

בינה מלאכותית בשירותי ממשל: הטמעת מערכות לקבלת החלטות מבוססות-אלגוריתם בשירותי הרווחה

ד"ר סיון תמיר

השתתפה בכתיבה: עו"ד לימור שמרלינג מגזניק

ליווי וניהול מקצועי (ג'וינט-אלכא): גב' נועה מילר

דצמבר 2020

Artificial Intelligence in Government: Implementing Algorithmic Decision-Making Systems in Welfare Services

Author: **Sivan Tamir**

Writing contributor: **Limor Shmerling Magazanik**

Professional project management (Joint-Elca): **Noa Miller**

המכון הישראלי למדיניות טכנולוגיה; שדרות רוטשילד 74-76, תל אביב יפו, 6521401;
info@techpolicy.org.il / www.techpolicy.org.il
ג'וינט-אלכא; גבעת הג'וינט, רופין 9, ירושלים, 95435;
<https://www.thejoint.org.il/elka>

© תשפ"א – 2020, המכון הישראלי למדיניות טכנולוגיה וג'וינט-אלכא

אודות המכון הישראלי למדיניות טכנולוגיה

המכון הישראלי למדיניות טכנולוגיה הוא מכון מדיניות וצוות חשיבה (think tank), הפועל לעיצוב מדיניות בסוגיות חברתיות, אתיות ומשפטיות מורכבות בעידן הטכנולוגי. המכון מהווה פלטפורמה לכינוס קהילות וגיבוש שיח בין בעלי עניין שונים ובכלל זה, ממשלה, תעשייה, אקדמיה וחברה אזרחית, לצורך יצירת ניירות עמדה, מאמרים, וכללי התנהגות מוסכמים (best practices) בתחומי מדיניות טכנולוגיה, וקידום פרקטיקות מעשיות ליישום הגנה על זכויות אדם תוך קידום חדשנות.

המכון מהווה את השלוחה הישראלית של 'הפורום לעתיד הפרטיות' (Future of Privacy Forum) [FPF] בארה"ב, מכון מחקר ומדיניות המתמחה בתחום הגנת הפרטיות וזכויות אדם, בהקשר של טכנולוגיות חדשניות. **ה-FPF נתמך** על ידי יותר מ- 200 חברות רב-לאומיות, העוסקות בתחומים שונים, כולל טכנולוגיה, אינטרנט, תקשורת, פיננסים, בריאות, קמעונאות, ועוד. בנוסף לחברות מן המגזר הפרטי, נתמך ה-FPF בידי קרנות מחקר ממשלתיות ופרטיות.

על חברי הוועדה המייעצת של המכון נמנים אנשי אקדמיה, נציגי ממשל ותעשייה ואנשי החברה האזרחית.

אודות הכותבות

ד"ר סיון תמיר מתמחה בתחום המשפט והביואתיקה, גנ-אתיקה (Genethics) ואתיקה של טכנולוגיות חדשניות. היא חוקרת בכירה ויועצת מדיניות במכון הישראלי למדיניות טכנולוגיה, ומרצה באקדמיה ומחוצה לה.

עו"ד לימור שמרלינג מגזניק מומחית בפיתוח וקידום מדיניות טכנולוגיה ומידע, ומתמחה בתחום הגנת הפרטיות, ובפיתוח וקידום מדיניות בתחומים מונעי-טכנולוגיה (technology-driven) ומידע. היא עומדת בראש המכון הישראלי למדיניות טכנולוגיה ועמיתת מחקר בכירה ב-Future of Privacy Forum.

אודות ג'וינט-אלכא

ג'וינט-אלכא מקדם ומאפשר ממשקים אפקטיביים, בין ובתוך המגזרים השונים (השלטון המרכזי והמקומי, המגזר השלישי והעסקי) העוסקים במתן שירותים ציבוריים וחברתיים.

אלכא מפתחת פתרונות לבעיות המושרשות עמוק במערכות הציבוריות ומגבילות את היכולת לספק שירותים חברתיים יעילים ומועילים. אתגרים אלה אינם ייחודיים לבעיה חברתית ספציפית, אלא מאפיינים מגוון בעיות ומשפיעים על מערכת השירות הציבורי כולה.

אלכא מתמקדת בארבעה אתגרי מערכת:

- דיגיטציה – הובלת טרנספורמציה דיגיטלית ופיתוח היכולת לאמץ חדשנות דיגיטלית במגזר הציבורי בתהליכי יצירה ואספקת שירותים חברתיים.
- מיצוי משאבים – פיתוח יכולת השלטון המקומי כדי למצות ולנצל באופן מלא ואפקטיבי משאבים שהוקצו על ידי השלטון המרכזי, ופיתוח האזוריות בישראל.
- שיתופי פעולה בין מגזריים – קידום שיתופי פעולה ושותפויות לשיפור השירותים הציבוריים בין ובתוך שלושת המגזרים.
- הבטחת איכות – קידום שקיפות והכללת כלל האוכלוסייה כדי להבטיח את איכות השירותים החברתיים, ששמים את האזרח במרכז ומספקים לאזרחים פתרונות הוליסטיים לצרכים חברתיים.

בנוסף, מובילה אלכא בפיתוח רשת של מנהיגות מתנדבת וקידום מעורבותם של אנשי עסקים במגזר החברתי.

ג'וינט ישראל הוא ה-R&D החברתי של ישראל, הפועל לקידום יכולתה של החברה בישראל לדאוג לרווחת כל אזרחיה, לפעול לצמצום פערים ולקדם חוסן וצדק חברתי. הארגון יוזם, מפתח ומטמיע שירותים, תפיסות ותכניות חדשניים ויעילים עבור אוכלוסיות פגיעות בישראל ופועל לקידום וחיזוק מעמדן. זאת ברוח עקרונות הג'וינט העולמי ובשותפות עם הממשלה, רשויות מקומיות, החברה האזרחית והמגזר העסקי.

תוכן העניינים

4	תקציר מנהלים
8	מורה נבוכים טכנולוגי בסיסי
8	מילון מונחים
9	המגמה: מאלגוריתם בסיסי לבינה מלאכותית.
10	רקע
13	מטרות מערכת ADM ומאפייניה
15	עקרונות בינה מלאכותית (AI Principles)
17	ההיבט המנהלי-ציבורי
17	הטמעת מערכות ADM בשירות הציבורי
20	אתגרי אימוץ מערכות ADM במגזר הממשלתי/ציבורי
20	אתגרים כלליים
20	אתגרי אימוץ הטמעה ויישום של טכנולוגיות מבוססות בינה מלאכותית במנהל הציבורי
23	יישום מערכות ADM בשירותי הרווחה
26	יישום מערכות ADM בשירותי הרווחה – קונפליקטים טיפוסיים.
26	קונפליקט בין המדינה לפרט – אחריות המדינה לרווחתו של הפרט ביישום מערכת ADM
26	קונפליקט בין המדיניות הממשלתית/מנהלית לבין משימתו החברתית ומחויבותו המוסרית של העובד הסוציאלי
28	יתרונות ותועלות מעשיים ואתיים בקבלת החלטות מבוססות-אלגוריתם
31	חסרונות וקשיים מעשיים ואתיים בקבלת החלטות מבוססות-אלגוריתם
31	זכויות פרט בסיכון.
33	קשיים אתיים אינהרנטיים לאופיין של מערכות ADM.
38	מסקנות והמלצות
38	המלצות כלליות
39	השלב המקדמי – טרום אימוץ מערכת ADM בשירות הציבורי
39	תנאי סף והצדקות לאימוץ מערכת בינה מלאכותית בשירותי הרווחה
40	תכנון מערכת ה- ADM או התאמתה ליישום בשירות הציבורי/שירותי הרווחה.
42	השלב המתקדם – עם אימוצה של מערכת ADM בשירות הציבורי
43	סיכום

תקציר מנהלים

רקע

היצף נתוני העתק המאפיין את תקופתינו, בשילוב עם טכנולוגיות דיגיטליות מתקדמות ופיתוחם של כלי ניתוח וניבוי מבוססי בינה מלאכותית, מבשרים באופן בלתי נמנע על הגעתו של עידן המשילות הדיגיטלית. כשהממשלה מאמצת ביטויים אלו של קדמה טכנולוגית, ומטמיעה אותם בתוך שירותיה לציבור האזרחים – חשוב להבטיח קיומו של אינטרס כפול: זה של הממשלה – ביעול עבודת מנגנון השירות לאזרח, וזה של האזרחים – בקבלת שירות משופר או מיטבי. אולם הבחירה (האמיצה והמאתגרת) של ממשלות בהטמעת מערכות בינה מלאכותית בשירות הציבורי, אינה מהווה אך ביטוי למדיניות הממשלתית של אימוץ חדשנות לטובת התייעלות, התאמה לקידמה וכיו"ב תועלות. היא מגיעה גם בתגובה לציפיית הציבור להשוואת רמת השירות, חדשנותו וממשקיו עם האזרח, לאלו המתקבלים מספקיות שירותים במגזר הפרטי.

סקירה זו, משקפת את המגמה ההולכת ומתגבשת היום בממשלים של מדינות מפותחות, כולל במדינת ישראל, המבטאת פתיחות כלפי רעיון הדיגיטציה של שירותים ממשלתיים והפיכתם ל"טכנולוגיים" יותר, מתוך הכרה ביתרונות אימוץ החדשנות ברמת הממשל, ובהנחה כי אלו עולים בקנה אחד עם האינטרס הציבורי. מסגרת הכללית של הסקירה היא יישומי טכנולוגיית הבינה המלאכותית בשירות הציבורי, תוך התמקדות במערכות לקבלת החלטות מבוססות-אלגוריתם (Algorithmic Decision-Making; להלן – "מערכות ADM"). מערכות אלו מפותחות לטובת שיפור ויעול תהליך קבלת החלטות, למגוון ייעודים במגזר הפרטי והציבורי. הן מסייעות לניתוח ותהליך קבלת ההחלטות האנושי ולעיתים, אף מחליפות את הגורם האנושי מקבל ההחלטות.

ההיבט המנהלי-ציבורי

חלק זה של הסקירה מוקדש לעצם הטמעת מערכות ADM, בשירות הציבורי בכלל ובשירותי הרווחה, בפרט. השירות הציבורי מהווה זירה חדשה יחסית של יישום בינה מלאכותית, והטמעת מערכות ADM בשירות הציבורי בכלל (בתחומי הבריאות, השיטור, המשפט, התחבורה, החינוך ועוד) נמצאת במגמת עלייה. פיתוח, יישום והטמעה של מערכות בינה מלאכותית מבוססות נתוני עתק בשירות הציבורי, מבוצע לרוב באמצעות מיקור-חוץ של שירותים ציבוריים למגזר הפרטי. את הרציונאל של תהליך דיגיטציה מתקדם זה, ניתן להבין באופן טוב יותר על רקע מרכזיותה של הטכנולוגיה כמחוללת שינוי ומקרבת קידמה, ולנוכח ההתפתחות שעוברת תפישת השירות הציבורי בשנים האחרונות. זו מוצאת את ביטויה בעיקר בפרדיגמה החדשה המתהווה בתחום ניהול השירות הציבורי – זו של המשילות הציבורית החדשה (New Public Governance), שעל יסודותיה נעמוד בסקירה.

יישום טכנולוגיות בינה מלאכותית במגזר הציבורי הוא תהליך רצוף אתגרי, זאת בעיקר בשל היותו חדש ולא-טבעי למערכת הממשלתית-ציבורית. אתגרים אלו, המפורטים בגוף הסקירה, באים לידי ביטוי במספר מישורים: הטכנולוגי, החוקי, האתי, והחברתי. אתגרים אחרים, ספציפיים יותר, עולים בהקשר של עצם מיקור החוץ של שירותי רווחה באמצעות התקשרויות עם ספקי טכנולוגיה חוץ-ממשלתיים, מסחריים, מן המגזר הפרטי. בעוד שמטרת יישום טכנולוגיות מבוססות בינה מלאכותית היא הבאת חדשנות טכנולוגית והתייעלות אל תוך השירות הממשלתי לאזרח, ובעוד מיקור החוץ מספק פתרון למגבלות וחסמים בפיתוחן של מערכות ה-ADM בתוך השירות הציבורי (כאלה הקשורים במומחיות טכנולוגית מוגבלת של מפתחים ובאוריינות דיגיטלית לא-מספקת של נותני השירות) – הדבר מעורר מספר אתגרים ספציפיים, כגון: בעיית פרטיות ואמון, וקושי בזיהוי האחראי לנזק-יציר-מערכת ADM, אליהם מתייחסת הסקירה.

יישום מערכות ADM בשירותי הרווחה

ככלל, נראה כי בשנים האחרונות, מערכת הרווחה מהווה אחת מזירות היישום הרחוקות יותר של הטמעת טכנולוגיות מונעות-נתונים (data-driven) ומבוססות בינה מלאכותית, למתן שירותי רווחה שונים. ניתן להצביע על שני תחומי יישום עיקריים של מערכות אלו בשירותי הרווחה, בעת הזו: האחד – תחום קצבאות הרווחה (קביעת זכאות לקצבאות; וזיהוי ואיתור הונאת קצבאות רווחה (welfare fraud)). השני – בהגנה על ילדים ואוכלוסיות בסיכון (איתור ילדים קורבן להתעללות וזיהוי המשפחות לצורך נקיטת התערבות מוקדמת; ומניעת ניצול אוכלוסיות מוחלשות והדרדרותן לעבריינות).

בשנים האחרונות, נעשים במדינות שונות נסיונות בהטמעת מערכות ADM במגוון שירותי רווחה, אשר הוכתרו במידות שונות, או מעורבות, של הצלחה וכישלון. בסקירה מתוארות מספר דוגמאות של נסיונות כאלה מבריטניה, דנמרק וארה"ב.

יתרונות ותועלות אל מול חסרונות קשיים מעשיים ואתיים בקבלת החלטות מבוססות-אלגוריתם

ליבת הסקירה היא זיהוי והצגת היתרונות והתועלות המעשיים, כמו גם החסרונות והקשיים, ביישום מערכות לקבלת החלטות מבוססות-אלגוריתם בשירות הציבורי. על היתרונות והתועלות המעשיים, נמנים: אובייקטיביות וניטרליות, יעילות, אחריות, פרסונליזציה, מניעת (או מזער) נזקים עתידיים, הגברת אמון הציבור, יתרונות מעשיים עבור העוסקים במלאכה – אנשי מערכת הרווחה, ועוד. החסרונות והקשיים הכרוכים בדיגיטציה של מערכת הרווחה ושימוש בכלים מבוססי-בינה-מלאכותית לצורך קבלת החלטות ביחס לבני-אדם, בייחוד בכל הקשור לזכויותיהם ולזכויותיהם של מטופלי רווחה, הינן בעלות השפעה שלילית פוטנציאלית על מימושן של מגוון זכויות פרט וביניהן: כבוד האדם, הזכות לאוטונומיה, הזכות לפרטיות, הזכות לבטחון סוציאלי, הזכות להגנות (זכותו של נתין האלגוריתם ליחס שוויוני וצדק על-ידי מערכת, הזכות להליך הוגן פרוצדוראלי), ועוד. בנוסף על אלו, קיימים גם קשיים אתיים אינהרנטיים לאופיין של מערכות ADM, המשליכים על מימוש זכויות הפרט המנויות לעיל, או על ההגנה עליהן, כגון: אי-הסברתיות (Non-explicability) ובעיית הקופסא השחורה ('Black box' problem), קושי בהוכחת נזק על-ידי הנפגע מיישום מערכת ADM, ואחריות ואחריותיות.

המלצות

המלצות כלליות

- ליווי הטמעת מערכת ADM בשירותי הרווחה בתכניות הנגשה של כישורים ותשתיות.** המציאות הנוכחית מצריכה קידום וחיזוק האוריינות הטכנולוגית והדיגיטלית בקרב עובדי השירות הציבורי בכלל ועובדי מערכת הרווחה, בפרט. עובדי מערכת הרווחה, שהכשרתם הסתיימה לפני המהפיכה הדיגיטלית, או במהלכה, ואף עובדי מערכת הרווחה הטריים יותר – לא הוכשרו להתמודד עם דיגיטציה של המערכת ועם עבודה לצד מערכות אוטומטיות לקבלת החלטות. תכניות הכשרה והנגשת תשתיות יסייעו הן בהתמודדות עם בעיית תת-הכשרתיות של עובדים (ביחס לכישורים הנדרשים לביצוע משימות בסביבות דיגיטליות חכמות) והן, למניעת ספק עצמי-מקצועי (דהיינו, ערעור הבטחון ביכולת ובניסיון המקצועיים) של עובדי המערכת הציבורית, לנוכח הטיית האוטומציה הקיימת.
- קביעת **טקסונומיה** – סיווג לסוגי מערכות ADM בשירות מערכת הרווחה בהן יעשה שימוש כאלגוריתם תומך-החלטה בלבד, וכאלה בהן ניתן להשתמש באלגוריתם מחליט.

3. **צמצום הטיית** במערכת הבינה המלאכותית ובקרב מיישמייה. לשם כך, הממשלה או הרשות הציבורית הרלבנטית, נדרשות לבחון דרכים ל'ניקוי' הנתונים ולהבטחת הגיון האנושי הרב ביותר, הן בשלב בניית המערכת ופיתוחה, והן בלב הטמעתה ויישומה בשירות הציבורי.

4. **בניית פרקטיקות רכש ממשלתיות אחריות, ייעודיות** לטכנולוגיות מבוססות בינה מלאכותית. לאחר שגוף ממשלתי קיבל החלטה לפיה פתרון מבוסס-בינה מלאכותית מהווה המענה הנכון לקושי או ללקונה בשירות הציבורי, מן הראוי לפעול בהתאם לפרקטיקת רכש סדורה ומותאמת, על מנת להבטיח שההליך עצמו, כמו גם המוצר הטכנולוגי הסופי, יהיו ראויים ומועילים. ממשלת ישראל יכולה לגבש פרקטיקות כאלה באופן עצמאי ומותאם למדיניות הציבורית המקומית, או להסתייע לשם הבניית פרקטיקות כאלו בהנחיות שפותחו על-ידי גופים שונים בעולם.

השלב המקדמי - טרום אימוץ מערכת ADM בשירות הציבורי

א. תנאי סף והצדקות לאימוץ מערכת בינה מלאכותית בשירותי הרווחה:

1. **בחינת כדאיות.** כשנבחנת אפשרות לאימוץ מערכת ADM והטמעתה בשירות הציבורי, על מקבלי ההחלטות לבחון את המרכיבים הבאים, אל מול החלופה האנושית הקיימת: עלויות הטכנולוגיה, עלויות הטמעתה (עלויות ישירות ועקיפות, כגון הכשרת עובדי מערכת הרווחה להבנת פעולת המערכת ותוצריה), עלויות תחזוקתה, התמודדות עם מערכת סיכונים ועלות הטעות.

2. **הכנת הצדקה לבחירת השימוש במערכת ADM.** נדרש תהליך חשיבה ושקלול פנים-מערכת, להערכת ההשפעה הפוטנציאלית של יישום מערכת ADM על 'לקוחות' השירות הציבורי - מטופלי הרווחה, במקרה שלפנינו. במסגרת זו, על קובעי המדיניות במערכת הרווחה לשקול: האם השפעת המערכת תהא מיידית, באילו תחומים היא עדיפה על פני המודל הקיים מבוסס-האדם, וכיצד היא עתידה לשפר את המערכת הקיימת.

3. **בחינת 'צדקת' האלגוריתם (Algorithmic Equity).** מדובר בבחינה פרואקטיבית מורכבת, המחפשת ודאות או בטחון לגבי הוגנות האלגוריתם וישרתו באופן פעולתו. לשם קיום בחינה מעין זו, ניתן לגייס מומחים כגון 'ציידי הטיית' (bias bounties), אשר יבחנו את עמידת האלגוריתם והמערכת כולה בעקרונות האתיקה של בינה מלאכותית, המתמקדים בצדק חברתי, הוגנות ואי-הפליה (שוויון, מניעת סטיגמטיזציה). לחילופין, ניתן להיעזר בכלים ייעודיים שונים שפותחו לבחינת צדקת האלגוריתם, כגון ה- Algorithmic Equity Toolkit, שפותח על-ידי האגודה האמריקנית לזכויות האזרח (ACLU), בשיתוף עם ה- Tech Fairness Coalition.

ב. תכנון מערכת ה- ADM או התאמתה ליישום בשירות הציבורי/שירותי הרווחה:

1. **אספקת אופציה שאינה דיגיטלית לשירות נתון.** במקרים בהם מיושמת מערכת מבוססת-בינה מלאכותית ככלי לשירות האזרח (להבדיל מכלי פנים-משרדי, הבא לסייע לפקידי השירות הציבורי בביצוע תפקידיהם), על הרשות לספק חלופה 'אנלוגית' לשירות, לטובת אותם אזרחים נעדרי נגישות טכנולוגית או בעלי אוריינות דיגיטלית נמוכה. הדבר מתבקש לטובת שמירת כבוד האדם של האזרח, מטופל מערכת הרווחה (במקרה שלפנינו), וחיוני על מנת לאפשר מימוש ומיצוי זכויות.

2. **תכנון ופיתוח עצמי, פנים-ארגוני.** פיתוח המערכת על-ידי המדינה, או הרשות המנהלית המבקשת להטמיעה, להבדיל ממיקור-חוץ של פיתוח המערכת לחברה חוץ-ממשלתית או מרכישת מערכת קיימת, תוך ביצוע התאמות מסויימות לצרכי השירות הציבורי הנתון. אף-על-פי שהדבר דורש

התגברות על חסם האוריינות הדיגיטלית במגזר הציבורי, לפיתוח מערכת ADM בתוך הרשות, או לחילופין – למעורבות גבוהה מצידה בפיתוחה, ישנו ערך רב וסיכויים מיטביים להתאמת המערכת לשירות הייעודי ולהצלחת יישומה.

3. **שיתוף אוכלוסיית היעד המקצועית** (פקידי הרווחה) בשלבי הפיתוח, להבטחת דרכי טיפול ההולמות את הקשיים, את היכולות והצרכים של אנשי הרווחה, המטפלים.

4. **שיתוף אוכלוסיית היעד הטיפולית** (מטופלי הרווחה) בתכנון המערכת ובעיצובה (שלבי הפיתוח), להבטחת רמת התאמת לבעיות האמיתיות של אותה אוכלוסייה.

5. **בחירה איכותנית של סוג הפלט (output) שמערכת ADM תספק:** האם מדד-סיכון (risk score) – שהינו סמכותני יותר ועשוי לאותת לגורם האנושי שהמכונה 'פיצחה' באופן מדויק יותר את רמת הסיכון הנשקפת למטופל הרווחה, מאשר הגורם האנושי, או – דו"ח ענייני, המייצר התראה במקרה של איתור סיכוי גבוה יחסית לסיכון ומותיר מרחב גדול יותר לשיקול הדעת של הגורם האנושי.

השלב המתקדם – עם אימוצה של מערכת ADM בשירות הציבורי

1. **שמירת מרכיב של נוכחות ואחריות אנושית בתהליך** ("Human-in-the-loop") הינה חיונית, במיוחד בזירת השירות הציבורי. מעורבות פקידי הציבור, אנשי מערכת הרווחה, בהחלטות המתקבלות על-ידי מערכות ADM, תאפשר או תבטיח רציפות של פעולת מערכת הרווחה במסגרת סמכותה, שמירה על רצף טיפולי, שליטה מסוימת, מעקב ופיקוח פנימי על פעילות המערכת האוטומטית, ועוד.

2. **גיבוש מערכות בקרה ויצירת תהליכים להערכה ביצועית** של הטכנולוגיה מבוססת-בינה מלאכותית. כינון מנגנוני פיקוח על פעילות מערכות אלה, מבטא אחריות ותקינות מנהלית, נכונות לבחינה עצמית ועשוי להבטיח, בין היתר, פעולת המערכת בהתאם לייעודה, שמירת זכויות הפרטים – נתיני המערכת, ועמידה בנורמות מנהל תקין.

3. **כינון מנגנון ערעור** על החלטות שירותי הרווחה, המתבצעות או נתמכות על-ידי מערכת ADM ויידוע האזרח על קיומו של המנגנון ודרכי הגישה אליו.

מורה נבוכים טכנולוגי בסיסי

מילון מונחים

אלגוריתם (Algorithm) – תהליך חד-משמעי, מורכב ממערכת הוראות וחוקים, המעבד קלט (נתונים; input) ומוציא פלט (מסקנות/החלטות; output), במטרה לפתור בעיה או קבוצת בעיות.

בינה מלאכותית (Artificial intelligence) – בינה מלאכותית היא מושג מאתגר להגדרה, אשר זוכה להצעות הגדרה מגוונות. להלן כמה מהן:

- ענף של מדעי המחשב, העוסק בהדמיית התנהגות אנושית על-ידי מחשבים.
- מערכת ממוחשבת המסוגלת לבצע משימות המצריכות בשגרה אינטליגנציה אנושית, בין היתר – קבלת החלטות.
- יכולתה של מכונה לחקות התנהגות אנושית אינטליגנטית.
- טכנולוגיה, מערכת אוטונומית, המסיקה מסקנות באמצעים סטטיסטיים ולומדת דפוסים מתוך נתוני עתק (big data) המוזנים לתוכה והמשמשים לאימונה, העולה בביצועיה על אלו של בני אנוש.¹

למידת מכונה (Machine learning) – תת-סוג, יישום של בינה מלאכותית, המבוסס על הרעיון שבהינתן גישה לנתונים, מכונות מסוגלות ללמוד בעצמן כיצד לבצע משימות שונות. מדובר למעשה במערכת אלגוריתמים המוזנת במידע מובנה והמסוגלת להשלים משימה, מבלי שתוכנתה לכך באופן מכוון.²

1. A.I. Wiki, **Artificial Intelligence (AI) vs. Machine Learning vs. Deep Learning**, Pathmind, <https://pathmind.com/wiki/ai-vs-machine-learning-vs-deep-learning>

ראה גם:

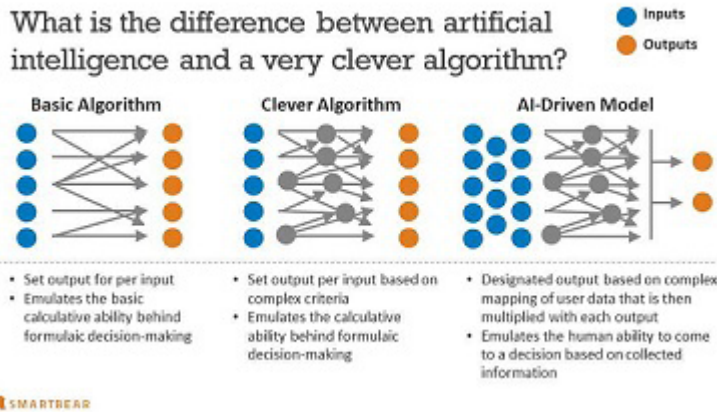
Bill Mitchell, **AI is a statistical-inference technology that learns by example**, Written evidence 11 British Computer Society; **A guide to using artificial intelligence in the public sector**, Government Digital Service (GDS) & the Office for Artificial Intelligence (OAI) [UK] (10 June 2019) <https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector>; **Artificial Intelligence and Public Standards** – A Review, Committee on Standards in Public Life [UK] (February 2020), p. 25; <https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-report>; OpenAI Charter: “highly autonomous systems that outperform humans at most economically valuable work”; <https://openai.com/charter/>.

2. Artificial Intelligence (AI) vs. Machine Learning, לעיל ה"ש 1. כפי שהגדיר זאת Arthur Samuel, אחד מחלוצי התחום, למידת מכונה היא:

“[a] field of study that gives computers the ability to learn without being explicitly programmed. Berend Berendsen, **What’s the difference between artificial intelligence, machine learning and algorithms?** Widgetbrain <https://widgetbrain.com/difference-between-ai-ml-algorithms/>

המגמה: מאלגוריתם בסיסי לבנינה מלאכותית

המעבר לקבלת החלטות תוך הסתייעות בטכנולוגיות מבוססות בינה מלאכותית, מהווה מעין קפיצת מדרגה בשיעור אקספוננציאלי, ביחס לשימושים מוכרים יותר באלגוריתם בסיסי, או אלגוריתם מורכב. אלגוריתם בסיסי מבטיח שקלט (input) מוגדר – יוביל לפלט (output) מוגדר, בהתאם ליכולת החישובית הבסיסית של קבלת החלטות נוסחאתית. אלגוריתם מורכב, הוא כזה המסוגל לייצר פלט מוגדר המבוסס על מערכת חוקים מורכבת, חישובים ופעולות שונות. גם כאן, מדובר בפעולה המחקק את היכולת החישובית שמאחורי קבלת החלטות נוסחאתית – אך כזו שהיא מורכבת יותר. לעומת השימוש באלגוריתמים כאלו לצורך קבלת החלטות, מערכת מבוססת בינה מלאכותית ניזונה מריבוי קלטים ואינה מגדירה מראש את הפלט שלה, אלא מייעדת (designates) אותו בהתבסס על מיפוי מורכב של נתוני משתמשים, באופן המשקף את יכולת קבלת החלטות האנושית בהתבסס על מידע שנאסף.³ התמונה הבאה,⁴ ממחישה היטב את ההבדל בין המערכות השונות:



3. Akshita Deora Puram, **Artificial Intelligence vs. a Clever Algorithm – What’s the Difference?** Altrends (January 8, 2019); "Basic Algorithm – Set output for per input; emulates the basic calculative ability behind formulaic decision-making"; "Clever Algorithm – Set output per input based on complex criteria; emulates the calculative ability behind formulaic decision-making"; "AI-Driven Model – Designated output based on complex mapping of user data that is then multiplied with each output; emulates the human ability to come to a decision based on collected information." <https://www.altrends.com/ai-software/software-development/artificial-intelligence-vs-a-clever-algorithm-whats-the-difference/>.

רקע

היצף נתוני העתק (big data) המאפיין את תקופתינו, בשילוב עם טכנולוגיות דיגיטליות מתקדמות ופיתוחם של כלי ניתוח וניבוי מבוססי בינה מלאכותית, מבשרים באופן בלתי נמנע על הגעתו של עידן המשילות הדיגיטלית. כשהממשלה מאמצת ביטויים אלו של קדמה טכנולוגית ומטמיעה אותם בתוך שירותיה לציבור האזרחים – חשוב להבטיח קיומו של אינטרס כפול: זה של הממשלה – ביעול עבודת מנגנון השירות לאזרח, וזה של האזרחים – בקבלת שירות משופר או מיטבי. אולם הבחירה (האמיצה והמאתגרת) של ממשלות בהטמעת מערכות בינה מלאכותית בשירות הציבורי, אינה מהווה אך ביטוי למדיניות הממשלתית של אימוץ חדשנות לטובת התייעלות, התאמה לקידמה וכיו"ב תועלות. היא מגיעה גם בתגובה לציפיית הציבור להשוואת רמת השירות, חדשנותו, וממשקיו עם האזרח, לאלו המתקבלים מספקיות שירותים במגזר הפרטי.⁵

המגמה ההולכת ומתגבשת היום בממשלים של מדינות מפותחות, כולל במדינת ישראל, היא כזו המבטאת פתיחות כלפי רעיון הדיגיטציה של שירותים ממשלתיים והפיכתם ל"טכנולוגיים" יותר, מתוך הכרה ביתרונות אימוץ החדשנות ברמת הממשל, ובהנחה כי אלו עולים בקנה אחד עם אינטרס הציבור. ביטוי לגישה זו ניתן למצוא בהחלטות ממשלה⁶ ואף ב"כללים מנחים לגיבוש הסדרים דיגיטליים" שהוצאו על-ידי היועץ המשפטי לממשלה⁷ בשלהי 2019, בהם נאמר כי: "המעבר לעולם הדיגיטלי משקף את מדיניות הממשלה, כפי שהובעה בשורה של החלטות ממשלה. הוא מעודד פיתוח והתייעלות ומסייע לממשלה לשפר את השירות לציבור ולשמור על האינטרס הציבורי." ביסודה של הנחיה זו עומדים שני עקרונות יסודיים: (1) **שקילות פונקציונלית**, המחייבת ניתוח של המטרות והשימושים של הסדרים פסיים, מסורתיים, וקביעה לגבי היכולת והאופן בהם ניתן יהיה להגשים את אלו, בסביבה אלקטרונית/דיגיטלית; ו- (2) **אי-העדיפות** להסדר פיסיו על-פני הסדר דיגיטלי, כביטוי לכך שהטכנולוגיה, לכשעצמה, אינה נחותה יותר, או שמגולמים בה סיכונים כבדים יותר מאשר הסדרים לא-טכנולוגיים. כך למשל, במשך שנים רבות הסתמכה המערכת המשפטית על שירותי מסירת דואר נייר פיסיו, באמצעות שירותי הדואר או פקידי מסירה שהוסמכו לכך. שירותי מסירה אלו כללו גם אישורים ברמה גבוהה, המעידים על מסירת הדואר לאדם פרטי או נציג של ארגון, באופן שניתן להסתמך משפטית על מסירה זו (לדוגמה, במקרה של מסירת הודעה שנקבע בחוזה שלאחריה נדרש צד לחוזה לעשות מעשה, כגון תיקון הפרה של החוזה; או במסירת הודעה על פתיחת תיק הוצאה לפועל בגין חוב של אדם והזמנתו לשלם את חובו או לטעון כנגדו בהליך משפטי). במקרים אלו, לעצם הוודאות של מסירת המסמך לאדם, ישנה משמעות רבה. בעת שאנו שוקלים מעבר למסירה וירטואלית, דיגיטלית, של הודעות אלו, עלינו לשקול האם המערכת הדיגיטלית שניצור תהיה מסוגלת למלא את התפקיד ששימשה המסירה הפיזית המסורתית ובאיזה אופן. זוהי השקילות הפונקציונלית.

עלינו גם להבטיח את אמינות והוגנות ההסדר הדיגיטלי, כחלופה להסדר הפיסיו, בהיבט של הגנה על זכויותיהם של אנשים והגנת אמינות המערכת בכללותה. יחד עם זאת, אין לשכוח שגם ההסדר הפיסיו לא היה חף מסיכונים כשל, כגון טעות, תקלה או זיוף, ושהסדר דיגיטלי אינו בהכרח נחות מהסדר פיסיו. זהו עיקרון אי-העדיפות. ההסדרים הפיזיים שהנהגנו במשך שנים ושעליהם ביססנו את הפעולות החברתיות, הכלכליות והמשפטיות שאנו עושים, עשויים יהיו לקבל מענה הולם ואולי אף טוב יותר, במרחב הדיגיטלי.

5. **Guidelines for AI Procurement**, World Economic Forum (September 2019), p. 4; available at http://www3.weforum.org/docs/WEF_Guidelines_for_AI_Procurement.pdf

6. ר' למשל, החלטת ממשלה 2733: אישור התכנית הדיגיטלית הלאומית, קידום המיזם הלאומי "ישראל דיגיטלית" ותיקון החלטות ממשלה, מיום 11.06.2017.

7. הנחיות היועץ המשפטי לממשלה 1.2500, "כללים מנחים לגיבוש הסדרים דיגיטליים" (10 אוקטובר 2019); <https://www.justice.gov.il/Units/YoezMespati/HanchayotNew/Seven/1.2500.pdf>

בסקירה זו, שמסגרתה הכללית היא יישומי טכנולוגיית הבינה המלאכותית בשירות הציבורי, נתמקד במערכות לקבלת החלטות מבוססות-אלגוריתם (Algorithmic Decision-Making; להלן – "מערכות ADM"). מערכות אלו מפותחות לטובת שיפור ויעול תהליך קבלת החלטות, למגוון ייעודים במגזר הפרטי והציבורי.⁸ הן מסייעות לניחוח ותהליך קבלת ההחלטות האנושי (מערכת תומכת-החלטה) ולעיתים, אף מחליפות את הגורם האנושי מקבל ההחלטות (מערכת מקבלת-החלטה).⁹

למעשה, ההבטחה הטמונה בהטמעת מערכות בינה מלאכותית בשירות הציבורי בכלל, ובשירותי הרווחה בפרט, רבה עד כדי כך שלבינה המלאכותית מיוחס פוטנציאל לחולל מהפיכה של ממש במתן שירותים ציבוריים. היא נתפשת כמבטיחה ייבוא של חדשנות ויעילות אל תוך השירות הציבורי, המתבטאת למשל בדברים הבאים: היכולת לעבד כמויות אדירות של מידע (נתוני עתק; big data) ולשלבן עם שיטות למידת מכונה להסקת מודלים סטטיסטיים – המאפשרים קבלת החלטות באופן מהיר ויעיל;¹⁰ הגדלת הדיוק והיעילות והפחתת הטיות אנושיות בקבלת החלטות; חסכון בעלויות בעקבות אוטומציה של תהליכים שונים; אפשרות הקצאה מחודשת של משאבים; צמצום זמני המתנה; והפחתה בנטלים המנהליים (אדמיניסטרטיביים).¹¹ יחד עם זאת, הוצאתו של הליך קבלת ההחלטות אל מחוץ להגמוניה האנושית, והחלפתו (גם אם חלקית) של שיקול-הדעת האנושי בזה של מכונה חכמה, עומדות בלב הרתיעה של החברה האנושית מפני טכנולוגיית בינה מלאכותית בעת הנוכחית, ממגוון טעמים שיפורטו בהמשך.

השירות הציבורי מהווה זירה חדשה יחסית של יישום בינה מלאכותית, והטמעת מערכות ADM בשירות הציבורי בכלל (בתחומי הבריאות, השיטור, המשפט, התחבורה, החינוך ועוד) נמצאת במגמת עלייה. בשנים האחרונות, ניכרת תופעה הולכת ורווחת של הטמעת טכנולוגיות מונעות-נתונים (data-driven) ומבוססות בינה מלאכותית, במתן שירותי רווחה שונים. התופעה הזו זכתה לכינוי 'מדינת הרווחה הדיגיטלית' ('digital welfare state'). היא נהייתה כה רווחת, שבבריטניה, כיום, אחת מתוך שלוש מועצות מקומיות עושה שימוש במערכות ADM בשירותי הרווחה שלה.¹² יחד עם זאת, חשוב לציין כי יישומים רבים מוטמעים בשדה הרווחה באופן נסיוני ומותנה במסגרת תכניות הרצה (pilots), והמחקר הייעודי לבחינת השלכות ואתגרי הטמעת בינה מלאכותית במגזר הציבורי, נמצא אך בראשיתו.¹³

8. Claude Castelluccia & Daniel Le Métayer, **Understanding algorithmic decision-making: Opportunities and challenges**, Panel for the Future of Science and Technology, EU, PE 624.261 (March 2019), pp. 3-4 [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU\(2019\)624261_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf).

9. Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter and Luchiano Floridi, **The ethics of algorithms: Mapping the debate**, Big Data & Society (July–December) 1, 3 (2016).

10. Ansgar Koene et al., **A Governance Framework for algorithmic accountability and transparency**, Panel for the Future of Science and Technology, EU, PE 624.262 (April 2019), p.3.

11. Bernd W. Wirtz, Jan C. Weyerer and Carolin Geyer, **Artificial Intelligence and the Public Sector—Applications and Challenges**, 42(7) International Journal of Public Administration, 596 (2019).

12. Sarah Marsh, **One in three councils using algorithms to make welfare decisions**, The Guardian (15 Oct 2019) <https://www.theguardian.com/society/2019/oct/15/councils-using-algorithms-make-welfare-decisions-benefits>

13. Wirtz et al., לעיל ה"ש 11, בעמ' 608.

אימוץ טכנולוגיות מבוססות בינה מלאכותית דורש התמודדות ראשונית עם אתגרים ייחודיים והתגברות על חסמים שונים. על אתגרים אלו נמנים, בקליפת אגוז: הצורך בהעלאת האוריינות הדיגיטלית/טכנולוגית של פקידי השירות הציבורי; קושי בבניית מומחיות בינה מלאכותית בקרב עובדי המגזר, והיעדר תקציב מספק לשכירת מומחים כאלה (קושי בהתחרות עם תנאי ההעסקה של המגזר הפרטי); היעדר גמישות באימוץ ויישום טכנולוגיות חדשות; ותרבות שמרנית, מאופיינת בספקנות טכנולוגית (technological skepticism) הדבקה בפרקטיקות ותהליכים מבוססים, תוך רתיעה מסיכונים לא הכרחיים.¹⁴ נוסף על כך, אימוץ ויישום מערכות ADM במערכת הרווחה, שלקוחותיה-מטופליה מהווים אוכלוסייה מוחלשת ופגיעה באופן אינהרנטי, מהווה אתגר כשלעצמו¹⁵ עבור המדינה, עבור שלוחיה – העובדים הסוציאליים, ועבור האזרחים, מטופלי הרווחה.

בשל כל אלה, דורשת הטמעת המערכת נקיטת גישת זהירות, הן בשלב שקילת אימוצה – אז תידרש בחינה מדוקדקת של הצדקת הצורך במערכת כזו בשירות המדובר ועריכת מאזן סיכון-תועלת הצפוי מיישומה¹⁶ והן, לקראת יישומה – אז תידרש בחינה מקדימה בסביבה מבוקרת (דהיינו, במסגרת תכנית הרצה), ניסויה באמצעות sandbox/testbed והטמעתה באופן מדורג, להקטנת פערי המידע והסיכונים שבאימוץ מערכת כגון זו.

14. Julian Torres Santeli & Sabine Gerdon, 5 **challenges for government adoption of AI**, World Economic Forum (16 Aug 2019) <https://www.weforum.org/agenda/2019/08/artificial-intelligence-government-public-sector/>; Joanna Redden, Lina Dencik and Harry Warne, **Datafied child welfare services: unpacking politics, economics and power**, Policy Studies 1, 14 (2020).

15. Santeli & Gerdon, לעיל ה"ש 14.
 16. ר' למשל בעניין זה, הנחיות היועץ המשפטי לממשלה "כללים מנחים לגיבוש הסדרים דיגיטליים", לעיל ה"ש 7, חלק ב', בעמ' 7-16, שם מוצע הליך בחינה מומלץ, המנחה את שיקול הדעת של הרשות המנהלית, בגיבוש הסדר דיגיטלי. ההליך כולל בחינת ההיבטים הבאים: זיהוי מטרות הכלליות של ההסדר הדיגיטלי; זיהוי התכליות של ההסדר הדיגיטלי; זיהוי הגורמים הנוגעים להסדר; בחינת סיכוני ההסדר (זיהוי הסיכונים, זיהוי עוצמת הפגיעה, זיהוי מידת היכולת לשאת בסיכונים או להקטינם); ביצוע ניהול סיכונים; בחינת הנטלים שבהם יכול לעמוד כל גורם המושפע מההסדר; בחירה בין הכלים הקיימים להגשמת כל תכלית; ובחינת התאמת ההסדר לדין והצורך לעגנו בחקיקה.

מטרות מערכת ADM ומאפייניה

כשניגשים להעריך את מידת הסיכון, או הקושי, הטמונים באימוצה ויישומה של מערכת ADM בשירותי הרווחה, קיימת חשיבות גם למטרה שלשמה עוצבה המערכת (ייעודה); לתפקיד שיועד לאלגוריתם ולמשקלו (היחסי) בקבלת החלטה; ולאוכלוסיית היעד לגביה תיושם המערכת. לפיכך, יש לבחון ולהעריך, בין היתר, את הנקודות הבאות:

- מידת **מעורבות הגורם האנושי** – האם בני אדם (ובפרט, אנשי המקצוע הרלבנטיים בשירות הציבורי) מעורבים בתהליך קבלת ההחלטות, אם לאו? האם מרגע הטמעת מערכת ה-ADM והפעלתה, פחתה המעורבות האנושית עד שנעלמה עם הפיכת המערכת לאוטונומית ואוטומטית, או שמא הופעל מודל המותיר בתמונה את הגורם האנושי ('Human-in-the-loop') ומאפשר שמירה על מידה מסויימת של שליטה ובקרה על תהליך קבלת ההחלטות האוטומטי? בהתאם למודל האחרון, הגורם האנושי מעורב בפעילות המערכת בעיקר כבקר וכמתבונן, המתערב רק כשהדבר הכרחי. ככזה, הוא נושא באחריות מוגבלת להחלטות תוצרי מערכת ADM. מנגד, קיימת גישה חלופית למעורבות הגורם האנושי, כזו המציעה שותפות אינטראקטיבית בין אדם למכונה ('Human-AI partnerships'), כך שהחלטות תוצר המערכת הן תוצר משותף של השניים.¹⁷ במצב זה, בהיעדר ישות משפטית עצמאית לאלגוריתם, האדם, שהיה שותף מלא לתהליך – נחשב לאחראי הבלעדי לתוצרי המערכת. למצב זה מיוחסים מספר יתרונות: צמצום/ביטול 'מרווח האחריות', אליו נתייחס בהמשך; הסרתם של פגמים פוטנציאליים באופן קבלת החלטות אנושיות והחלטות מכונה (ואולי אף ריפוי צולב של פגמים אלו); והגברת האיכות והדיוק של ההחלטות.
- האם מדובר באלגוריתם '**תומך החלטה**' (algorithm-assisted decision making) דהיינו, ההחלטה מבוצעת ע"י גורם אנושי המסתייע באלגוריתם בשיקול הדעת שלו, או ב'אלגוריתם מחליט' (algorithmic decision-making), שאז ההחלטה מתקבלת על-ידי האלגוריתם?
- האם הכלי המתוכנן – **פרו-אקטיבי או ריאקטיבי?** ככלל, כלי פרו-אקטיבי חודרני יותר למרחב האוטונומי האישי-פרטי של מטופל הרווחה ומופעל ללא הזמנה, יוזמה, או ידיעה של המטופל. לעומתו, כלי ריאקטיבי אולי פועל ללא הזמנה בפועל של המטופל, אך נדמה כי נכנס לפעולה במעין הזמנה בכח, לפי צורך, ועל מנת להגן על אינטרס של המטופל או הסביבה, המצויים בסיכון. ההבדל בין השניים (פרו-אקטיבי/ריאקטיבי), משפיע על מידת הלגיטימציה הנחזית ביישומם.
- **מי היא אוכלוסיית היעד** (מיהם 'נתיני-האלגוריתם')? במקרה שלפנינו, אוכלוסיית היעד של מטופלי רווחה היא אוכלוסייה פגיעה באופן אינהרנטי כמעט. אולם יש לזכור, כי אוכלוסייה זו איננה עשויה מקשה אחת ויש שיימצאו בה פרטים שאינם מוחלשים בשלל הכשירויות והמאפיינים שלהם, אלא רק בהיבטים מסויימים, באופן המצריך הגנה פטרנליסטית פחותה בעניינם. יחד עם זאת, בתוך אלו, ישנן אוכלוסיות שלעולם יחזו כרגישות, כגון אלו של קטינים וחסרי ישע. כשזהו קהל היעד של מערכת ADM, עלינו לבחון אותה ביתר זהירות. כמו כן, ל(א-) מעורבות הגורם האנושי מיוחס משקל רב יותר כשעסקינן באוכלוסיות רגישות אלה.

17. Committee on Standards in Public Life, **Artificial Intelligence and Public Standards**, לעיל ה"ש 1, בעמ' 21-22.

שאלות נוספות הקשורות במאפייני המערכת ומחייבות בחינה והערכה, הן השאלות הבאות:

- באיזה אופן תיושם המערכת, ועל-ידי מי?
- מי היה מעורב בפיתוח האלגוריתם? האם הצוות היה מגוון דיו? האם כלל נציגי אנשי מקצוע, או אוכלוסיית היעד?
- מהו המידע שעל בסיסו אומן ושבזו הוזן האלגוריתם? האם היה נקי (יחסית) מהטיות?
- באיזה אופן תובטח פרטיות נושאי המידע שפרטיהם מצויים במערכת?
- האם נתונה לפרט אפשרות ערעור על החלטות המערכת?
- האם ואם כן, באיזו מידה – ניתן להסביר למושא החלטת המערכת, את אופן פעולת האלגוריתם (‘black box’; explicability)?

בהמשך הסקירה נתייחס למרבית הסוגיות העולות משאלות אלו.

עקרונות בינה מלאכותית (AI Principles)

לצורך ניתוח התועלת, הסיכון והיחס שביניהם בהקשר של טכנולוגיות בינה מלאכותית באופן כללי, ולשם התמודדות עם קונפליקטים שונים יצירי-הטכנולוגיה, מתפתח בשנים האחרונות במסגרתו של שיח אקדמי ומקצועי ער ודינמי, אגד של עקרונות אתיים מנחים לפיתוח ויישום טכנולוגיות בינה מלאכותית.

העקרונות מפותחים בשלב זה בעיקר על-ידי ארגונים בינ"ל¹⁸, המגזר הממשלתי¹⁹, ועל-ידי חברות הטכנולוגיה העוסקות בפיתוח טכנולוגיות אלו, ביישומן ובשיווקן (Google²⁰, Microsoft²¹, IBM²²) ואח', כחלק ממהלכי אסדרה עצמית (self-regulation) בנושא; ומהווים בסיס לאסדרה כללית נרחבת ומקיפה יותר של גופים ומדינות. זוהי ליבת העקרונות המוסכמים:

1. **כיבוד חוקי-האדם** ומעורבות **הגורם האנושי**: מערכות AI יפותחו באופן המכבד את שלטון החוק, זכויות האדם וערכים דמוקרטיים. כמו כן, יכללו מנגנוני הגנה, כגון: הבטחת מעורבותו של הגורם האנושי במצבים מסויימים.
2. **הטבה**, קידום רווחה, שימור הכבוד ושמירה על כדו"א: יצירת טכנולוגיית בינה מלאכותית המועילה לאנושות, לרווחת הפרט, ולסביבה.
3. **אי-עשיית (ומניעת) נזק**: הקפדה על בטיחות בינה מלאכותית באופן שיבטיח מניעת פגיעה בחירויות הפרט ובפרטיותו; והמנעות משימוש לרעה בטכנולוגיות AI (למשל באופן העלול לסכן הבטחון הלאומי). זאת, מתוך **אחריות מוסרית** (moral responsibility) ו**אחריותיות** (accountability) של מפתחים ומעצבי הטכנולוגיה להשלכות חברתיות ומוסריות של יישומי המערכת.
4. **אוטונומיה** – הכח/סמכות להחליט (האם להחליט): בפיתוח בינה מלאכותית יש לקדם את האוטונומיה של בני האדם, תוך כדי **הגבלת האוטונומיה של מערכות ממוחשבות**.
5. **צדק חברתי**: הבטחת גישה **שוויונית** לתועלות טכנולוגיות הבינה המלאכותית; אזהרה מפני **הסיכון להטיה** בבסיסי נתונים המשמשים לאימונה; הבטחה שהשימוש בטכנולוגיות בינה מלאכותית ייצר

18. **What are the OECD Principles on AI?** OECD <https://www.oecd.org/going-digital/ai/principles/>; **Ethics guidelines for trustworthy AI**, European Commission (EU) (8 April 2019) <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

19. ר' למשל, ועדת משנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית – דין וחשבון (אוגוסט 2019); רונה ממשלת ניו-זילנד לשימוש ממשלתי באלגוריתמים: האמנה הממשלתית, עליה חתמו בשלב הראשוני 19 סוכנויות ממשלתיות, מתחייבת לשקיפות לגבי אופן קבלת החלטות ממשלתית על-ידי אלגוריתם; הסברתיות; הנגשת מידע אודות התהליכים והנתונים המצויים בשימוש; וזיהוי הטיית אלגוריתמים וניהולן. בנוסף, הוקדשה תשומת לב מיוחדת לשמיעת עמדתן של אוכלוסיות מיעוט, במיוחד זה המאורי, בעניין איסוף נתונים אודותיהם ולהתייעצות עם קבוצות העשויות להיות מושפעות מכך. האמנה אף מכוננת מנגנון ערעור על החלטות שהתקבלו על-ידי אלגוריתם;

Charlotte Graham-McLay, **New Zealand claims world first in setting standards for government use of algorithms**, The Guardian (27 Jul 2020)

<https://amp-theguardian-com.cdn.ampproject.org/c/s/amp.theguardian.com/world/2020/jul/28/new-zealand-claims-world-first-in-setting-standards-for-government-use-of-algorithms>.

20. **Responsible AI Practices**, GoogleAI <https://ai.google/responsibilities/responsible-ai-practices/>.

21. **Guidelines for Responsible AI**, Microsoft AI <https://www.microsoft.com/en-us/ai/responsible-ai-resources?activetab=pivot1%3aprimar4>

22. **Trusted AI for Business**, IBM <https://www.ibm.com/watson/ai-ethics/>

תועלות אותן ניתן לחלוק בחברה; שימור המגוון (diversity) החברתי והתרבותי. כל אלה – לטובת קידום השגשוג ושימור הסולידריות בחברה.

6. היכולת והאפשרות להסביר (Explainability – 'הסברתיות'): שקיפות וגילוי אחראי לגבי אופן פעולת מערכות בינה מלאכותית, אשר יאפשרו הבנתן (כיצד עובדת טכנולוגיית ה-AI?) ואת היכולת להסביר את תוצריהן, בצירוף עם **אחריותיות** (מי אחראי לתוצר הטכנולוגיה?). אלו יבטיחו אפשרות לערער על תוצרי המערכת, באופן שיבטיח צדק בדמותו של הליך הגן.

העקרונות הללו נועדו להגן על האינטרסים של בני-האנוש ועל זכויות פרט מגוונות, לנוכח שילוב מערכות בינה מלאכותית במציאות חיינו. בהקשרים מסוימים, מהווים אלו אף בסיס ליצירתן של זכויות ייחודיות חדשות (de novo), ייעודיות לתקופה הנוכחית ואף צופות פני העתיד. העקרונות ישמשו אותנו בהמשך לניתוח התועלות והקשיים/סיכונים שבהטמעת טכנולוגיות מבוססות בינה מלאכותית בשירות הציבורי, ובשירותי הרווחה – בפרט.

ההיבט המנהלי-ציבורי

מבלי להיכנס לעוביו של המשפט המנהלי בהקשר שלפנינו, הסוגיה נשוא סקירה זו מזמינה התייחסות מיוחדת לזירת ההטמעה של מערכות ADM – בשירות הציבורי, על אתגרי היישום והאימוץ הייחודיים למגזר זה.

הטמעת מערכות ADM בשירות הציבורי

תהליכי דיגיטציה בממשל הציבורי בכלל, והקמת מאגרי נתונים ואפשרויות קישוריות בין מאגרים ציבוריים בפרט – הינם כאלה המשפרים משילות (Governance). יחד עם זאת, ישנם כאלה – וביניהם, הדווח המיוחד (special rapporteur) מטעם האו"ם על עוני קיצוני וזכויות אדם – הסבורים כי אנו מצויים על סיפו של עידן **המשילות הדיגיטלית** ומבקרים במיוחד את הפרטת שירותי הרווחה ואת האוטומציה שלהם. מצב זה, בו מושקעים סכומי עתק מכספי הציבור במערכות אוטומטיות מבוססות בינה מלאכותית, באלגוריתמים של חיזוי, בבניית מודלים לסיכון, ובמערכות ביומטריות, לשם יישוםם בזירת הרווחה – מכונה, כפי שצינו ברקע הדברים: 'מדינת הרווחה הדיגיטלית' ('digital welfare state'), שבה השתנה ללא הכר אופיה של ההגנה החברתית.²³

יישום והטמעה של מערכות בינה מלאכותית מבוססות נתוני עתק בשירות הציבורי, מהווה תהליך דיגיטציה מתקדם שאת הרצינות שלו ניתן להבין באופן טוב יותר על רקע השינויים שעוברת תפישת השירות הציבורי בשנים האחרונות.

עצם מיקור-החוץ של שירותים ציבוריים למגזר הפרטי (דהיינו, הפרטתם), כמו גם שיתוף האזרח בתהליכים בשירות הציבורי, הינם ביטוי לפרדיגמה החדשה המתהווה בתחום ניהול השירות הציבורי – זו של **המשילות הציבורית החדשה** (NPG – New Public Governance).

הפרדיגמה החדשה מעדכנת פרדיגמות קודמות – את זו המסורתית, של 'האדמיניסטרציה הציבורית' (Public Administration) המבטאת מודל של שליטה היררכית-אנכית, לפיו המדינה, באמצעות השירות הציבורי, היא הספקית הבלעדית של 'מוצרים ציבוריים' (public goods); ואת זו הטרייה והמאוחרת יותר – של ה'ניהול הציבורי החדש' (NPM – New Public Management), שערערה למעשה על בלעדיותה של המדינה כספקית השירות הציבורי לאזרח, על-ידי שילוב המגזר העסקי והמגזר השלישי כספקים, לשם יצירת תחרותיות והגברת/השגת יעילות כלכלית בפעולת המערכת הציבורית, לטובת האזרחים (לקוחותיה).²⁴

23. Philip Alston (Australia), Special Rapporteur on extreme poverty and human rights, **Report of the Special rapporteur on extreme poverty and human rights** A/74/48037, ohchr.org (11 October 2019); Philip Alston (Australia), Special Rapporteur on extreme poverty and human rights, **World stumbling zombie-like into a digital welfare dystopia, warns UN human rights expert**, UN, ohchr.org (17 October 2019) <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25156>; Nicolas Kayser-Bril, **UN special rapporteur on digital technology and social protection denounces a 'human rights free-zone'**, Algorithm Watch (October 16, 2019) <https://algorithmwatch.org/en/un-special-rapporteur-on-digital-technology-and-social-protection-denounces-a-human-rights-free-zone/>.

24. גלעד בארי 'המשילות הציבורית החדשה' ג'וינט ישראל – המכון למנהיגות וממשל (11.2014) www.theinstitute.org.il/files/11.2014-articles/meshilut.pdf.

המשילות הציבורית החדשה מושתתת על ארבעה יסודות עיקריים:²⁵

1. תהליך מדיניות מכליל, המבטא ויתור על הגמוניה של המדינה בתחום, על-ידי **שיתוף פעולה מקיף בין המגזר הציבורי, המגזר הפרטי, המגזר השלישי והאזרחים**;
2. **מיצוב האזרח** (משתמש השירות הציבורי) **במרכז**, על-ידי **שיתופו האקטיבי** בעיצוב השירות הציבורי ובהערכתו. יסוד זה מביא לשיפור השירות ולהטמעת חדשנות בו. מגמה כזו, של שיתוף הציבור בעיצוב תהליכים ציבוריים אשר ישפיעו על חייו, על חיי קהילתו ועל כלל האזרחים, מתכתבת היטב עם, או מצויה על הרצף של, תהליך **הדמוקרטיזציה** המתרחש בשנים האחרונות ביחס להנגשתן וזמינותן של טכנולוגיות שונות לאזרחים.²⁶
3. **שיתוף פעולה מקיף** במסגרת **מערכות יחסים ארוכות טווח**, שבמרכזן אִמון בין השותפים השונים.
4. **מרכזיותה של הטכנולוגיה** (בעיקר טכנולוגיות מידע ותקשורת, אגירת נתוני עתק ועיבודם ועוד) כמחוללת שינוי ומקרבת קידמה, על-ידי הטמעתה באופן נרחב ואינטנסיבי בשירות הציבורי. נדמה, כי יסוד זה הוא המאתגר ביותר במודל המשילות הציבורית החדשה. זאת, לנוכח קשיים וחסמים טיפוסיים להטמעת טכנולוגיות מתקדמות במערכת ותיקה, שמרנית, מסורבלת ולא גמישה, כגון זו של המגזר הציבורי,²⁷ עליהם נעמוד בהמשך.

המשילות הציבורית החדשה היא, כאמור, פרדיגמה צעירה ומתגבשת. לא ברור באיזה מקום על רצף הפרדיגמות האמורות מצויה ישראל בעת הנוכחית, אך ניתן להניח כי היא מצויה בשלב המעבר שבין הניהול הציבורי החדש (NPM) לבין המשילות הציבורית החדשה (NPG). זאת, בין היתר, משום שלפחות בסוגייה שלפנינו – בשעה שישנה הזמנה למגזר הפרטי ולמגזר השלישי להשתתף בתהליך לבחינת פתרונות וחלופות יעילים כלכלית (NPM), לא ניכר עדיין שיתוף אקטיבי משמעותי של האזרח (NPG) בבחינת יישום מערכות ADM בשירותי הרווחה. השלב המעברי של הרפורמה (שכאמור, ייתכן שבו מצויה ישראל), מתאפיין לרוב באימוץ חלקי של הפרדיגמה החדשה, תוך הותרת חלקים מפרדיגמת הניהול הציבורי החדש על כנם, במעין תהליך של 'היברידיזציה' (hybridization) או 'שיכוב' (layering).²⁸

25. שם, בעמ' 5-6.

26. דמוקרטיזציה זו באה לידי ביטוי ביתר שאת, בתחום ניהול הבריאות ואוטונומיית הרבייה של הפרט: החל מערכות לבדיקת הריון, דרך ערכות לזיהוי זיהום חיידיקי בגרון, וכלה בבדיקות גנטיות הזמינות ישירות לצרכן, ערכות לעריכת גנום עצמית ועוד – כל אלה נגישות באופן ישיר לאזרחים ומייתרות תיווך של אנשי מקצוע.

27. לעיל ה"ש 24, בעמ' 8-9.

28. שם, בעמ' 8.

מיקור החוץ, של שירותי רווחה באמצעות התקשרויות עם ספקי טכנולוגיה חוץ-ממשלתיים, מסחריים, מן המגזר הפרטי – שמטרתו הבאת חדשנות טכנולוגית והתייעלות אל תוך השירות הממשלתי לאזרח – מעורר מספר אתגרים ספציפיים:

- **חוסר ידע/הבנה של פקידי הממשל ושלוחיו** – מיישמי המערכת ומפעיליה – לגבי **אופן הפעילות של האלגוריתמים** ומגוון הסיכונים שהם מציבים.²⁹
- **טשטוש הגבולות שבין הציבורי לפרטי**. כך למשל, בעוד שמבחינת המדינה מדובר במטופלי רווחה – אוכלוסייה שהיא רגישה ומחלשת ושחלקה מצוי בסיכון, עבור הספקים החוץ-ממשלתיים מדובר ב'לקוחות' – מעמד ניטראלי שעבור חלקם, לפחות, אינו מחייב התייחסות רגישה במיוחד, או מחויבות פטרנליסטית כמו זו של המדינה.
- **בעיית פרטיות ואמון**, כתוצאה מן הצורך להנגיש מידע ונתונים אישיים אודות פרטים לגופים חוץ-ממשלתיים, לשם אימון והפעלת מערכות ה-ADM. בעיה זו עשויה להיות מוחרפת במצב של מיקור-חוץ של שירותי רווחה, כשאוכלוסיית היעד הינה כאמור פגיעה ומחלשת מטבעה, או לכל הפחות – בעלת נקודות תורפה אותן עלולים לנצל לרעה גורמים בעלי כוונות זדוניות (למשל, למטרות הונאה בתחום קצבאות הרווחה השונות). אוכלוסייה זו תלויה פעמים רבות בשירות שניתן לה, ונדרשת למידה לא מבוטלת של אמון במערכת ובנותני השירות מטעמה, לשם הבטחת מיצוי זכויותיה והסעדים להם היא זקוקה, בצורה אפקטיבית. אמון זה הוא שברירי, ולא בנקל ניתן יהיה לשקם אמון פגוע של מטופל רווחה, בשל פגיעה בפרטיותו או בזכויות ואינטרסים אחרים שלו, על-ידי מערכת ADM שבמיקור-חוץ.
- **קושי בזיהוי האחראי (או האחראים) לכשל, או נזק**: דהיינו, האם תהא זו הממשלה כגורם מזמין, תומך ומיישם של המערכת, או שמא החברה המפתחת את מערכת ה-ADM? או שאולי, יתכן גם האלגוריתם עצמו – רעיון המכונה 'algorithmic accountability'³⁰?

29. Kate Crawford & Jason Schultz, **AI Systems As State Actors**, 119(7) Columbia L. Rev., 1941, 1942 (2019), available at <https://columbialawreview.org/content/ai-systems-as-state-actors/>.

30. Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin and Rhema Vaithianathan, **Toward Algorithmic Accountability in Public Services – A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-Making in Child Welfare Services**, CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Paper No.: 41, 1 (May 2019); <https://doi.org/10.1145/3290605.3300271>.

אתגרי אימוץ מערכות ADM במגזר הממשלתי/ציבורי

כפי שצינו בתמציתיות ברקע לדברים, יישום טכנולוגיות בינה מלאכותית במגזר הציבורי הוא תהליך רצוף אתגרי – כלליים, וכאלה הייחודיים למנהל הציבורי, בעיקר בשל היותו חדש ולא-טבעי למערכת הממשלתית-ציבורית. להלן פירוט האתגרים, בחלוקה לקטגוריות:

אתגרים כלליים

אתגר-על: תרגום חקיקה, רגולציה מורכבת, ועמימות או כוונות פתוחות משפטיות – לקוד, אלגוריתם, כך שתוצריו יוכלו לשקף את אלו באופן יעיל, מדויק ונאמן לרוח הדברים ולכוונת המחוקק, או קובעי המדיניות. זוהי סוגייה מעט מופשטת, המצריכה בחינה עיתית, יזומה, של קורלציות בין החקיקה הממשלתית והנחיות הרשות המנהלית, לבין אופן פעולת האלגוריתם. נדמה שהמפתח להשגת תוצאה מיטבית כאן, הוא תיווך מדויק ואמין של הרגולציה והמדיניות שמאחוריה, למפתחי האלגוריתם. הדבר אף מדגיש את חשיבות שיתוף נציגי קובעי המדיניות ועובדי הרשות המנהלית בתכנון האלגוריתם, בשלב הפיתוח ובפיקוח ארוך-טווח על תוצריו. אתגר העל הזה קשור עם אתגר מעשי-יישומי כללי יותר, שבתרגום המדיניות המצויה והעתידי לבי האלגוריתמים עצמם, למנגנוני אסדרה וסטנדרטים מציאותיים ישימים ע"י התעשייה.³¹

אתגרי אימוץ הטמעה ויישום של טכנולוגיות מבוססות בינה מלאכותית במנהל הציבורי

ניתן להצביע על אתגרי יישום טכנולוגיות מבוססות בינה מלאכותית במנהל הציבורי, במספר מישורים: הטכנולוגי, החוקי, האתי, והחברתי.³²

המישור הטכנולוגי

אלו הם אתגרי האימוץ והחסמים ליישום הטכנולוגיה במישור זה:

- סוגיות של בטיחות בינה מלאכותית – הבטחה שמערכת מבוססת בינה מלאכותית היא אמינה, אינה יוצאת משליטה אנושית, צפויה דיה ואינה גורמת לתוצאות שלא כוון אליהן (תואמת את ייעודה המתוכנן), ושפעילות המערכת ותוצריה תואמים ערכים אנושיים.
- איכות הנתונים/המערכת – 'חכמת' המכונה תלויה ישירות באיכות הנתונים המוזנים אליה. איסוף, אחסון ושימוש בנתונים רלבנטיים וחפים מהטיות הינם בעלי חשיבות מכרעת להטמעה נכונה ומוצלחת של מערכות בינה מלאכותית במגזר הציבורי.
- שילוב (אינטגרציה) – שאיפה להטמעה הרמונית של מערכת מבוססת בינה מלאכותית בשירותי ציבורי נתון, באופן שהמשק בין האדם המפעיל את המערכת לבין המערכת, יהיה חלק, יעיל ומועיל.

31. Mittelstadt et al., לעיל ה"ש 9, בעמ' 13;

Michele Gilman, **AI algorithms intended to root out welfare fraud often end up punishing the poor instead**, The Conversation (February 15, 2020) <https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625>.

32. Wirtz et al., לעיל ה"ש 11.

- התמחות והתמקצעות – המחסור במומחי בינה מלאכותית בקרב עובדי המגזר הציבורי, מקשה על הטמעתן של מערכות בינה מלאכותית בזירה זו ומעכב אותה.
- היתכנות כלכלית – (אי-)יכולתו של המגזר הציבורי לשאת בעלויות פיתוח, הטמעה, תפעול ותחזוקה של המערכת, כמו גם בעלויות שינבעו משגיאות המערכת (כאלה הקשורות בחבויות משפטיות, למשל).
- התגברות על חסמים שמקורם בתרבות הארגונית ובאופיו היותר שמרני ופחות גמיש של השירות הציבורי. למשל, מדיניות רכש קיימת, שאינה עדכנית ומתאימה לטכנולוגיות בינה מלאכותית.³³

המישור החוקי

משילות בינה מלאכותית מצויה עדיין בחיתוליה, הן ברמת האסדרה המקומית (לאומית) והן ברמת האסדרה הגלובאלית. מכך נגזרים אתגרי הניהול והבקרה של הטכנולוגיה, על השלכותיה החברתיות והכלכליות. על מגוון הקשיים המשפטיים הנדרשים למענה חוקי, שלחלקם נתייחס בהמשך, נמנים:

- מעמדה החוקי של בינה מלאכותית – למשל, האם זו תוכר בעתיד כישות משפטית עצמאית ('digital'/'electronic personality')?
- בעיית ההסברות וה'קופסא השחורה' של החלטות ופעולות מערכות בינה מלאכותית.
- ייחוס אחריות משפטית לנזק יציר-בינה-מלאכותית ומרווח האחריותיות.
- פרטיות ואבטחה של הנתונים המוזנים למערכת, ושל תוצריה.

המישור האתי

אתגרים במישור זה, נובעים בעיקר מן המתח ביחסי אדם-מכונה. בולט במיוחד בהקשר זה, החשש ש'הגולם קם על יוצרו'. חששות אלה, נוגעים בעיקר להשלכות היישום הבאות של מערכות ADM:

- הבדלים בקבלת החלטות על-ידי אדם ועל-ידי מערכת בינה מלאכותית – תהליך קבלת החלטות האנושי מושפע לא רק מעקרונות רציונאליים, אלא גם ממערכת ערכים מוסרית, מרגשות וממודעות שקשה לתרגם אל תוך האלגוריתם. עובדה זו יכולה להשפיע על הבדלים בין השיפוט הערכי המבוצע על-ידי השניים. ייתכן אף, שעם הזמן, יפתחו מערכות בינה מלאכותית מערכת ערכים משלהן ובכך ידגישו את חוסר התאימות עם מערכת הערכים האנושית. החשש אותו מעלים אתיקנים בהקשר זה, הוא שבעתיד, תיהפך הבינה המלאכותית לעצמאית לגמרי וששיקול הדעת של המכונה יעמוד בסתירה לזה האנושי.
- תוצרים לא מוסריים, או מפלים, של מערכת בינה מלאכותית, שיש בהם כדי לגרום לאי-שוויון וחוסר הוגנות ביחס לנשואי החלטות מערכת הבינה המלאכותית.

33. World Economic Forum, **Guidelines for AI Procurement**, לעיל ה"ש 5, בעמ' 4.

המישור החברתי

במישור זה, המתח בין אדם למכונה מוחשי אף יותר. כאן מודגש החשש (האוניברסלי) שמא המכונה תהווה בשלב מסוים תחליף לאדם. דבר זה בא לידי ביטוי בהיבטים הבאים:

- שינויים בכח העבודה, בעיקר כאלה המטמיעים אוטומציה בתהליכים שונים, העלולים להוביל לייתור עבודה אנושית והחלפתה על-ידי טכנולוגיות בינה מלאכותית.
- (אי-) קבלה חברתית של הבינה האנושית ו(אי-) אמון חברתי במערכות כאלה.
- מטאמורפוזת (התמרה), שינוי עמוק ויסודי ביחסי-הגומלין שבין אדם למכונה (H2M) ובין מכונה למכונה (M2M).

יישום מערכות ADM בשירותי הרווחה

זירת הרווחה, מהווה אחד מתחומי היישום הרווחים יותר של מערכות ADM בשירות הציבורי. ניתן להצביע על שני תחומי יישום עיקריים של מערכות אלו, בעת הזו:

1. קצבאות רווחה
 - קביעת זכאות לקצבאות
 - זיהוי ואיתור הונאת קצבאות רווחה (welfare fraud) במגוון תחומים: סיעוד, אבטלה, שיכון ועוד.
2. הגנה על ילדים ואוכלוסיות בסיכון
 - איתור ילדים קורבן להתעללות וזיהוי המשפחות, לצורך נקיטת התערבות מוקדמת.
 - מניעת ניצול אוכלוסיות מוחלשות (ילדים ואנשים צעירים) והדרדרותן לעבריינות.

בשנים האחרונות, נעשים במדינות שונות נסיונות בהטמעת מערכות ADM במגוון שירותי רווחה. אלו הוכתרו במידות שונות, או מעורבות, של הצלחה וכישלון. להלן מס' דוגמאות:

בריטניה

בבריטניה הוטמעה מערכת ADM בשירותי הרווחה במספר רשויות מקומיות, על מנת לספק הכוונה אוטומטית בתביעות זכאות (להטבות בשיכון ומיסי מועצה), למניעת התעללות בילדים ולהקצאת מקומות בבתי-הספר. השימוש במערכות אלו הופסק בשל חששות הנוגעים לפרטיות ואבטחת מידע, אי-היכולת של פקידי הרשות להבין את אופן פעולת המערכות, הקושי של תושבים בתקיפת ההחלטות האוטומטיות יצירות-המערכת, ולאחר שניתוח החיזוי/ניבוי (predictive analytics) על-ידי המערכת, גרם לעיכוב לא מוצדק בתשלומי רווחה (כשזו זיהתה תביעות בסיכון נמוך, ככאלה המצויות בסיכון גבוה).³⁴

למרות זאת, במקומות אחרים בבריטניה הוטמעו מערכות ADM בהצלחה. כך למשל, במועצה המקומית Hillingdon, בדרום מערב לונדון, הוטמעה מערכת העושה שימוש בניתוח חיזוי למניעת ניצולם של ילדים וצעירים פגיעים. המערכת מנתחת מידע המתקבל ממקורות שונים (משטרה, מערכות הרווחה והבריאות, שירותי נוער, בתי-ספר וקבוצות בקהילה), לזיהוי אזורים גיאוגרפיים בסיכון גבוה. זיהוי זה מאפשר למועצה הפניית שירותים אינטנסיביים לתמיכה בנוער לאזורים אלו, על מנת להסיט אנשים צעירים ממסלול של אלימות ועבריינות.³⁵

34. Marsh, לעיל ה"ש 12.

35. Lynn Eaton, **Is it right to use AI to identify children at risk of harm?**, The Guardian (18 Nov 2019) <https://www.theguardian.com/society/2019/nov/18/child-protection-ai-predict-prevent-risks>

דנמרק³⁶

בשלהי 2018, הוכרזה בדנמרק – אחת המדינות הנהנות מאמון ציבורי גבוה בממשל ומתכניות רווחה מהחזקות ביותר – רפורמה דיגיטלית במגזר הציבורי, אשר עוגנה בחקיקה. הרפורמה נדמית מרחיקת לכת בכך שמאפשרת לחברות פרטיות (כגון: בנקים וחברות ביטוח) גישה לנתונים ציבוריים, כולל נתונים אישיים אודות אזרחים, מתוך רצון לקדם את האסטרטגיה הממשלתית לדיגיטציה.³⁷

בין תכניות הדיגיטציה, ניתן היה למצוא מערכת ADM אוטומטית-למחצה שיושמה למעקב המונים (mass surveillance) וליצירת פרופיל (profiling) של אזרחים מובטלים, תוך התבססות על מידע אישי אגרטיבי שהצטבר במאגרי מידע דיגיטליים ציבוריים שונים. מטרתה: זיהוי סיכון לאבטלה ארוכת-טווח בקרב אזרחים. יש הסבורים, שיישום המערכת נועד, בין היתר, לשמש כאמצעי לחסכון בעלויות ממשלתיות. הניסיון ליישם כלי זה ליצירת פרופיל/מיפוי המוני, באופן שיטתי, לא עלה יפה. בין היתר זכתה היוזמה לביקורת על היעדר שקיפות ממשלתית לגבי הנתונים בהם עושה שימוש האלגוריתם והמשקל המיוחס לו בהערכה הכוללת.

כמו כן, הוטמעה באופן נסיוני באחת הרשויות המקומיות (Gledsaxe), ובהמשך ברשויות נוספות, מערכת העושה שימוש בלמידת מכונה ובאלגוריתמים ליצירת פרופיל לזיהוי ילדים המצויים בסיכון להתעללות, תוך שימוש במידע אישי המצוי במאגרי מידע דיגיטליים שונים ומקושר למס' זיהוי אישי (מזהה ייחודי). המטרה הייתה איתור פרואקטיבי של ילדים פגיעים בשל נסיבות חברתיות, טרם אלו הראו סימנים המצביעים על צרכים מיוחדים. המערכת אפשרה לרשויות לאתר את המשפחות לגביהן התקבלה התראה (flagging), לצורך התערבות מוקדמת. התהליך כלל גם בחינה אל מול קבוצת ביקורת של משפחות שאינן סובלות מבעיות סוציאליות. בהתאם לתקנות הגנת המידע האירופאיות (ה – 38GDPR), עיבוד מידע אישי שכזה הצריך בסיס חוקי להצדקתו. בעקבות זאת, ובעקבות ביקורת ציבורית על הגישה הנרחבת למידע אישי שהתקבל מסוכנויות ממשלתיות שונות והשימוש בו לבניית מדדי סיכון המבוססים על מודל של הקצאת נקודות המגלמות סיכון (לדוגמא: אבטלה = 500 נק'; מחלת נפש = 3,000 נק'; החמצת תור לרופא = 1,000 נק'), הושעתה לפי שעה התכנית ליישומה של המערכת הניסיונית. אולם, הממשלה הדנית מתכננת להרחיב השימוש בשירותים מבוססי-אלגוריתם בשדה הרווחה ברחבי המדינה.³⁹ גם אם השימוש במערכת ספג ביקורת מטעמים שאינם קשורים באופן פעילות המערכת, או ייעודה, עצם הביקורת הציבורית כנגד השימוש במידע, מדגימה את הצורך בקיומו של דיון ציבורי רחב ושקוף בעת שימוש במידע אישי נרחב ורגיש במערכות אלו.

36. Jacob Mchangama & Hin-Yan Liu, **The Welfare State Is Committing Suicide by Artificial Intelligence**, Foreignpolicy.com (December 25, 2018)

<https://foreignpolicy.com/2018/12/25/the-welfare-state-is-committing-suicide-by-artificial-intelligence/>

37. Catrine S. Byrne & Julia Sommer, **Is the Scandinavian Digitalisation Breeding Ground For Social Welfare Surveillance?** DataEthics (May 27, 2019)

<https://dataethics.eu/is-scandinavian-digitalisation-breeding-ground-for-social-welfare-surveillance/>

38. See Regulation (EU) 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 (General Data Protection Regulation), OJ L 119, 4.5.2016, p. 1–88, Art. 6; <https://gdpr-info.eu/art-6-gdpr/>.

39. Mchangama & Liu, לעיל ה"ש 36.

<https://foreignpolicy.com/2018/12/25/the-welfare-state-is-committing-suicide-by-artificial-intelligence/>; Brigitte Alfter, **Report – Automating Society [Denmark]**, Algorithm Watch (January 29, 2019) <https://algorithmwatch.org/en/automating-society-denmark/>

ארה"ב

מספר מדינות בארה"ב וביניהן, פלורידה, קונטיקט, אוקלהומה ועוד, יישמו מערכת ADM שמטרתה זיהוי ואיתור ילדים המצויים בסיכון להתעללות (גרימת פגיעה משמעותית, או מוות), באמצעות ניתוח חיזוי. במדינת אילינוי⁴⁰ למשל, 'הציפה' המערכת פרטיה של 4,100 ילדים, שזוהו בהתאם לתוצאת הסיכון (risk score) שהותאמה להם כמצויים בסיכון לפגיעה או למוות, בהסתברות של 90% ומעלה. יחד עם זאת, המערכת נכשלה בניבויים של מקרי מוות שהתרחשו בפועל, של ילדים לגביהם התקיימו מספר חקירות בגין התעללות. בעקבות תוצאות אלו, הופסק השימוש בכלי זה.

במדינת ארקנסו, יושם בשירותי הרווחה בשנת 2016, אלגוריתם שפותח על-ידי חברה פרטית. משימת האלגוריתם: החלפת ביקורי בית של אחיות להערכת הצורך בשעות סיעוד ביתי, עבור אנשים עם מוגבלויות שבחרו לקבל שירותי סיעוד בביתם, כחלופה לאשפוז במוסדות סיעודיים, על-ידי מערכת אוטומטית. מטרת יישום האלגוריתם: התייעלות כלכלית (שכשלעצמה, מהווה שיקול יעד חיוביים של הרשות המנהלית, המבטא אחריות ונאמנות כלפי האזרחים בניצול כספי ציבור) וקבלת החלטות חפות מהטיות אנושיות (העשויות לפגוע בשוויון). בדיעבד התברר, כי מערכת ה-ADM פעלה שלא כשורה וכי קביעות האלגוריתם, שהיו שרירותיות, חסרות הגיון ואבסורדיות לעיתים (כגון קביעה שקטוע כף רגל אינו סובל מבעיות רגליים, במה שזכה לכינוי 'algorithmic absurdities'), השפיעו לרעה על כמחצית ממבוטחי הבריאות של Medicaid במדינה. אי-הכרה בזכאות לשעות סיעוד, או הפחתה בשעות סיעוד מוכרות, הביאה לכך שמבוטחים רבים בעלי מוגבלויות משמעותיות, נותרו ללא גישה למזון, תרופות או שירותים במשך שעות ארוכות, כתוצאה מקביעות המערכת. המערכת גם לא היטיבה להבחין בין מצבים קשים מורכבים, וקטלגה אותם בצורה שגויה שלא זיכתה את המבוטחים בהיקף הסיוע הדרוש להם בפועל.

השימוש במערכת ה-ADM הופסק בהוראת ביהמ"ש הפדראלי, בעקבות תביעה שהוגשה נגד רשות הרווחה על-ידי קבוצת אנשים עם מוגבלויות מארקנסו, שנפגעו מקביעת מערכת ה-ADM. במהלך המשפט, הודו פקידי הסוכנות הממשלתית כי אינם יודעים להסביר את אופן עבודת המערכת, ואף נעדרו את המומחיות לאתר בה שגיאות. הודאה זו אינה מפליאה, לנוכח העובדה שכותבי האלגוריתם וספקי התוכנה עצמם, לא ידעו להסביר כיצד פועל האלגוריתם (או שלא ניאותו לספק הסבר כזה).⁴¹

הניסיון המצטבר עד כה (שאינו נרחב, בהינתן שיישומי טכנולוגיות אלו בשירות הציבורי, מצויים בחיתוליהם), מעיד על ליקויים שהדיגיטציה של מערכות רווחה ויישומן של טכנולוגיות שונות – המשמשות, בין היתר, לאוטומציה, ניבוי, זיהוי, מעקב, שירות מרחוק, איתור וענישה – הולידו. הליקויים כללו: צמצום מאגר המוטבים, ביטולם של שירותים שונים, צמצום משמעותי בתקציב הרווחה, הכנסת התניות מכבידות ופולשניות למיניהן, ועוד.⁴² יחד עם זאת, יש להניח כי ככל שהטמעת מערכות ADM תתרחב, ולקחים מיישומיהן יילמדו ויהוו בסיס לשיפורים והתאמות – התועלות משימושים אלו תבואנה לידי ביטוי באופן נרחב יותר.

40. Emily Keddell, **Algorithmic Justice in Child Protection: Statistical Fairness, Social Justice and the Implications for Practice**, 8 Social Sciences, 281 (2019); doi:10.3390/socsci8100281; Matt Leonard, **Illinois drops predictive analytics for at-risk children**, GCN.com (Dec 11, 2017); <https://gcn.com/articles/2017/12/11/illinois-drops-predictive-analytics-children-risk.aspx>.

41. Ryan Calo & Danielle Keats Citron, **The Automated Administrative State: A Crisis of Legitimacy**, Emory Law Journal (forthcoming) 1-2, 23-27 (2020).

42. Philip Alston, **World stumbling zombie-like into a digital welfare dystopia**, לעיל ה"ש 23.

יישום מערכות ADM בשירותי הרווחה – קונפליקטים טיפוסיים

ניתן להצביע על שני קונפליקטים מרכזיים, המתגלעים עם יישום מערכות ADM בשירותי הרווחה: האחד, בין המדינה לפרט; והשני, בין המדינה לבין שלוחיה – עובדי מערכת הרווחה.

קונפליקט בין המדינה לפרט – אחריות המדינה לרווחתו של הפרט ביישום מערכת ADM

האחריות כלפי האזרחים בתחומים כגון: בטחון, בריאות, חינוך ורווחה, נחה על המדינה באופן בלעדי (למעט היבטים מסויימים של אחריותו האישית של הפרט, לרווחתו הוא). למרות זאת, במדינת ישראל מספר שחקנים פעילים בזירת שירותי הרווחה. אלו מוענקים בפועל, הן על-ידי הממשלה – באמצעות משרד העבודה והשירותים החברתיים, הן על-ידי השלטון המקומי – על-ידי מחלקות הרווחה ברשויות המקומיות והן, על-ידי המגזר שלישי – דרך עמותות שונות.

יישום והטמעה של מערכות ADM פרי פיתוחן של חברות במגזר הפרטי, בשירותי הרווחה, מכניסה שחקן נוסף לזירה מבוזרת זו. גורם נוסף, בעל השפעה לא מבוטלת על עצם ואופן מתן שירותי הרווחה לפרטים הזקוקים להם, הוא האלגוריתם עצמו. במיוחד, משמדובר באלגוריתם לומד (למידת מכונה) ההופך למשפיע עצמאי בזירה זו.

ככלל, החלפת שיקול הדעת הבלעדי של הנציג האנושי של השירות הציבורי, בזה של אלגוריתם (באופן חלקי או מלא), מהווה סטייה מסויימת מן התפישה המסורתית של אחריות המדינה כלפי אזרחיה. כך למשל, יישום מערכות ADM בשירותי הרווחה, מביא במקרים מסויימים לכך שהאזרח נדרש ליתן דין וחשבון לרשויות הרווחה, על מנת להוכיח את עצם זכאותו לקבלת שירותי רווחה שונים, באופן שמהווה היפוך הנטל המקובל של אחריותיות הנדרשת לרוב מן המדינה כלפי אזרחיה, בתחום זה.

קונפליקט בין המדיניות הממשלתית/מנהלית לבין משימתו החברתית ומחוייבותו המוסרית של העובד הסוציאלי

מעבר לאחריות המדינה והאחריות המנהלית של הרשות הציבורית, עומדת גם האחריות האישית של שלוחיה, העובדים הסוציאליים ופקידי הסעד, המחוייבים מקצועית לרווחתו של מטופל הרווחה, תוך כפיפות לחוק (בראש ובראשונה – לחוק העובדים הסוציאליים, תשנ"ו – 1996; להלן – "חוק העובדים הסוציאליים") ולכללי האתיקה המקצועית.

חוק העובדים הסוציאליים הבא להסדיר את העיסוק המקצועי, הוא בעל מטרה כפולה, כפי שנקבע בס' 1 לחוק:

1. "קידום רווחת הפרט והחברה בישראל תוך שמירה על כבוד האדם ועקרון השוויון;

2. שמירה על רמה מקצועית והתנהגות הולמת של העוסקים במקצוע העבודה הסוציאלית."

[ההדגשות אינן במקור – ס.ת.]

יישום תהליכי דיגיטציה בשירותי הרווחה ושימוש במערכות ADM, צפויים לאתגר את ערכי החוק, כמו גם את הערכים העומדים בבסיס הדיסציפלינה של העבודה הסוציאלית, ככזו המחוייבת לאידיאל של צדק חברתי ועוד.⁴³

כך למשל, ייתכנו מקרים בהם יישום מערכות ADM לצד מעורבות אנושית והפעלת שיקול הדעת האנושי, או כחלופה לזה, יעמדו בניגוד למהותה הבסיסית של העזרה הסוציאלית, שעל הגשתה והנגשתה אמונים העובדים הסוציאליים. ואכן, מחקר איכותני שנעשה בארה"ב הביא את עמדתם (חששם) של אנשי המקצוע בתחום הרווחה, לגבי אפשרות קבלת החלטות על-ידי אלגוריתם בלבד. זאת, במנותק מהקשר וללא מעורבות גורם אנושי, על איכויותיו הייחודיות והחשובות לקבלת הכרעות בתחום. עמדת אנשי המקצוע הדגישה כי החלטות בתחום רווחת הילד, למשל, דורשות הבנה של ההקשר, וכי זו מתאפשרת באמצעות פרשנות אנושית וקשר אנושי בלבד.⁴⁴

בזירה המקומית, ערכי המקצוע הללו מגולמים בין היתר, בעקרונות קוד האתיקה החדש של העובדים הסוציאליים⁴⁵ בישראל. מתן טיפול ושירות רווחה בגישה מרוחק, או הכנסת מכונה לתהליך, שתסייע בעיבוד מידע ותדריך קבלת החלטות או שתייצרן לבדה, ללא מעורבות גורם אנושי, עשויים להציב קשיים במימוש העקרונות הבאים המוצאים את ביטויים בסעיפי הקוד [ההדגשות אינן במקור – ס.ת.]:

- **כבוד המטופל, אוטונומיה, הבנתיות:** "ישתמשו בשפה ברורה ומובנת לעדכן את הלקוחות במטרת השירות, בסיכונים הקשורים לשירות, במגבלות המוטלות על השירות ובחלופות סבירות לשירות, ובזכות הלקוחות לסרב או לסגת מהסכמתם."⁴⁶
 - **הסמכה למסירת מידע פרטי; מידתיות; הגבלת/ייחוד מטרה:** "העובדים הסוציאליים ימסרו לאחרים מידע על אודות לקוחותיהם רק כאשר החוק או הנסיבות המקצועיות יחייבו זאת, ורק במידה הנדרשת למטרה שלשמה תעשה החשיפה."⁴⁷
 - **מימוש/כיבוד אוטונומיה - הזכות לדעת:** "העובדים הסוציאליים יעדכנו את הלקוחות במידת האפשר על חשיפת המידע הסודי אודותיהם ועל ההשלכות האפשריות של פעולה זאת".
 - **מרכיב הקשר האנושי:** "מתוך הבנה, כי יחסים בין אנשים הם כלי חשוב לשינוי, יראו העובדים הסוציאליים את בני האדם להם הם נותנים שירות, כשותפים פעילים בתוך תהליך הסיוע ויחתרו לממש שותפות זו לקידום מטרות מקצועיות."⁴⁸
- מימוש מחוייביות אלו של העובדים הסוציאליים, במצב בו מיושמת מערכת ADM כחלק ממדיניות הממשלה/הרשות בתחום הרווחה, עשוי להיות מאתגר אף בשל מאפייניה הייחודיים של מערכת כזו והקשיים האינהרנטיים הנלווים לה, כפי שיפורט בהמשך בפרק 5.

43. Keddell, לעיל ה"ש 40.

44. Brown et al., לעיל ה"ש 30, בעמ' 8.

45. קוד האתיקה המקצועית של העובדים הסוציאליים בישראל, **האגודה לקידום העבודה הסוציאלית בישראל** (2018).

46. שם, בעמ' 13.

47. שם, בעמ' 14.

48. שם, בעמ' 11.

יתרונות ותועלות מעשיים ואתיים בקבלת החלטות מבוססות-אלגוריתם

בפרק זה נציג את היתרונות שביישום מערכת ADM בשירות הציבורי, ונפרט על תועלותיה בהקשר שלפנינו.

- **יעילות.** מערכות ADM מספקות שירות מהיר, מתאים וממוקד ומאפשרות ניצול מיטבי של משאבי כח אדם, תוך ניתובם למשימות דחופות יותר, או לכאלה בהן מעורבות הגורם האנושי היא הכרחית. ראה הרחבה על כך בהמשך הפרק, בסעיף על יתרונות מעשיים עבור העוסקים במלאכה.

- **אחריות.** יש שיטענו, שדווקא השימוש במערכות ADM והנכונות ליישם את תוצרי המערכת, משקפים גישה אחראית, המבטאת הכרה באוזלת היד של המערכת בנסיבות מסוימות, או בכשלונה להתמודד עם צרכים של פרטים מרובים, בדרך אחידה, שוויונית ועקבית. שימוש במערכות כאלה, מבטיח לכאורה התמודדות באופן רצוי כזה, וניחן אף ביכולת 'כיסוי' אפקטיבית יותר של פרטים זנוחים בפוטנציה (בבחינת 'המערכת רואה את כולם'). מערכות מבוססות בינה מלאכותית יכולות להוות כִּרְ נִסוּי להדמיית מערכות מורכבות, דבר המאפשר לקובעי-המדיניות להתנסות באפשרויות מדיניות שונות ואיתור השלכות לא רצויות או בלתי מכוונות לפני שמתחייבים לאמצן.⁴⁹ התנסות כזו משקפת אחריות וגישת זהירות, המצופה מנאמן הציבור.

- **מקסום תועלות.** שימוש במערכות אוטומטיות מבוססות בינה מלאכותית, עשוי לספק מידע ותחזיות מדוייקות יותר, אשר יספקו תוצאות טובות יותר. בהתאמה, הדבר צפוי להשיא רווחה גדולה, איכותית וממוקדת יותר, למספר גדול יותר של נזקקים, בעלי צרכים מרובים ומגוונים. יתרונותיה של המערכת יבואו לידי ביטוי במיוחד מקום בו עץ ההחלטה ברור, אינו נתון לפרשנות ואינו מותיר מרחב לשיקול דעת (או שמותיר מרחב מצומצם ביותר לכך), והפעולה היא פעולה חזרתית/מחזורית. במקום כזה, ניתן להסתמך יותר על המערכת, להפחית עד מאוד את מעורבות הגורם האנושי ולהפיק את מירב התועלת האפשרית מפעולת המערכת האוטומטית, תוך הידרשות מינימלית לסיכונים הנובעים מפעילותה.

- **אובייקטיביות וניטראליות.** גופים ממשלתיים וציבוריים מחוייבים באובייקטיביות בנקיטת פעולות ובקבלת החלטות. ביטוי של דבר הוא, שהחלטותיהם של אלו והקווים המנחים את פעילותיהם, יהיו חסרי פניות, נקיים מהטיות וכאלה שאינם נגועים באפליה כלפי אדם, או מגזר זה או אחר. על פניו נדמה, כי להחלטות מערכת ADM, תוצרי-מערכת-אוטומטית ופרי אלגוריתם מדעי, פוטנציאל לעמוד בסטנדרט של אובייקטיביות שגורם אנושי לעולם לא יוכל לספק. אולם, כנגד הנחה זו מושמעת ביקורת, כפי שניתן לראות להלן, בפרק 5.

- **צדק חלוקתי.** קידום שוויון, אי-הפליה וחלוקת תועלות באופן צודק; כל אלה, בזכות אובייקטיביות המערכת.

— **צדק חברתי** הטמעת מערכות ADM בשירותי רווחה ברמת השלטון המקומי, מקדמת הרמוניזציה באמצעות השוואת תנאים, ותורמת לצמצום פערים ברמת ואיכות השירותים להם זוכים תושבי רשויות מקומיות שונות. כך למשל, ניתן למנוע מקרים בהם פקידי רווחה במועצות מקומיות שונות,

49. GDS & OAI, **A guide to using artificial intelligence in the public sector**, לעיל ה"ש 1, בעמ' 9.

יקבלו החלטות שונות (באופן שרירותי, או מטעמים שאינם ענייניים) לגבי מקרים שנסיבותיהן דומות או זהות, ובכך ייפגע השיוויון בין האזרחים מקבלי השירות.

- פרסונליזציה.** שיפור השירותים הציבוריים באמצעות התאמה אישית מדויקת יותר שלהם לנסיבותיו האישיות של האזרח, לנתוניו ולצרכיו האישיים וזאת, בהתבסס על מידע מדויק ומפורט יותר.⁵⁰ ההבנה כי אנשים נבדלים זה מזה באישיותם וביכולת ההתמודדות שלהם עם מצבי סיכון ומצוקה, ולפיכך אף נדרשים לסעד שונה – מזמינה למעשה אספקת מגוון שירותי רווחה באמצעות מערכות ADM. מערכות חכמות אלו מסוגלות, על-ידי ניתוח ועיבוד כמויות גדולות של נתונים, להתאים לכל מטופל רווחה את השירות המתאים לו, או לזהות את הסיכון הנשקף לו מעצם המצאו במצב מסויים, בצורה מיטבית. הסיכוי/סיכון שמטופל פוטנציאלי 'יחמוק מתחת לרדאר' של מערכת כזו, בהשוואה ליכולת הזיהוי והתשומות המוגבלים של הגורם האנושי בשירותי הרווחה, גם הוא נמוך יותר.
- מניעת (או מזעור) נזקים עתידיים.** התראה (flagging), איתור וגילוי מוקדם של פרטים המצויים בסיכון, אשר לא היו נחשפים אילולא אותרו על-ידי התראות המערכת. מחסור בכח אדם במערכת הרווחה, בשילוב עם יכולת אנושית מוגבלת (בהשוואה לזו של מערכת אוטומטית) לעיבוד כמויות מידע אדירות והסקת מסקנות מהירה, מצביעים על יתרון ביישום מערכות ADM למתן התראה על סיכון לפרטים, מבחינת היעילות שבדבר. לגבי דיוק ההתראות, ההשוואה בין זה של המערכת הטכנולוגית לבין זה של הגורם האנושי היא מורכבת ורבודה, ולכן קשה במצב הידע הנוכחי לקבוע בדיוק מי מאלו נוטה יותר לטעות. גורם נוסף בעל השפעה בהיבט זה, הוא מרכיב ההיקף הפוטנציאלי של השפעת השגיאה, הרחב ביותר במקרה של מערכת ADM, כפי שיפורט בהמשך.
- הגברת אמון הציבור.** בעקבות הניטרליות הלכאורית של ההחלטות המתקבלות על-ידי מערכות ADM, ואם יישום המערכת ילווה גם בשקיפות טכנית ואחרת, ניתן לצפות כי שימוש במערכות ADM יגביר את אמון הציבור במערכת הציבורית.
- יתרונות סביבתיים וחברתיים.** רשות מנהלית מחוייבת בהחלטתה ובפעולתה לשקול שיקולי תכלית, יעילות, משאבים מצויים, ושאר שיקולים מעשיים.⁵¹ ואכן, יישום מערכות מבוססות בינה מלאכותית בשירות הציבורי עשוי להצמיח תועלות מעשיות לסביבה ולמערכת כולה, במגוון היבטים ורבים, כגון: יצירת פתרונות עבור בעיות חברתיות אוניברסליות מאתגרות (השפעה חברתית חיובית), על ידי רתימת בינה מלאכותית לדבר;⁵² והתייעלות, למשל על-ידי חסכון בתשומות המשאב האנושי, המאפשרת חסכון כלכלי והפניית כח אדם לביצוע משימות אחרות, מהן יפיק תועלת אישית ומקצועית כמתואר מטה.

50. GDS & OAI, **A guide to using artificial intelligence in the public sector**, לעיל ה"ש 1, בעמ' 9.

51. יצחק זמיר, "שיקול הצדק בהחלטות מנהליות", **משפט וממשל** ז תשס"ה, 623-667, בעמ' 623.

52. GDS & OAI, **A guide to using artificial intelligence in the public sector**, לעיל ה"ש 1, בעמ' 9.

יתרונות מעשיים עבור העוסקים במלאכה – אנשי מערכת הרווחה:⁵³

- **הקלה בעומס.** למשל, על-ידי אוטומציה של מטלות חזרתיות וגוזלות-זמן, שהינן מתסכלות ולא מתגמלות עבור פקידי הרווחה. אוטומציה כזו, תאפשר לפקידי הרווחה לעבור מעיסוק במשימות שגרתיות, להתמקדות במשימות רבות-ערך ומתגמלות יותר, כאלו שהינן גם מורכבות יותר ומחייבות הפעלת שיקול דעת אנושי, הפגנת רגישות וחמלה וכיו"ב.
- **הקצאה מחדש של משאבים.** תיעול הפנאי, להכשרה מחדש של כח האדם באופן המיטבי עבור המערכת והעובד.
- **העצמת עובדי הרווחה.** העבודה בסביבה דיגיטלית מתקדמת, ושיתוף הפעולה של פקידי הרווחה עם טכנולוגיות תומכות, דוגמת מערכת ה-ADM, יש בה כדי להעשיר ולהעצים את העובד המקצועי ואת יכולותיו וכישוריו. התפישה החלופית (לזו המקובלת, המבטאת חשש מייטור העובד האנושי עם הטמעת טכנולוגיות בינה מלאכותית בתחומי תעסוקה שונים), הרואה בבינה מלאכותית כלי מסייע ומכירה בערך זה, מכונה: Augmented Intelligence (אינטליגנציה מועצמת).⁵⁴ תפישה זו, מבקשת להצביע על התרומה שבעבודה המשותפת של אדם ומכונה, זה לצד זה (למשל, שיתוף פעולה עם רובוטים), ליכולות הקוגניטיביות של העובד האנושי. זוהי מעין תגובת-נגד לדוקטרינה הטוענת שלהטמעת מערכות בינה מלאכותית בסביבת העבודה האנושית ישנה השפעה שלילית של הפחתת-כישורים (deskilling) עבור העובד האנושי. זוהי גם תגובה לתופעה לפיה, עובדים אנושיים העובדים בסביבה בה מוטמעת מערכת מבוססת בינה מלאכותית, נוטים להגיב בהטיית אוטומציה (automation bias). הטיה זו מייחסת ערך רב יותר לתוצרי אוטומציה ובינה מלאכותית, ובהתאמה – הסתמכות-יתר על החלטה תוצר-המערכת, מבלי להרהר אחר ההגיון שבה או לערער על דיוק חישוביה. משמעות ההטיה היא, שמכוחה מתערער אמונם של העובדים האנושיים ביכולותיהם המקצועיות (professional confidence).⁵⁵

53. Rainer Binder, Paolo Pellegrini, Gaurav Gujral, **Intelligence-Powered Human Services**, Accenture (2017) <https://www.accenture.com/acnmedia/PDF-65/Accenture-AI-in-Social-Services-POV-FINAL.pdf#zoom=50>;
 GDS & OAI, **A guide to using artificial intelligence in the public sector**, לעיל ה"ש 1, בעמ' 9.

54. **Augmented Intelligence**, Gartner Glossary, Gartner <https://www.gartner.com/en/information-technology/glossary/augmented-intelligence>; Daniel Araya, **3 Things You Need To Know About Augmented Intelligence**, Forbes (Jan 22, 2019) <https://www.forbes.com/sites/danielaraya/2019/01/22/3-things-you-need-to-know-about-augmented-intelligence/#4741fe693fdc>; **Artificial Intelligence vs. Augmented Intelligence**, AMA <https://www.ama-assn.org/amaone/augmented-intelligence-ai>.

55. Redden et al., לעיל ה"ש 14, בעמ' 14.

חסרונות וקשיים מעשיים ואתיים בקבלת החלטות מבוססות-אלגוריתם

דיגיטציה של מערכת הרווחה ושימוש בכלים מבוססי בינה מלאכותית לצורך קבלת החלטות ביחס לבני-אדם, בייחוד בכל הקשור לזכויותיהם ולזכאויותיהם של מטופלי רווחה, הינם בעלי השפעה שלילית פוטנציאלית על מימושן של מגוון זכויות פרט. לאתגרים המעשיים הייחודיים לאימוץ מערכות ADM בשירות הציבורי, אותם הבאנו בתחילתו של נייר זה, מתווספים קשיים אתיים העולים ממאפיינים הטיפוסיים למערכות ADM (ומערכות מבוססות בינה מלאכותית, בכלל), כמו גם סיכון של מגוון זכויות פרט, כפי שיפורט להלן.

זכויות פרט בסיכון

זכויות האדם והאזרח העשויות להיות מושפעות לרעה מיישום מערכות ADM, או להימצא בסיכון כתוצאה מכך, הינן מגוונות. נמנות עליהן, בין היתר: הזכות לכבוד, הזכות לאוטונומיה, הזכות לפרטיות, הזכות לשוויון (ליחס שוויוני ושאינו מפלה על-ידי האלגוריתם); הזכות להגדרה עצמית (במקרים של הגדרות סטטוס לזכאות לשירותים שונים – מיהו נכה/מוגבל? מיהו ילד המצוי בסיכון? וכיו"ב); הזכות לבטחון סוציאלי, הזכות לרווחה/בריאות (במקרים של אי-עמידה בקריטריונים לזכאות לשירותים שהוגדרו על-ידי המערכת); הזכות להיות מוגן מפני התערבות צדדים שלישיים; הזכות לצדק חברתי (המבקשת להבטיח חלוקת תועלות באופן צודק והוגן בין פרטים בחברה); וזכויות אדם ספציפיות של ילדים (מטופלי רווחה בפועל, או בפוטנציה)⁵⁶. עצם הפגיעה באותן זכויות תלויה במהות שירות הרווחה לגביו מיושמת המערכת, בנסיבותיו האישיות של הפרט, נתין-המערכת, ובסוג ההחלטות המתקבלות על-ידי האלגוריתם שלה (אלגוריתם מסייע, או אלגוריתם מחליט).

מספר זכויות פרט הנתונות לסיכון פוטנציאלי על-ידי מערכת ADM, מעניינות במיוחד בהקשר שלפנינו:

- **הזכות להוגנות והזכות לצדק (מהותי ופרוצדוראלי).** זכותו של נתין האלגוריתם ליחס צודק וללא משוא פנים על-ידי מערכת (ככל שמערכת מסוגלת לחקות התנהגות אנושית כזו). הזכות לצדק כוללת את הזכות לקבול נגד החלטה ומחייבת לשם כך הבנת אופן קבלת ההחלטה בעניינו של מושאה, על מנת שיוכל לתקוף אותה באופן קונקרטי וביעילות. מימושה של זכות זו מחייב שקיפות, הן לגבי עצם קבלת החלטה באופן אוטומטי על-ידי מערכת ADM והן, לגבי אופן פעילות האלגוריתם וסוג המידע שהוזן אליו. האחרון, מהווה למעשה את דרישת ההסברתיות (explainability).
- **הזכות להליך הוגן פרוצדוראלי.** זכות זו, היא המעניינת אותנו בעיקר בהקשר של קבלת החלטות מבוססות אלגוריתם, שכן עצם יישום מערכת ADM מאתגר את הליך קבלת ההחלטות המסורתית, עורך בו שינויים, וכרוך באי-ודאותם של מושאי-ההחלטה ביחס ליכולתם ליהנות מהליך שיבטיח מימוש

56. זכויות אלו מעוגנות באמנה לזכויות הילד משנת 1989, אותה אשררה ישראל בשנת 1991, ומוחלות במפורש גם על רשויות מנהליות ועל מוסדות רווחה. אגד הזכויות הזה בא להבטיח הגנה על טובת הילד בפעולות הננקטות על-ידי גופים אלו, במגוון היבטים, כגון: הזכות להשמע בהליכים מנהליים הנוגעים אליו, והזכות להיות מוגן מפני חדירה לפרטיותו; ראה אמנה בדבר זכויות הילד, כ"א 31, 221 (נפתחה לחתימה ב- 1989) (אושרה ונכנסה לתוקף ב- 1991);

available at <https://www.ohchr.org/en/professionalinterest/pages/crc.aspx>.

זכויותיהם בהקשרים נשואי ההחלטה. הזכות להליך הוגן היא למעשה אגד זכויות פרוצדורליות, הבאות להגן על כבודו וחירותו של הפרט. על אגד זה נמנות: זכות הטיעון; הזכות להישמע בתגובה להחלטות מנהליות המתקבלות בעניינו של אדם ומשפיעות על זכויות ואינטרסים יסודיים שלו; הזכות להיבחן על-ידי מקבל החלטות הוגן, בלתי-מוטה, ועצמאי; הזכות לקבל הסבר/נימוק להחלטה – לצורך הבנתה וערעור עליה; וזכות הערעור על ההחלטה. ישנם מלומדים הסוברים, כי בהאצלת סמכות הרשות לקבל החלטות – למערכות אוטומטיות, הזכות להליך הוגן, במובנה המקובל, עוברת שינוי ייעודי והופכת כפי שכבר הוזכר לעיל, לזכות ל"הליך הוגן טכנולוגי".⁵⁷

• **הזכות שלא להיות נתון להליך קבלת החלטות אוטומטי בלבד – זכות ייעודית חדשה, יצירת ה - (GDPR) (General Data Protection Regulation), רגולציית הפרטיות העדכנית של האיחוד האירופי [Art. 22].** ישראל עשויה לאמץ זכות זו, במסגרת תיקוני חקיקה עתידיים בתמיכה במאמצי ישראל לשמר את הכרת האיחוד האירופי בה, כמדינה המעניקה רף גבוה של הגנה על מידע אישי של פרטים.^{58,59} זכותו של נושא המידע שלא להיות נתון להליך קבלת החלטות אוטומטי בלבד, כזה שאין בו מעורבות גורם אנושי והנטול שק"ד אנושי, באה להבטיח מפני שרירותיות ההחלטה פרי-מערכת אוטומטית. דהיינו, יש לאפשר לפרט התמודדות מול גורם אנושי וקבלת מענה אנושי, אותו ניתן לתקוף בתנאים מוכרים וידועים ובכלים סבירים ודמוקרטיים (כאלה המצויים בידיו של כל אזרח). מזכות זו נגזרות הזכויות הבאות:

— **הזכות לקבל התערבות אנושית.** מענה מגורם בעל שיקול דעת אנושי, הפועל על-פי מערכת ערכים מקובלת, צפויה וודאית באופן יחסי (מנגד, יש שיאמרו שהחלטה מחושבת של מערכת אוטומטית, תהיה ודאית ועקבית יותר).

— **הזכות להביע עמדה**

— **הזכות לערער על החלטה (של מערכת אוטומטית).** הזכות להתנגד להחלטה ולתקוף אותה. מימושה של זכות כזו, מצריך כתובת מוסמכת, אפקטיבית, לשמיעת ובחינת התנגדותו/ערעורו של פרט, בגין החלטה שהתקבלה בעניינו באמצעות מערכת ADM.

יודגש, כי לא כל הליך דיגיטציה, או כל יישום של מערכת ADM בשדה הרווחה, ישפיעו בהכרח על כל מגוון הזכויות המנויות לעיל. לעיתים יושפעו רק חלק מן הזכויות האמורות. אך גם השפעה אפשרית על חלק מהן, מחייבת התייחסות ושקילת תועלות מול סיכונים, בבואה של המדינה, או הרשות, לאמץ מערכת כזו אל שירותי הרווחה.

57. ראה: Calo & Citron, לעיל ה"ש 41, בעמ' 2-1, 27-23; Danielle Keats Citron, **Technological Due Process**, 85 WASH. U. L. J. 27-23, 2016 (REV. 1249).

58. Regulation (EU) 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 (General Data Protection Regulation), OJ L 119, 1 (4.5.2016).

ישראל אמנם אינה כפופה במישורין לרגולציה ה-GDPR, אך זכתה בעבר למעמד של Adequacy (מדינה שאינה חברה ב-EU, אך הוכרה כמספקת הגנה נאותה למידע אישי של פרטים ולפיכך, מותר להעביר אליה לעיבוד מידע על אזרחי האיחוד האירופי), ביחס לדיקטיבה האירופית הקודמת להגנת הפרטיות. ככל שישאל תפעל להמשך הכרה זו בה, צפוי שהיא תיחדש את דיניה לעקרונות החדשים שנוספו ב-GDPR ביחס לאזרחי האיחוד האירופי.

59. Mittelstadt et al., לעיל ה"ש 9, בעמ' 14.

קשיים אתיים אינהרנטיים לאופיין של מערכות ADM

להלן מספר מאפיינים בעייתיים של מערכות ADM, והשלכות טיפוסיות של יישומן, המשליכים על מימוש זכויות הפרט המנויות לעיל, או ההגנה עליהן:

- **אי-הסברתיות (Non-explicability) ובעיית הקופסא השחורה ('Black box' problem).** האלגוריתמים הטיפוסיים המעסיקים כיום את החשיבה האתית בתחום, הם כאלה שפעולותיהם קשות לחיזוי אנושי, או שלוגיקת קבלת החלטות שלהם מורכבת להסבר, לאחר מעשה.⁶⁰ הקושי בחיזוי, מאתגר את הרעיון של אחריות מיישם מערכות ADM כלפי מושפעי-המערכת. אתגר ההסברתיות הוא כה גדול ומורכב, שלעיתים אפילו מפתחי האלגוריתם עצמם, או מומחים בתחום, יתקשו להסביר את אופן פעולתו של אלגוריתם. הדבר נכון במיוחד, מרגע שזה יצא ל'דרך עצמאית' על-ידי למידת מכונה, כ- (semi-autonomous) (creation).⁶¹ יחד עם זאת, בדו"ח שפורסם לאחרונה בבריטניה, מצאה הוועדה הממשלתית - **Committee on Standards in Public Life**, כי:

"Fears over 'black box' AI, however, may be overstated, and the Committee believes that explainable AI is a realistic goal for the public sector."⁶²

באותו דו"ח, תואר שיח עם חברות פרטיות המפתחות מערכות מעין אלה ולפיו, בעוד קיימת בידהן היכולת להפוך את מוצריהן ליותר 'ניתנים להבנה', הן לעיתים רחוקות נתבקשו לעשות כן על-ידי אלו האמונים על רכש המערכות עבור השירות הציבורי. משמעו של דבר, אם כך, שלפחות בחלק מן המקרים ניתן **לאיין את בעיית הקופסא השחורה, או לצמצם את אי-ההסברתיות** על-ידי דרישה מוקדמת **לשקיפות טכנית**, דהיינו בשלב תכנון ופיתוח מערכת ה- ADM, או בשלב הזמנתה על-ידי פקידי השירות הציבורי. יתרה מכך, לנוכח העובדה שאופן קבלת החלטות על-ידי גורם אנושי הוא דבר קשה להסבר לכשעצמו, יש הסבורים שדווקא במצב בו שיקול הדעת האנושי ייתמך על-ידי ADM, תוך ליווי בשקיפות טכנית והסברה מסויימת (שתספק, בין היתר, בהירות לגבי המשתנים המשפיעים על קבלת החלטה כזו או אחרת) – ניתן יהיה להגביר את הפתיחות והאחריות בשירות הציבורי.⁶³

— **מהי רמת ההסברתיות הנדרשת?** מבחינה מהותית, לצורך קביעת רמת ההסברתיות הראויה, יש להבחין בין **החלטות בעלות השפעה משמעותית** על זכויות אזרח, בהן מעורבת מערכת ADM, כגון אלה הקובעות זכאות לשירות ציבורי, רמות סיכון וכיו"ב, לבין **החלטות בעלות השפעה נמוכה** על זכויות אלה, הקשורות בתפקוד המערכת וביעילותה. ככל שמדובר בהחלטה בעלת השפעה עמוקה ונרחבת יותר על זכויות ואינטרסים של נשואי השירות הציבורי, כך נדרשת רמת הסברתיות גבוהה ומעמיקה יותר לרציונאל שהדריך את קבלת ההחלטה בנסיבותיו של מקרה ספציפי. בהתאמה, ככל שמדובר בהחלטה בעלת השפעה רדודה על הפרט, כזו הבאה לשרת צרכים פנים-מערכתיים, כך תהא רמת ההסברתיות הנדרשת נמוכה וכללית יותר.⁶⁴

60. שם, בעמ' 3.

61. שם, בעמ' 11.

62. Committee on Standards in Public Life, **Artificial Intelligence and Public Standards**, לעיל ה"ש 1, בעמ' 6.

63. שם, בעמ' 23-24.

64. שם, בעמ' 24.

— **הסברתיות עבור מי?** שאלה חשובה מסדֵר שני המצריכה חשיבה, היא למי יש לספק את ההסברים האמורים, מתי ובאיזה אופן. ההנחה המובלעת צריכה להיות, שההסברים לפעילות מערכות ADM בשירות הציבורי מכוונים לציבור הרחב, ונדרשים להיות זמינים ומפורסמים באמצעים נגישים מקובלים (למשל, באתרי האינטרנט של אותם שירותי ממשל).

— **פגיעה באוטונומיה.** אי-ההסברתיות ובעיית הקופסא השחורה, מאתגרות גם את האוטונומיה של הפרט, הזכאי לוודאות, שקיפות והבטחת עקביות בקבלת החלטות המשפיעות, בדרך זו או אחרת, על מימוש זכויות ואינטרסים שלו. אי-היכולת של הפרט לסרב להיות נתון לשיקול דעת אלגוריתמי על-ידי מערכת ה-ADM, פוגע גם הוא באוטונומיה שלו. זאת במיוחד, כשהחלטת האלגוריתם היא כזו השוללת זכויות או מצמצמת זכויות באופן המשפיע על אינטרסים שונים של הפרט.

— **קושי בהוכחת נזק על-ידי הנפגע מיישום מערכת ADM.** אי-ההבנה של אופן פעולת האלגוריתם, דרך קבלת ההחלטה או המשתנים שנלקחו בחשבון במהלכה, בצירוף פערים אפיסטמיים (פערי ידע) – משפיעים על יכולת הנפגע, הפרט מושא-ההחלטה, להוכיח נזקים יצירי-האלגוריתם. הקושי בתקיפת החלטות מערכת ADM, במיוחד כאלה הקובעות זכאות לשירותים, או של מערכות ADM מעין-שיפוטיות, עשוי למנוע ממושא ההחלטה יכולת אפקטיבית להוכיח שהראשונה פגעה במגוון זכויות אישיות וקנייניות שלו ולגרום, כאמור לעיל, לפגיעה בזכות להליך הוגן (אי-צדק פרוצדוראלי). מדובר בבעיה מסדר שני, תוצר אי-ההסברתיות ובעיית הקופסא השחורה. ניתן לטעון כי בעיה זו מוחרפת במיוחד, כשמדובר באוכלוסייה מוחלשת ופגיעה (אם באופן בסיסי, ואם מכורח הנסיבות), כאוכלוסיית הרווחה.

• **אחריות ואחריותיות.** ברי, כי ככל שהשליטה האנושית באופן קבלת החלטות מצטמצמת, עד שכמעט מתאיינת (במקרים בהם מערכת מבוססת בינה מלאכותית ותהליך קבלת החלטות – נהיים אוטומטיים יותר), כך יהיה קשה יותר לייחס לגורם האנושי אחריות על תוצאות פעילות המערכת,⁶⁵ הן מבחינת מהות החלטות המתקבלות והן, מבחינת הדרך בה התקבלו.

כפי שצויין בפרק האתגרים לעיל, ייתכן שבעתיד מערכות בינה מלאכותית יוכרו בחוק כישות משפטית הנושאת גם בחובות, ומתוקף כך – ניתן יהיה לייחס למערכת עצמה אחריות להחלטות שהתקבלו על-ידיה, או לנזקים-יצירי-המערכת. להכרה כזו יהיו כמובן השלכות חברתיות ומשפטיות מרחיקות לכת. עד אז, ייחוס האחריות יהיה תוצר של ניתוח מורכב, שיידרש בראש ובראשונה ל**זיהוי** מגוון השותפים לתהליך ול**קביעת חלקם היחסי בו**, כדי לבסס אחריות. מצב זה הופך את ייחוס האחריות לבלתי-ודאי, שכן ייתכן שלא נדע להצביע באופן יעיל על כל האחראים בפועל. האם יהיו אלו מתכנני המערכת, יצרניה, הספקים הפרטיים, או שמא הממשלה – המזמינה, מטמיעה ומיישמת המערכת בשירות הציבורי לאזרח? זאת, בנוסף על הקושי ההוכחתי המשמעותי בייחוס אחריות לתוצאת האלגוריתם. מצב כזה מזמין בהכרח 'מרווח אחריותיות' (accountability gap) – כשאיש, או מכונה, לא נמצאים אחראים להחלטות המערכת, דבר המותיר את האזרח נושא ההחלטה ללא סעד או יכולת ערעור עליה.

שאלת מרווח האחריותיות עולה ביתר שאת, בשעה שממשלות פונות בהדרגה ל- third-party vendors של כלי בינה מלאכותית, על מנת שאלו יספקו את הארכיטקטורה האלגוריתמית לשירותים ציבוריים, בין היתר בתחום הרווחה. משמעות הדבר היא, שמערכות ADM תורמות למנגנון קבלת החלטות הממשלתי, מבלי

65. שם, בעמ' 20.

שיפעל לצידן מנגנון של חבות (liability) או אחריותיות (accountability).⁶⁶ באופן זה, נופלות המערכות לתוך מרווח האחריותיות, ומותירות את האזרח חסר כתובת ישירה, מובהקת, הנושאת כלפיו באחריות ספציפית לנזק יציר מערכות ADM. על מנת למנוע תוצאה זאת, על ממשלות העושות שימוש במערכות אלו לקבל שאינן יכולות לפטור עצמן מאחריות לפעולת המערכת, הגם שהן נעדרות יכולת אמיתית להבין את אופן פעולתה או להבינה באופן מלא.

• **הטיה.** כנגד הנחת אובייקטיביות (מדעיות) המערכת והחפות מהטיות הטמונה בבסיסה, מושמעת ביקורת^{68,67} המבקשת לסתור את אותה הנחה. יכולתה של המערכת לשמור על אובייקטיביות, תלויה לחלוטין באובייקטיביות המידע המוזן אליה ושעל בסיסו היא מאומנת. וזהו בדיוק עקב אכילס של אובייקטיביות מערכת בינה מלאכותית.

נתוני העתק הזמינים למערכת, כוללים הטיות באופן אינהרנטי. המידע מבוסס, בין היתר, על מאגרים שמלכתחילה אספו מידע באופן סלקטיבי, על מחקרים שבוצעו באוכלוסייה סלקטיבית,⁶⁹ ועל דוגמאות מן החיים האמיתיים שהוטמעה בהן העדפה מפלה (למשל, בהעדפת מועמדים ממין זכר בקבלה לעבודה).⁷⁰ ניסיונות ל'טיהור' הנתונים על-ידי הסרת מאפייני מידע מסויימים העשויים לגרום להטיות, כגון מין או גזע, נכשלו עד כה. זאת, אם בשל העובדה שאותן הטיות התקיימו ממילא במאפיינים מייצגים סמוכים (proxy characteristics), כגון אזור מגורים, או בשל העובדה שהסרת מאפיינים כאלו, עלולה לגרום להטיות אחרות כנגד פרטים או מגזרים אחרים. כך לדוגמה, הסרת מאפיין המין בחיזוי מסוכנות לרצדיביזם (מועדות חוזרת לפשיעה), עשויה להיטיב עם גברים, אך באותה העת לפגוע בנשים, המועדות פחות לפשיעה חוזרת אך עשויות לקבל ציון סיכון גבוה יותר בשל החרגת הרלבנטיות של מין.

נראה שהתגברות על הטיות בנתונים, למשל על-ידי הפיכת מאפיינים מסויימים לסמויים מן האלגוריתם, אינה מבטיחה בהכרח קבלת תוצרי מערכת הוגנים יותר ותידרש עדיין לאמצעים טכניים וטכנולוגיים פרואקטיביים, לזיהוי ההטיות הקיימות (ר' להלן דוגמאות בפרק ההמלצות, בסעיף בחינת 'צדקת' האלגוריתם).

ואף אם נצליח להתגבר על הטיות המידע, נידרש להתגבר על מכשול נוסף, אינסטרומנטלי – זה של ההטיות הקיימות **באנשים, מפתחי** (בעיקר, מתכנתי) **המערכת** עצמם. אלה לרוב אינם מגוונים דיים, מגיעים מפלח מסויים של האוכלוסייה (מעמד סוציאקונומי בינוני-גבוה, השכלה דומה וממוצא אתני לא מגוון) ולפיכך – לוקים בהטיות דומות. ההטיות האנושיות משועתקות אל תוך השימוש בנתונים, במישורן (על-ידי מתכנתי המערכת ומפתחיה), או בעקיפין (באמצעות הנתונים בהם נעשה שימוש לאימון המערכת והזנתה) ומשפיעות על ההחלטות תוצר-המערכת.

66. Crawford & Schultz, לעיל ה"ש 29, בעמ' 1942.

67. Philip Alston, **Report of the Special rapporteur on extreme poverty and human rights**, לעיל ה"ש 23, בעמ' 4.

68. Annette Zimmermann, Elena Di Rosa, and Hochan Kim, **Technology Can't Fix Algorithmic Injustice**, Boston Review (January 09, 2020); <http://bostonreview.net/science-nature-politics/annette-zimmermann-elena-di-rosa-hochan-kim-technology-cant-fix-algorithmic>.

69. כך למשל, מרבית המידע המצטבר ממחקרים קליניים בבני-אדם סובל מחוסר גיוון (הגורם להטיה מובנית), לאחר שנאסף במשך שנים בבעיקר ממשותפים ממוצע אירופאי (Caucasians), בעשור השלישי והרביעי לחייהם.

70. Jeffrey Dastin, **Amazon scraps secret AI recruiting tool that showed bias against women**, Reuters (October 10, 2018)

<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>; Karen Hao, **Facebook's ad-serving algorithm discriminates by gender and race**, MIT Technology Review (April 5, 2019)

<https://www.technologyreview.com/2019/04/05/1175/facebook-algorithm-discriminates-ai-bias/>

יחד עם זאת, ייתכן בהחלט שכמות ההטיות המתקיימות במגוון הרב של הגורמים האנושיים המקבלים אותן, במצב המסורתי בו השירות הציבורי אינו מסתייע בבינה מלאכותית – אינה פחותה מזו של ההטיות הקיימות בהחלטות מערכת ADM. שהרי באופן טבעי, אופן חשיבתם של בני אנוש מושפע משפיע הטיות שמקורן במגוון רחב של השראות חיצוניות ונטיות מחשבתיות (פרי חינוך, תרבות, ניסיון חיים, אקלים אידיאולוגי/פוליטי/דתי/חברתי, ועוד). נראה, כי אנו נדרשים למחקר אמפירי נוסף שיסייע בהשוואת משקלה של ההטיה האנושית בקבלת החלטות על-ידי גורם אנושי, אל מול משקלן של הטיות במידע ובפיתוח, בקבלת החלטות באמצעות מערכת ADM. דהיינו, אילו החלטות מוטות יותר, באופן יחסי.

על רקע זה, הולכת ומתגבשת ההבנה, כי אחד האמצעים לשיפור בעיית אי-האובייקטיביות במידע, בנוסף על 'ניקוי' הנתונים מהטיות, הוא התגברות על בעיית היעדר המגוון האנושי בקרב מפתחי מערכות ה-ADM. המצב הוגדר היטב במשפט הבא:

“[w]hile data bias may create discrimination, a lack of diversity will facilitate it”⁷¹

• **אפליה על-ידי אלגוריתם.** כאשר מערכת ADM מייצרת החלטות שונות, בנוגע למושאי-החלטה דומים, מטעמים ששיפוט אנושי היה מגדירם כ'טעמים שאינם ענייניים', היא גורמת לאפליה. מקורה של אפליה (לא מכוונת) שכזו, היא כאמור בהטיות המובנות במידע בו היא מוזנת ועל בסיסו מאומנת, או בהטיות של המפתחים עצמם.^{72,73}

דוגמא לכך, ניתן למצוא בשימוש בטכנולוגיות של שיטור מנבא (predictive policing) לחיזוי של אזורים בהם צפויה היארעות גבוהה של פשיעה. המערכת מאומנת על נתוני פשיעה קיימים המקשרים התנהגות עבריינית עם רקע גזעי מסויים, או עם אזור מסויים בעיר, ואלו אף מוזנים לתוכה. הנתונים גורמים למערכת להתמקד בהגברת מאמצי שיטור ואכיפה באזורים אלו, או באנשים מרקע גזעי מסויים, וליצירת הנחות סטריאוטיפיות מפלות לגביהם, מראש. כך קרה גם במערכת האלגוריתמית - COMPAS, שסייעה לשופטים במערכת המשפט הפלילית בארה"ב בחיזוי סיכון לרצדיביזם, בשלב גזירת העונש.⁷⁴ המערכת התבררה כמפלה ומוטה גזעית, ככזו המצביעה על סיכון מוגבר לרצדיביזם בעניינם של נאשמים ממוצא אפריקאי-אמריקאי ועל סיכון נמוך – כשמדובר בנאשמים 'לבנים', גם אם האחרונים כבר הוכיחו מועדות כזו בפועל. מערכות ADM אחרות, בתחום התעסוקה, התגלו כמוטות לטובת גברים והביאו להפליה על בסיס מגדרי.⁷⁵

טיעון זה, בדבר אפליה על-ידי אלגוריתם, סותר את ההנחה לפיה ADM היא פעולה ניטראלית, אובייקטיבית ומדעית ביסודה, המבוצעת על-ידי מערכת אוטומטית ולפיכך – חפה מהטיות.

— **סטיגמטיזציה.** סימון בהקשר שלילי, או הכתמה של קבוצה/קהילה, בעקבות קביעות של מערכות ADM. פגיעה כזו עשויה לפגוע בכבוד האדם, בזכות להיות נתון ליחס שווה, להסב נזקים אישיים וחברתיים, ולגרום לאובדן הזדמנויות (בתעסוקה, חינוך, דיור ועוד).

71. **Artificial Intelligence and Public Standards**, Committee on Standards in Public Life, לעיל ה"ש 1, בעמ' 27.

72. KNJ Macnish, **Unblinking Eyes: The Ethics of Automated Surveillance**, 14 (2) Ethics and Information Technology, 151, 158 (2012); Zimmermann et al. 60 לעיל ה"ש 60.

73. Mittelstadt et al., לעיל ה"ש 9.

74. Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin, **How We Analyzed the COMPAS Recidivism Algorithm**, ProPublica (May 23, 2016)

<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>; Ed Yong, **A Popular Algorithm Is No Better at Predicting Crimes Than Random People**, The Atlantic (January 17, 2018) <https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/>

75. Samuel Gibbs, **Women less likely to be shown ads for high-paid jobs on Google, study shows**, The Guardian (8 Jul, 2015)

<https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>.

- **אי-אמון.** שילובם של הגורמים הבאים – **חוסר יכולת להבין** את אופן קבלת ההחלטות; **מחסור במידע לגבי שימושי** הרשות המנהלית בבינה מלאכותית; **אי-שקיפות** כפולה: הן לגבי עצם קבלת החלטות על-ידי מערכת ADM⁷⁶, והן לגבי אופן פעולת האלגוריתם/ המערכת; **ואי-ודאות** לגבי תוצאות יישומה – עשוי להוביל ל**אי-אמון** בהחלטות המתקבלות על-ידי מערכת ADM וכתוצאה מכך, ברשות המנהלית המיישמת אותה.

- **שגיאות.** מערכות אוטומטיות, המתוכננות על-ידי גורם אנושי ומוזנות על-ידי מידע, אינן חפות משגיאות. מערכת ה- ADM המיושמת ככלי ניבוי, עשויה לייצר גם החלטות וקביעות שהינן false-positive או false-negative.

— **היקף השפעת השגיאה:** שגיאה בהחלטת אלגוריתם, ובכלל זה החלטה מוטית או נגועה באפליה, היא בעלת השפעה נרחבת ומקיפה בהרבה, מזו של שגיאת הגורם האנושי. בדו"ח מ- 2018 של המכון AI Now שבאוניברסיטת New York, צוין כי: “While individual human assessors may also suffer from bias or flawed logic, the impact of their case-by-case decisions has **nowhere near the magnitude or scale that a single flawed automated decision-making system can have across an entire population** [ת.].”⁷⁷ [ההדגשה אינה במקור – ס.ת.]”

קביעה, או הערכה זו, נדרשת להילקח בחשבון משנשקלת הטמעת מערכת ADM בתחומים רגישים כמו שדה הרווחה ובשקילת סוג המערכת המיושמת – מערכת מחליטה, או מערכת תומכת החלטה. כאשר מזהה שגיאת מערכת עמוקה במיוחד או בעלת היקף נרחב במיוחד, או במקרה של שגיאות חוזרות מסוג זה, מן הראוי שפקידי הציבור יתערבו בפעולת המערכת, יבקרו את פעולתה ויגבילו את סמכויותיה.

- **פגיעה בפרטיות.** מערכות ADM מאומנות על נתונים אישיים, וגופי נתונים שלמים כאלה מוזנים לתוכן לצורך הפעלת האלגוריתם. כאמור לעיל, במקרה של מיקור-חוץ של פיתוח מערכת ADM או יישומה, במסגרת מקבלות חברות מן המגזר הפרטי גישה לנתונים אישיים המוחזקים בידי הרשות השלטונית, החשש לפרטיות נושאי המידע – גובר. כאשר מדובר בנתונים אישיים של מטופלי רווחה, המהווים 'מידע רגיש' הנוגע לצנעת חייו של הפרט, בהתאם לחוק הגנת הפרטיות, התשמ"א – 1981,⁷⁸ הסיכון לפרטיות גובר אף יותר ועל הרשות יהיה לנקוט באמצעי אבטחת מידע מוגברים, כנדרש בדין, לשם הבטחת פרטיות נושאי המידע – מטופלי הרווחה. הפגיעה בפרטיות עשויה להתממש גם במקרה של הפרת עקרון צמידות המטרה (purpose limitation) – עקרון ליבה מהותי בהקשר של הגנת הפרטיות, במיוחד בהיבט של אוטונומיית הפרט לגבי מידע שמקורו בו. איסוף המידע הראשוני אודות מטופלי הרווחה, נעשה למטרה או מטרת ספציפיות, המעוגנות תחת הסמכה חוקית. יש להבטיח שגם המטרות החדשות (המשניות) להן ישמש המידע באמצעות עיבודו על-ידי הבינה המלאכותית, תעלינה בקנה אחד עם אותן מטרות מקוריות להן הוסמכה הרשות השלטונית או המנהלית בחוק, ולא תחרוגנה מהן.

76. “Agencies continue to struggle with how to give meaningful notice about a computer’s decisions. Despite a decade of experience, for example, we have not yet figured out how to provide notice about automated decisions”; Calo & Citron, לעיל ה"ש 41, בעמ' 27.

77. Lord Clement-Jones – Co-Chair of the All Party Parliamentary Group On AI, **The government’s approach to algorithmic decision-making is broken: here’s how to fix it**, NSTech (18 Feb. 2020) <https://tech.newstatesman.com/guest-opinion/algorithmic-decision-making>

78. ראה חוק הגנת הפרטיות, תשמ"א – 1981, ס"ח תשמ"א מס' 1011 מיום 11.3.1981, עמ' 128, ס' 7.

מסקנות והמלצות

המלצות כלליות

1. **ליווי הטמעת מערכת ADM בשירותי הרווחה, בתכניות הנגשה של כישורים ותשתיות.**⁷⁹ המציאות הנוכחית מצריכה קידום וחזוק האוריינות הטכנולוגית והדיגיטלית בקרב עובדי השירות הציבורי בכלל ועובדי מערכת הרווחה, בפרט. עובדי מערכת הרווחה, שהכשרתם הסתיימה לפני המהפיכה הדיגיטלית, או במהלכה, ואף עובדי מערכת הרווחה הטריים יותר – לא הוכשרו להתמודד עם דיגיטציה של המערכת ועם עבודה לצד מערכות אוטומטיות לקבלת החלטות. כפי שציינו לעיל, לנוכח הטמעת מערכות ADM עשויים למצוא עצמם עובדים אלו במצב של תת-הכשרות, דהיינו ברמת הכשרה נמוכה מדי ביחס לזו הנדרשת לביצוע המשימה, או שהכשרתם המקורית תהא מיושנת ובלתי-רלבנטית לדרישות המקצועיות הנוכחיות. לכן, גם מטעמים אלו, ישנה חשיבות רבה להשקעה בהיבט זה, למניעת ספק עצמי-מקצועי (דהיינו, ערעור הבטחון ביכולת ובניסיון המקצועיים), לנוכח הטיית האוטומציה הקיימת.
2. קביעת **טקסונומיה** (סיווג) לסוגי מערכות ADM בשירות מערכת הרווחה, בהן יעשה שימוש באלגוריתם תומך-החלטה בלבד, וכאלה (לאחר שקילה של האם הדבר ראוי בכלל במפגש עם מטופלי רווחה וצרכיהם) בהן ניתן להשתמש באלגוריתם מחליט. כך למשל, בפיתוח ויישום מערכות ADM לצורך הערכות סיכון מנבאות – רצוי יהיה שאלו יועדו לשמש כמערכות תומכות החלטה ולא כמערכות מחליטות, המחליפות את שיקול הדעת של אנשי הרווחה.⁸⁰
3. **צמצום הטיית** במערכת הבינה המלאכותית ובקרב מיישמה. לשם כך, הממשלה או הרשות הציבורית הרלבנטית, נדרשות לבחון דרכים ל'ניקוי' הנתונים ולהבטחת הגיון האנושי הרב ביותר, הן בשלב בניית המערכת ופיתוחה, והן בלב הטמעתה ויישומה בשירות הציבורי.⁸¹
4. **בניית פרקטיקות רכש ממשלתיות אחראיות, ייעודיות** לטכנולוגיות מבוססות בינה מלאכותית. לאחר שגוף ממשלתי קיבל החלטה לפיה, פתרון מבוסס-בינה מלאכותית מהווה המענה הנכון לקושי או ללקונה בשירות הציבורי, מן הראוי לפעול בהתאם לפרקטיקת רכש סדורה ומותאמת על מנת להבטיח שההליך עצמו, כמו גם המוצר הטכנולוגי הסופי, יהיו ראויים ומועילים. ממשלת ישראל יכולה לגבש פרקטיקות כאלה באופן עצמאי ובהתאמה למדיניות הציבורית המקומית, או להסתייע לשם הבניית פרקטיקות כאלו בהנחיות האוניברסליות לכאורה, שפותחו על-ידי ה – World Economic Forum Centre for the Fourth Industrial Revolution, ביחד עם קהילת בעלי-עניין.⁸² בהנחיות אלו מגולמות מספר המלצות להנחיית אנשי הרכש והמדיניות בשירות הציבורי בדיוק הזמנות להגשת הצעות (RFPs) ובהתאמתן לצורכי השירות הממשלתי הנתון; בין היתר, באופן שיסביר לספקי טכנולוגיות הבינה המלאכותית את הציפיות והצרכים הממשלתיים ממערכת כזו. כך למשל, הוצע להגדיר מראש את תועלות השימוש במערכת מבוססת בינה מלאכותית לציבור, תוך הערכת הסיכונים

79. Philip Alston, **Report of the Special rapporteur on extreme poverty and human rights**, לעיל ה"ש 23, בעמ' 14.

80. Redden et al., לעיל ה"ש 14, בעמ' 12.

81. Committee on Standards in Public Life, **Artificial Intelligence and Public Standards**, לעיל ה"ש 1, בעמ' 59.

82. World Economic Forum, **Guidelines for AI Procurement**, לעיל ה"ש 5.

הצפויים מכך; לעבוד עם צוות מגוון ורב-תחומי; להתמקד בנורמות שקיפות ובמנגנוני אחריותיות אלגוריתמית; ליישם תהליך להבטחת קשר שוטף עם ספק טכנולוגיית הבינה המלאכותית, גם לאחר השלמת הרכישה והטמעת הטכנולוגיה, לצורכי העברת ידע והכשרת עובדי השירות הציבורי וביצוע הערכות סיכון ארוכות-טווח, ועוד.

ככלל, בגיבושו של הליך של אישור / הסמכה (certification)⁸³ לרכישת הטכנולוגיה, לבחינתה ולאמוצה בשירות הציבורי, יש ערך עבור קיומו של מנהל תקין, לעידוד שקיפות, להגברת בטחון עובדי הרשות הציבורית בעיסוקם בטכנולוגיות מבוססות-בינה מלאכותית, ולהגברת אמון הציבור ברשות המנהלית.

השלב המקדמי – טרום אימוץ מערכת ADM בשירות הציבורי

תנאי סף והצדקות לאימוץ מערכת בינה מלאכותית בשירותי הרווחה

- 1. בחינת כדאיות.** כשנבחנת אפשרות לאימוץ מערכת ADM והטמעתה בשירות הציבורי, על מקבלי ההחלטות לבחון את המרכיבים הבאים, אל מול החלופה האנושית הקיימת: עלויות הטכנולוגיה, עלויות הטמעתה (עלויות ישירות ועקיפות, כגון הכשרת עובדי מערכת הרווחה להבנת פעולת המערכת ותוצריה), עלויות תחזוקתה, התמודדות עם מערכת סיכונים ועלות הטעות (false- negatives/positives)⁸⁴.
- 2. הכנת הצדקה לבחירת השימוש במערכת ADM.** נדרש תהליך חשיבה ושקלול פנים-מערךתי, להערכת ההשפעה הפוטנציאלית של יישום מערכת ADM על נתיני השירות הציבורי – מטופלי הרווחה, במקרה שלפנינו. במסגרת זו, על קובעי המדיניות במערכת הרווחה לשקול: האם השפעת המערכת תהא מידתית, באילו תחומים היא עדיפה על פני המודל הקיים מבוסס-האדם, וכיצד היא עתידה לשפר את המערכת הקיימת.
- **מתי נוותר על אימוץ מערכת ADM בשירות הציבורי?** בהיעדר תועלת ברורה לציבור (הכללי, או הרלבנטי) ביישומה; ו/או כשהפגיעה הפוטנציאלית בזכויות פרטים כתוצאה מאימוצה, אינה הכרחית, או מידתית,⁸⁵ לא ניתן יהיה להצדיק אימוץ מערכת ADM ומן הראוי יהיה לוותר על כך.
- 3. בחינת 'צדקת' האלגוריתם (Algorithmic Equity).** מדובר בבחינה פרואקטיבית מורכבת, המחפשת ודאות או בטחון לגבי הוגנות האלגוריתם וישרתו באופן פעולתו. לשם קיום בחינה מעין זו, ניתן לגייס אנשים מומחים כגון 'ציידי הטיות'⁸⁶ (bias bounties), אשר יבחנו את עמידת האלגוריתם והמערכת כולה בעקרונות האתיקה של בינה מלאכותית, המתמקדים בצדק חברתי, הוגנות ואי-הפליה. כמו כן, לאחרונה פותחו כלים שונים, ייעודיים, לבחינת צדקת האלגוריתם, כגון ה- Algorithmic Equity

83. Meeri Haataja, **How certification promotes responsible innovation in the algorithmic age**, TechTalks (May 28, 2020); <https://bdtechtalks.com/2020/05/28/autonomous-intelligent-systems-certification-ieee/>

84. **A Guide to Using Artificial Intelligence in the Public Sector**, the Office for Artificial Intelligence (OAI) & Government Digital Service (GDS) [UK] (Jan. 2020), p. 12.

85. **Artificial Intelligence and Public Standards**, Committee on Standards in Public Life, לעיל ה"ש 1, בעמ' 58-57.

86. ראה למשל:

[Khari Johnson, AI researchers propose 'bias bounties' to put ethics principles into practice](https://venturebeat.com/2020/04/17/ai-researchers-propose-bias-bounties-to-put-ethics-principles-into-practice/), venturebeat.com (April 17, 2020); <https://venturebeat.com/2020/04/17/ai-researchers-propose-bias-bounties-to-put-ethics-principles-into-practice/>

Toolkit, שפותח על-ידי האגודה האמריקנית לזכויות האזרח (ACLU), בשיתוף עם ה-Tech Fairness Coalition.⁸⁷ הכלי מיועד לקבוצות בקהילה, לצורך זיהוי טכנולוגיות ADM המופעלות בשירות הממשלה לצורכי מעקב או קבלת החלטות באמצעות אלגוריתם, על מנת להבין את אופן פעולתן, ולבחון את יעילותן, השפעתן והפיקוח המבוצע באמצעותן. אולם, יש להניח שכלי כגון זה יכול, במידה כזו או אחרת, לעבוד גם בשירות הממשלה לצורך בחינה-עצמית של טכנולוגיות ה-ADM אותן היא מבקשת להטמיע וליישם, ולהציף בשלב מוקדם, טרם אימוצן, בעיות וקשיים שונים. לבחינה כזו ניתן לייחס יתרון כפול: הן בהגברת בטחונם המקצועי של אנשי הרווחה העושים שימוש במערכת והן, בהגברת אמון הציבור במערכת.

תכנון מערכת ה-ADM או התאמתה ליישום בשירות הציבורי/שירותי הרווחה

1. **אספקת אופציה שאינה דיגיטלית לשירות נתון.** במקרים בהם מיושמת מערכת מבוססת-בינה מלאכותית ככלי לשירות האזרח (להבדיל מכלי פנים-משרדי, הבא לסייע לפקידי השירות הציבורי בביצוע תפקידיהם), על הרשות לספק חלופה 'אנלוגית' לשירות, לטובת אותם אזרחים נעדרי נגישות טכנולוגית או בעלי אוריינות דיגיטלית נמוכה. הדבר מתבקש לטובת שמירת כבוד האדם של האזרח, מטופל מערכת הרווחה (במקרה שלפנינו), וחיוני על מנת לאפשר מימוש ומיצוי זכויות.

2. **תכנון ופיתוח עצמי, פנים-ארגוני.** פיתוח המערכת על-ידי המדינה, או הרשות המנהלית המבקשת להטמיעה, להבדיל ממיקור-חוץ של פיתוח המערכת לחברה חוץ-ממשלתית או מרכישת מערכת קיימת, תוך ביצוע התאמות מסוימות לצרכי השירות הציבורי הנתון.

אף-על-פי שהדבר דורש התגברות על חסם האוריינות הדיגיטלית במגזר הציבורי, לפיתוח מערכת ADM בתוך הרשות, או לחילופין – למעורבות גבוהה מצידה בפיתוחה, ישנו ערך רב להבטחת סיכוי מיטבי להתאמת המערכת לשירות הייעודי ולהצלחת יישומה. מעבר לכך, הדבר אף עולה בקנה אחד עם תכנית הדיגיטציה הלאומית ('ישראל דיגיטלית'),⁸⁸ ועם יוזמות לאומיות משלימות לקידום השירות הדיגיטלי לאזרחים,⁸⁹ שאינן מצריכות בהכרח פנייה למיקור-חוץ, לשם הגשמתן.

87. **Algorithmic Equity Tool Kit**, ACLU Washington
<https://www.aclu-wa.org/AEKit>.

88. ראה החלטת ממשלה 2733: קידום המיזם הלאומי "ישראל דיגיטלית", לעיל ה"ש 6.

89. ראה למשל, החלטת ממשלה מס' 2097 "הרחבת תחומי פעילות התקשוב הממשלתי, עידוד חדשנות במגזר הציבורי וקידום המיזם הלאומי "ישראל דיגיטלית" (10 אוקטובר, 2014); **התכנית להאצת השירותים הדיגיטליים לציבור ולקידום הלמידה הדיגיטלית**, טיוטת הצעות ההחלטה שעניינן שינויים מבניים בתכנית הכלכלית לתקציב המדינה הקרוב (5 יולי, 2020).

דרכים להתמודדות אד-הוק עם מכשול האוריינות הדיגיטלית כוללות, בין היתר, יצירת תמריצים עבור מפתחים איכותיים אשר תאפשר גיוסם למגזר הציבורי, או העסקת מומחים חיצוניים כיועצים לפרוייקט ייעודי שכזה.

אלו הם היתרונות העיקריים שבפיתוח מערכת ADM על-ידי המדינה/הרשות:⁹⁰

- הבטחת שליטה מוחלטת של המדינה/רשות על תהליך הפיתוח.
- פרסונליזציה של המערכת (custom-made) משלב התכנון, להבטחת התאמה אופטימלית לשירות הציבורי הנתון.
- שמירת מאגרי הנתונים, עליהם מאומנת מערכת ה-ADM והמוזנים לתוכה, בתוך מערכת השירות הציבורי והבטחת פרטיות נושאי מידע.
- המנעות מבעיית הבנתיות (black box) בקרב אנשי השירות הציבורי המיישמים המערכת ונעזרים בה (למרות שלשם כך תידרש גם הכשרה פנים-גופית של כל אנשי השירות האמורים, בכל הרמות והתפקידים הרלבנטיים, שהרי רק מעטים מתוך הגוף הציבורי יהיו מעורבים בפועל בפיתוח המערכת ועיצובה).

3. **שיתוף אוכלוסיית היעד המקצועית** (פקידי הרווחה) בשלבי הפיתוח – בין אם זה נעשה על-ידי המדינה/הרשות המנהלית, ובין אם על-ידי גוף חיצוני, באמצעות מיקור-חוץ – להבטחת דרכי טיפול ההולמות את הקשיים, את היכולות והצרכים של אנשי הרווחה המטפלים.⁹¹ שיתוף זה חיוני להצלחת יישום המערכת.

4. **שיתוף אוכלוסיית היעד הטיפולי** (מטופלי הרווחה) בתכנון המערכת ובעיצובה (שלבי הפיתוח), להבטחת התאמה אופטימלית לבעיות האמיתיות של אותה אוכלוסייה.⁹²

5. **בחירה איכותנית של סוג הפלט (output) שמערכת ADM תספק:** האם **מדד-סיכון** (risk score) – שהינו סמכותני יותר ועשוי לאותת לגורם האנושי שהמכונה 'פיצחה' באופן מדויק יותר את רמת הסיכון הנשקפת למטופל הרווחה, מאשר הגורם האנושי, או – **דו"ח ענייני**, המייצר **התראה** במקרה של איתור סיכוי גבוה יחסית לסיכון ומותר מרחב גדול יותר לשיקול הדעת של הגורם האנושי.⁹³

90. ראה, לדוגמה, לקחים מפיתוחה של ה- Bristol's Integrated Analytical Hub ויישומה על-ידי מועצת העיר Bristol, אנגליה;

Redden et al., לעיל ה"ש 14, בעמ' 10.

91. Philip Alston, **Report of the Special rapporteur on extreme poverty and human rights**, לעיל ה"ש 23, בעמ' 14.

92. שם.

93. ראה לדוגמה מערכת מנבאת סיכון שיושמה במועצת Hackney באנגליה, אשר פותחה בעזרת תובנות שהתקבלו מעובדים סוציאליים, והפלט של הערכת הסיכון, תוצר-המערכת, מסופק כדו"ח, במקום כמדד/תוצאת סיכון (risk score); Redden et al., לעיל ה"ש 14, בעמ' 12.

השלב המתקדם – עם אימוצה של מערכת ADM בשירות הציבורי

1. **שמירת מרכיב של נוכחות ואחריות אנושית בתהליך** ('Human-in-the-loop') הינה חיונית, במיוחד בזירת השירות הציבורי. מעורבות פקידי הציבור – אנשי מערכת הרווחה, בהחלטות המתקבלות על-ידי מערכות ADM תאפשר, או תבטיח, בין היתר:⁹⁴
 - רציפות של פעולת מערכת הרווחה במסגרת סמכותה.
 - שמירה על רצף טיפולי [גורם אנושי-מקצועי מחליט ← מערכת ADM תומכת החלטה/מחליטה ← גורם אנושי-מקצועי מסתייע (מערכת תומכת החלטה) / גורם אנושי מבקר את תוצרי המערכת (מערכת מחליטה)].
 - שליטה מסויימת, מעקב ופיקוח פנימי על פעילות המערכת האוטומטית.
 - הערכת ההחלטות המתקבלות על-ידי המערכת, או באמצעותה, באופן שוטף.
 - המנעות (מלכתחילה) מבעיות וריפוי פגמים (בדיעבד), במענה מהיר להופעתם.
 - אחריותיות של הגוף הציבורי על החלטות תוצרי מערכת ADM.
2. **גיבוש מערכות בקרה ויצירת תהליכים להערכה ביצועית** של הטכנולוגיה מבוססת-בינה מלאכותית.⁹⁵ כינון מנגנוני פיקוח על פעילות מערכות אלה מבטא אחריות ותקינות מנהלית, נכונות לבחינה עצמית, ועשוי להבטיח, בין היתר:
 - פעולת המערכת בהתאם לייעודה.
 - שמירת זכויות הפרטים – נתיני המערכת.
 - עמידה בנורמות מנהל תקין (כגון: סמכות, סבירות, מידתיות, תום לב, שוויוניות, היעדר שיקולים זרים, עקביות).
3. **כינון מנגנון ערעור** על החלטות שירותי הרווחה, המתבצעות או נתמכות על-ידי מערכת ADM ויידוע האזרח על קיומו של המנגנון ודרכי הגישה אליו. יצירת מנגנון כזה עולה בקנה אחד עם זכותו של הפרט מושא החלטת מערכת ADM להליך הוגן, ויחד עם השקיפות לגבי פעילותו – תגביר את אמון הציבור בשירות הציבורי בכלל, ובמערכות ADM, בפרט.

⁹⁴ Artificial Intelligence and Public Standards, Committee on Standards in Public Life, לעיל ה"ש 1, בעמ' 20.

⁹⁵ שם, בעמ' 61.

סיכום

עניינה המהותי של הסוגייה שלפנינו, הוא במפגש בין המגזר הציבורי לבין המהפיכה התעשייתית הרביעית, זו המביאה לפתחנו בין היתר, את טכנולוגיית הבינה המלאכותית. על פניו, מדובר במיזוג לא-טבעי בין שני דברים בעלי דינאמיקה מנוגדת: הבינה המלאכותית מונעת-החדשנות, והמגזר הציבורי המסורבל (מתוקף גודלו ונתק קיומו), השמרני והגמיש פחות, מעצם טבעו. אך זהו בדיוק השילוב המוזר הזה, שהינו מושך, מעורר מחשבה ומזמין בחינה.

ואכן, סקירה זו התמקדה במפגש אחד כזה, שבין מערכת הרווחה החברתית, לבין מערכות ADM. במסגרתה, נסקרו בהרחבה היבטים שונים של יישום מערכות ADM בשירות הציבורי, ובכללם: אתגרי אימוץ והטמעה של מערכות מבוססות בינה מלאכותית בשירות הציבורי; מגוון התועלות הפוטנציאליות שיישום כזה עשוי להצמיח ומשמעותן החברתית; והקשיים האתיים והסיכונים הכרוכים בדבר. בסיכומה של הסקירה הובאה מערכת המלצות מדיניות בנושא.

הקורא חד-העין יכול היה לזהות לאורך כל הנייר, את המתח האינהרנטי לסוגייה – זה שבין תפקידו ושליחותו של האדם (העובד הסוציאלי), לבין אימוץ המכונה כחלופה אנושית. זהו מתח אופייני לשיח, הבא לידי ביטוי בדיון האקדמי-אתי-מדינותי שבין מאמצים נלהבים של בינה מלאכותית ואוטומציה, לבין ספקני הטכנולוגיה (technology skeptics), החוששים מכך שהדבר מבשר על תחילתו של עידן האלגורקסיה (algocracy) – שליטת האלגוריתם והקוד הדיגיטלי בחיינו.

יש להניח, שכפי שקורה לרוב בפועל – מהלך האימוץ של טכנולוגיות אלה בשירות הציבורי יבטא מדיניות מתונה מזו של נלהבי האימוץ, אך אמיצה מזו של ספקני הטכנולוגיה. לו היינו צריכים להצביע על המרכיבים החיוניים ביותר להצלחתה של הטמעת מערכות ADM בשירות הציבורי, בפרט בשירותי הרווחה, נדמה כי היו אלה שקיפות והבנתיות של אופן פעולת המערכת, שלא ניתן להפריז בערכן לקידום אמון הציבור, כמו גם הבטחת אחריותיות של הרשות הציבורית כלפי ציבור המטופלים מקבלי השירות. לצד זאת, יש לנו אינטרס חברתי לשמר ולחזק את תחושת השליחות החברתית, הרגישות והמחויבות החברתית המאפיינות את אלו שבחרו בעבודה הסוציאלית ובהגשת שירותי רווחה, כייעוד וכמקצוע. על מנת לעשות כן, עלינו להבטיח את שיתופם הפעיל והשוטף של העוסקים במלאכה בפיתוח מערכות אלו, להבטיח את הבנתם (באופן בסיסי) את אופן פעולת המערכת, ובאמצעות זאת – לקדם את אמונם שלהם בה.

דווקא בתחום כמו זה של הרווחה החברתית, בו הקשר האנושי הבין-אישי מהווה גורם שאין-בלתו, נראה כי מערכות ADM לא יוכלו להוות תחליף שלם והולם לגורם האנושי. אולם בד בבד עם כך, ייתכן מאוד שהעוסקים במלאכה, אם יזכו להכשרה נאותה והנחיה מלווה, יוכלו להיתרם רבות מהטמעה זהירה, מדודה ונכונה – של מערכות כאלה. הטמעת מערכות ADM תספק תמיכה במשימות המערכת, תאפשר לשלוחיה לבצע את עבודתם באופן אחראי ומשמעותי ותקדם את השירות לו זכאים מטופלי הרווחה.