

## זכויות אדם בבינה מלאכותית?

## על המתווה האמריקאי למגילת זכויות בבינה מלאכותית ומשמעותו

עמיר כהנא\* ודפנה דרור-שפוליאנסקי\*\*

באוקטובר 2022 פרסמה המחלקה המייעצת לנשיא ארצות-הברית בנושאי מדע וטכנולוגיה (The Office of Science and Technology Policy (OSTP) את "המתווה למגילת זכויות בבינה מלאכותית – הפעלת מערכות אוטומטיות עבור העם האמריקאי" (The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People),<sup>1</sup> אשר מוגדר כ"ספר לבן".<sup>2</sup> המתווה מציע קווי מתאר לפיתוח מדיניות ופרקטיקות להגנה על זכויות אזרחיות ולקידום ערכים דמוקרטיים בבנייה, פיתוח וממשל של מערכות אוטומטיות העשויות להשפיע על זכויות הציבור, על הזדמנויות, או על גישה לצרכים חיוניים.<sup>3</sup> ברשימה זו נבקש לסקור בקצרה את המתווה ואת משמעותו בתחום האסדרה המשפטית של בינה מלאכותית.

## הרקע לפרסום המתווה

בשנים האחרונות מתקיים בעולם שיח ער על העקרונות האתיים שיש להחיל בשימוש ובפיתוח מערכות בינה מלאכותית, ועל הדרך הראויה לאסדרה של המערכות האלו. מערכות בינה מלאכותית הוטמעו בשנים האחרונות במספר תחומים, ונמצאות בשימוש נרחב במערכות ממשלתיות וציבוריות,<sup>4</sup> לרבות בשירותן של סוכנויות אכיפת החוק והביטחון,<sup>5</sup> בהקצאת משאבי רווחה,<sup>6</sup> וגם בזירה המשפטית.<sup>7</sup>

**אזכור הרשימה:** עמיר כהנא ודפנה דרור-שפוליאנסקי "זכויות אדם בבינה מלאכותית? על המתווה האמריקאי למגילת זכויות בבינה מלאכותית ומשמעותו" **אתר משפט ועסקים** (18.5.2023) <https://runilawreview.org/2023/05/18/cahane-dror-shpoliansky>.

\* דוקטורנט בפקולטה למשפטים, האוניברסיטה העברית בירושלים; חוקר במרכז לערכים ולמוסדות דמוקרטיים, המכון הישראלי לדמוקרטיה; ועמית מחקר, [Three Generations of Digital Human Rights, European Research Council \(ERC\) Research Project](#).

\*\* דוקטורנטית בפקולטה למשפטים, האוניברסיטה העברית בירושלים; דוקטורנטית אורחת בפקולטה למשפטים, אוניברסיטת טורונטו; ועמיתת מחקר, [Three Generations of Digital Human Rights, European Research Council \(ERC\) Research Project](#).

1 The White House, Office of Science and Technology Policy, [Blueprint for an AI Bill of Rights: Making Automated Systems work for the American people](#) (להלן: AI BOR).

2 שם, בעמ' 2.

3 שם, בעמ' 3.

4 U.N. Secretary-General, [Extreme Poverty and Human Rights](#), U.N. Doc. A/74/493 (General Assembly, Oct. 11, 2019).

5 למשל, באמצעות טכנולוגיות לחיזוי פשיעה המסייעות בהקצאה של משאבי משטרה. ראו Andrew Guthrie Ferguson, [Crime](#).

6 Andrew G. [Mapping and the Fourth Amendment – Redrawing "High-Crime Areas"](#), 63 HASTINGS L.J. 179 (2011).

7 Ferguson, [Policing Predictive Policing](#), 94 WASH. U. L. REV. 1109 (2017).

אדם, מכונה, מדינה: לקראת אסדרה של בינה מלאכותית, חלק 2.3.2 (המכון הישראלי לדמוקרטיה, צפוי להתפרסם ב-2023).

6 ראו, לדוגמה, את מערכת SyRI ההולנדית – Max Vetzo, [The Netherlands-Algorithmic Fraud Detection System Violates Human Rights—The Case of SyRI](#), 3 PUB. L. 650 (2021).

7 (המכון הישראלי לדמוקרטיה 2022). 56–54.

משפטנים מסתייעים במערכות אלגוריתמיות כדי להעריך את הסתברותן של החלטות שיפוטיות. כך, למשל, חברת ההזנק הקנדית

Benjamin Alarie, Anthony Niblett & Albert [Using Machine Learning to Predict Outcomes in Tax Law](#), 58 CAN. BUS. L.J. 231 (2016); Benjamin Alarie, Anthony

H. Yoon, [How Artificial Intelligence will Affect the Practice of Law](#), 68 U. TORONTO L. J. 106 (2018). נוסף על

כך, מדינות ברחבי העולם מתחילות לשקול להיעזר בבינה מלאכותית במלאכת השפיטה. כך, למשל, אסטוניה הכריזה על כוונתה

להשתמש בבינה מלאכותית כשופט הדין בתביעות קטנות (שסכומן לא עולה על 7,000 אירו), בכפוף לזכות ערעור על ההחלטה

הממוכנת בפני שופט אנושי. Eric Niiler, [Can AI Be a Fair Judge in Court? Estonia Thinks So](#), WIRED (March 25, 2019).

במסגרת זו, פורסמו בשנים האחרונות מסמכי אתיקה רבים,<sup>8</sup> בין היתר על ידי הנציבות האירופית,<sup>9</sup> ה-OECD<sup>10</sup> ולאחרונה גם בישראל.<sup>11</sup> זאת ועוד, שורה של הצעות חוק בעולם, שחלקן כבר נכנסו לתוקף, מאותות כי שיח המדיניות בתחום הבינה המלאכותית עובר מחשיבה על עקרונות אתיים לאסדרה בפועל. דוגמה בולטת לניסיון לאסדר בינה מלאכותית בחקיקה היא הצעת חוק הבינה המלאכותית של הנציבות האירופית (AIA),<sup>12</sup> ונוסף לה ישנם ניצני חקיקה בסין,<sup>13</sup> בברזיל,<sup>14</sup> בקנדה<sup>15</sup> ועוד.<sup>16</sup>

בשנת 2019 פרסם ממשל טראמפ את הצו הנשיאותי ל"שמירת ההובלה האמריקאית בבינה מלאכותית" המתווה את האסטרטגייה האמריקאית בתחום.<sup>17</sup> הצו הנשיאותי דאז לא כלל את המונח "אתיקה" בתוכו אלא התרכז בצורך בפיתוח בינה מלאכותית ובהובלה טכנולוגית בתחום.<sup>18</sup> בהתאם להוראות הצו, משרד הניהול והתקצוב בבית הלבן פרסם בשנת 2020 חוזר הכולל עקרונות למדיניות אסדרה של בינה מלאכותית (OMB), שמטרתו הייתה בעיקרה להנחות סוכנויות פדרליות באשר לשיקולים הנוגעים לאסדרה של התחום באופן שיקדם חדשנות, הסרת חסמים והגנה על ערכים אמריקאיים, כמו פרטיות, חירות, וזכויות אדם.<sup>19</sup>

בנאום בשנת 2021 ציין אנתוני בלינקן, מזכיר המדינה בממשל ביידן, כי "משימתנו היא לנסח וליישם חזון משכנע על אודות שימוש בטכנולוגיה באופן שישרת את אנשינו, יגן על האינטרסים שלנו ויתמוך בעקרונות הדמוקרטיים שלנו".<sup>20</sup> בהתאם, המתווה למגילת זכויות אדם בבינה מלאכותית, אשר פורסם לאחרונה, מהווה מהלך הצהרתי חשוב העשוי להיות אבן-דרך במדיניות אסדרת הבינה המלאכותית של ארצות-הברית בעתיד. גם תהליך ניסוח המתווה, אשר ארך תקופה

- 8 רביה שוקלת להשתמש בטכנולוגיות דומות. ראו Milica Stojanovic, *Serbia Eyes Artificial Intelligence in Courts, but Experts See Dangers*, BALKANINSIGHT (Jan. 25, 2021). איגודים מקצועיים בבריטניה וגורמים באקדמיה העלו מספר הצעות להפחתת העומס השיפוטי באמצעות מערכות בינה מלאכותית. ראו YADONG CUI, ARTIFICIAL INTELLIGENCE AND JUDICIAL MODERNIZATION 23 (2020). להרחבה ראו כהנא ושוורץ אלטשולר, לעיל ה"ש 5.
- 9 לסקירות כלליות של מסמכי אתיקה ראו Anna Jobin, Marcello Ienca & Effy Vayena, *The Global Landscape of AI Ethics Guidelines*, 1 NATURE MACHINE INTELLIGENCE 389 (2019); Yi Zeng, Enmeng Lu & Cunqing Huangfu, *Linking Artificial Intelligence Principles*, ARXIV PREPRINT ARXIV:1812.04814 (2019); Thilo Hagendorff, *The Ethics of AI Ethics: An Evaluation of Guidelines*, 30 MINDS & MACHINES 99 (2020); Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Christopher Nagy & Madhulika Srikumar, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI* (Berkman Klein Center for Internet & Society, Research Publication No. 2020-1, 2020).
- 10 *AI HLEG, ETHICS GUIDELINES FOR TRUSTWORTHY AI* (European Commission, April 8, 2019).
- 11 *OECD Recommendation of the Council on Artificial Intelligence*, OECD/Legal/0449 (May 22, 2019). (להלן: OECD (2019)).
- 12 *ועדת משנה של המיום הלאומי למערכות נכונות בנושא אתיקה ורגולציה של בינה מלאכותית דין וחשבון* (נובמבר 2019). *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 206 final (April 21, 2021) (להלן: AIA).
- 13 *Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022*, DIGICHINA (Jan. 10, 2022). להרחבה על תקנות מערכות ההמלצה האלגוריתמיות הסיניות ראו Gilad Abiri & Xinyu Huang, *The People's (Republic) Algorithms*, 12 NOTRE DAME J. INT'L & COMP. L. 17 (2022).
- 14 *Projeto De Lei N° 21-A de 2020* (Braz.). לתרגום לאנגלית של ההצעה ראו Walter Gaspar, *Non-Official Translation Of The Brazilian Artificial Intelligence Bill, N. 21/2020*, CYBERBRICS (Oct. 25, 2021).
- 15 *Directive on Automated Decision-Making* (Government of Canada 2019); *Bill C-27: An Act to Enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to Make Consequential and Related Amendments to Other Acts* 85–100 (1st Sess., 44th Parl., 1st Reading, June 16, 2022).
- 16 *Algorithmic Accountability Act of 2022*, H.R. 6580, 117th Cong. (2nd Sess., 2022).
- 17 *Maintaining American Leadership in Artificial Intelligence*, Exec. Order No. 13859 (2019).
- 18 תקציב 2019 של ארצות-הברית התייחס לבינה מלאכותית כאחד מ"תחומי העניין העיקריים" בתחום המחקר והפיתוח. ראו OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, BUDGET OF THE UNITED STATES GOVERNMENT, FISCAL YEAR 2019, 36 (2018).
- 19 Memorandum for the Heads of Executive Departments and Agencies, *Guidance for Regulation of Artificial Intelligence Applications*, Exec. Off. of the President, Off. of Mgmt. & Bud., M-21-06 (Nov. 17, 2020).
- 20 *Secretary Antony J. Blinken at the National Security Commission on Artificial Intelligence's (NSCAI) Global Emerging Technology Summit*, U.S. DEPARTMENT OF STATE (July 13, 2021). (התרגום של המחברים).

ממושכת וכלל התייעצות עם שלל גורמים ובתוכם ארגוני חברה אזרחית, חברות טכנולוגיה בולטות, והציבור הרחב,<sup>21</sup> עשוי להצביע על כך שזו רק יריית הפתיחה לקראת צעדי אסדרה עתידיים נוספים בהשראת עקרונות אלו. זאת ועוד, כפי שיפורט להלן, פרסום המסמך האמריקאי על רקע מהלכי החקיקה של האיחוד האירופי עשוי להצביע על שאיפה לקדם אסדרה אמריקאית של בינה מלאכותית ובכך למנוע דומיננטיות אירופית בתחום, כפי שקרה בעבר בנוגע לדיני הגנת הפרטיות האירופיים.<sup>22</sup>

## עיקרי המתווה

המתווה האמריקאי מציע עקרונות ופרקטיקות נלוות שמטרתן לספק קווים מנחים לעיצוב, לשימוש ולפריסה (deployment) של מערכות אוטומטיות להגנה על זכויות "הציבור האמריקאי" (The American People) בבינה מלאכותית. המתווה כולל שלושה רבדים: (1) עקרונות יסוד – חמישה עקרונות למערכות אוטומטיות: בטיחות ויעילות; הגנה מפני אפליה; פרטיות במידע; הודעה והסבר; וחלופה אנושית; (2) הסבר מפורט על חשיבותו של כל אחד מהעקרונות, כולל דוגמאות מהתנהלות של מערכות אוטומטיות שפועלות בשוק כבר כיום; (3) פירוט כיצד ניתן ליישם את העקרונות בפרקטיקה. להלן נפרט על העקרונות שמונה המתווה.<sup>23</sup>

(1) **מערכות בטוחות ויעילות:**<sup>24</sup> פרטים צריכים להיות מוגנים ממערכות אוטומטיות בלתי בטוחות או לא אפקטיביות. בפיתוחן נדרש להתייעץ עם בעלי עניין ומומחים מגוונים על מנת לזהות סיכונים או השלכות אפשריות של המערכת. מערכות צריכות לעבור בדיקות טרם פריסתן, זיהוי ומיתון סיכונים, וניטור מתמשך שמבטיח שהן בטוחות ואפקטיביות. תוצאות אפשריות של בדיקות כאמור הוא הפסקת פעילות המערכת.

(2) **הגנה מאפליה אלגוריתמית:**<sup>25</sup> פרטים לא צריכים להיות מושאים של אפליה אלגוריתמית, ועיצובן של מערכות צריך שיהיה הגון (equitable). על מעצבים, מפתחים ומפעילים של מערכות אוטומטיות לנקוט אמצעים על מנת להגן על פרטים וקהילות מפני אפליה אלגוריתמית של קבוצות מוגנות.

זהו סעיף מעניין במיוחד שכן נראה כי לשון הסעיף מפנה זרקור אל עבר המעצבים, המפתחים או המפעילים (deployer) של מערכות אוטומטיות. באופן שאינו משתמע לשתי פנים, המתווה קובע כי על המעצבים והמפתחים לנקוט גישה פרואקטיבית להגנה מפני אפליה במערכות אלה. גישה זו צריכה לכלול, בין היתר, הערכת הגינות (equity assessment) כחלק מעיצוב המוצר; הגנה מפני אפליה עקיפה (proxies) בגין נתונים דמוגרפיים; עיצוב מערכות אוטומטיות באופן המבטיח

21 AI BOR, לעיל ה"ש 1, בעמ' 4.

22 ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD 78–81, 131–169 (Oxford Univ. Press 2020); He Li, Lu Yu & Wu He, *The Impact of GDPR on Global Technology Development*, 22 J. GLOBAL INFO. TECH. MGMT. 1 (2019); Cedric Ryngaert & Mistale Taylor, *The GDPR as Global Data Protection Regulation?*, 114 A. J. INT'L. L. 5 (2020); Graham Greenleaf, *Global Data Privacy Laws 2021: Despite COVID Delays, 145 Laws Show GDPR Dominance*, 169 PRIVACY L. & BUS. INT'L REPORT 1 (2021). לעמדות שונות ביחס להגמוניה העתידית של האסדרה האירופית בתחום הבינה המלאכותית ראו CHARLOTTE SIEGMANN & MARKUS ANDERLJUNG, [THE BRUSSELS EFFECT AND ARTIFICIAL INTELLIGENCE: HOW EU REGULATION WILL IMPACT THE GLOBAL AI MARKET](#) (Centre for the Governance of AI, Aug. 2022); Alex Engler, [The EU AI Act will have Global Impact, but a Limited Brussels Effect](#), BROOKINGS (Jun. 8, 2022).

23 AI BOR, לעיל ה"ש 1, בעמ' 5–7.

24 שם, בעמ' 15–22. השוו עם הסקירה של Fjeld, Achten, Hilligoss, Nagy & Sri Kumar, לעיל ה"ש 8, בעמ' 37. מסמך של ועדת האתיקה הגרמנית למידע מתייחס, בתוך המונח הרחב "ביטחון", להיבטים של אבטחת סייבר, הגנת מידע, בטיחות פיזית ונפשית ולהיתכנות של נזקים קניטיים-ממשיים בשל פעילות מערכות בינה מלאכותית. ראו [OPINION OF THE DATA ETHICS COMMISSION](#) (Oct. 22, 2020).

25 AI BOR, לעיל ה"ש 1, בעמ' 23–29. כל (100%) המסמכים שדגמו Fjeld, Achten, Hilligoss, Nagy & Sri Kumar, לעיל ה"ש 8, וכ-76% מהמסמכים שדגמו Jobin, Lenca & Vayena, לעיל ה"ש 8, מתייחסים להוגנות. השוו ל-AI HLEG, לעיל ה"ש 9, בעמ' 18.

שהיינה נגישות לאנשים בעלי מוגבלויות, וכן תכלול הערכת נבדלות (disparity-assessment).<sup>26</sup> המשמעות היא כי מערכות אוטומטיות צריכות לעבור בחינה המוודאת שאינן מפלות, וזאת ביחס פרטני לאוכלוסיות שעשויות להיות מופלות מחמת גזע, צבע, אתניות, מגדר, לידה, נטייה מינית, מידע גנטי, ועוד.<sup>27</sup> המתווה אף מציין כי הערכת נבדלות צריכה להיעשות לא רק בשלב בעיצוב והפיתוח אלא גם על בסיס קבוע כשהמוצר בשימוש. הערכה זו תצטרך להיעשות על ידי מי שמשמש במוצר.<sup>28</sup>

הגדרת האחריות של המעצבים והמפתחים, וכן הדרישה לפיקוח ובדיקה שוטפת של המשתמשים במערכות, מרחיבות למעשה את האחריות הפוטנציאלית לבעלי עניין רבים בשרשרת האספקה של מערכת מבוססת בינה מלאכותית, ונדמה כי במידה מסוימת הדבר מבטא הכרה בכך שהמנגנונים הקיימים כיום להגנה על זכויות אדם – דוגמת חוקי איסור אפליה – עשויים להיות מוגבלים כאשר מדובר באפליה המוטמעת במערכת אוטומטית ואשר קשה לאתרה.<sup>29</sup> המתווה עצמו מכיר בקושי זה, ומציין כדוגמה מערכת אלגוריתמית בשימוש מערכת הבריאות שהתברר שהיא מפלה באופן עקבי אזרחים כהי עור.<sup>30</sup>

בהתחשב בעובדה שמערכות ציבוריות רבות בארצות-הברית כבר מושקות ומשתמשות במערכות אוטומטיות, החל ממערכות להקצאת זכויות סוציאליות<sup>31</sup> ומערכות לזיהוי ילדים במצוקה הנדרשים להתערבות שירותי הרווחה,<sup>32</sup> וכלה במערכות להערכת מסוכנות של עצורים המיושמת ב-46 מדינות בארצות-הברית,<sup>33</sup> ולנוכח הביקורת על מערכות אלו כלוקות בהטיות אלגוריתמיות מובנות, והרגישות הפוליטית בארצות-הברית לסוגיות של אפליה על בסיס אתני – נדמה שהמהלך מתבקש.

**(3) פרטיות במידע:**<sup>34</sup> מערכות אוטומטיות צריכות להיות מעוצבות לפי עקרונות של הנדסת פרטיות (privacy by design), ומפתחיהן צריכים להתבסס ככל האפשר על הסכמה של מושאיהן לאיסוף, לשימוש, לגישה, להעברה ולמחיקה של מידע אישי. שימוש במידע אישי רגיש יהיה בכפוף להגנות מוגברות. פרטים וקהילות צריכים להיות מוגנים ממעקב בלתי מבוקר, וטכנולוגיות מעקב צריכות להיות כפופות לפיקוח מוגבר הכולל לכל הפחות הערכה בטרם פריסתן של סיכונים ואיומים לפרטיות ולחירויות אזרח. אין לבצע מעקב רציף בתחומי בהקשרים בהם טכנולוגיות מעקב עשויות להגביל זכויות, הזדמנויות או גישה.

26 AI BOR, לעיל ה"ש 1, בעמ' 26.  
27 לקוד המוניציפלי המנהלי של העיר ניו-יורק נתוספה לאחרונה הוראה שלפיה השימוש במערכות המקבלות החלטות או המלצות מבוססות אלגוריתמים על העסקת או קידום עובדים יותנה בהליכי בקרת הטיות (bias audit) עצמאיים. ראו N.Y.C. Admin. Code § 20-870–20-874. ראו גם Airlie Hilliard et al., *Regulating the Robots: NYC Mandates Bias Audits for Ai-Driven Employment*, HOLISTIC AI (2022) *Decisions*.  
28 AI BOR, לעיל ה"ש 1, בעמ' 27.  
29 Dafna Dror-Shpoliansky, *From Digital Ethics to Human Rights by Design – a Constructive Approach to Incorporate International Human Rights Law in AI Technology* (Presented at the European Society of International Law, Aug. 2022) (עותק שמור אצל המחברים).  
30 AI BOR, לעיל ה"ש 1, בעמ' 29.  
31 *Cahoo v. SAS Analytics Inc.*, No. 18-1296 (6th Cir. 2019); Sonia M. Gipson Rankin, *The Midas Touch: Atuahene's "Stategraft" and Unregulated Artificial Intelligence*, 98 N.Y.U. L. Rev. Online 225 (2023).  
32 VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR*, cha. 4 (2018).  
33 ראו, לדוגמה, *State v. Loomis*, 881 N.W.2d 749, 754 (Wis. 2016).  
34 AI BOR, לעיל ה"ש 1, בעמ' 30–39. 97% מהמסמכים שדגמו Fjeld, Achten, Hilligoss, Nagy & Sri Kumar, לעיל ה"ש 8, וכ-56% מהמסמכים שדגמו Jobin, Ienca & Vayena, לעיל ה"ש 8, מתייחסים לפרטיות.

(4) **יידוע והסבר**:<sup>35</sup> פרטים זכאים לדעת שמערכת אוטומטית נמצאת בשימוש,<sup>36</sup> ולהבין כיצד ואיך היא תורמת לתוצאות ההחלטה שמשפיעה עליהם. מעצבים, מפתחים ומפעילים של מערכות אוטומטיות נדרשים להסביר בשפה פשוטה כיצד המערכות שלהם מתפקדות ומה תפקיד האוטומציה בהן. על ההסבר להיות תקף, בעל משמעות ושימושי לכל מי שנדרש להבין את המערכת.

(5) **חלופה אנושית**:<sup>37</sup> לפי עיקרון זה לאדם צריכה להינתן האפשרות להחלטה אנושית במקום החלטה אוטומטית במקרים הראויים לכך. המקרים ייקבעו בהתבסס על ציפייה סבירה בכל מקרה לגופו, כשהדגש הוא בהבטחת נגישות רחבה והגנה על הציבור מפני השפעות פוגעניות במיוחד. עוד מצוין במסגרת עיקרון זה כי במקרים מסוימים חלופה אנושית תהיה חובה הנדרשת לפי חוק. במקרים אחרים, נדרש שלפרט תהיה גישה לשיקול דעת אנושי (וסעד שיינתן באמצעות עתודה אנושית) וכן הליך טיפול מסלים (escalation process), אם התגלו שגיאה או כשל במערכת אוטומטית, או במקרה שבו הפרט רוצה לערער או למחות בכין השפעותיה עליו. שיקול דעת או עתודה אנושית צריכים להיות נגישים, הוגנים, יעילים, וכן להיות מנוהלים ומלווים בהכשרות מתאימות. דגשים נוספים נדרשים ביחס להחלטות אוטומטיות ביחס למה שהוגדר כ"תחומים רגישים", קרי, החלטות במסגרת מערכת המשפט הפלילי, תעסוקה, חינוך ובריאות; כאשר לתחומים אלה, נדרש יהיה כי המערכות תהיינה מותאמות למטרה, תספקנה גישה משמעותית לפיקוח, וכן תכלול הכשרות לפרטים הבאים באינטראקציה עם מערכות אלו. כמו כן, מערכות אלו צריכות לכלול שיקול דעת אנושי בהחלטות המוגדרות "סיכון גבוה" או "מזיקות". על הדיווחים באשר לפיקוח האנושי על מערכות אלו להיות זמינים לציבור.<sup>38</sup>

בכך מתכתב המתווה האמריקאי עם ה-AIA המוצע על ידי הנציבות האירופית, שבו מוגדרת קטגוריה של מערכות אוטומטיות בסיכון גבוה הכוללות מערכות בשימוש מערכת חינוך, תעסוקה, משפט פלילי ועוד.<sup>39</sup> ואולם, לפי ה-AIA לא תינתן זכות אוטומטית להחלטה של גורם אנושי במקרים אלו אלא המערכות האלו יהיו כפופות למערכת ניהול סיכונים בכל שלבי מחזור החיים שלהן.<sup>40</sup> בהקשר זה, נדמה כי המתווה האמריקאי אף רחב יותר מהזכות להחלטה של גורם אנושי לפי ה-GDPR, לפיה בכפוף למקרים מסוימים עומדת לפרט הזכות שלא להיות נתון להחלטה המבוססת כולה על החלטה אוטומטית, כשהחלטה זו יש תוצאות משפטיות בנוגע אליו או שמשפיעות עליו משמעותית.<sup>41</sup> לעומת זאת המתווה מרחיב את מסגרת המקרים שבהם צריכה להינתן אפשרות להחלטה של גורם אנושי, ואף מפרט על טיב "העתודה האנושית" והמצופה ממנה במסגרת הליך קבלת החלטות שיעבור לידיה.

יש לתת את הדעת על כך שבמסגרת הדוגמאות שהמתווה מפרט להחלטות אוטומטיות העשויות להצדיק חלופה אנושית,

35 AI BOR, לעיל ה"ש 1, בעמ' 40-45. 94% מהמסמכים שדגמו Fjeld, Achten, Hilligoss, Nagy & Srikumar, לעיל ה"ש 8, וכ-87% מהמסמכים שדגמו Jobin, Ienca & Vayena, לעיל ה"ש 8, מתייחסים לשקיפות. ראו פרל ושוורץ אלטשולר, לעיל ה"ש 6.

36 על הזכות לאינטראקציה מיוחדת עם בינה מלאכותית ראו ס' 52(1) ל-AIA; AI HLEG; לעיל ה"ש 9, בעמ' 18; [THE TORONTO DECLARATION: PROTECTING THE RIGHT TO EQUALITY IN MACHINE LEARNING](#), sec. 18 (2018; Deutsche Telekom, [Guidelines for Artificial Intelligence](#), sec. 4 (Nov. 5, 2018).

37 עקרונות של חירות ואוטונומיה, ובתוכם עקרונות המעמידים את האדם במרכז, נכללים בחלק ממסמכי המדיניות והאתיקה של בינה מלאכותית אשר פורסמו על ידי מדינות שונות בשנים האחרונות. 69% מהמסמכים שדגמו Fjeld, Achten, Hilligoss, Nagy & Srikumar, לעיל ה"ש 8, כוללים התייחסויות להיבטים של שליטה אנושית בטכנולוגיה; וכ-40% מהמסמכים שדגמו Jobin, Ienca & Vayena, לעיל ה"ש 8, מתייחסים לחירות ואוטונומיה.

38 AI BOR, לעיל ה"ש 1, בעמ' 46.

39 ס' 6 והתוספת השלישית ל-AIA.

40 ס' 8-9 ל-AIA.

41 [Commission Regulation 2016/679 of 27 Apr. 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 \(General Data Protection Regulation\)](#), Christopher Kuner, Lee A. Bygrave & Christopher Docksey, *Background* כמו כן ראו (להלן: GDPR) 2016 O.J. (L 119) 1, art. 22 and Evolution of the EU General Data Protection Regulation (GDPR), in THE EU GENERAL DATA PROTECTION REGULATION (GDPR) 523-525 (2020).



נמנות גם החלטות הנעשות במסגרת סינון תכנים אוטומטי (content moderation),<sup>42</sup> הנפוץ ברשתות חברתיות. אף שהמתווה אינו מחייב, ברור שסינון תכנים היא פעולה הנעשית באופן כמעט בלעדי על ידי מתווכי התוכן, ועל כן צעד זה עשוי להתפרש כקריאת תיגר על המדיניות המקלה והבלתי-מתערבת שהממשל האמריקאי נקט עד כה ביחס לאחריותם של מתווכי תוכן.<sup>43</sup>

## המשמעות המשפטית של המתווה

מבחינה משפטית גרידא, התוכנית האמריקאית אינה דבר חקיקה ואף לא טיוטת הצעה לחקיקה; מדובר במתווה בלתי מחייב ובלתי אכיף. זוהי למעשה הצהרת עקרונות, שנראה כי מילולית היא נראית ממוסגרת בהקשר הלאומי באמצעות פנייה לעם האמריקאי (For the American People).<sup>44</sup>

עם זאת, קשה להתעלם מהחשיבות ההצהרתית שבמסמך. ראשית, הטרימינולוגיה שנבחרה לכותרת המתווה – המשתמשת במונח "מגילת זכויות" (Bill of Rights) – מסמנת באופן ברור שלצד ההזדמנויות והפוטנציאל החיובי שיש לבניה מלאכותית, היא גם טומנת בחובה סכנות לזכויות אדם, ולכן יש להבטיח שזכויות וחירויות הפרט לא ייפגעו מפעולתן של מערכות אוטומטיות מבוססות בינה מלאכותית. נדמה שזהו צעד מתבקש לנוכח הספרות המחקרית הענפה המצביעה על כך שפעמים רבות מערכות אלגוריתמיות פועלות בצורה מפלה, בהיעדר שקיפות, מובנות ונגישות, באופן המערער עקרונות יסוד של מנהל דמוקרטי ובתוכם עקרון השקיפות והחובה לנמק החלטות.<sup>45</sup>

בחירה צורנית נוספת במסמך שעליה ניתן להצביע היא שימוש בגוף שני בתיאור העקרונות שלו.<sup>46</sup> באמצעות פנייה במישרין אל הקורא האנושי מגלם המתווה את עקרון האדם במרכז (human centric), שלפיו על מערכות אוטומטיות להיות תומכות ומעצימות בני אנוש, ולהיות מעוצבות בצורה שתכלול נורמות כמו שלטון החוק, זכויות אדם, ערכים דמוקרטיים, הגנה על פרטיות, והיעדר אפליה.<sup>47</sup>

זאת ועוד, השימוש של המתווה בשיח זכויות, להבדיל משיח של עקרונות אתיים, ובמיוחד בארצות-הברית שעד כה נקטה ככלל מדיניות של אי-התערבות בתחום הרגולציה של בינה מלאכותית, עשוי לסמן תפנית משמעותית. בעוד שעקרונות אתיים הם עמומים, רחבים ופתוחים לפרשנות בלתי מחייבת, זכויות – אם במישור הלאומי או הבינ-לאומי – הן מערכת נורמטיבית מוגדרת, המכילה נורמות מוכרות ומבוססות. זכויות הן גם לרוב אכיפות משפטית ולכן בעלות משמעות חזקה יותר מעקרונות אתיים בעלמא.<sup>48</sup> אף על פי שהמעבר לשפה של זכויות במתווה האמריקאי מוגבל, שכן משאנו מקלפים את הכותרת "מגילת זכויות" מתגלה כי המתווה מפרט "עקרונות" (principles), ולא "זכויות" (right to), הרי שבעצם ההצבה של המינוח מגילת זכויות לצד מערכות אוטומטיות, יש כדי לאותת לכל הפחות על שינוי מגמה במדיניות האסדרה של בינה מלאכותית.

42 AI BOR, לעיל ה"ש 1, בעמ' 53.

43 Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L.J. 639 (2014).

44 ראו, לדוגמה, AI BOR, לעיל ה"ש 1, בעמ' 1, 2, 31 ו-36.

45 Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305 (2019).

46 ראו, לדוגמה, AI BOR, לעיל ה"ש 1, בעמ' 6: "You should know how and why an outcome impacting you was determined by an automated system, including when the automated system is not the sole input determining the outcome"

47 *Human-Centred Values and Fairness (Principle 1.2)*, OECD.AI POLICY OBSERVATORY

48 ראו Karen Yeung, Andrew Howes & Ganna Porgebna, *AI Governance by Human Rights-Centered Design, Deliberation, and Oversight: An End to Ethics Washing*, in *THE OXFORD HANDBOOK OF ETHICS OF AI* 76 (Markus D. Dubber, Frank Pasquale & Sunit Das eds., Oxford Univ. Press 2020).

## הטבעת חותם במישור הבין-לאומי?

בשנים האחרונות נראה כי אירופה היא שחקן משמעותי בתחום האסדרה של טכנולוגיות חדשות. אם בהשפעה של ה-GDPR משנת 2016 או בזו של דברי חקיקה מאוחרים יותר, כגון טיוטת ה-AIA, ה-EU Digital Service Act<sup>49</sup> או הכרזת הנציבות האירופית על זכויות דיגיטליות (Declaration on Digital Rights and Principles)<sup>50</sup>. נוסף עליהם, מועצת אירופה, גוף אירופי חשוב נוסף, פרסמה לאחרונה "קווי יסוד" לבינה מלאכותית.<sup>51</sup> לפיכך, ייתכן שחרף השימוש בלשון הלאומית (American Public) יש בפרסום המתווה הצהרה אמריקאית שלפיה גם ארצות-הברית מבקשת להטביע חותם בפיתוח המדיניות הגלובלית בתחום, וזאת חרף השלבים המתקדמים בחקיקת ה-AIA והסבירות שבשלב זה ארצות-הברית "איחרה את הרכבת".

החקיקה האירופית המסתמנת בתחום הבינה המלאכותית ומסמכי המדיניות שעליה היא מתבססת נעדרים טרמינולוגיית זכויות. אומנם, מסמך המומחים שהתווה את הבסיס ל-AIA מבסס אנליטית את העקרונות האתיים הנמנים בו על זכויות יסוד אירופיות, במטרה לייצר מסגרת אתית לבינה מלאכותית אמינה (trustworthy) הנשענת עליהן, אך אלו אינם מנוסחים כזכויות אדם.<sup>52</sup>

לבסוף, יצוין כי בעת כתיבת שורות אלה פורסם מסמך מדיניות רגולציה ואתיקה בתחום הבינה המלאכותית בישראל.<sup>53</sup> בדומה למתווה האמריקאי, גם מסמך המדיניות הישראלי אינו דבר חקיקה. מדובר בעקרונות אתיים מנחים ביחס למדיניות אסדרה של מערכות בינה מלאכותית.<sup>54</sup> בעוד שחלק מהעקרונות המוצעים בו זהים לאלו המנויים במתווה האמריקאי, המסגור שלהם נותר כעקרונות אתיים מומלצים בלבד. יתר על כן, המסמך הישראלי נועד להתוות את האופן שבו תיבחן אסדרה של בינה מלאכותית במגזר הפרטי;<sup>55</sup> בעוד שהמתווה האמריקאי פונה הן למגזר הפרטי והן למגזר הציבורי.<sup>56</sup>

## סיכום

המתווה האמריקאי אינו מהווה מסמך מחייב כשלעצמו אלא בעל תוקף הצהרתי גרידא. זוהי גם עיקר הביקורת המכוונת אליו.<sup>57</sup> בפועל, הוא משאיר את אסדרת הבינה המלאכותית בארצות-הברית כפי שהייתה, וזאת למרות העובדה שהמערכות האלו כבר מהוות תפקיד משמעותי בחייהם של פרטים, בין אם הם אזרחי ארצות-הברית ובין אם לאו. עם זאת, במידה

*Proposal for a Regulation of the European Parliament and of the Council on the Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, COM (2020) 825 final (Dec. 15, 2020) 49

*EUROPEAN DECLARATION ON DIGITAL RIGHTS AND PRINCIPLES FOR THE DIGITAL DECADE* (2023) 50

*CAHAI – Ad Hoc Committee on Artificial Intelligence*, COUNCIL OF EUROPE 51

ראו AI HLEG, לעיל ה"ש 9. 52

**מסמך מקצועי: עקרונות מדיניות, רגולציה ואתיקה בתחום הבינה המלאכותית** (נספח למסמך המדיניות, משרד המשפטים, ייעוץ וחקיקה, משרד החדשנות המדע והטכנולוגיה (2022). 53

שם, בעמ' 102–109. עקרונות אלו שואבים את השראתם מהמלצות ה-OECD. ראו שם, בעמ' 103. 54

שם, בעמ' 16: "מדיניות הרגולציה והאתיקה בתחום הבינה המלאכותית, המוצעת במסמך זה, נועדה להתוות את האופן שבו תבחן הממשלה את השאלה האם יש לקבוע כללי התנהגות לפיתוח ושימוש בבינה מלאכותית במגזר הפרטי ואת האופן לעשות כן. זאת, במטרה לקדם פיתוח ושימוש בטכנולוגיות בינה מלאכותית ואף להפחית חסמים משפטיים ורגולטוריים הניצבים בפניהן, ובה בודד לצמצם פגיעות אפשריות בזכויות יסוד ואינטרסים ציבוריים אגב פיתוח ושימוש בבינה מלאכותית". 55

ראו AI BOR, לעיל ה"ש 1, בעמ' 8–9. 56

ראו *6 Reactions to the White House's AI Bill of Rights: The Nonbinding Principles are Being Both Celebrated and Vilified*, IEEE SPECTRUM (Oct. 14, 2022). 57

מסוימת, המעבר מרטוריקה של אתיקה לזכויות,<sup>58</sup> מכיר בכך שמערכות בינה מלאכותית אינן מהוות טכנולוגיה "ניטרלית", אלא שיש בכוחן לסכן חירויות וזכויות יסוד. מעבר לקריאה לקובעי מדיניות להיות מודעים לכך ולפעול בהתאם לעקרונות המפורטים במתווה, המתווה עשוי להיות בבחינת סימן לבאות המאות על שילוב שיח של זכויות אדם ועל מדיניות רגולציה הדוקה יותר בתחום הבינה המלאכותית.

---

58 [THE GLOBAL AI INDEX](#) (Tortois). בהקשר זה מעניין לשים לב כי חרף הדירוג הגבוה של ארצות-הברית ושל ישראל באינדקס הבינה המלאכותית הגלובלי (במקומות הראשון והחמישי בהתאמה), שתיהן נשרכות מאחור בקטגוריית האסטרטגיה הממשלתית (במקומות ה-17 וה-45 בהתאמה). ייתכן שפרסום התוכנית למגילת הזכויות בבינה המלאכותית (*AI BOR*, לעיל ה"ש 1) ומסמך העקרונות למדיניות רגולציה בבינה מלאכותית (מסמך מקצועי: עקרונות מדיניות, רגולציה ואתיקה בתחום הבינה המלאכותית, לעיל ה"ש 53) יחזק את דירוגן באינדקס.