברים לגבי שני מקרים שנראים דומים והובילו לתוצאה שונה. המסמך המלווה מציין כי להסברתיות יש מחירים ועלויות ולעיתים מערכות בעלות תוצאות יותר מדויקות הן מורכבות יותר להסבר. בהתאם לכך יש לבחון את היחס שבין עלות לתועלת לפי כל הנסיבות, כגון למשל רגישות ההחלטה והשימוש.[265]

יצוין כי גם בהתאם לדיני הפרטיות בישראל, כאשר פעילות הבינה המלאכותית מבוססת על איסוף או עיבוד מידע אישי, ייתכנו חובות ספציפיות בנושא זה.[266]

עיקרון השקיפות וההסברתיות רלוונטי גם ליחסים שבין יצרן בינה מלאכותית לבין ארגון המשתמש בה, בהם עשוי להיות פער מידע אודות מאפיינים אלה. סגירת פער מידע זה בהתאם לעיקרון השקיפות וההסברתיות מאפשר לארגון המשתמש לעמוד מצדו בחובות החלות עליו כלפי מי שמושפע מהחלטותיו. בנוסף, קיום עיקרון השקיפות וההסברתיות מעצים את הארגון המשתמש בשיח עם הארגון היצרן, ומאפשר לו לנהל שיח עשיר יותר לגבי מאפייניה ואופן השימוש הנכון בה, חלוקת הסיכונים ביחס אליה ומעקב אחר השימוש בה. השקיפות וההסברתיות מאפשרת לארגון המשתמש גם להתאים את אופן השימוש והבקרה שלו על השימוש, ובכלל זה עמידה בעיקרון השקיפות וההסברתיות שלו כלפי מי שמושפע מאופן השימוש שלו בטכנולוגיה ועמו הוא בא במגע.

ה. אמינות, עמידות, אבטחה ובטיחות

בפיתוח ושימוש במערכות בינה מלאכותית יש להביא בחשבון את הצורך בכך שמערכות אלה תהיינה אמינות, מאובטחות ובטיחותיות, כך שבתנאי שימוש רגילים, שימוש צפוי או שימוש שגוי או תנאים מסוכנים אחרים, הן יפעלו כראוי ולא יהוו סיכון בטיחותי בלתי סביר. לשם כך יש לנקוט אמצעים סבירים בהתאם לתפיסות מקצועיות מקובלות של ניהול סיכונים על מנת לצמצם סיכוני בטיחות ואבטחת מידע לכל אורך מחזור החיים של מערכות בינה מלאכותית.

עיקרון זה משקף את הסיכונים הנובעים מטכנולוגיות המידע. כמפורט לעיל בפרק "סוגיות ואתגרים המתעוררים בקשר לבינה המלאכותית", הניסיון המצטבר בתחום אבטחת המידע והגנת הסייבר מלמד על החשיבות של התמודדות עם הסיכונים האינהרנטיים הנובעים משימוש במחשבים ותקשורת, ועם החשש לשימוש לרעה בידי גורמים זדוניים המנצלים "חולשות" בטכנולוגיה. כאמור, מאפיינים אלה רלוונטיים גם לתחום הפיתוח והשימוש בבינה מלאכותית.

[264] OECD, AI in Society, p. 93.
[265] OECD, AI in Society, p. 95.
[266] ראו: "חובת יידוע במסגרת איסוף ושימוש במידע אישי", לעיל ה"ש 27; כן ראו דוח בינה מלאכותית בשירותים פיננסים עמ' 42-43 על סקירה על הוראות החקיקה האירופית.

לצד סיכונים אלה, הבינה המלאכותית חשופה לסיכונים ייחודיים הקשורים באופן אימון האלגוריתם, החשש מפני מניפולציה שתבוצע כלפיו, וסיכון הנובע מאופן השימוש בה. היקף השימוש הצפוי בבינה מלאכותית, במוצרים שכיום אין להם יכולת חישובית (כגון מכוניות ומוצרים מסוגים שונים), מרחיב את היקף הרלוונטיות של סוגיות בטיחות הנובעות מסיכונים ממוחשבים. כך, כל עוד מדובר במוצרים שיש בהם רכיבים פיזיים בלבד, הסיכון הבטיחותי מהם נובע מאופן פעולתם הפיזי. שילוב רכיבים ממוחשבים מייצר סיכון מסוג חדש. זאת, בפרט שבינה מלאכותית עשויה ללמוד ולהשתנות תוך כדי ההפעלה שלה לאחר שלב הייצור.

על מנת לקדם את הוודאות והיכולת לנהל את סיכוני הבטיחות, יש חשיבות בהעברת המידע הנדרש על מאפייני הבטיחות בין הגורמים השונים המשתמשים בבינה מלאכותית, על מנת לאפשר ניהול סיכונים אחיד. בנוסף, בדומה לפיתוח תוכנה, יש מקום לכך שהליך עיצוב ופיתוח בינה מלאכותית יאפשר בחינה ובדיקה לאחור, באמצעות תיעוד.

לצד חשיבותו העצמאית שתוארה לעיל, עקרון זה תומך ביכולת לבחון את קיום העקרונות הנוספים. זאת, משום שקיום תהליכי פיתוח ובדיקה מהימנים של טכנולוגיה מאפשרים גם מעקב טוב יותר אחרי אופן קיום העקרונות הנוספים המפורטים בפרק זה.

ו. <u>אחריותיות</u>

מפתחי בינה מלאכותית, מפעיליה או המשתמשים בה יגלו אחריות לתפקודה התקין, ולקיום העקרונות האחרים בפעילותם, בין היתר בשים לב לתפיסות ניהול סיכונים מקובלות ולאפשרויות הטכנולוגיות הזמינות.

עיקרון האחריותיות נועד לבסס את קיומו של "גורם אחראי" לבינה מלאכותית, שעשוי להיות המפתח או המפעיל של מערכת מבוססת בינה מלאכותית. בהתאם לעקרון המוצע, מערכות מבוססות בינה מלאכותית נוצרות בידי אדם, ולכן כברירת מחדל האדם הוא שאחראי להן.[267] בכך עיקרון האחריותיות מאפשר גם את העיקרון של מעורבות אנושית בבינה המלאכותית.

אחריות זו משמעה בין היתר הצורך לנקוט באמצעים לתפקודה התקין, לפעול על מנת לצמצם את החשש לפגיעה בזכויות יסוד ובאינטרסים ציבורים, ולשאוף לקיום העקרונות שתוארו לעיל. אופן ההתמודדות עם הסיכונים ויישום העקרונות מבוסס על אמצעים סבירים ויכולת צפיות סבירה, לפי מידת ההתפתחות הטכנולוגית. העיקרון משרת גם את היכולת לבצע בקרה בידי גורם בלתי תלוי, במקרים הרלוונטיים, בנוגע לאופן הפיתוח של בינה מלאכותית והשימוש בה.

עיקרון זה מבטא אמון בגורמים האמורים, ומאפשר להם לפעול בתנאי שינהגו כמי שאחראיים לבינה המלאכותית. הוא מאפשר לגשר בין העקרונות המפורטים לעיל, לבין החדשנות המבוצעת בפועל בידי ארגונים המפתחים או משתמשים בבינה מלאכותית.

[267] Joanna J. Bryson, The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation, in: *The Oxford Handbook of Ethics of AI (Markus D. Dubber, Frank Pasqule, Sunit Das (ed.), 2020,* (Hereinafter: Bryson), *p. 4-5*.

עיקרון זה משקף את התפיסה הבסיסית שלפיה ראוי כי מי שיוצר את הסיכון יתמודד איתו ויהיה אחראי לו, בלי להחצין את הסיכון לצדדים אחרים. בנוסף, עיקרון האחריותיות ותפיסה הוליסטית לגבי ניהול הסיכונים בארגון תורמת להעלאת החוסן הארגוני והיכולת להשתמש במלוא התועלות הטכנולוגיות.[268] לעיקרון האחריותיות זיקה הדוקה לתפיסות מקובלות של ניהול סיכונים כמפורט בפרק "התוויית שפה אחידה באמצעות מסגרת מומלצת (וולונטרית) לניהול סיכונים" להלן.

6.3. <u>מיסוד מוקד ידע ותיאום ממשלתי להסדרת בינה מלאכותית</u>

מוצע להפקיד בידי גורם ממשלתי אחד את ריכוז ותיאום סוגיית ההסדרה של בינה מלאכותית. ישנה חשיבות לכך שגורם זה יפתח מומחיות מקצועית וטכנולוגית, ויהיה בעל ראייה רוחבית ואחריות לקידום האינטרס הכלל-משקי בעידוד הטכנולוגיה והשגת היעדים הממשלתיים.

גורם זה יעסוק ביישום מדיניות רגולציה ואתיקה זו, וגיבוש המלצות לעדכונה לפי הצורך; בייעוץ למשרדי הממשלה ולרגולטורים בגיבוש מדיניות ורגולציה לגבי בינה מלאכותית; ובהנגשת מידע וכלים לשימוש אחראי בבינה מלאכותית, ובכלל זה כלי ניהול סיכונים, עבור הממשלה והציבור.

השימושים בבינה מלאכותית מגוונים והיא מוטמעת כמעט בכל ענף במשק, ובכלל זה בריאות, תחבורה, פיננסים, חקלאות וחינוך. גיוון השימושים הוא השיקול המכריע שבגינו מוצע לעיל במסגרת מדיניות הרגולציה שלא לאמץ חקיקת מסגרת רוחבית שתסדיר את כל השימושים בבינה מלאכותית, אלא שפיתוח ושימוש בבינה מלאכותית ייעשה ברמת הענף שבו היא נדרשת על בסיס צורך קונקרטי ובהובלת הרגולטור האמון עליו (ראו לעיל בעניין מדיניות הרגולציה המוצעת).

ואולם, לנוכח הבסיס המשותף של הבינה המלאכותית, ובפרט כשמדובר בסוגיות רוחב כגון אחריותיות, אפליה, מעורבות אנושית או הסברתיות – מוצע לעיל כי ההסדרה הענפית תיעשה תוך פעולה לאור מדיניות ממשלתית אחידה באמצעות תיאום. תיאום זה יאפשר לשמור על הקוהרנטיות של ההסדרה הענפית, למנוע סתירות ולייצר ודאות ואחידות, וכן יאפשר לממשלה להוביל מהלכים מתואמים ורחבים להתמודדות עם אתגרים ושימושי בינה מלאכותית חדשניים חוצי-ענפים.

על רקע זה, יש חשיבות רבה להיבטים הארגוניים והמוסדיים, שיסייעו בתיאום בין הגורמים הממשלתיים השונים הפועלים בתחום זה. לפיכך, ועל מנת לחזק ולשמר תיאום זה, ובדומה למקובל במדינות שונות ברחבי העולם, מוצע למסד גורם ממשלתי שישמש כמרכז ומתאם לתחום ההסדרה של בינה מלאכותית עבור כל הממשלה. בנוסף, תיאום מובנה זה מאפשר להתגבר על האתגר הטמון בכך שלעיתים קרובות טכנולוגיות בינה מלאכותית מעוררות סוגיות המטופלות על ידי כמה רגולטורים שונים, מה שעלול להקשות על הרגולטורים להגיב לטכנולוגיות באופן המצריך תיאום עם רגולטורים אחרים.[269]

---

[268] ראו ממצאים מסקר גרטנר בנושא זה, https://blogs.gartner.com.

[269] Practical Guidance on Agile Regulatory Governance to Harness Innovation (hereafter the "Practical Guidance") C(2021)99/ADD1;
באופן כללי, לגבי אתגר התיאום בין רגולטורים, ראו: יובל רויטמן "הרפורמה הרגולטורית: בגין הגלוי לסמוי", 425, 429 **משפט חברה ותרבות, מסדירים רגולציה – משפט ומדיניות** (ישי בלנק, דוד לוי-פאור ורועי קרייטנר, עורכים, 2016).

ועדת המשנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית הממליצה גם היא "להקים מנגנון תיאום פנים ממשלתי על-משרדי. כל זאת על מנת ליצור מדיניות אחידה, ברורה וקוהרנטית, בין כלל משרדי הממשלה". בנוסף, החלטת ממשלה 212 הטילה את הובלת הטיפול בנושא בידי שרת החדשנות, המדע והטכנולוגיה, ובאמצעות משרדה.

מעבר למשימות התיאום והריכוז, מוצע כי הגורם האמור ישקוד על הנגשת מידע טכנולוגי ומקצועי למשרדי הממשלה ולרגולטורים הרלוונטיים, ובפיתוח כלים לעיצוב מדיניות ביחס לבינה מלאכותית כגון כלי לניהול סיכונים לשימוש הרגולטורים (ראו הצעה בעניין זה להלן). באשר למידע, הקצב המהיר של ההתפתחות הטכנולוגית והתפתחות הרגולציה והתקינה במדינות רבות ובארגונים בינלאומיים, מייצר צורך בגורם מקצועי קבוע שיהווה מוקד ידע, יעביר את המידע הרלוונטי בין הגורמים השונים העוסקים בנושא בתוך הממשלה וישתף בתובנות הדדיות. כך לדוגמה, ארגונים שונים בעולם מקדמים עתה כללי התנהגות קונקרטיים להתמודדות עם בינה מלאכותית, במטרה לתרגם את העקרונות שפיתחו לכללי פעולה מעשיים. על רקע זה, והעבודה הנמרצת של גופי תקינה שונים לתקינה בתחום הבינה המלאכותית, יש חשיבות לכך שיהיה מוקד ידע ממשלתי המכיר את התקנים, ומאפשר להתאים את התפתחות הרגולציה הישראלית אליהם.

לבסוף, מוצע כי חלק מהפעילות של הגורם האמור תהיה מכוונת גם כלפי גורמים פרטיים. בהקשר זה הוא יוכל לסייע, בין היתר, בהעלאת המודעות לעקרונות האתיים שהוצעו לעיל ויישומם, ובהנגשת מידע ופיתוח כלים לשימוש וולונטרי על ידי גורמים פרטיים המעוניינים בכך.

6.4. <u>הקמת פורום רגולטורים ופורום לשיתוף ציבור לתחום הבינה המלאכותית</u>

מוצע למסד פורום מקצועי פנים-ממשלתי הכולל נציגי רגולטורים ומומחים לטכנולוגיה, מדיניות ומשפט, על מנת לקדם את התיאום ולדון בסוגיות משותפות; וכן פורום הכולל נציגי תעשייה, אקדמיה, ארגוני חברה אזרחית והציבור הרחב, כדי לדון בסוגיות של הסדרת בינה מלאכותית.

<u>בחלק הראשון</u>, מוצע לקבוע "שולחנות עגולים" שיתקיימו באופן עיתי או לפי הצורך בין רגולטורים וגורמי ממשלה רלוונטיים, בדגש על תחומים עתירי "חיכוך", במטרה לדון בנושאים שעל סדר היום ולבחון את הצורך בעדכון מדיניות הרגולציה לתחום הבינה המלאכותית ממבט מאקרו. במסגרת זאת ניתן יהיה לקיים למידת עמיתים בנוגע להתפתחויות הטכנולוגיות בתחום, לגישות חדשניות לרגולציה ולתובנות מהשטח. בדרך זו, הפורום יסייע באחידות יישום מדיניות רגולציה ואתיקה זו. בנוסף, ההתפתחויות המהירות בתחום הטכנולוגי מחייבות בחינה עיתית של המלצות המדיניות, ולכן מוצע שמטרה נוספת של הפורום האמור תהיה לדון בהיבטים אלה באופן קבוע.

<u>בחלק השני</u>, מוצע לקבוע פורום נוסף עם נציגי תעשייה, אקדמיה וחברה אזרחית, וכן בעלי עניין חוץ-ממשלתיים רלוונטיים נוספים, במטרה לאתר את התחומים העיקריים שבהם יש צורך בהתאמה של הרגולציה או מדיניות הרגולציה בתחום הבינה המלאכותית, כדי להסיר חסמים לפעילות כלכלית או חברתית רציפה ויעילה או לגבש מענים מותאמים לצרכים.

תחום הבינה המלאכותית מובל בידי התעשייה והאקדמיה ולכן יש חשיבות לעדכון ושיח רציף בינם לבין גורמי הממשלה. במסגרת פורום זה, ניתן יהיה לדון בהתפתחויות הטכנולוגיות, בשאלה אם כלי רגולציה מסוימים נותנים מענה לאתגרים השונים כמו גם לעקרונות שפורטו לעיל ובשאלות נוספות. כמו כן, ניתן יהיה לקיים שמיעת ציבור רציפה שתסייע לממשלה ולרגולטורים השונים

לגבש מדיניות ולבחון אותה מעת לעת. יתרון נוסף של פורום זה ושל קיום שיח פתוח ולומד, הוא בהיותו מגביר שקיפות וככזה מסייע אף להגברת אמון הציבור והוודאות הרגולטורית.

יצוין, כי מיסוד פורום קבוע תואם גם את המלצות ה-OECD לגבי רגולציה גמישה, וכי פורמט שולחנות עגולים הוצע ככלי כללי לתיאום בין רגולטורים על ידי צוות רגולציה חכמה.[270] עוד יצוין, כי ניתן להפקיד את מלאכת ריכוז פורומים אלה בידי מוקד הידע והתיאום הממשלתי שיוקם.

6.5. מיפוי השימושים בבינה מלאכותית והאתגרים הנלווים להם בענפים מוסדרים

מוצע כי גורמי הממשלה והרגולטורים הרלוונטיים יפעלו להבנה ומיפוי של השימושים הקונקרטיים במערכות מבוססות בינה מלאכותית הנעשים על ידי המפוקחים בענף המוסדר שעליו הם אמונים; האתגרים, החששות והסיכונים הכרוכים בכך; והמענים האפשריים.

כאמור, בשנים האחרונות חל גידול משמעותי בהיקף השימושים במערכות מבוססות בינה מלאכותית והן מוטמעות כמעט בכל ענף במשק. חלק ניכר מהענפים שבהם נעשה שימוש במערכות מבוססות בינה מלאכותית מוסדר באמצעות רגולציה ענפית או רוחבית (כגון רפואה ובנקאות או תחרות ופרטיות) וכפוף לרגולטור ענפי או רוחבי (כגון משרד הבריאות והפיקוח על הבנקים בבנק ישראל או רשות התחרות והרשות להגנת הפרטיות). לעיתים ההסדרה כוללת אף בחינה מראש של הרגולטור או קבלת רישיון מהרגולטור כתנאי לפעילות.

עם זאת, בשיח ראשוני שקיים הצוות הבין-משרדי שעסק ברגולציה ואתיקה במהלך גיבוש התכנית הלאומית לבינה מלאכותית עם רגולטורים שונים, עלה כי יש מקום לעריכת מיפוי של שימושים משמעותיים במערכות מבוססות בינה מלאכותית על ידי גורמים מפקחים בתחום שעליו הם אמונים ובהתאם לסמכותם לפי דין; של הסוגיות, החששות והסיכונים המתעוררים לגבי שימושים אלה, בין היתר לאור האתגרים שנמנו במסמך זה (אפליה, מעורבות אנושית, הסברתיות, גילוי, אמינות עמידות, אבטחה ובטיחות, אחריותיות ופרטיות); והמענים האפשריים.

ככל שהדבר אפשרי, ובמגבלות הדין והסמכות, ניתן יהיה להפיק תועלת רבה מפרסום לציבור של ממצאי המיפוי ומסקנות גורמי הממשלה והרגולטורים בעקבותיו. בנוסף או לחלופין, ניתן יהיה לשתף זאת עם מוקד הידע והתיאום הממשלתי שיוקם, שיוכל לסייע לרגולטורים למקד את השאלות ולייצר מהממצאים תובנות משמעותיות. בנוסף, על רקע ממצאי המיפוי יוכל מוקד הידע והתיאום הממשלתי לייצר תמונת מצב רחבה עבור הממשלה, על אודות היקף השימוש בבינה מלאכותית ואופן ההתמודדות עם סיכונים לזכויות יסוד ואינטרסים ציבורים, כמו גם על חסמים לקידום בינה מלאכותית.

יודגש כי אין בהצעה זו כדי לקרוא להתערבות רגולטורית בפועל שתשפיע על היכולת לפתח או להשתמש בבינה מלאכותית כשהדבר אינו נדרש, אלא להבנת תמונת המצב במבט צופה פני עתיד.

[270] דוח צוות הבין-משרדי לרגולציה חכמה, לעיל ה"ש 245, עמ' 118.

6.6. <u>מעורבות אקטיבית בפיתוח הרגולציה והתקינה בפורומים בינלאומיים</u>

מוצע כי גורמי ממשלה הפועלים במסגרת פורומים וארגונים בין-לאומיים, או עומדים בקשרי עבודה עם מדינות מפותחות אחרות בתחום זה יפעלו יחד לקידום מדיניות רגולציה ואתיקה מאוזנת ותואמת למדיניות רגולציה זו וליעדים הממשלתיים. זאת, בשים לב להחלטת ממשלה 212, שהטילה על שרת החדשנות, המדע והטכנולוגיה "להוביל את שיתופי הפעולה הבין-לאומיים האזרחיים בתחום, לייצג את ישראל בפורומים בין-לאומיים אזרחיים, לרבות פורום השרים של ארגון ה-OECD".

ארגונים בינלאומיים ומדינות רבות בעולם עוסקים במדיניות ביחס לבינה מלאכותית. בין הארגונים נמנים לדוגמה ה-OECD, UNESCO, CAI (ועדה של מועצת אירופה), GPAI והאיחוד האירופי.[271] עם המדינות המובילות נמנות למשל ארה"ב, בריטניה, קנדה, יפן וסינגפור. בנוסף, נעשית עבודה נמרצת בארגוני תקינה בינלאומית כמו ISO ו-IEEE לפיתוח תקנים ל"בינה מלאכותית אחראית".[272] אלה, ורבים אחרים, עוסקים בהסדרת בינה מלאכותית, ויש להניח כי לתוצרי העבודה שלהם תהיה השפעה גלובלית, כולל על הנעשה בישראל. בנוסף, ישראל משתתפת בחלק מהפורומים המערבים ארגונים בינלאומיים ומדינות אלה ומקיימת שיח עמם.

על רקע זה, בכפוף להחלטת ממשלה 212 המצוטטת לעיל, מוצע כי גורמי הממשלה הרלוונטיים ימשיכו לפעול באופן אקטיבי, בין היתר, במסגרת פורומים אלה לקידום מדיניות רגולציה ואתיקה מאוזנת ותואמת למדיניות רגולציה זו וליעדים הממשלתיים. בד בבד, מוצע כי ההשתתפות בפורומים אלה תשמש גם ככלי מרכזי ללמידה מניסיונן של המדינות השונות, להעמקת הידע בישראל בתחומים אלה, ולהבנת המגמות הבינלאומיות המסתמנות.

6.7. <u>התוויית שפה אחידה באמצעות מסגרת מומלצת (וולונטרית) לניהול סיכונים</u>

מוצע לפתח או לאמץ כלי אחיד לניהול סיכונים ביחס לשימוש בבינה מלאכותית, שייצור שפה משותפת בין גורמי הממשלה והרגולטורים ובינם לבין גורמים פרטיים. השפה האחידה תסייע למגזר הפרטי להעריך את הסיכונים הנלווים לשימוש מסוים בבינה מלאכותית, וכן לרגולטורים לבחון את הסיכונים הכרוכים בשימוש בה בתחום שעליו הם אמונים ואת הצורך בהתערבות.

באופן כללי, תפיסות ניהול הסיכונים נועדו למנוע סיכונים לזכויות יסוד ואינטרסים ציבוריים. העקרונות האתיים מסייעים בהכוונת הארגונים לגבי סוגי הסיכונים הנפוצים הנובעים משימוש בבינה מלאכותית, ב"מחזור החיים" של הפיתוח והשימוש בבינה מלאכותית.

---

[271] ארגון המאגד נציגי ממשלות, מדענים, חברות טכנולוגיה, ארגוני חברה אזרחית וארגונים בינ"ל, שמטרתו לקדם את שיתוף הפעולה הבינ"ל בנושא ולהוות מרכז ידע לסוגיות מובילות של בינה מלאכותית, במטרה לקדם מערכות בינה מלאכותית אמינות. ראו: https://gpai.ai.

[272] לסקירה על התקינה בארגון ISO ובגופי התקינה האירופים ראו:
Nativi S. (DG JRC), DE NIGRIS S. (DG JRC), AI Watch: AI Standardization Landscape State of Play and link to the EC proposal for an AI Regulatory framework, European Commission, 2021, https://op.europa.eu;
לסקירה על קידום התקינה בארה"ב ראו:
NIST, AI Risk Management Framework, https://www.nist.gov/itl/ai-risk-management-framework.

תפיסות ניהול סיכונים מהוות מכנה משותף למדיניות בינה מלאכותית במדינות המפותחות. תפיסת ניהול הסיכונים בבינה מלאכותית נשענת על תפיסות ניהול סיכונים מקובלות (שנועדו להתמודד עם אי ודאות שנוצרת כתוצאה מפעילות כלשהי),[273] תפיסות ניהול סיכונים לניהול טכנולוגיות המידע,[274] ועל ההקשר המשפטי והחברתי הקונקרטי בו מופעלת הבינה המלאכותית.

המחקר המלווה להמלצות ה-OECD מסכם באופן כללי את עקרונות-העל של ניהול סיכונים הנדרש בכל "מחזור החיים" של בינה מלאכותית:[275] (1) הגדרת יעדים – הגדרת יעדים, פונקציות או תכונות של מערכת הבינה המלאכותית; (2) זיהוי בעלי ענין וגורמים מעורבים; (3) הערכת סיכונים – הערכה של ההשפעות הפוטנציאליות, הן התועלות והן הסיכונים לבעלי העניין והגורמים המעורבים; (4) צמצום סיכונים – התאמה של אסטרטגיות צמצום סיכונים לסוג הסיכונים; (5) יישום האמצעים לצמצום סיכונים; (6) בקרה, הערכה ומשוב.[276]

חלק משמעותי בניהול הסיכונים הוא קיום תהליכים סדורים ומתועדים, המאפשרים לבחון את אופן יישום העקרונות בסיטואציה שבה נעשה שימוש במערכות מבוססות בינה מלאכותית. בנוסף, תהליכים אלה מאפשרים להתחקות אחר תהליך הפיתוח והשימוש, ולבקר היבטים אלה במידת הצורך כדי להפיק לקחים, לתקן החלטות שגויות, ולשפר את הליך קבלת ההחלטות, באופן רציף. עקב כך הם מדגישים תיעוד, עבודה בהתאם לכללים מקובלים בתחום וקיום נהלים.

עקרונות אלה הם כלליים ומהווים מכנה משותף רחב בתחום ניהול הסיכונים. לצד זאת ניתן לראות במדינות העולם ובגופי התקינה תפיסות המרחיבות היבטים אלה.

טיוטת החקיקה האירופית כוללת שילוב של דרישות ניהול סיכונים שמוצע לקבוע בחוק, והשלמה של דרישות אלה באמצעות תקינה מתאימה. סעיף 9 לטיוטת החקיקה של האיחוד האירופי כולל פירוט של הרכיבים הנדרשים למנגנון ניהול סיכונים עבור מערכות בינה מלאכותית המסווגות

---

[273] המחלקה העוסקת בבינה מלאכותית ב- OECD מבססת את עבודתה על תפיסות ניהול הסיכונים המקובלות בסדרת תקני ISO בנושא ניהול סיכונים:
ISO, ISO 31000, *Risk management – Guidelines*, https://www.iso.org
"*ISO 31000, Risk management – Guidelines, provides principles, a framework and a process for managing risk. It can be used by any organization regardless of its size, activity or sector.*
Using ISO 31000 can help organizations increase the likelihood of achieving objectives, improve the identification of opportunities and threats and effectively allocate and use resources for risk treatment.
However, ISO 31000 cannot be used for certification purposes, but does provide guidance for internal or external audit programmes. Organizations using it can compare their risk management practices with an internationally recognized benchmark, providing sound principles for effective management and corporate governance.";
ISO 31000 - Risk management, https://www.iso.org/publication
"ISO 31000 provides direction on how companies can integrate risk-based decision making into an organization's governance, planning, management, reporting, policies, values and culture. It is an open, principles-based system, meaning it enables organizations to apply the principles in the standard to the organizational context.";
התקן אומץ כתקן ישראלי: ת"י 31000 חלק 1, ניהול סיכונים: הנחיות למימוש התקן הישראלי ת"י 31000-ניהול סיכונים-עקרונות וקווים מנחים, https://www.sii.org.il.
תקנים אלה שימשו את ה- OECD גם בגיבוש המלצות לניהול סיכונים **בתחום הגנת הסייבר**, OECD, Digital Security Risk Management, https://www.oecd.org.

[274] ראו למשל את סדרת תקני ISO בתחום ניהול סיכוני סייבר:
ISO/IEC 27001 Information security management, https://www.iso.org.
הדגש של תקנים אלה הוא במערכת ניהול טכנולוגיית המידע כדי לנהל את סיכוני האבטחה.
התקן אומץ כתקן ישראלי:
ת"י - ISO/IEC 27001 מערכת ניהול אבטחת מידע, https://www.sii.org.il/he/iso-27001.

[275] OECD, AI in Society, p. 95-96.

[276] שם.

ב"סיכון גבוה".[277] לצד זאת, מוצע בטיוטת החקיקה, לגבי שימושים מסוימים, כי ארגון שעומד בתקינה שנקבעה בהתאם לחקיקה יהיה בעל "חזקת" עמידה בהוראות החוק.[278] מכאן יוצא שהנחת העבודה היא שהתקינה תהיה מפורטת יותר, ותאפשר מצד אחד ודאות לגבי אופן עמידה בה, ומצד שני לארגונים שיבחרו לעמוד בדרישות ניהול הסיכונים החקוקות בדרך אחרת, לעשות זאת.[279] טיוטת החקיקה כוללת גם דרישה לביצוע "בדיקת התאמה" ("conformity assessment") לחקיקה למערכות ב"סיכון גבוה".[280]

בעקרונות לאמנת מועצת אירופה (ראו פרק "מועצת אירופה וקידום אמנה בנושא בינה מלאכותית") יש שילוב ניהול סיכונים, ולצד זאת מוצע גם "תסקיר השפעה אלגוריתמי"[281] ו"עיצוב טכנולוגי ונהלי מראש".[282] השימוש בתסקיר השפעה נועד לחשוף את הסיכונים האפשריים גם לזכויות המהותיות, לנהל שיח לגביהם, לקבוע דרכי התמודדות ואת הסיכון שנותר לאחר מכן, ולקבל הכרעה ערכית לגבי השאלה אם הסיכון לאחר נקיטת אמצעי ההתמודדות הוא סביר.

תפיסות ניהול סיכונים באות לידי ביטוי בשני מובנים מרכזים. <u>מובן ראשון</u> עוסק בנקודת מבטו של הרגולטור. תפיסות ניהול סיכונים ביחס לבינה מלאכותית משמשות גורמי ממשלה ורגולטורים, לגיבוש רגולציה ומדיניות רגולציה. מובן זה מקושר גם לחוק עקרונות האסדרה, בכך שהוא משמש כלי עבודה לבחינת ההצדקה להתערבות בפעילות המגזר הפרטי. כך, בהתאם לתפיסת ניהול סיכונים, ההצדקה לקביעת כללי התנהגות לפיתוח ולשימוש בבינה מלאכותית גוברת במקום שבו הסיכון לזכויות יסוד או אינטרס ציבורי הוא גבוה. בהתאמה, ככל שהסיכון נמוך יותר, ההצדקה להתערבות מצטמצמת, וכללי ההתנהגות מהווים המלצה לאופן ההתמודדות הארגוני עם הסיכון.

<u>מובן שני</u> עוסק בנקודת מבטם של גורמים פרטיים המפתחים או משתמשים בטכנולוגיות מבוססות בינה מלאכותית. מובן זה עוסק בתפיסות מקובלות לניהול סיכונים (כמפורט להלן), בהתאם לעקרונות האתיים, כדי לאתר את הסיכונים הנובעים מבינה מלאכותית לזכויות יסוד ואינטרסים ציבוריים, ולנקוט באמצעים טכנולוגים ונהליים לצמצם אותם לרמה מקובלת. ביצוע האמור בצורה שיטתית גם מאפשר להפיק מידע בצורה אחידה על הבינה המלאכותית ואופן ניהול הסיכונים שלה. מידע זה רלוונטי לגורמים נוספים הקשורים בשימוש בבינה המלאכותית, כגון שותפים עסקיים או לקוחות, הנדרשים לקבל החלטות ולנהל את הסיכונים הקשורים בפעילות שלהם בנושא זה.

הממשק בין שני מובנים אלה הוא שתפיסת ניהול הסיכונים מסייעת להפעלת סמכות הרגולטור בין היתר כדי לבחון האם ההתמודדות הארגונית הוולונטרית מספקת מענה הולם לסיכון לפגיעה

---

[277] EU Draft AI Regulation, Article 9 "Risk Management System".

[278] EU Draft AI Regulation, Article 42 " Presumption of conformity with certain requirements".

[279] לפרשנויות על אודות תהליכי העבודה המפורטים הנדרשים כדי לעמוד בחקיקה האירופאית ראו: Floridi, Luciano and Holweg, Matthias and Taddeo, Mariarosaria and Amaya Silva, Javier and Mökander, Jakob and Wen, Yuni, capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act (March 23, 2022). Available at SSRN: https://ssrn.com/abstract or http://dx.doi.org/10.2139/.

[280] EU Draft AI Regulation, Article 43; ליחס שבין "בדיקת התאמה" לתפיסה המוכרת בחקיקת הפרטיות של "תסקיר השפעה על הפרטיות", ראו: Katerina Demetzou, Introduction to the Conformity Assessment under the draft EU AI Act, and how it compares to DPIAS, Future of Privacy Forum, 12.08.2022, https://fpf.org/blog.

[281] CAHAI section 45-51.

[282] CAHAI 2021, section 57-62.

בזכויות יסוד או באינטרסים ציבוריים. בנוסף, תפיסת ניהול הסיכונים מהווה שפה משותפת לשיח בין הרגולטור לבין הארגון.

בהערת אגב יצוין, כי בתחומים אחרים של ניהול סיכוני טכנולוגיית מידע, כגון בתחום הגנת הסייבר ובתחום ההגנה על הפרטיות, התפתחו תקנים ושיטות עבודה שמטרתם ניהול שיטתי ורציף של הסיכון ומניעת התממשותו.

כמצוין לעיל, יש תועלת בקביעת שפה משותפת וכלי עבודה דומים לצורך קידום המדיניות הרגולטורית באופן אחיד, ויישום העקרונות על ידי גורמים פרטיים בענפי המשק השונים. כלי אחד ליצירת השפה המשותפת הוא העקרונות האתיים המוצעים לעיל (לפיו למשל הרגולטור והמפוקח מכירים שניהם את החשש לאפליה ויודעים מהי הסברתיות). כלי נוסף להתוויית שפה משותפת, שמוצע כאן, הוא תפיסת ניהול הסיכונים (שלפיה הם יוכלו למשל לבחון באופן דומה מה נחשב לשימוש בסיכון גבוה או נמוך, ובאיזה סוג של שימוש ניכרת שכיחות גבוהה של סיכונים).

לקיומה של שפה משותפת גם במסגרת ניהול הסיכונים מספר יתרונות, ובכלל זה, קידום שפה אחידה להתמודדות עם טכנולוגיה מסוימת; קידום תפיסה קוהרנטית לגבי פעילות שחוצה תחומי סמכות רגולטוריים; שמירה על זיקה בין תפיסות ניהול סיכונים בתחום הבינה המלאכותית להוראות אחרות בתחום ניהול הסיכונים, כדי להקל על ארגונים בנטל ניהול הסיכונים החל עליהם; אפשרות פיתוח תחום בודקי הסיכונים באופן אחיד; סיוע לרגולטורים בבחינת התחומים הדורשים את התערבותם; והתאמה לכללים המתקדמים בשווקים של המדינות המפותחות. שפה משותפת תאפשר גם לארגון להעביר את המידע הנדרש לרגולטור לצורך בחינת הצורך בהתערבות.

כפי שתואר לעיל, טיוטת הרגולציה של האיחוד האירופי קובעת במפורש את רמות הסיכון השונות. היא מבחינה בין מערכות ב"סיכון בלתי מקובל", "סיכון גבוה", "סיכון מוגבל", ו"סיכון מינימלי", ומגדירה את התהליך לסיווג המערכות לתוך קטגוריות אלה. גם ה-OECD פעל לייצר תבחינים לסיווג מערכות בינה מלאכותית, אשר נועדו למפות את הרכיבים השונים במערכות בינה מלאכותית שעשויים להצביע על מידת רגישותן. ואולם, בשונה מהשימוש בתבחינים באיחוד האירופי, אלה של ה-OECD אינם כוללים הכרעה ערכית לגבי מהי רמת סיכון גבוהה.

על רקע זה, מוצע כי במסגרת תפיסה אחידה לניהול סיכונים, תיקבע שיטה הכוללת תבחינים אחידים למיון מערכות בינה מלאכותית. עם זאת, מוצע כי הקביעה הנורמטיבית לגבי שילוב התבחינים העוברים את הסף שעשוי להצדיק התערבות רגולטוריות, תיעשה על ידי כל רגולטור ברמה הענפית ובהתאם לסביבה המשפטית והרגולטורית החלה. בהקשר זה יודגש כי מוקד הבחינה אינו בהכרח הטכנולוגיה ככזו, אלא סוג השימוש בה. זאת, משום שמדובר בטכנולוגיה הניתנת ליישום בהקשרים שונים.

לאור האמור לעיל, מוצע לפתח או לאמץ כלי אחיד לניהול סיכונים, שישמש הן את גורמי הממשלה והרגולטורים והן גורמים פרטיים, לשם הערכת הסיכון ביחס לשימושים מסוימים במערכות מבוססות בינה מלאכותית.