

# מודל ליצירת שקיפות אלגוריתמית

גדי פרל | תהילה שוורץ אלטשולר

אוגוסט 2022



המכון הישראלי  
לדמוקרטיה

הצעה  
לסדר  
47



המכון הישראלי  
לדמוקרטיה

# מודל ליצירת שקיפות אלגוריתמית

גדי פרל | תהילה שוורץ אלטשולר

הצעה לסדר 47

אוגוסט 2022

A Model for Algorithmic Transparency  
Gadi Perl | Tehilla Shwartz Altshuler

עיצוב הסדרה והעטיפה: סטודיו Alfabees  
ביצוע גרפי: רונית גלעד, ירושלים  
הדפסה: גרפוס פרינט, ירושלים

מסת"ב 8-397-519-965-978 ISBN

© כל הזכויות שמורות למכון הישראלי לדמוקרטיה (ע"ר), 2022  
נדפס בישראל תשפ"ב/2022

אין לשכפל, להעתיק, לצלם, להקליט, לתרגם, לאחסן במאגר ידע, לשרר או לקלוט בכל דרך או  
אמצעי אלקטרוני, אופטי או מכני או אחר – כל חלק שהוא מהחומר בספר זה. שימוש מסחרי מכל  
סוג שהוא בחומר הכלול בספר זה אסור בהחלט אלא ברשות מפורשת בכתב מהמו"ל.

**המכון הישראלי לדמוקרטיה (ע"ר)**

רח' פינסקר 4, ת"ד 4702, ירושלים 9104602  
טל': 02-5300888  
אתר האינטרנט: [www.idi.org.il](http://www.idi.org.il)

**להזמנת ספרים:**

החנות המקוונת: [www.idi.org.il/books](http://www.idi.org.il/books)  
דוא"ל: [orders@idi.org.il](mailto:orders@idi.org.il)  
טל': 02-5300800

כל פרסומי המכון ניתנים להורדה חינם, במלואם או בחלקם, מאתר האינטרנט.

## המכון הישראלי לדמוקרטיה

---

המכון הישראלי לדמוקרטיה הוא מוסד עצמאי אי-מפלגתי, מחקרי ויישומי, הפועל בזירה הציבורית הישראלית בתחומי הממשל, הכלכלה והחברה. יעדיו הם חיזוק התשתית הערכית והמוסדית של ישראל כמדינה יהודית ודמוקרטית, שיפור התפקוד של מבני הממשל והמשק, גיבוש דרכים להתמודדות עם אתגרי הביטחון מתוך שמירה על הערכים הדמוקרטיים וטיפוח שותפות ומכנה משותף אזרחי בחברה הישראלית רבת הפנים.

לצורך מימוש יעדים אלו חוקרי המכון שוקדים על מחקרים המניחים תשתית רעיונית ומעשית לדמוקרטיה הישראלית. בעקבותיהם מגובשות המלצות מעשיות לשיפור התפקוד של המשטר במדינת ישראל ולטיפוח חזון ארוך טווח של תרבות דמוקרטית נכונה לחברה הישראלית ולמגוון הזהויות שבה. המכון שם לו למטרה לקדם בישראל שיח ציבורי מבוסס ידע בנושאים שעל סדר היום הלאומי, ליזום רפורמות מבניות, פוליטיות וכלכליות ולשמש גוף מייעץ למקבלי ההחלטות ולציבור הרחב.

המכון הישראלי לדמוקרטיה הוא זוכה פרס ישראל לשנת תשס"ט על מפעל חיים – תרומה מיוחדת לחברה ולמדינה.

## תוכן העניינים

---

7	תקציר
9	פרק 1. רגולציה של מערכות בינה מלאכותית
14	פרק 2. בינה מלאכותית ובעיית העכירות
16	פרק 3. שקיפות אלגוריתמית במענה לעכירות אלגוריתמית
18	פרק 4. השקיפות הקלסית
26	פרק 5. שקיפות אלגוריתמית ומשמעויותיה
38	פרק 6. המודל המוצע: מודל כפול – שקיפות ואיתנות מדעית
50	פרק 7. היתרונות והחסרונות של המודל המוצע
54	פרק 8. המודל המוצע ויישומו
60	פרק 9. סיכום



## תקציר

מערכות מבוססות בינה מלאכותית יכולות לספק פתרונות יעילים בסביבות מורכבות ובתחומים שדרשו עד כה כוח אדם מיומן ושבידיו הופקד שיקול דעת לביצוע המשימה. אלא שלצד התועלות הברורות, מקרים אחדים שהתפרסמו לאחרונה – כגון דירוג אשראי מפלה והטיות אלגוריתמיות בתכנים ברשתות החברתיות – מעלים חשש הן מפני תפקוד לקוי בשל הסתמכות על מאגרי נתונים מוטים או מפלים והן מפני הקושי לאתר תקלות ולתקןן בשל אופי ה"קופסה השחורה" וחוסר השקיפות – ה"עכירות" המובנית – בדרך פעולתן של מערכות בינה מלאכותית.

במענה לבעיית העכירות חוקרים ורגולטורים מעודדים הגברת חובות של "שקיפות אלגוריתמית", בכוונה לספק למשתמשים מידע על אודות דרך פעולתן של מערכות אלגוריתמיות. ברם, לא ברור למה הכוונה ב"שקיפות אלגוריתמית" זו. הכתיבה הקיימת אינה מספקת מענה ברור למהות השקיפות ולדרישות שיש להציב כדי לממשה. פער זה יוצר אי־בהירות באשר לחובות שיוטלו בפועל, פוגם בצפייה לוודאות בעניין רמת האחריות והחובות שיוטלו ועלול להביא גם לפגיעה בזכויות המשתמשים וגם באינטרסים של היצרנים.

שקיפות היא מונח שמקורו במשפט המינהלי, והתמרתו לעולם האלגוריתמי דורשת הבהרה. סקירה של הספרות הקיימת בתחום מעלה כ־13 פרשנויות שונות למושג, אחדות חופפות זו לזו אך אחרות שונות זו מזו. יש הרואות בשקיפות האלגוריתמית מרכיב של חובת ההסברתיות, בעוד אחרות רואות בה מושג רחב יותר, הכולל גם חובות גילוי נאות צרכניות וגם חובות לשקף מידע על מטרות התוכנה והשפעתה על זכויות.

מחקר זה נועד להשלים פער פרשני זה, והוא מציע שני מסלולים חלופיים למימוש חובת השקיפות האלגוריתמית – המסלול הקלסי; ומסלול האיתנות התהליכית המדעית.

המסלול הראשון מתמקד במקורות השקיפות הקלסית ודורש מתן הסבר רציונלי לסיבה שבשלה מערכת אלגוריתמית סיפקה פלט מסוים ולא אחר – בדומה למצב

הקיים במשפט המינהלי, הרואה בשקיפות אמצעי מקדמי המאפשר למערכות הבקרה לבחון את איכותן של החלטות ואת דרכי קבלתן. המסלול השני מאפשר מתן הוכחה מתמטית-סטטיסטית לתקינות המערכת האלגוריתמית ולביסוס הנתונים שעליו היא מסתמכת. כאשר מפעיל המערכת יכול להוכיח שהיא נבנתה על סמך מאגר נתונים מתאים סטטיסטית ונקי מהטיות, ושהאלגוריתם מותאם למשימה ומופעל באופן שוויוני – יהיה בכך תחליף להסבר הרציונלי.

ברירת המחדל תהא המסלול הראשון, ואולם במקרים שבהם בשל מגבלות טכנולוגיות מובנות במערכות בינה מלאכותית המערכת אינה מסוגלת לספק את המענה, וכאשר ההשפעה של האלגוריתם על הזכויות איננה משמעותית, ניתן יהיה להסתפק במסלול השני.

כדי לממש את חובות השקיפות באמצעות שני מסלולים אלה, אנו ממליצים להטמיע שינויים רגולטוריים. מאחר שמערכות בינה מלאכותית עתידות לפעול במספר רב של שווקים, שלחלק מהם רגולציה ייעודית, כגון שוק התרופות והתחבורה, אנו ממליצים כי הבחירה בין שני המסלולים תיבחן לגופה על ידי הרגולטורים הייעודיים.



# פרק 1

## רגולציה של מערכות בינה מלאכותית

בשנים האחרונות ניתן לראות עלייה בשימושים במערכות המבוססות על בינה מלאכותית – החל בסינון של צריכת תכנים ברשתות החברתיות, המשך בחלוקת כספי רווחה וכלה בדירוגי אשראי בנקאיים. הפוטנציאל החיובי של מערכות אלו ברור: הן יכולות לעבוד באופן מדויק ומהיר בסביבה רבת

משתנים, ביעילות רבה יותר וללא הכשלים המתרחשים בעת קבלת החלטות על ידי בני אדם.<sup>1</sup> מנגד, עולים חששות בדבר ליקויים בתפקוד של מערכות אלו בגלל התבססותן על מערך נתונים מוטה או שגוי; נטייתן להעצים הטיות קיימות; והיעדר היכולת לאתר תקלות בתפקודן בשל הקושי להבין כיצד הגיעו לתוצריהן.<sup>2</sup> חששות אלו התממשו לאחרונה בכמה מקרים ברחבי העולם, בין השאר במידע ראשוני בעניין קביעות מפללות של דירוגי אשראי;<sup>3</sup> תפקוד לקוי בחלוקת כספי רווחה באירופה – ובכלל זה באנגליה ובהולנד;<sup>4</sup> מעורבות בהטיות בחירות ברשת

1 על היחרונות של מערכות אלו ראו בהקדמה לספרו של ראיין אבוט: "Artificial intelligence (AI) is doing more than ever before, and often doing it cheaper, faster, and better than people are." RYAN ABBOTT, THE REASONABLE ROBOT: ARTIFICIAL INTELLIGENCE AND THE LAW (2020)

2 Matthew Hutson, *The Opacity of Artificial Intelligence Makes It Hard to Tell When Decision-Making is Biased*, 58 (2) IEEE SPECTRUM 40 (2021). על הנזק החברתי של מערכות אלו ראו הערת הדווח לאו"ם בנוגע לזכויות אדם: 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)

3 Nikita Aggarwal, *The Norms of Algorithmic Credit Scoring*, 80 (1) THE CAMBRIDGE LAW JOURNAL 42 (2021)

4 על אירוע האפליה בהולנד ראו למשל Justin Burack, *Addressing Algorithmic Discrimination in the European Union*, A PATH FOR EUROPE, 24 November, 2020

החברתית פייסבוק;<sup>5</sup> הכרעות בנוגע למדיניות סינון תוכן ברשת החברתית עם השפעות רחב בעייתיות;<sup>6</sup> ועוד. בד בבד, עולים חששות לגבי העתיד, למשל בכל הנוגע לאחריות לתאונות שיתרחשו ברכבים אוטונומיים, לטעויות במערכות לאבחון רפואי ועוד.<sup>7</sup> לכן, בצד ההכרה בתועלות הצפויות משימוש במערכות בינה מלאכותית ובמערכות מבוססות למידת מכונה, מתרבה העיסוק המחקרי בהתמודדות עם הנזקים שמערכות אלה עלולות לגרום – בין בשל מאפיינים אינהרנטיים של תהליך הלמידה לגווניו, בין בשל פעולה לא תקינה, ובין בשל הפעלה לא מבוקרת של המערכות על ידי גורמים זדוניים או רשלניים.

לפיכך לנוכח החששות והסימנים הראשונים לתקלות במערכות אלגוריתמיות החלו להישמע קולות הקוראים לרגולציה על מערכות בינה מלאכותית, אשר תבטיח פעולה של מערכות אלו באופן שיקדם אינטרסים ציבוריים, בהם הגנה על זכויות אדם, הבטחת הפעילות התקינה של המערכות והתאמה לעקרונות משפטיים קיימים. ספרות אקדמית ומקצועית וכן מסמכי מדיניות מדינתיים ועל-מדינתיים מציעים מודלים לפיקוח על פיתוח מערכות בינה מלאכותית ועל השימוש בהן.<sup>8</sup> העיסוק ברגולציה של בינה מלאכותית מופיע בארבע סוגות של מסמכים:

5 Karen Hao, *Facebook's Ad-Serving Algorithm Discriminates by Gender and Race*, MIT TECHNOLOGY REVIEW, April 5, 2019; Information Commissioner's Office of the United Kingdom, "Investigation into the Use of Data Analytics in Political Campaigns," 11 July, 2018

6 ועדת ביקורת שהוקמה על ידי פייסבוק על מנת להתמודד עם טענות אלו פרסמה החלטות בנושא אשר מעודכנות באחר חברה פייסבוק.

7 ABBOTT, לעיל ה"ש 1, פרק 3, בעמ' 50-70.

8 Jessica Morley, Luciano Floridi, Libby Kinsey, & Anat Elhalal, *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices*, 26 (4) SCIENCE AND ENGINEERING ETHICS 2141 (2020)

(1) מחקרים אקדמיים של משפטנים ושל אנשי מדעי החברה – בהם ניתן למנות את בריינט ווקר סמית (Walker Smith), ראיין אבוט (Abbott) ווירג'יניה דיגנום (Dignum).<sup>9</sup>

(2) מסמכי מדיניות של מכוני מחקר אקדמיים וממשלתיים, כמו למשל מסמך העמדה של ארגון המדינות המתועשות (OECD); מסמך של קבוצת מומחים מטעם הנציבות האירופית; מסמך מומחים שנכתב עבור הממשל הפדרלי בארצות הברית; ניתוח והמלצות מאת ארגון התקינה הבינלאומי IEEE; פעילות מכון המחקר Partnership on AI; סקירה מטעם Tech Policy Institute בישראל; והדין וחשבון של ועדת משנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית.<sup>10</sup>

(3) ניירות מדיניות של חברות מסחריות העוסקות בפיתוח, אשר משמשים מעין הצהרה על כוונותיהן העתידיות בכל הנוגע למוצריהן – כגון זה של גוגל, מייקרוסופט ו-IBM.<sup>11</sup>

Bryant Walker Smith, *How Governments Can Promote Automated Driving*, 9 VIRGINIA DIGNUM, RESPONSIBLE ; לעיל ה"ש 1; ABBOTT ;47 NML REV. 99 (2017) ARTIFICIAL INTELLIGENCE: HOW TO DEVELOP AND USE AI IN A RESPONSIBLE WAY (Springer Nature, 2019)

Recommendation of the Council on Artificial Intelligence, OECD 10 Legal Instruments, 2019; High-level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, European Commission, Directorate-General for Communications Networks, Content and Technology, April 8, 2019; David Freeman Engstrom, Daniel E. Ho, Catherine M. Sharkey, & Mariano-Florentino Cuéllar, "Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies," NYU School of Law, Public Law Research Paper, No 20-54, 2020; Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, IEEE, 2019; Partnership on AI סיון תמיר "בינה מלאכותית בשירותי ממשל: הטמעת מערכות לקבלת החלטות מבוטסות-אלגוריתם בשירותי הרווחה" ISRAELI TECH POLICY INSTITUTE; ועדת משנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של בינה מלאכותית – דין וחשבון (אוגוסט 2019).

Responsible AI Practices, GoogleAI; Guidelines for Responsible AI, 11 Microsoft AI; Trusted AI for Business, IBM

(4) מסמכים מדינתיים ורבי-מדינתיים הכוללים הצעות חקיקה. בהקשר זה נדגיש במיוחד את טיוטת החקיקה שפרסם האיחוד האירופי באפריל 2021 בנוגע לשימושים של מערכות מבוססת בינה מלאכותית באיחוד.<sup>12</sup>

העיסוק העיקרי במסמכים אלה הוא בסוגיית חוסר הבהירות (opacity), חוסר השקיפות ו"הקופסה השחורה" של מערכות בינה מלאכותית, שבה דנים כותבים ומקבלי החלטות כבר שנים אחדות. החשש הוא מפני "מכונה" שתחליף את שיקול הדעת האנושי בלי שנוכל להבין כיצד היא פועלת ומהי דרך קבלת ההחלטות שלה. מדובר בחשש שיש לו היבטים רציונליים ומיסטיים כאחד, והוא תקף בכל הקשר של שימושים במערכות לומדות.

כפי שנראה להלן, הטענה הרווחת היא שכדי להתגבר על חוסר הבהירות של מערכות לומדות והסכנות הפוטנציאליות הטמונות בהן, נדרשת "שקיפות אלגוריתמית". כלפי חברות הטכנולוגיה התביעה היא: הסבירו לנו כיצד האלגוריתמים שלכן עובדים, בדיוק כפי שנדרש מחברת קוקה קולה לחשוף את המתכון הסודי שלה; וכלפי ממשלות: בטרם תתירו למשטרה להשתמש במערכות זיהוי פנים, הסבירו לנו כיצד פועל בהן האלגוריתם.

ואולם "שקיפות" היא מושג שמקורו בדין המינהלי, בהקשר של היחסים שבין הפרט (או הציבור) לבין השלטון. יתרה מזו, הכתיבה העוסקת בצורך ב"שקיפות אלגוריתמית" לרוב אינה מגדירה במדויק את המושג, ולכן הדרישה בה לשקיפות כזאת היא הצהרתית בלבד. המצדדים ב"שקיפות אלגוריתמית" אינם מבהירים באיזה חלק של תהליך הפעולה של מכונה לומדת היא נדרשת, ואיזו רמה של

Proposal for a Regulation of the European Parliament and of the 12  
Council Laying Down Harmonised Rules on Artificial Intelligence  
(Artificial Intelligence Act) and Amending Certain Union Legislative  
Acts, COM (2021) 206 final, 21 April 2021; ישראל התייחסה ראשונית לנושא  
מופיעה במסמך של מרכז המחקר והמידע של הכנסת - רועי גולדשמיט "הטיה אלגוריתמית  
במערכות בינה מלאכותית". ראו גם פרסום של ממשלת ניו זילנד בנושא Trustworthy  
AI in Aotearoa (March 2020); Charlotte Graham-McLay, *New Zealand Claims  
World First in Setting Standards for Government Use of Algorithms*, THE  
GUARDIAN, 27 Jul., 2020

שקיפות נדרשת בשימושים שונים המבוססים על בינה מלאכותית, כך שבפועל, המושג "שקיפות אלגוריתמית" נשאר בלתי שקוף.

כדי להשלים את הפער, מסמך זה ינסה לבאר את המושג "שקיפות אלגוריתמית". לשם כך נסביר תחילה בקצרה כיצד פועלות מערכות בינה מלאכותית; לאחר מכן נציג את "בעיית העכירות" של מערכות לומדות כפי שהיא מופיעה בספרות ואת הטיעונים בזכות שקיפות במענה לבעיה זו; נמשיך בסקירה של ההקשרים שבהם מופיע העיסוק בשקיפות ובהגדרות השונות למושג. אגב הסקירה, נסביר את הבעיות בהגדרות השונות ואת הקשיים שנוצרים כתוצאה מכך.

לאחר מכן נציג מודל למימוש חובת השקיפות האלגוריתמית. המודל מבוסס על הטענה שבמקביל לחובה לספק הסבר רציונלי ומובן להחלטה ספציפית, יש צורך ליצור מסלול חלופי שיאפשר הוכחת "איתנות מדעית". כלומר, חזקת תקינות של פעולת המערכת הלומדת במקרים שבהם אי-אפשר לספק הסבר רציונלי שמממש את הדרישה ל"שקיפות אלגוריתמית" במובנה הפשוט. בהתאם למודל שנציע, נפרט מהי מהות השקיפות שיש לדרוש בעת שיווק מוצרים מבוססי בינה מלאכותית, וכיצד היא צריכה להשתנות בהתאם לשימושים השונים של המוצר ולטכנולוגיות הבינה המלאכותית השונות העומדות בבסיס האלגוריתם בכל אחד משלבי פעילות המערכת הלומדת. כדי לתקף את המודל נבחן כיצד יישום של מבחן האיתנות המדעית היה יכול לסייע בשני מקרי בוחן – האחד של מערכות לאיתור הונאות רווחה בהולנד, והשני בקביעת מדדי הצלחה בתהליך הרגולציה של כלי רכב אוטונומיים.

## פרק 2

### בינה מלאכותית ובעיית הענירות

בטרם נציג את בעיית הענירות נבהיר תחילה, במגבלות מסמך שאיננו מתחום מדעי המחשב, מהן מערכות בינה מלאכותית, ובעיקר מהו הפלט שהן מפיקות.<sup>13</sup>

בתהליך המכונה "למידת מכונה" תוכנה מסוגלת לייצר בעצמה רצף של כללים על

בסיס מאגר נתונים שעומד לרשותה ובהתאם להיזון חוזר שניתן על ידי מתכנת או גורם אחר. פעולת למידת המכונה הבסיסית ביותר מייצרת, לאחר למידה, "עץ החלטות". עץ החלטות הוא בעצם רצף של החלטות של "אם" ו"אז", עד לקבלת תוצאה סופית. יתרונה של שיטה זו שהיא עקיבה וברורה להבנה. דע עקא, שיטת למידה זו מוגבלת לפעולות שהן לינאריות באופיין ושניתן להמיר אותן לעץ החלטות.

כאשר מדובר בביצוע משימות מורכבות יותר, יש צורך להפעיל מתודולוגיות למידת מכונה מתוחכמות יותר, כגון רשתות נוירונים. רשתות נוירונים נקראות כך משום שהן מדמות את מבנה המוח האנושי, המייצר קשרים בין רצפטורים ונוירונים ומסתמך עליהם כדי לבסס פעילות. בלמידת מכונה מסוג רשת נוירונים המתכנת מזין לתוכנה מאגר נתונים ומאפשר לאלגוריתם היזון חוזר המשקף מתי הצליח בביצוע הפעולה ומתי לא. הצלחה מתגמלת ומבססת את קשר הנוירונים, ולאורך זמן התוכנה מבססת מספיק קשרים כאלה, עד שהמתכנת סבור שהיא הגיעה לשיעורי הצלחה גבוהים דיים. בתוצאה הסופית מתקבל רצף של פקודות אשר מראה הצלחה סטטיסטית במשימה. מתודולוגיה זו, שהיא הנפוצה מתוך שיטות הלמידה המתקדמות, יעילה מאוד אך החיסרון שיש בה לענייננו הוא התבססותה על קורלציה ולא על סיבתיות.

נוסף על כך, הליך הייצור של אלגוריתמים דורש משאבים ניכרים משני סוגים: ראשית, יש צורך במאגר נתונים גדול, עשיר ומטוב, אשר יאפשר לתוכנה להגיע מהר ככל האפשר להצלחה סטטיסטית במשימתה. ככל שהפעולה הנלמדת מורכבת יותר, כך נדרש מאגר נתונים גדול יותר. לכן לא מפתיע שהשחקנים המצליחים ביותר בתחום הבינה המלאכותית הם שחקנים שממילא יש בידיהם מאגרי מידע גדולים. המשאב השני הוא פעולה אנושית לצורך מתן היזון חוזר למכונה על מנת שהיא תוכל ללמוד אם הצליחה או לא. פעולה זו היא יקרה וגם מועדת לטעויות, ולכן חברות ענק עתירות משאבים, שבאפשרותן לממן את ההשקעה הכבדה אך המניבה בטווח הארוך, נהנות מעדיפות גם בהיבט זה.

בשל קשיים אלה פותחו דרכים אשר מנסות לעקוף את הקושי בהיזון החוזר האנושי. אחת השיטות לעשות זאת, המתאימה במקרים מסוימים, היא לימוד אדוורסרי של מערכות בינה מלאכותית. למשל, יצירת מערכת אחת שמטרתה לפרוץ צופן, ומולה יצירת מערכת אחרת שמטרתה הפוכה – לייצר צופן שלא ניתן לפיצוח. בכל סבב בין המערכות הן משתכללות זו מזו, ללא צורך בהיזון חוזר אנושי.

כאמור, הקושי המשמעותי בכל הנוגע לרשתות נוירונים קשור להיותן "קופסה שחורה" ולהיעדר היכולת להסביר באופן נקודתי כיצד נוצר הפלט המסוים שאליו הגיעו. המתכנת אינו יודע את משמעות הקשרים שנוצרו וכיצד התוכנה הגיעה להצלחה, כלומר אינו יודע מדוע התוצר הסופי והפלט של התוכנה הם כפי שהם. במובן זה, על אף ההשוואה בין רשתות נוירונים לבין פעולת המוח האנושי, הדרך שבה התוכנה הגיעה להצלחה אינה סיבתית, והיא תוצר של פעולות סטטיסטיות, חלקן רנדומליות. התוכנה אינה מסוגלת להפיק פלט שמסביר כיצד ומדוע התקבלה תוצאה או החלטה מסוימת ולא אחרת.

## פרק 3

### שקיפות אלגוריתמית במענה לעכירות אלגוריתמית

בתחום הרגולציה של מערכות בינה מלאכותית, ובפרט בתחום האתיקה, יש עיסוק נרחב במרכיב העכירות המובנה במערכות אלו,<sup>14</sup> המקשה את הבנת פעולתן ואת הבקרה על תפקודן התקין. כאמצעי לפתרון בעיית חוסר בהירות זה במערכות קבלת החלטות אלגוריתמיות ולקידום אחריותיות (accountability) מוצע בספרות המושג "שקיפות

אלגוריתמית". התפיסה הרווחת היא שעכירות יוצרת קשיים באיתור הגורם האחראי במקרה של תקלות, והיא עלולה להעצים את הנזק הנגרם מפעילותה של המכונה, מאחר שהמשתמשים אינם מסוגלים להבין מה היה צריך להיות השימוש הנכון או לאתר את המקור לתקלה.

סוגיית השקיפות האלגוריתמית עולה כמעט בכל המסמכים העוסקים ברגולציה של מערכות אלגוריתמיות. כפי שעולה ממסמך מרפז של מכון ברקמן-קליין באוניברסיטת הרוורד, למעלה מ-92% מהמסמכים העוסקים ברגולציה ואתיקה של בינה מלאכותית מזכירים את עקרון השקיפות, ויש לו חלק מכריע במארג הפתרונות המוצע בכתיבה הקיימת.<sup>15</sup> מטה-הסתכלות נוספת על מסמכי מדיניות אתיים, שעשו אנה ג'ובין (Jobin) ושותפיה,<sup>16</sup> העלתה אף היא שלמרות הקושי לבודד עיקרון מאחד לכל מסמכי המדיניות, המושגים של שקיפות, צדק, הגינות, היעדר היזק, אחריות ופרטיות מופיעים בלמעלה ממחצית מ-84 מסמכי המדיניות שנבדקו. עם זאת, מעניין שהמושג "שקיפות אלגוריתמית" נעדר משני מסמכים מרכזיים: מכתיבתו האקדמית של אבוט, שבספרו *The Reasonable*

14 להסבר על עכירות אלגוריתמית ראו DIGNUM, לעיל ה"ש 9, בעמ' 59.

15 Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, & Madhulika Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," Berkman Klein Center Research Publication No. 2020 (Jan. 15, 2020)

16 Anna Jobin, Marcello Ienca, & Effy Vayena, *The Global Landscape of AI Ethics Guidelines*, 1 (9) NATURE MACHINE INTELLIGENCE, 389 (2019)



*Robot* שקיפות מזכרת בדרך אגב בלבד,<sup>17</sup> וממסמך מדיניות העוסק בבינה מלאכותית של חברת גוגל, שנמנעת מלציין את המושג "שקיפות" ובמקום זאת בוחרת בהגדרה המצומצמת יותר של "תקשור" (communication) מגבלות התוכנה למשתמשים בשלבי הייצור השונים,<sup>18</sup> ובכך נעסוק בהמשך.

הבעיה היא שחרף העיסוק הנרחב במושג "שקיפות אלגוריתמית", המושג עצמו "עכור" ואינו נהיר דיו. ההגדרות של המושג, מיקומו ביחס לערכים אחרים, זיהוי השחקנים שמוטלת עליהם חובת השקיפות, זהות השחקנים שכלפיהם תתקיים השקיפות, ומיקומן של החובות לאורך שרשרת הערך או חיי המוצר – כולם פתוחים לפרשנות ולמחלוקת. לתפיסתנו, ללא הגדרה מדויקת של המושג עולה החשש ששקיפות לא תגשים את ייעודה. יתרה מזו, עצם השימוש במושג שקיפות כעיקרון לא ברור יסכל את השקיפות ככלי לעידוד התנהלות אתית תקינה ויטיע בידי מי שמבקשים להותיר את האתיקה כעניין הצהרתי בלבד (מה שמכונה בספרות, *ethics washing, ethics shopping, ethics dumping*, *ethics shrinking* ואף *ethics lobbying*).<sup>19</sup>

ממסמכי המדיניות הקיימים עולה שהכוונה במושג "שקיפות אלגוריתמית" היא ליישם את עקרון השקיפות הקלטי על מערכות טכנולוגיות. כדי להבין אם הדבר אפשרי, נגדיר בקצרה מהי שקיפות במובנה הקלטי, מהן הנמקותיה, ומהם הקשיים ביישומה.

17 ABBOTT, לעיל ה"ש 1.

18 GoogleAI, לעיל ה"ש 11.

19 Luciano Floridi, Josh Cows, Monica Beltrametti, et al., *AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, 28 (4) MINDS AND MACHINES 689 (2018)

## פרק 4

### השקיפות הקלסית

על פי ההגדרה של מילון מרים וובסטר, דבר הוא "שקוף" (transparent) כאשר הוא עונה על אחד מהתנאים הבאים:<sup>20</sup>

- a: free from pretense or deceit: frank
- b: easily detected or seen through: obvious
- c: readily understood
- d: characterized by visibility or accessibility of information especially concerning business practices

מרכיבים אלו של קלות הגילוי, ההבנה והנגישות מוצאים את ביטויים גם בעולם המשפטי. שקיפות מופיעה במקור בספרות המשפטית העוסקת במינהל תקין בעיקר בהקשר של חובת מערכות הממשל והליכי החקיקה להיות פתוחים לציבור ולביקורת מצד מערכת המשפט. כך היא מופיעה, למשל, אצל קולין פורלצה (Porlezza) כערך מרכזי במשפט המינהלי.<sup>21</sup> ה-OECD מציין את השקיפות, האחריות והשיתופיות כשלושה עקרונות מרכזיים במינהל תקין של מערכות השלטון.<sup>22</sup> השקיפות נתפסת בספרות גם כעיקרון דמוקרטי מרכזי,

Transparency, [www.merriam-webster.com](http://www.merriam-webster.com) 20

Colin Porlezza, *Data Journalism and the Ethics of Open Source: 21 Transparency and Participation as a Prerequisite for Serving the Public Good*, GOOD DATA, 189 (29 Theory on Demand, Angela Daly, Kate S. Devitt, & Monique Mann eds., 2019); ראו גם דפנה ברק ארז "המשפט המנהלי והמאבק בשחיתות שלטונית" משפטים ל"ז 667 (תשס"ז).

OECD, "European Principles for Public Administration," SIGMA 22 Papers, No. 27, 8 (1999)

ומקובל לומר שברירת המחדל של השלטון צריכה להיות שקיפות.<sup>23</sup> גם במיזמי "ממשל פתוח" ברחבי העולם שקיפות מוזכרת כערך יסודי.<sup>24</sup>

בהקשר של חובות שלטוניות, שקיפות מתמקדת בשני מאפיינים מרכזיים: האחד, החובה להעמיד מידע שלטוני לפני הציבור; השני, החובה להנגיש את המידע באופן שיהיה מובן לציבור.<sup>25</sup> במחקרה של שוורץ אלטשולר בנוגע לממשל זמין, היא מוסיפה לכך אף חובה לייצר זכות לציבור להשתמש במידע שמועמד לרשותו.<sup>26</sup> בעידן המידע התביעה לשקיפות משתכללת במובנים של העמדת מאגרי המידע של המדינה לרשות הציבור; פרסום חקיקה והחלטות שלטוניות ברשת האינטרנט; ואף פיתוח פלטפורמות דיגיטליות לצורך מעקב אחר תהליכים שלטוניים, כגון שימוש בתקציב המדינה.<sup>27</sup>

במגזר הפרטי חובות השקיפות מתקיימות במגוון הקשרים. ראשית, בדומה למגזר הציבורי – חובות אלו מוטלות על גופים המספקים שירותים ציבוריים או נתפסים כגופים דומהותיים.<sup>28</sup> שנית, מוטלות חובות מעולם הגנת הצרכן, כגון סימון מוצרים, גילוי נאות כלפי צרכנים על מאפייני המוצר, תנאי התשלום, האספקה והאחריות כלפי המוצר.<sup>29</sup> נושא הגילוי הצרכני רחב ועמוק וחל גם על עניינים כגון גודל התווים, בהירות השפה, כמות הפרטים שמוסרים לגבי מהות

23 Tal Z. Zarsky, *Transparent Predictions*, 4 U. Ill. L. Rev., 1503 (2013)

24 לדוגמה, ראו הכרזת ממשל אובמה בנושא: Barack Obama, "Transparency and Open Government," Memorandum for the Heads of Executive Departments and Agencies, 2009

25 תהילה שוורץ אלטשולר מדיניות ממשל פתוח בישראל בעידן הדיגיטלי (מחקר מדיניות 91, המכון הישראלי לדמוקרטיה 2012).

26 שם, עמ' 76.

27 לעניין זה ראו למשל הניסיונות לפתיחת מאגרים משטחיים הנוגעים לאכיפה כחלק מהליכי שקיפות: Transparency and Accountability at the Frontlines of Justice: Police Data Transparency, OPEN GOVERNMENT PARTNERSHIP

28 אסף הראל גופים ונושאי משרה דומהותיים 47 (מהדורה שנייה, 2019).

29 חוק הגנת הצרכן, התשמ"א-1981, סעיף 4.

העסקה, אורך הטקסט וסוגיות של הסכמה המותאמת לסוגי האוכלוסיות. שלישית, בדיני החברות קיימות חובות המיועדות לייצר אמון וביטחון במסחר הבורסאי ולעודד השקעות. כאלה הן, למשל, חובות השקיפות החלות על חברות הנסחרות בבורסה – בדמות פרסום דוחות עיתיים הכוללים מידע על רווחי החברה, אירועים משמעותיים בחברה, ובכלל זה כל פתיחה בחקירה פלילית נגד מי מנושאי המשרה בחברה בקשר לפעילות הנוגעת לעבודתו. נורמות שקיפות מרחיבות, בעיקר במדינות המתפתחות, מומלצות ככלי אפקטיבי לעידוד השקעות חיזוניות, במטרה ליצור אמון כלפי משקיעים זרים.<sup>30</sup> חובות שקיפות מוטלות גם על גופים פרטיים מפוקחים, למשל חובות שמטיל המפקח על הבנקים – לפרסם מידע בנוגע לאחזקותיהם, התפלגות הרווחים, דיווחי עמלות ועוד.

בניגוד לשקיפות שלטונית, שבה ברירת המחדל היא שקיפות, ויש צורך בטעמים חזקים כדי לעצור הנגשה של מידע, במגזר הפרטי איזון האינטרסים שונה, ועיקרון השקיפות מתחרה על בסיס שווה באינטרסים נוגדים כגון סודות מסחריים, זכויות פרטיות וקניין רוחני.<sup>31</sup> כך למשל מתקיים היום דיון בדבר חובות השיתוף בין רכבים מקושרים עם רכבים אחרים בכביש בסביבתם, אל מול הזכויות לפרטיות וההגבלות על השיתוף ברגולציה של האיחוד האירופי.<sup>32</sup>

ביסוד עקרון השקיפות עומדות כמה הצדקות נורמטיביות, פוליטיות וכלכליות. נסקור אותן בקצרה.

Belay Seyoum & Terrell Manyak, *The impact of Public and Private Sector Transparency on Foreign Direct Investment in Developing Countries*, 5 (3) CRITICAL PERSPECTIVES ON INTERNATIONAL BUSINESS 187 (2009)

Zarsky, לעיל ה"ש 23. 31

Araz Taeihagh & Hazel Si Min Lim, *Governing Autonomous Vehicles: Emerging Responses for Safety, Liability, Privacy, Cybersecurity, and Industry Risks*, 39 (1) TRANSPORT REVIEWS, 103 (2019) 32

## הצדקות לשקיפות

### 1. שקיפות כמקדמת אחריותיות

שקיפות נתפסת כשלב מקדים והכרחי ביצירת אחריותיות של מקבלי ההחלטות. לפי תפיסה זו, ללא שקיפות אי אפשר לבצע בקרה, וללא בקרה לא ניתן לאתר את הגורם האחראי על מנת שיתקן את מעשיו. במשפט הישראלי שקיפות ככלי ליצירת בקרה מופיעה בהקשר של חוק חופש המידע,<sup>33</sup> אך גם בהקשרים של חובות פרסום מכרזים וחובות פרסום של רשות המיסים. היגיון זה נכון גם במינהל הציבורי האמריקאי, שבו שקיפות ואחריותיות מופיעות בצמידות גבוהה.<sup>34</sup> שקיפות בהקשר הזה, כפי שמציינים מורלי ואחרים (Morley),<sup>35</sup> נתפסת כמרכיב מרכזי ביצירת אמון בין היחיד לשלטון, ולכן יש לה תפקיד משמעותי בדיון המינהלי, ובפרט בסוגיות של הנמקת החלטות שיפוטיות ומינהליות ושקיפות של מסמכי תקציב.<sup>36</sup>

### 2. שקיפות ככלי להגברת מעורבות הציבור

שקיפות היא מרכיב מרכזי בהגברת מעורבותו של הציבור בהליכי השלטון. על פי הצדקה זו, פרסום פעולות השלטון מביא לכך שהציבור מודע לתהליכים, מגיב בהתאם ואף פועל לתיקון טעויות כאשר הוא מצליח להבחין בהן. בפן השלילי, מעורבות ציבורית יכולה להוביל לביוש (שיימינג) של מי שביצעו פעולות שאינן

33 בג"ץ 3751-03 יוסי אילן נ' עיריית תל-אביב-יפו, פ"ד נט(3) 817.

34 Carolyn Ball, *What is Transparency*, 11 (4) PUBLIC INTEGRITY, 297 (Fall 2009); הופעה של מילים אלו בצמידות אפשר לראות ב-Federal Funding Accountability Act משנת 2006 וב-Open Government Act משנת 2007.

35 Morley, Floridi, Kinsey, and Elhalal, *לעיל* ה"ש 8.

36 Carol Harlow & Richard Rawlings, *Proceduralism and Automation: Challenges to the Values of Administrative Law*, THE FOUNDATIONS AND FUTURE OF PUBLIC LAW (in Honour of Paul Craig), 275 (Elizabeth Fisher, Jeff King, & Alison Young eds., 2019)

בהכרח שליליות,<sup>37</sup> וכן לחיזוק כוחם של החזקים מקרב הציבור המפעילים לחץ לא פרופורציונלי על מקבלי ההחלטות.

### 3. שקיפות ככלי לעידוד יעילות כלכלית

לנוכח התועלות שנוצרות בעקבות פרסום הליכי השלטון לפני האזרחים, הוספת מידע מצדדים שלישיים והגדלת שיתוף הפעולה של הציבור בעקבות אמון מוגבר שהשקיפות סייעה ליצור, שקיפות מייצרת יעילות כלכלית.<sup>38</sup> כך למשל, מחקר אמפירי הוכיח לאחרונה כי הגברת השקיפות לא רק שאינה מאיטה הליכי חקיקה, אלא אף מסייעת בהאצתם.<sup>39</sup> קישור בין שקיפות לבין יעילות גבוהה יותר של מוסדות השלטון היה מרכזי בתפיסת "המשילות הציבורית החדשה" (New Public Management) בבריטניה בתחילת המאה ה-21:

המשילות הציבורית החדשה פירושה שינוי בערכים: מאידיאלים קודמים של שירות ציבורי לעבר, בתקופה של הידוק חגורה הולך וגובר, אידיאולוגיה של מכונה "משומנת ומיומנת". כמו שחזה ה־OECD, מודרניזציה תוביל ליותר פתיחות, אחריותיות ובקרה יעילה, שיסופקו על ידי ביצועים משופרים של השירות הציבורי.<sup>40</sup>

### 4. שקיפות ככלי לקידום האוטונומיה

שקיפות היא גם נגזרת של הזכות לאוטונומיה ולחופש הביטוי,<sup>41</sup> משום שאזרח זכאי לקבל מידע ולהבין את הרציונל העומד מאחורי ההחלטות שמתקבלות בעניינו. השקיפות ככלי לקידום אוטונומיה מוצאת את ביטוייה, למשל, ברגולציה

37 Zarsky, לעיל ה"ש 23.

38 Sara Hagemann & Fabio Franchino, *Transparency vs. Efficiency? A Study of Negotiations in the Council of the European Union*, 17 (3) EUROPEAN UNION POLITICS 408 (2016)

39 ראו Hagemann & Franchino, שם.

40 Harlow & Rawlings, לעיל ה"ש 36, בעמ' 9-10 (כל התרגומים במסמך זה הם שלנו, אלא אם כן צוין אחרת).

41 Zarsky, לעיל ה"ש 23, בעמ' 1545.

45/2001 של הנציבות האירופית, העוסקת בהגנה על היחיד במסגרת איסוף מידע, ובחובה 13(d) לחשוף לפני היחיד לא רק את המידע על החלטות המתקבלות בעניינו, אלא גם את הלוגיקה שמאחוריהן.<sup>42</sup>

לעומת ההגיונות החשובים שצוינו לעיל התומכים בעקרון השקיפות, לעיתים שקיפות מביאה לתוצאה ההפוכה מזו שביקשה להשיג. הבנה של התהליכים שהשקיפות הופכת בהם לחרב פיפיות חשובה, והיא תשרת אותנו בהמשך כשנססה לבחון את ההטמעה של רעיון השקיפות האלגוריתמית.

## מגבלות השקיפות

### 1. השקיפות כמחזקת חזקים

הנגשת מידע מורכב שלצורך הבנתו דרושה מומחיות מחייבת התאמה ופישוט, שאם לא כן העמדתו לרשות הציבור מחזקת את מי שמחזיקים בידיהם כוח פרשנות, והיא אינה תורמת לשקיפות השלטונית. כך למשל נמצא שהעמדה של מידע פיננסי ללא הנגשתו סייעה לבעלי מקצוע שהצליחו לייצר תובנות מתוך המידע, בעוד אזרחים "רגילים", שהמידע לא הוגש להם, לא נהנו מפירות השקיפות.<sup>43</sup>

Article 13(d), Regulation 45/2001/EC of the European Parliament and 42 of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data, OFFICIAL JOURNAL OF THE EUROPEAN COMMUNITIES L 8/1, 2001

Zarsky, לעיל ה"ש 23. 43

## 2. שקיפות אינה משיגה את המטרה של עידוד מעורבות ציבורית ומבזבזת משאבים

מקרים שבהם פרסום המידע לא הוביל למעורבות הציבור ממחישים את המגבלות של שקיפות: כך למשל, פרסום משכורות של בכירים בתחום הפיננסים;<sup>44</sup> וכשלים בחובות גילוי בדיני חוזים.<sup>45</sup>

## 3. שקיפות עלולה להוביל למרוץ לתחתית

הוראות שקיפות ללא תרבות של שקיפות עלולות להוביל לתוצאה הפוכה. כך למשל, חובת כתיבת פרוטוקול לישיבות הביאה להגברת ההיוועצות במסדרונות; ויצירת חובות תיעוד הביאה לעיתים לניסוחם של מסמכים לקוניים בלבד.<sup>46</sup>

## 4. שקיפות עלולה להביא לירידה באמון הציבור ולא לעלייה בו

מקרים שהשקיפות חושפת בהם חסרונות של מערכת פועלת – אם בחשיפה של כשלים שקיימים בה, ואם בהפניית תשומת הלב לתכונות השליליות של האנשים שפעילותם נחשפת – עלולים להוביל לכך שהציבור יאבד את האמון במערכות אלו, לכחות בטווח הקצר, בשל הנזקים השליליים המידיים של החשיפה.

לסיכום חלק זה נבחר כי התמרה של עקרון השקיפות הקלסית ויישומה בתחום מערכות הבינה המלאכותית דורשים, תחילה, השלמות בכל הנוגע להגדרת המושג "שקיפות אלגוריתמית", היקפו והיקף חובות היישום שיוטלו על הגופים שיחויבו במימושו. לשם כך יש צורך לענות על שאלות כגון:

44 ראו דיון בנושא אצל יונתן ארליך "שכר מנהלים בכירים בחברות ציבוריות: רגולציה ומנגנוני פיקוח" מרכז המחקר והמידע של הכנסת, 2006.

45 OMRI BEN-SHAHAR & CARL E. SCHNEIDER, MORE THAN YOU WANTED TO KNOW: THE FAILURE OF MANDATED DISCLOSURE (2016)

46 על התפיסות בקרב רשויות בנוגע לחופש המידע ראו תהילה אלטשולר ויונתן ארבל מידע רוצה להיות חופשי: הטמעת חוק חופש המידע בישראל פרק 8, 84 (מחקר מדיניות 74, המכון הישראלי לדמוקרטיה, 2008).



- מהי המטרה ששקיפות אלגוריתמית באה לשרת?
  - האם מדובר במטרות זהות למטרות השקיפות הקלסית?
  - האם שקיפות זו נועדה לאפשר בקרה אתית או לשמש כלי אכיפה משפטי?
- מאילו סוגי גופים נדרשת שקיפות אלגוריתמית לעומת שקיפות שלטונית?
  - מהן החובות שנגזרות מעקרון השקיפות?
  - האם שקיפות אלגוריתמית משמעה חשיפת האלגוריתם, או די במתן פלט סטטיסטי?
- מהן אמות המידה שלפיהן נגדיר שקיפות אלגוריתמית "טובה"?

## פרק 5

### שקיפות אלגוריתמית ומשמעויותיה

כאמור, על אף ריבוי המסמכים העוסקים במדיניות רגולציה ואתיקה של בינה מלאכותית, על אף השכיחות של המושג "שקיפות" בספרות העוסקת בכך ועל אף ליבה של הסכמה בדבר הצורך בהטלת חובות שקיפות, אין קונצנזוס באשר להגדרת המושג עצמו או באשר לשלבים ולעוצמה שבהם יוטלו חובות שקיפות

כלפי מפתחים. כפי שנראה להלן, יש הסבורים כי חובות אלו צריכות להיות מוטלות כלפי כל המשתמשים, ואילו אחרים מגבילים זאת לשחקנים מסוימים בשלבים השונים; יש המוסיפים לשקיפות גם חובות של גילוי נאות בתקשורת עם התוכנה, על מנת שזו תזדהה כ"לא־אנושית", ואילו אחרים רואים בשקיפות מעין חובה חזית באשר למהות העסקה, ותו לא.

כותבים שונים מעניקים למושג "שקיפות אלגוריתמית" משמעויות שונות ופורטים אותה למושגים נלווים שונים, בהתאם לתפיסתם לגבי (1) חשיבות המושג, (2) מיקומו בשרשרת הערך של הייצור ו־(3) המבנה הרגולטורי שהם מציעים. על הגדרת המושג משפיעה גם זווית ההסתכלות של הכותבים: אם מהכיוון של חובות היצרנים מול משתמשי הקצה, ואם מהכיוון של תכליות שונות של המושג – תפיסת השקיפות כאמצעי להשגת אמון או כתכלית בפני עצמה. לתוך המושג שקיפות מוטמעים, בהתאם לתפיסת העולם של כל כותב, מושגים מעולם הגילוי הנאות הצרכני ומעולם המשפט המינהלי והצורך בשקיפות של מעשי השלטון. לעיתים מושגים אלה משמשים בערבוביה. בחלק זה נסקור את המשמעויות שניתנות למושג "שקיפות אלגוריתמית" בספרות, בהצעות חקיקה ובמסמכי מדיניות.

## 1. שקיפות כחובת יצרן

ארגון התקינה הבינלאומי (IEEE) מגדיר שקיפות מזווית ההסתכלות של החובות המוטלות על יצרן התוכנה כ"חובה להפוך את החלטות התוכנה לבנות גילוי":  
 "The basis of a particular A/IS decision should always be discoverable"<sup>47</sup>.  
 כאשר המסמך פורט את חובות השקיפות, עולות המשמעויות הבאות, שהן רחבות יותר מגילוי של נימוקי ההחלטות:

- **עקיבות** (traceability) – שקיפות בנוגע להליך יישום התכנון ההנדסי של התוכנה. חובה זו אמורה לאפשר איתור של הנורמות שעמדו בבסיס הכנסת המידע וגם של הטיית במידע.
- **וידואיות** (verifiability) – יצירת לוג אשר יאפשר מעקב אחר הליך קבלת ההחלטות שבוצע על ידי התוכנה.
- **עיצוב הגון** (honest design) – עיצוב ממשק המשתמש באופן שיאפשר למשתמש להבין את פעילות המערכת ולקבל את המידע הנכון.
- **בהירות** (intelligibility) – חובה להבטיח הבנה אנושית של מה שאירע בפעולתו של האלגוריתם ומדוע.

ועדת המומחים מטעם הנציבות האירופית מגדירה גם היא את השקיפות מזווית ההסתכלות של החובות המוטלות על היצרן, אך מדגישה גם את החשיבות של שקיפות ביצירת אמון. לכן הוועדה מגדירה שקיפות באופנים הבאים:

- **הסבריות** (explicability) – החובה שמידע על פעילות התוכנה יובן על ידי הצד המשתמש. מונח זה כולל שקיפות של תהליך קבלת ההחלטה; תקשור נכון של יכולות המערכת ומטרותיה; וחובת הסבריות של ההחלטה שמתקבלת על ידי המכונה כלפי מי שמושפעים ממנה.<sup>48</sup>

47 IEEE, לעיל ה"ש 10, בעמ' 11.

48 European Commission, לעיל ה"ש 10, בעמ' 15.

בהמשך נכללות גם הגדרות של "עקיבות" ו"הסברתיות" (explainability) שהוסברו לעיל, אבל מופרטת גם הדרישה הזאת:

• **תקשור** (communication) שמשמעה שכאשר מדובר במערכת בינה מלאכותית המתקשרת עם בני אדם יש חובה להצהיר על כך במפורש.<sup>49</sup> דרישה זו קשורה לגילוי נאות כלפי הצרכן ומתכתבת עם חקיקת הפרטיות האירופית,<sup>50</sup> הכוללת חובות גילוי נאות כאשר מערכת מקבלת החלטה אוטומטית בעניינו של הפרט.<sup>51</sup>

## 2. שקיפות כזכות של המשתמש

שלא כמו התפיסות המדגישות את החובות היצרניות, תפיסות אחרות מתייחסות לשקיפות ככלי להשגת אמון של המשתמשים במערכות, ולכן הן מדגישות את ההיבט של הסברתיות. למשל, ה־OECD מגדיר שקיפות כחובה לספק הסבר "משמעותי", בהתאם להקשר וברמה מקצועית גבוהה.<sup>52</sup>

בדומה, דיגנום רואה שקיפות באופן כפול – הן כחובת גילוי הנוגעת למידע שביסוד האלגוריתם ולשימוש במערכת והן כחובת גילוי לתהליך קבלת ההחלטה וליכולת להבין ולבקר תהליך זה: "שקיפות מציינת את היכולת לתאר, לבדוק ולשחזר את המנגנונים שבאמצעותם מערכות בינה מלאכותית מקבלות החלטות

49 שם, בעמ' 18.

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) 50

Article 22, Regulation (EU) 2016/679 on the GDPR 51

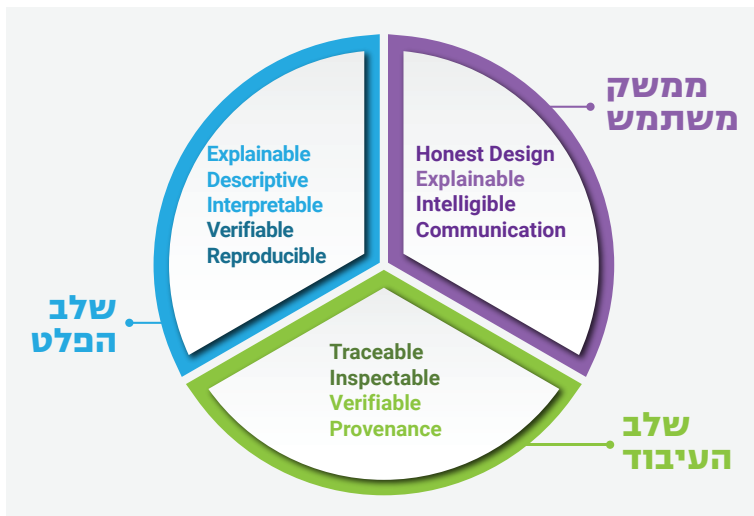
OECD, לעיל ה"ש 10, בעמ' 8, פסקה 1.3. 52

ולומדות להתאים את עצמן לסביבתן, ואת המקור והדינמיקה של הנתונים המשמשים את המערכת והמיוצרים על ידה".<sup>53</sup>

כלומר, דיגנום שמה את הדגש על חובות העברת מידע שיאפשרו למשתמשים לבדוק ולשחזר את דרך פעולתן של מערכות מבוססות בינה מלאכותית, וכן יאפשרו להם לאמת את המידע ואת הדרך שהתוכנה השתמשה בה.

**תרשים 1**

המשמעויות השונות שניתנו על ידי הכותבים שסקרנו למושג "שקיפות", לפי רצף פעולת התוכנה



\* בגוון הכהה יותר - פרוצדורה; בגוון הבהיר יותר - מהות.

התרשים מלמד שהמושג פתוח לפרשנויות רבות – כל פרשנות מדגישה מרכיב מסוים בשלבי פעילותו של האלגוריתם ומטילה בפועל חובות שונות על היצרן. ניתן לראות שללא קונקרטיזציה, השקיפות האלגוריתמית עשויה להיות רחבה מדי או צרה מדי ולא להתפרש על פני מכלול הפעילות של המערכת הלומדת.

### **3. שקיפות ככלי עצמאי או כחלק ממארג כלים**

ההבדלים בפרשנות למושג "שקיפות אלגוריתמית" תלויים גם בשאלה אם רואים בה ערך עצמאי או אמצעי להשגת מטרות אחרות, כגון אחריותיות. להבדלים אלו יש השלכות בכל הקשור לאיזון בין חובת השקיפות לבין אינטרסים אחרים שייתכן שיישומה נוגד – כגון זכויות קניין יצרניות, הגנת פרטיות ועוד.

את הראייה הרחבה ביותר של המושג "שקיפות" ניתן למצוא במסמך "Ethically Aligned Design (EAD) של ארגון התקינה הבינלאומי. במסמך נקבעים שלושה עקרונות אתיים מרכזיים שלפיהם תעוצב ותפעל כל מערכת מבוססת בינה מלאכותית: על המערכת – (1) לפעול בכפוף לזכויות האדם האוניברסליות; (2) לפעול בהתאם להקשר הפוליטי של המידע העומד לרשותה; (3) ולהיות "איתנה" מבחינה טכנולוגית (robust).<sup>54</sup> כדי לעמוד בעקרונות אלו של עיצוב מבוסס אתיקה וערכים מפרטים בארגון התקינה שמונה עקרונות פעולה, שהחמישי שבהם הוא שקיפות. שקיפות מוגדרת בנפרד מאחריותיות, אף שהיא כלי למימושה, בשל הקשרה הרחב יותר.<sup>55</sup>

גישה שבה לשקיפות יש מקום מרכזי, אך כחלק מהמושג "אחריותיות", ניתן למצוא במסמך המסכם של ועדת המומחים מטעם הנציבות האירופית שפורסם באפריל 2019.<sup>56</sup> במסמך זה ההצדקה לשקיפות מגיעה בעיקר מהצורך

54 Ethically Aligned Design, לעיל ה"ש 10, בעמ' 10.

55 שם, בעמ' 29.

56 European Commission, לעיל ה"ש 10.

ב"הסברתיות". המסמך מונה שלושה מרכיבים עיקריים אשר חייבים להתקיים לאורך כל שלבי הפעולה של מערכת מבוססת בינה מלאכותית: (1) הדרישה לחוקיות, היינו התאמה למערכת החקיקה והרגולציה הקיימת; (2) הדרישה שמערכות אלו יפעלו באופן אתי; (3) והדרישה שמערכות אלו יהיו איתנות (robust) הן ברמה הטכנולוגית והן בבחינת השפעותיהן החברתיות.<sup>57</sup>

לעומת זאת, לפי דיגנום שקיפות היא חלק ממכלול שהיא מכנה "Art of AI" – אמצעי אתי לעיצוב ההחלטות של מערכות בינה מלאכותית (ethics in design). אף שהיא מבחינה בין אחריותיות, הסברתיות ושקיפות, היא משתמשת בביטוי Art כמכלול וראה בשלושת הערכים מקשה אחת שצריכה להיות מיושמת בעת עיצוב המוצר ככלי הכרחי להבטחת עמידתו בחובות אתיות ובערכי החברה.<sup>58</sup> גם מורלי ואחרים רואים בשקיפות אמצעי להשגת אינטרסים של מועילות, אייצירת נזק, שמירה על האוטונומיה האנושית, צדק והסברתיות.<sup>59</sup>

בספרות ניתן לראות גם מקרי ביניים שהשקיפות מופיעה בהם כעיקרון ראשי, אך עם זאת עדיין בסמיכות לעקרונות אחרים ולא באופן עצמאי. במסמך של Tech Policy Institute, העוסק בשימוש באלגוריתמים בתחום הרווחה, השקיפות וההסברתיות מופיעות יחד כעקרונות ראשיים, מתוך תפיסת השקיפות ככלי לבקרה ולהגברת האמון וגם ככלי הכרחי לקיומו של הליך הוגן וצודק כשמדובר בתוכנות בינה מלאכותית.<sup>60</sup> אותו ערבוב של אחריותיות, הסברתיות ושקיפות ניתן למצוא גם במסמך ועדת המשנה של המיזם הלאומי למערכות נבונות, בכפילות באשר להגדרה של חובת השקיפות. מצד אחד, בתחילת המסמך השקיפות מוגדרת כערך נפרד, והוועדה מגדירה אותה במפורש כמכילה חובה של הנגשת מידע הן על תהליך יצירת המערכת הן על דרך קבלת ההחלטה על ידה. הוועדה

57 שם, בעמ' 9.

58 DIGNUM, לעיל ה"ש 9, בעמ' 52.

Jessica Morley, Anat Elhalal, Francesca Garcia, Libby Kinsey, 59  
 Jakob Mökander, & Luciano Floridi, *Ethics as a Service: A Pragmatic  
 Operationalisation of AI Ethics*, 31 (3) MINDS AND MACHINES 1 (2021)

60 תמיר, לעיל ה"ש 10, בעמ' 16.

אף מסבירה כי חוסר השקיפות, שהוא חלק ממאפייני הבינה המלאכותית, הוא גורם משמעותי ביצירת חוסר אמון במכונה, ומגדירה את הנושא כעניין שיש להציב לו פתרון. מצד שני, במסמך עצמו השקיפות יונקת את הנמקוטיה מהצורך ביצירת אחריותיות. כך, לפי הוועדה על מנת לייצר אחריותיות נדרשות שקיפות, הסברתיות, אחריות וניהול סיכונים.<sup>61</sup> חוסר הבהירות של מושג השקיפות נוגע לפיכך גם לתפיסות השונות לגבי מיקומה הגאומטרי של החובה עצמה, בין כערך עצמאי ובין כנגזרת של ערכים אחרים.

#### **4. איזון בין שקיפות לבין אינטרסים מתחרים**

ניתן ללמוד על הפרשנויות השונות למושג "שקיפות אלגוריתמית" גם מן הגישות השונות ביחס לאיזון שיש לקיים בין שקיפות לבין אינטרסים מתחרים, ובפרט האיזון בין שקיפות לבין יעילות וחדשנות ובין שקיפות לבין הסברתיות, במקרה שהן נמצאות בתחרות ביניהן. יש המאזנים בין חובות השקיפות לבין שמירה על יעילות ועל היכולת לממש חדשנות, בעוד אחרים מעניקים לשקיפות (ולהסברתיות) מעמד עליון.

במסמך של Tech Policy Institute נכתב כי בהינתן חוסר הבהירות המובנה של טכנולוגיית בינה מלאכותית, במקרה שתוכנה איננה מסוגלת לספק הסברים, יש להגביל את שימושה במקום שהחלטותיה משפיעות על זכויות של בני אדם.<sup>62</sup> לעומת זאת, במסמך מדיניות של תכנון תוכנות מבוססות בינה מלאכותית שהתפרסם באתר חברת IBM אין הכרה במגבלה בשימושים, אף שיש מערכות שמובהר כי איראפשר לייצר לגביהן שקיפות.<sup>63</sup>

61 ועדה משנה של המיזם הלאומי למערכות נבונות, לעיל ה"ש 10, בעמ' 4 ו-13. נציין עם זאת שהוועדה לא הגדירה שקיפות כפי שנעשה במסמכי מדיניות אחרים.

62 ראו תמיר, לעיל ה"ש 10, בעמ' 31.



הרציונל לשקיפות לתפיסת דיגנום הוא היותה כלי הכרחי שמאפשר תקשורת בין השחקנים המעורבים ביצירת המערכת. היא מתמקדת באחריותיות ובאחריות כמוקדים עיקריים, ולכן, בניגוד לוועדת המומחים הקובעת כי יש לאזן בין ביצועים לבין הסברתיות, טוענת כי במקומות שבהם ההסברתיות נפגעת יש להעדיף אינטרסים של שקיפות.<sup>64</sup> תקינת ה-IEEE נותנת לשקיפות מעמד עליון, ולעומתה ועדת המומחים של הנציבות האירופית מבטאת מודעות למגבלה הטכנית של מערכות בינה מלאכותית לספק הסברתיות באופן שיובן לבני אדם. לפיכך היא מציינת כי נדרש לשקלל ולאזן בין הצורך בהגברת ההסברתיות של המערכת על חשבון פגיעה בדיוק בשל הוספת מרכיב זה.<sup>65</sup>

ניתן לראות שקשת הפרשנויות למושג שקיפות מביעה תפיסות שונות בעניין האיזונים הנדרשים בין חובת השקיפות לבין חובות אחרות, כגון הסברתיות, וכן לגבי החובות המשתמעות מכך. כל אלה מחזקים את הטענה בדבר הצורך בעיגון פרשנות ברורה ומקובלת למושג.

## 5. הבעייתיות בהגדרת המושג "שקיפות"

### אלגוריתמית" כפי שהיא מתבטאת בהצעות חקיקה

העכירות שבמושג "שקיפות אלגוריתמית" באה לידי ביטוי בולט בשני מקרים שנעשה בהם ניסיון לפרוט את המושג לחובות חוקיות. להלן נתייחס לטיוטת הצעת חוק הבינה המלאכותית של הפרלמנט האירופי, שפורסמה באפריל 2021, וכן לחלק במסמך המדיניות של ארגון התקינה הבינלאומי העוסק בעיצוב אתי של אלגוריתמים.

64 DiGNUM, לעיל ה"ש 9. החרשים המופיע בעמ' 53 מסייע בהמחשת התפיסה המשולבת של שלושה מרכיבים אלו.

65 DiGNUM, שם, בעמ' 26 ו-33.

## א. טיוטת הצעת החוק של הפרלמנט האירופי

טיוטת הצעת החוק של הפרלמנט האירופי מאפריל 2021, שמטרתה ליצור מסגרת הרמונית לחקיקת בינה מלאכותית ולתיקונם של דברי חקיקה קיימים ברחבי האיחוד האירופי,<sup>66</sup> היא משום קונקרטיזציה של מסמכי המדיניות שאזכרו לעיל. הטיוטה בחרה בגישה של הטלת מגבלות רגולטוריות בהתאם לסוגי השימושים המיועדים למערכת המבוססת על בינה מלאכותית. לפיכך החקיקה מקטלגת את השימושים לכאלה שהיא מגדירה ב"סיכון גבוה", ואלה יהיו כפופים לרגולציה מחמירה ובין השאר לחובות רישוי,<sup>67</sup> ולכאלה שאינם זקוקים לרגולציה, וזו תחול עליהם באופן וולונטרי בלבד.<sup>68</sup> טיוטת החקיקה של האיחוד האירופי משקפת שקלול של ההגדרות השונות, והיא מתייחסת לחובות השקיפות בשלושה מישורים:

### **איזון בין שקיפות לבין זכויות אחרות**

סעיף 3.5 במבוא להצעת החוק קובע כי חובות השקיפות שיוטלו במסגרת טיוטת החקיקה המוצעת לא יפגעו בקניין הרוחני של משווקי התוכנה, אבל מכיר באפשרות שהן יפגעו בחירות העסקית ובזכויות של חופש אומנותי.

מאחר שסביר להניח שדרישות שקיפות כלשהן יצטרכו לבוא על חשבון קניין רוחני, מדובר באמירה בעייתית שמעדיפה, אפריורית, זכויות יוצרים על פני שקיפות. עם זאת, לא ברור שזו הייתה הכוונה המפורשת, וייתכן שהדבר יתקן בהמשך.<sup>69</sup>

66 European Parliament, לעיל ה"ש 12.

67 ראו שם, בעמ' 41.

68 שם, בעמ' 10.

69 על היחס באיחוד האירופי לזכויות קניין רוחני אגב פיתוח מערכות מבוססות בינה מלאכותית ראו דוח בנושא אשר אומץ על ידי פרלמנט האיחוד: European Parliament, *Intellectual property rights for the development of artificial intelligence technologies*, 2020/2015(INI) בסעיף 8 הדוח "stresses the importance of streaming services: מציינ כי האיחוד האירופי: being transparent and responsible in their use of algorithms, so that

## **רמות שונות של חובת שקיפות, לפי הסכנה הצפויה מהשימוש המיועד לתוכנה**

חובות השקיפות יחולו בשני מקרים:

- (1) מערכות המוגדרות ב"סיכון גבוה" – מערכות שמשפיעות על זכויות אדם, מצב פיננסי ועוד. בעניין מערכות אלו החקיקה מציינת במפורש כי המשתמשים צריכים להיות בעלי היכולת לפרש את הפלט של המערכת.
- (2) סעיף 5.2.4 להצעת החוק מציינ כי שקיפות מוגברת תחול על מערכות העונות להגדרות האלה: יש להן ממשק עם בני אדם; נועדו לאתר רגשות או תחושות; עלולות לבצע מניפולציה על החלטות אנושיות. במקרים כאלה תחול חובה ליידע את המשתמשים בדבר היכולות של המערכת.

### **היקף החובה והשלבים השונים שבהם יש ליישם אותה**

בסעיף 13 להצעת החוק מפורטות חובות התייעוד והיידוע, הכוללות את הרכיבים האלה:

- (1) זהות היצרן ופרטי קשר להשגתו.
- (2) מאפיינים ומגבלות של התוכנה והפעלתה, ובכלל זה:
  - מטרת התוכנה
  - רמת הדיוק
  - תוצאות צפויות
  - התפקוד ביחס לקבוצות המטרה
  - כאשר רלוונטי – מידע לגבי סוג המידע ואמינותו.
- (3) התאמות שבוצעו כדי לתת מענה לרמת הסיכון.
- (4) הבקרה האנושית העומדת לרשות מפעיל התוכנה.
- (5) תוחלת החיים של התוכנה והתחזוקה הנדרשת להפעלתה.

---

access to cultural and creative content in various forms and different languages as well as impartial access to European works can be better guaranteed"

## ב. מסמך המדיניות של ארגון התקינה הבינלאומי

עמדה מפורטת כחות אך מקיפה יותר בנקודות ההתייחסות שלה ניתן למצוא במסמך המדיניות של ארגון התקינה. לפי מסמך זה, חובות השקיפות צריכות להביא בחשבון הן את המשתמש הסופי, הן את היצרן והן את הרגולטורים. המסמך גם מבקש להשית חובות של העמדת מידע למי שיידרשו לחקור אירועים של תקלה, אם מטעם המדינה או מטעם היצרן. המסמך דורש לפחות שלוש רמות של שקיפות שיש לכלול בכל תוכנה:<sup>70</sup>

- (1) מהזווית של המשתמש – יצירת מעין "כפתור" שבלחיצתו התוכנה תספק הסבר מדוע בוצעה כל פעולה.
- (2) תיעוד של הליכי האשרור והעמדתם למפתחים אחרים וליצרנים לשם בדיקת המוצר.
- (3) יצירת "לוגים" ואספקת חיישנים שישמרו באופן מאובטח את המידע על פעילות התוכנה – לשם בחינת פעולתה בעת תקלה.

המסמך קורא ליצירת מדדים כמיתים לבחינת שקיפות המוצר, אך לא מפורטות בו הדרכים לקביעתם.<sup>71</sup>

## 6. העכירות של השקיפות האלגוריתמית – סיכום

### ביניים

כפי שראינו עד כה, לנוכח העלייה בשימושים במערכות המבוססות על בינה מלאכותית והחששות מפני תקלות בתפקודן, עולה הצורך ביצירת מערכת כללים מסודרת שבאמצעותה ייושמו עקרונות השקיפות בעולם האלגוריתמי. ואולם

70 Ethically Aligned Design, לעיל ה"ש 10, בעמ' 28.

71 שם, בעמ' 27.

על יצירת כללים אלו מקשה הגיוון הגדול בספרות באשר לעצם המשמעות של המושג "שקיפות אלגוריתמית", ובעקבות זאת באשר להדגשה של חובות שונות ובשילבים שונים של התהליך.

יש הסבורים כי מימוש השקיפות האלגוריתמית צריך להיעשות באמצעות הטלת חובות שנועדו לאפשר למשתמש שליטה והבנה במתרחש סביבו; ואילו אחרים שמים את הדגש על מבחן התוצאה ודורשים כי המשתמש ידע להעריך את איכות השירות שניתן לו על ידי התוכנה ויתאפשר לו להשיג על החלטותיה בעת הצורך. אחדים אף מרחיבים את חובת השקיפות האלגוריתמית כך שתכלול חובות גילוי צרכניות. כן ישנה אי־בהירות באשר להיקף החובות כאשר הטכנולוגיה נמצאת בשימוש במגזר הפרטי, בגופים דומהותיים, או כאשר מדובר בשימוש על ידי גוף ממשלתי כחלק מפעולותיו.

ישנה מחלוקת גם בנוגע לאיזון בין חובת השקיפות האלגוריתמית לבין חובות אחרות שיחולו על מוצרים מבוססי בינה מלאכותית. לכך יש השלכות על מקרים שיכולת הבקרה נפגעת בהם כאשר התוכנה איננה מסוגלת לספק מענה לדרישות המוצעות במסמכי המדיניות ובהצעות החקיקה, כמו בטיוטת ההצעה של הפרלמנט האירופי. במצב דברים זה המושג "שקיפות אלגוריתמית" נותר עכור, הן בהיעדר הגדרה ממצה למושג, הן בכל הנוגע להיקף תחולתו, והן בנוגע ליישום מעשי של הדרישות השונות של שלל הכותבים ובהתמודדות עם מקרים שבהם חסמים טכנולוגיים לא יוכלו לתת מענה לדרישות אלו. לא ברור גם מהן הדרישות מיצרן של מערכות אלגוריתמיות, ומה מעמדן של דרישות אלו במקרים שבהם הטכנולוגיה הקיימת לא תוכל לעמוד בהן – אם השימושים של טכנולוגיות אלו יהיו מוגבלים יותר ובמקרים מסוימים היצרנים אף ייאלצו לנטוש את פיתוחן.

## פרק 6

### המודל המוצע: מודל נפול - שקיפות ואיתנות מדעית

היתרונות בשימוש בטכנולוגיות מבוססות בינה מלאכותית אינם נתונים במחלוקת, בייחוד במקרים שהן יכולות להחליף בהם באופן יעיל שיטות קיימות או לחסוך הכשרה יקרה, זמן, הטיית ושגיאות. עם זאת, כדי להימנע מפגיעה במשתמשים כפרטים, ומתוך התחשבות בהקשרים חברתיים רחבים יותר, יש לוודא כי הליך הפיתוח והשימוש

בטכנולוגיות אלו ייעשה מתוך הקפדה על הגנה על זכויות אדם, ובראשן הזכות לשוויון, הזכות לפרטיות וחובות מתחום המשפט המינהלי.

כפי שתואר לעיל, אחת מן הערובות להגנה על זכויות אלו היא הטלת חובת השקיפות. אלא שבהקשר האלגוריתמי נתקלנו עד כה בשתי בעיות:

(1) מקרים שבהם לא ניתן לייצר שקיפות אלגוריתמית, שכינינו כאן מקרי "עכירות אלגוריתמית". מדובר בעיקר במערכות הפועלות כרשתות נוירונים, ויש להניח שבעתיד יהיו אתגרי שקיפות נוספים. אם תפיסת המוצא היא שגם במקרים של עכירות אלגוריתמית, ללא שקיפות במובנה הקלסי, מוצדק לפתח מערכות לומדות ולא לאסור לגמרי פיתוח כזה, נדרשים נתיבים חלופיים שיאפשרו הגנה על זכויות ואינטרסים.

(2) עצם ההגדרה של חובות השקיפות האלגוריתמית במסמכים שונים היא עמומה, ולכן אינה מאפשרת החלה מהותית של חובה זו על שחקנים שונים ואינה יוצרת ודאות לשחקנים בשוק. אופן היצירה של מערכת מבוססת בינה מלאכותית דורש כי יתקיימו הוראות ברורות בכל הקשור לאופן ניהולה, גיוון המידע הנדרש והליכי התייעוד כבר בשלבי איסוף המידע הראשוניים, ולא רק בשלבי ההמשך המאוחרים כאשר האלגוריתם כבר נוצר. ללא כללים ברורים, יוצרי אלגוריתמים יתקשו לחזות את הדרישות הצפויות מהם, דבר שיוביל לכשלים בחישוב העלויות ויפגע ביעילות המוצר הסופי.

בשנים הקרובות אנו צפויים להאצת הליכי החקיקה והרגולציה של מערכות מבוססות בינה מלאכותית – וניצנים לכך ניתן לראות באיחוד האירופי, בטיזת הצעת החוק שכבר הזכרה לעיל, וגם בסין,<sup>72</sup> בארצות הברית<sup>73</sup> ואף בישראל.<sup>74</sup> סביר להניח שחקיקה זו תכלול חובות שקיפות שיוטלו על השחקנים השונים – יצרנים, משווקים, מפעילים ומשתמשי קצה.

לאור המגרעות של המושג "שקיפות" שתוארו לעיל, אנו מציעים כאן כאמצעי בקרה מודל שקיפות המבוסס על התפיסה הקלסית של שקיפות, אבל שמציג, נוסף על חובות ההנמקה המסורתיות, חלופה המותאמת למגבלות הטכנולוגיות של מערכות אלגוריתמיות שאינן מסוגלות לספק הסבר רציונלי לפלט נקודתי. המודל מציע אפוא שני מסלולים חלופיים למימוש חובת השקיפות:

(1) מסלול ברירת המחלל ידרוש הסברתיות ובדיקה של התוכנה ויחייב את יצרן האלגוריתם לספק פלט שיוכל לנמק מדוע התקבלה החלטה מסוימת ולא אחרת.

(2) המסלול השני ידרוש הוכחה של "איתנות התהליך", כלומר שמהלך הייצור נעשה באופן מדעי מוכח, ומתוך הקפדה על מימוש ערכים שייבחרו. מהלך הייצור נוגע לשלושה מקטעים בשרשרת הייצור: החל במידע שמוכנס למערכת, דרך עיבוד המידע וכלה בשלב יישום האלגוריתם לצורך קבלת תוצר. מסלול זה מספק מענה לקושי הטכנולוגי של מערכות אלגוריתמיות מסוימות, שעל אף יעילותן הסטטיסטית אינן מסוגלות לספק פלט שעולה לכדי הסבר רציונלי. הוא מיועד למקרים שבהם אין אפשרות לספק פלט לפי המסלול הראשון, אבל עדיין יש נחיצות חברתית בשמירה על חובות שקיפות, כדי שלא למנוע פיתוח ושימוש בטכנולוגיות מסוימות. לכן הוא כולל חובת הוכחה שבניית התוכנה עומדת בקריטריונים מדעיים מחמירים ושיישומה הוא שוויוני.

China's New AI Governance Initiatives Shouldn't Be Ignored, CARNEGIE 72  
 ENDOWMENT FOR INTERNATIONAL PEACE

Alex Engler, The EU and U.S. Are Starting to Align on AI Regulation, 73  
 BROOKINGS, 1 Feb, 2022

74 דוח ועדת בינה מלאכותית ומדע הנתונים, פורום תל"ם, מרץ 2021.

על פי הצעתנו, המודל ישולב במסגרות חקיקתיות שיאמצו את הפרשנות למושג "שקיפות" ואת המסלולים החלופיים שבו. לתפיסתנו, גם בתקופת הביניים, עד שהחקיקה תנוסח, המודל יכול להעניק בהירות לגבי חובות השקיפות האתיות ולאפשר רגולציה עצמית.

## 1. רציונל המודל: המסלול השני משקף את הדין

### הקיים

המסלול השני במודל המוצע שואב השראה מהדין המינהלי, שבמסגרתו מדיניות תיחשב סבירה אם הדרך שבה התקבלה ההחלטה הייתה תקינה. בביקורת על החלטה מינהלית בית המשפט אינו בוחן "מה היה מחליט בית המשפט בנסיבות דומות, אלא אם החלטת המינהל עומדת בפני הביקורת, בהיותה החלטה שרשות מינהלית סבירה היתה יכולה לקבל".<sup>75</sup> בית המשפט אינו נכנס בנעליו של הגורם המקצועי מקבל ההחלטה, ואינו בוחן אם התקבלה ההחלטה הטובה ביותר בנסיבות העניין.

במקום זאת, כדי להצדיק התערבות, העותר נדרש להוכיח התקיימותן של עליות מוגדרות, המתמקדות בהליך קבלת ההחלטה ולא בטיב ההחלטה עצמה. בית המשפט בודק אם הובאו לפני הגורם מקבל ההחלטה כל העובדות, אם שקל את כל השיקולים, אם פעל באופן שוויוני ובתום לב, ואם נבחנו חלופות פוגעניות פחות כלפי האזרח. במקרה שהחלטת גורמי המקצוע מסתמכת על חוות דעת מקצועית מוסמכת, בית המשפט מהסס להתערב בהחלטה.<sup>76</sup>

75 בג"ץ 2920/94 אדם טבע ודין – אגודה ישראלית להגנת הסביבה, עמותה רשומה נ' המועצה הארצית לחכנון ולבנייה, פ"ד נ(3) 441, עמ' 16.

76 לפירוש על הביקורת בנוגע להחלטה המינהלית ראו דפנה ברק־ארז משפט מינהלי כרך ב, שער רביעי: "שיקול הדעת המינהלי" (2010).



מודל האיתנות המדעית-הליכית מציע גם הוא הליך בחינה הדומה למבחני המשפט המינהלי, הקובעים כי החלטה שהתקבלה בסמכות, אחרי שקילת השיקולים ומתוך הסתמכות על בחינת חלופות – תעמוד במבחן הביקורת.

אומנם קיימים הבדלים ברורים בין המקרים, אולם אותו עיקרון הבוחן את תקינות ההליך חלף בחינת תקינות התוכן של כל החלטה עומד גם ביסוד המודל המוצע כאן. תפיסת היסוד של הביקורת השיפוטית על ההליך המינהלי היא שהביקורת על תקינות הליך קבלת ההחלטה מאפשרת בקרה איכותית, המוודאת שמירה על זכויות האזרח ובלי להחליף את שיקול הדעת של אנשי המקצוע.<sup>77</sup> יתר על כן, המודל המוצע שואב השראה מאמירתו המפורסמת של השופט הולמס על המשפטן העתידי, שהוא סטטיסטיקן או כלכלן,<sup>78</sup> וכן מן המבחנים שהחילו בתי המשפט למשל על קבילותן של ראיות פורנזיות והדרישה שאלו יוצגו בידי מומחים מתאימים, על סמך מתודולוגיות מקובלות בעולם המדעי.<sup>79</sup> כפי שראיות טכנולוגיות המובאות לפני בית המשפט מוצגות בידי מומחים, כך במודל המוצע תוגדל המעורבות של סטטיסטיקאים, מהנדסי מחשב ומתמטיקאים, והם יהיו אלה שעליהם יוטל להראות כי האלגוריתם מתוכנן ופועל על פי סטנדרטים מדעיים מקובלים.

77 יש להפריד בין שאלת עצם הצורך במחן הסבר לבין הצורך בפישוט ההסבר כדי להחאימו לצרכנים שייחכן שיחקשו בהסבר מסובר מדי. אנו סבורים כי שאלה זו היא שאלה לדיון בשלב מאוחר יותר לאחר הבהרת עצם הצורך בקיומו של מודל המספק הסבר, או הסבר חלופי, ומדובר בעניין לבחינת הרגולטור הייעודי. על כך ראו Fabiana Di Porto & Nicoletta Rangone, *Cognitive-Based Regulation: New Challenges for Regulators?* RIVISTA DI DIRITTO PUBBLICO ITALIANO, COMUNITARIO E COMPARATO (forthcoming)

"For the rational study of the law the blackletter man may be the 78 man of the present, but the man of the future is the man of statistics and the master of economics." Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457 (1897)

Orin S. Kerr, "Digital Evidence and the New Criminal Procedure," 79 CYBERCRIME: DIGITAL COPS IN A NETWORKED ENVIRONMENT, 221-246 (Jack Balkin, James Grimmelman, Eddan Katz, Nimrod Kozlovski, Shlomit Wagman, & Tal Zarsky eds., 2007)

## 2. שלבי המודל: איתנות מדעית של האלגוריתם לפי מבחן משפטי-סטטיסטי-מתמטי משולב

המבחן המוצע במסלול השני של המודל יחול לאורך כל שלבי פיתוח התוכנה – החל בבחינת סף משפטית של הגדרת מטרות התוכנה, דרך בחינה סטטיסטית של איכות הנתונים שהחזנו למאגר שהיא מושתתת עליו והתאמתם ליעדי התוכנה, וכלה בבדיקה מתמטית או אלגוריתמית בכל הנוגע לעיבוד נתונים אלו לשם ייצור הפלט. בשורות הבאות נפרט את שלושת שלבי התהליך המוצע, שהעמידה בהם תשמש חותמת לתקינות התוכנה ותוכל להיות מוצגת במקום נימוק רציונלי לכל פלט בנפרד.

### א. מבחן הסף המשפטי – מטרות התוכנה

מבחן הסף המשפטי יעסוק בפירוט של פונקציית המטרה של התוכנה, ובבחינה אם מטרה זו היא חוקית בנסיבות שבהן יש כוונה להשתמש בה. לדוגמה, במקרה של פיתוח אלגוריתם שתכליתו לזהות צרכנים הנוטים לקבלת החלטות רכישה פזיזות כדי לשווק להם מוצרים – קיים ספק בנוגע לחוקיות המטרה, גם אם יוכחו יעילותו של האלגוריתם ויישומו באופן שוויוני. הטלת חובה לפרט מהי פונקציית המטרה תאפשר לנהל דיון אם המערכת עומדת בדרישות החוק, גם כאשר מטרתה היא הגברת יעילות בלבד.<sup>80</sup> נוסף על כך, הדבר יאפשר דיון ציבורי בקביעת סדרי העדיפויות של מטרות המערכות בהתחשב גם בעקרונות חברתיים, בדומה למשל לביקורת שהושמעה על ידי דווח מועצת זכויות האדם של האו"ם בדבר הדגשת יתר של איתור הונאות בתשלומי רווחה על חשבון שיפור הנגישות לתושבים הנזקקים לתשלומים אלה.<sup>81</sup>

80 תיחכן טענה שפירוט מטרות התוכנה הוא כשלעצמו סוד מסחרי. טענה זו תהיה מוגבלת במקרים שבהם בדיקת מטרות התוכנה נועדה לברר את חוקיותה.

81 ראו הערת הדווח לאו"ם בנוגע לזכויות אדם – Alston, לעיל ה"ש 2.

## ב. איתנות סטטיסטית

בשלב השני יהיה על יצרן האלגוריתם או המוצר להוכיח כי מאגר הנתונים שעליו מבוססות החלטות הסטטיסטיות של התוכנה עומד בדרישות שייקבעו בהתאם לשימוש המיועד. כך למשל יהיה עליו להוכיח כי מאגר הנתונים מתאים לקבוצת היעד שעליה הוא מתכוון לפעול – הן בגודל המדגם והן בהתאמת האוכלוסייה ומאפייניה. נוסף על כך, יהיה עליו להוכיח כי המודל נקי מהטיות שיש בהן כדי לפגוע בחוקיותו – כגון מידע שעלול להביא לתוצאות שונות בקרב קהלי יעד שונים, על בסיס גזע או מין, גם אם אלה לא מופיעים במפורש בתוך המאגר. שלב זה ישפיע על כל מהלך ייצור האלגוריתמים, שכן הוא יחייב הכללת מומחי תוכן מטעם היצרן בכל השלבים של ניהול הליך ייצור התוכנה. כן הוא יחייב את יצרני התוכנה לתעד את כל הפעולות ששימשו להזנת התוכנה, מאחר שמומחים יצטרכו לבסס את חוות הדעותיהם על מידע על הנתונים שהוכנסו למאגר. דבר זה יחייב את היצרנים המעוניינים לשווק את מוצריהם לשימושים שפוטנציאל הפגיעה בזכויות שלהם גבוה להקפיד, כבר בשלבים הראשונים, על עבודה מדעית מסודרת.

## ג. התאמת האלגוריתם למטרות שהוגדרו

בשלב השלישי יהיה על היצרן להוכיח כי האלגוריתם המיושם בפועל מתאים למשימה שהוא נדרש לבצע, הן ברמה התאורטית והן ביישום בפועל. בשלב זה יהיה על היצרן לספק מידע בנוגע למגבלות של יכולת הניבוי והקטגוריזציה של האלגוריתם בשימוש המיועד, ובכלל זה מידע על ה־false positive ו־false negative<sup>82</sup>, וכן לגבי אורך חיי התוכנה והצורך לעדכן אותה מעת לעת.<sup>83</sup> דרישה

82 דוגמה למחקר מקיף שבדק, בדיעבד, מהימנות של חוכנות לבדיקת אלגוריתמים לזיהוי פנים הוא מחקר של מכון החקנים בארצות הברית. הובאו בחשבון זיהויים שגויים – הן ככשל לזוהות דמיון והן בזיהוי דמיון היכן שלא היה. ראו Patrick Grother, Austin Hom, Mei Ngan, & Kayee Hanaoka, "Face Recognition Vendor Test (FRVT) Part 7: Identification for Paperless Travel and Immigration," NIST Interagency/Internal Report (NISTIR) – 8381

83 לעניין חובות עדכון התוכנה ראו בסעיף 13 בטיוטת החקיקה האירופית המוצעת (לעיל ה"ש 12).

זו דומה לדרישת ההתוויה הנהוגה כיום בעולם אישור התרופות, הקובעת תהליך וידוא שהתרופה מתאימה לא רק לצורך השימוש שלו אושרה לכתחילה, אלא בכל תחום אחר שבו יש כוונה להשתמש בה. היצרן אף יצטרך לדווח לרשויות הרלוונטיות ולשקף למשתמשים עצמם את מגבלות האלגוריתם בכל הקשור לניבוי ולקטגוריזציה.

### 3. בחירה בין שני מסלולי השקיפות: על פי מבחן ההשפעה הפוטנציאלית על הזכויות

---

עד כה סקרנו את מרכיבי המודל המוצע, שכאמור מורכב משני מסלולים חליפיים. הבחירה בין מסלולי השקיפות (המטילים חובת שקיפות בעוצמות שונות) תהיה על פי קריטריון ההשפעה הפוטנציאלית של האלגוריתם והמוצר על זכויות חוקתיות.

(1) כאשר הפגיעה הפוטנציאלית חמורה, שקיפות תתפרש כחובה לספק הנמקה רציונלית לבחירה הספציפית של האלגוריתם. כך למשל, במקרים של הרשעות לפי הדין הפלילי ובהחלטות שיש בהן כדי לפגוע באופן מהותי בחופש התנועה – כגון הוראות בידוד, צווי עיכוב יציאה מהארץ ועוד – תידרש הנמקה באופן זהה לזה שהיה מקובל עד כה במערכת המשפט, ללא התחשבות בכך שהגורם מייצר ההחלטה הוא אלגוריתם. מנגד, תוכנות להחלטות לגבי מתן אשראי יוכלו להשתמש במסלול החליפי, וכך גם תוכנות העוסקות במחיקת פרסומים ברשתות החברתיות. תוכנות מסחריות שנועדו לאפשר חוויית לקוח או תוכנות פיננסיות שאינן עוסקות במתן זכויות אשראי, כגון המלצות משכנתה, לא יחויבו בחובות שקיפות לכתחילה (אך עמידתן במבחנים תהיה תנאי בדיעבד, לשם בדיקה של עמידת המשווק בחובות לפי דיני הנזיקין בעת סכסוך משפטי). בדומה, אלגוריתמים שיש להם פוטנציאל לפגיעה פיזית בבני אדם יצטרכו לעמוד בתנאי שקיפות מחמירים. אבל מרגע שהיצרן יוכיח כי הסיכוי לפגיעה פיזית אגב שימוש במוצר נמוך מאוד, הוא יוכל לפעול לפי המסלול החליפי.

(2) תובא בחשבון גם זהותו של מי שיפעיל את התוכנה – אם מדובר בגוף שלטוני ומדינותי, גוף דו־מהותי<sup>84</sup> או גוף פרטי. במקרה שהשימוש בתוכנה הוא מעין פעולה שלטונית, ההשפעה החוקתית תיחשב חמורה יותר, והמבחנים יוחמרו בהתאם. ככל שהפגיעה הפוטנציאלית תפחת, כך תתאפשר הבחירה במסלול החליפי של איתנות מדעית. כך למשל, כשמדובר בזכויות בנייה ובזכויות פיננסיות כגון מתן אשראי.

להתמקדות בעוצמה הפוטנציאלית של הפגיעה בזכות חוקתית כמדד לבחירה בין המסלולים יש שני יתרונות. ראשית, בדיקה פרטנית של הפגיעה הפוטנציאלית של האלגוריתם מאפשרת גמישות ומונעת ניצול לרעה של לקונות רגולטוריות שנוצרות בדרך כלל בממשק שבין משפט וטכנולוגיה. כך למשל, הדבר יאפשר הימנעות מחסר רגולטורי כפי שהתקיים במשך שנים בתחום הרשתות החברתיות,<sup>85</sup> עד להתערבות שאירעה בשלב מאוחר ולאחר שנוצר נזק.<sup>86</sup> שנית, התמקדות זו עדיפה ממדדים אחרים כגון הבחנה על בסיס ענף במשק או על בסיס התמקדות בזכות המוגנת בלבד. התמקדות משקית עלולה ליצור עיוותים רגולטוריים במקרים שבהם לאותו אלגוריתם ישנם שני שימושים פוטנציאליים, כל אחד בשוק אחר.

בתחום הרפואי, למשל, תחילה יחולו החובות המוטלות על ידי הרגולטורים העוסקים בהבטחת איכות של מוצרים בתחום זה. אלו ממילא נוטים להיות

84 בשלב זה הכוונה היא לגופים דו־מהותיים, כפי שהם מוגדרים בדין הישראלי. ראו ע"א 91-294 חברת קדישא גחש"א "קהילת ירושלים" נ' ליונל אריה קסטנבאום, (מו"ר) 464 (1992). כיום מתנהל ויכוח בעניין מעמדן של רשתות חברתיות לנוכח מקומן כזירה חברתית לחלופת רעיונות. על כך ראו נועה מור "רשתות חברתיות מקוונות כזירות לעיצוב זכויות והקצאתן: לקראת החלתן של חובות מן המשפט הציבורי" *דין ודברים* יד (2019).

85 על התפתחות הרגולציה, למשל במקרה של פייסבוק, ניתן לראות במאמרו של רוטם מדזיני – Rotem Medzini, *Enhanced Self-Regulation: The Case of Facebook's Content Governance*, *NEW MEDIA & SOCIETY* (February 2021)

86 על היקפי הנזקים בעקבות שיח פוגעני ראו למשל Mainack Mondal, Leandro Araujo Silva, & Fabricio Benevenuto, *A Measurement Study of Hate Speech in Social Media*, *PROCEEDINGS OF THE 28TH ACM CONFERENCE ON HYPERTEXT AND SOCIAL MEDIA* 85-94 (July 2017)

מחמירים יותר לנוכח ההיסטוריה של התפתחות הרגולציה בתחום. בשלב השני יהיה ניתן להבחין בין אלגוריתמים שיש להם השפעה משמעותית על זכויות, כגון אלגוריתמים שימשו לאבחון מחלות, לבין אלגוריתמים שהשימוש בהם עלול ליצור פגיעה פוטנציאלית מחמירה פחות, כגון אלגוריתמים למיקוד המלצות לרפואה משלימה.

נוסף על כך, ההתמקדות בזכות החוקתית ללא התחשבות בעוצמת הפגיעה עלולה להביא לכשל רגולטורי וליצירת חסמים לתוכנות מסחריות שהנזק מהן מינימלי. כך למשל, תוכנה לזיהוי פנים המשמשת רשות אכיפת חוק לעומת תוכנה דומה המשמשת באפליקציה חינוכית להתאמת איפור. ההתמקדות בזכות המוגנת – כגון פגיעה בכבוד האדם בשל אפליה פוטנציאלית בתוכנת זיהוי פנים – תחייב יצרנים לייצר פתרונות טכנולוגיים שאין להם הצדקה כלכלית או חוקתית בהשוואה למקרים שבהם פעולה מתבצעת בעולם הלא-אלגוריתמי. בפתרון המוצע כאן, אנו שואבים השראה משיח הזכויות החוקתיות, שבו ככל שהפגיעה החוקתית חמורה יותר, ונוגעת יותר לליבת הזכויות, כך הדרישות לקשר בין האמצעי שנבחר למידת הפוגעניות שלו הופכות מחמירות יותר.<sup>87</sup>

#### **4. היקף חובת התייעוד במסלול האיתנות התהליכית**

מסלול האיתנות התהליכית המדעית יכלול חובה לייצר תיעוד של ההליך וגם לשקף מידע זה למשתמשים. היקף חובת התייעוד שיידרש וכן המועד שבו יצטרך היצרן לספק את המידע בעניין התייעוד – טרם השיווק של מוצר או לאחריו – ישתנו בהתאם לנסיבות העניין, על פי השוק הרלוונטי וקריטריון עוצמת הפגיעה בזכות החוקתית.

87 אהרון ברק פרשנות במשפט כרך שלישי: פרשנות חוקתית, שער ה', 708 (תשנ"ד).

מאחר שאין כוונתנו ליצור חלופה למערכות בקרה מגזריות קיימות, אנו מציעים מערכת הטמעה דו־שלבית של המודל. היקף חובות התייעוד יהיה כפוף תחילה לחובות הקיימות ממילא בשוק הרלוונטי, במקרה שאלו מחמירות יותר. כך בנוגע לחובות פיקוח בנקאיות, חובות רישוי העוסקות בכלי ירייה, תרופות ועוד. בשלב השני יותנה היקף חובות התייעוד בעוצמת הפגיעה הפוטנציאלית בזכויות ויתקיים הן לגבי מועד תחולת השקיפות והן לגבי עוצמת המבחנים המדעיים. ככל שהפגיעה הפוטנציאלית בזכויות תהיה משמעותית יותר, המבחנים המדעיים יהיו מחמירים יותר, וחובת אספקת המידע תחול מראש ותהיה תנאי לשיווק האלגוריתם. במקרים שבהם הפגיעה תקטן, ניתן יהיה לאפשר את אספקת חובת התייעוד רק במקרה של סכסוך משפטי, והמבחנים שיושנו יותאמו לעוצמת הפגיעה.

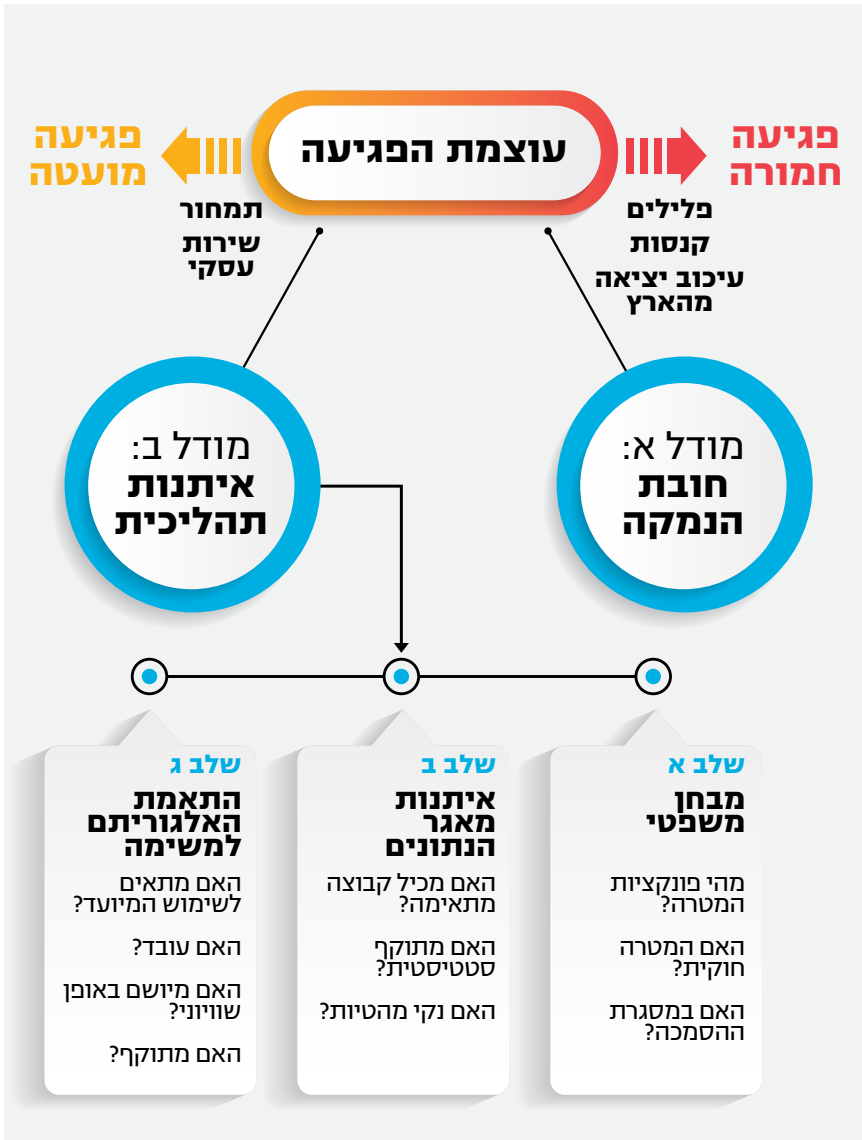
יובהר כי לשם הגנה על סודות מסחריים, היצרן לא בהכרח יצטרך לספק את הנתונים עצמם לכל משתמש, וניתן יהיה להסתפק, על פי בחינה שוקית, בחשיפת החומרים לצד שלישי מוסכם או לרגולטור, על מנת שזה יאשר מראש את העמידה בתנאים.<sup>88</sup>

88 סוגיית הקניין הרוחני תופסת מקום מרכזי בטיוטת החקיקה האירופית המוצעת (לעיל ה"ש 12). סעיף 3.5 להקדמה מציין במפורש:

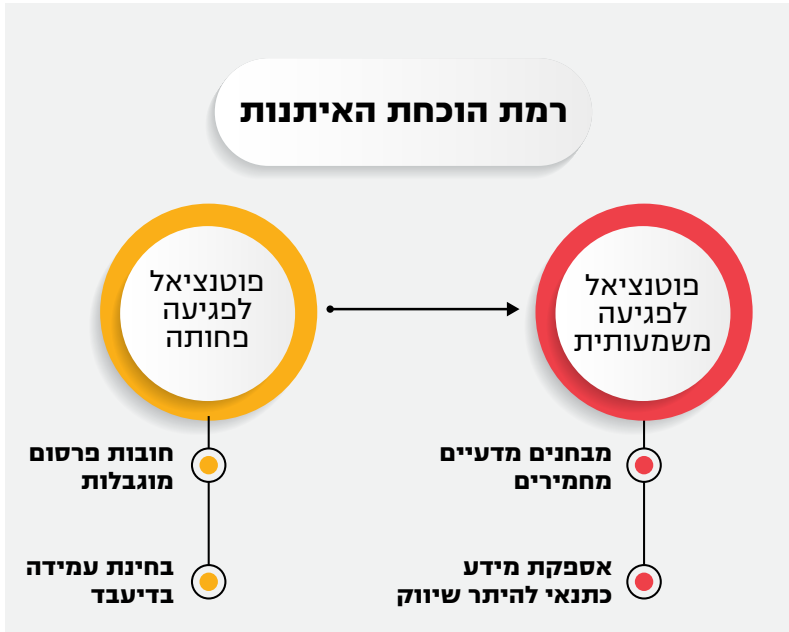
The increased transparency obligations will also not disproportionately affect the right to protection of intellectual property (Article 17(2)), since they will be limited only to the minimum necessary information for individuals to exercise their right to an effective remedy and to the necessary transparency towards supervision and enforcement authorities, in line with their mandates. Any disclosure of information will be carried out in compliance with relevant legislation in the field, including Directive 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure. When public authorities and notified bodies need to be given access to confidential information or source code to examine compliance with substantial obligations, they are placed under binding confidentiality obligations.

2 חרשים

שני המודלים המוצעים ליישום שקיפות אלגוריתמית







## פרק 7

### היתרונות והחסרונות של המודל המוצע

לתפיסתנו, המודל המוצע משקף איזון ראוי בין הגנה על ערכים זכויות לבין הבטחת חדשנות ויעילות בפיתוח טכנולוגיות. המודל נותן מענה לקשיים של טכנולוגיות מבוססות בינה מלאכותית לספק שקיפות והסבר רציונלי ובד בבד עומד על עקרונות חשובים של הגברת מעורבות ויצירת בקרה. המודל יסייע ביצירת הוודאות הדרושה למפתחים מתוך שמירה על פרטיות ועל גמישות.

#### 1. יתרונות המודל המוצע

##### א. מענה למקרים שבהם לא ניתן לספק הסבר רציונלי

המודל נותן מענה לבעיה טכנולוגית משמעותית של היעדר יכולתן של טכנולוגיות אלגוריתמיות מסוימות לספק הסבר רציונלי נקודתי לכל החלטה. היעדר יכולת זו עשוי להביא למניעת השימוש בטכנולוגיה, גם במקרים שבהם היא מסוגלת לספק מענה טוב יותר, הן בהגנה על זכויות והן בכל הנוגע לעלות הכלכלית, אל מול הפתרונות הקיימים.

##### ב. הגדרה בהירה ושימה לשקיפות

המודל נותן מענה לקשיי היישום שהיו עד עתה בכל הקשור לחובות השקיפות, כפי שהופיעו בהצעות המגוונות של כותבים שונים. הפתרון מציע צעדים ברורים בכל הקשור לחובות בנוגע לאיסוף המידע, השימוש בו והפלט שיווצר על ידי התוכנה. הדרישות בכל אחד משלושת השלבים ניתנות לתרגום לשפת מעשה ולמומחים שיידרשו בכל שלב.

##### ג. סיוע ליצירת ודאות

הצעדים הנדרשים במודל ניתנים לצפייה מראש, דבר המסייע ביצירת ודאות אצל היצרנים. מצד אחד, המודל מחיל חובות מחמירות בכל הקשור למקרים

שבהם קיימת פגיעה פוטנציאלית משמעותית, אך מצד שני מונע העמסת עלויות מיותרות במקרים שבהם השימושים הצפויים אינם בעייתיים. היצרן יוכל להתאים את העלויות ולפתח מוצר בהתאם.

#### **ד. המודל משקף תפיסות של Ethics by Design**

המודל משקף תפיסות של Ethics By Design הדורשות מעורבות של שיקולים אתיים וחוקיים כבר בשלב עיצוב המוצר ולא רק בשלבים המאוחרים של שיווק.<sup>89</sup> הכנסת שיקולים של אתיקה כבר בשלב העיצוב מגבירה את היכולת של היצרנים להכניס את השינויים הנדרשים כאשר העלות נמוכה, ולא באמצעות תיקונים יקרים בדיעבד.

#### **ה. גמישות**

המודל המוצע גמיש בהטלת החובות ובהתאמתן לשוק ולמוצר. כאשר מוצר אינו פוגע בזכויות, הדרישות נמוכות יותר, ולפיכך עלויות הפיתוח שלו זולות יותר.

#### **ו. יצירת בקרה יעילה וקידום אמון**

המודל נותן מענה הן לצורכי המשתמשים והן לצורכי היצרנים. הוא מעניק למשתמשים במערכות אלגוריתמיות את הידיעה שעל היצרנים הוחלו חובות שמטרתן להגן על זכויותיהם, גם במקרים שהאלגוריתם אינו מספק להם הנמקה נקודתית, או שאין להם היכולת הפרקטית להתחקות אחר כל החלטה שהתקבלה בעניינם. מבחינת היצרנים, המודל מעניק ודאות שאינה קיימת כיום. במקום בחינה בדיעבד אם התוכנות שלהם עומדות במבחני הפטיקה, הם יכולים לעגן את הדרישות המוטלות עליהם בבדיקות מדעיות, ולפיכך מתאפשר להם לתמחר מראש בהתאם את המוצרים ולהביא ליעילות כלכלית.

89 ראו למשל Caroline Whitbeck, *Ethics as Design: Doing Justice to Moral Problems*, 26 (3) HASTINGS CENTER REPORT, 9 (1966)

## 2. חסרונות המודל המוצע

יש לתת את הדעת גם לכמה חסרונות של המודל. ראשית, המודל מנסה לתת מענה לקשיים טכנולוגיים הנובעים ממערכות אלגוריתמיות הנמצאות בשימוש כעת,<sup>90</sup> ודבר זה משפיע על אורך החיים הפוטנציאלי שלו. היה והטכנולוגיה תשתנה ויפותחו מערכות בינה מלאכותית שמסוגלת לתת הסבר רציונלי להחלטות שלהן, מסלול של איתנות תהליכית יהפוך לאות מתה. עם זאת, כרגע נראה שמדובר במודל שנותן מענה לפיתוחים טכנולוגיים שסביר שיופיעו בעתיד, והסיכון שתפותח טכנולוגיה חדשה אינו מצדיק את השארת המצב ללא פתרון.

שנית, הפטור מהחובה לספק הסבר רציונלי עם יצירת קיומו של מסלול איתנות תהליכית עלול לבטל את התמריץ לפיתוח טכנולוגיות הנמקה חדשות – גם אם פיתוח כזה אפשרי. דבר זה עלול להנציח חיסרון טכנולוגי, במקום לעודד את הפיתוח של הפתרון. כך למשל, אכיפת חובות בטיחות כלפי יצרני הרכב הן שעודדו את היצרנים לפתח את הטכנולוגיות באופן יעיל יותר, דבר שספק אם היה מתרחש אילולא הוטלו חובות אלו מלכתחילה.<sup>91</sup> נציין שמדובר בהשערה תאורטית, ויש המטילים ספק בכך שהחובות הרגולטוריות הן אלו שאכן העלו את רמת הבטיחות ברכבים.<sup>92</sup> בהשאלה למקרה שלנו, דווקא יצירת ההליך החלופי של איתנות מדעית-תהליכית תשפר את ההגנה על זכויות.

שלישית, המודל מחייב הכרעה בירוקרטית שעשויה לגרור אחריה ביקורת שיפוטית, על כל העלויות הכרוכות בכך. מלבד זאת, בדומה לכל אכיפה

Bert-Jaap Koops, *Should ICT Regulation Be Technology-Neutral?*, 90 STARTING POINTS FOR ICT REGULATION. DECONSTRUCTING PREVALENT POLICY ONE-LINERS, IT & LAW SERIES, Vol. 9, 77 (Bert-Jaap Koops, Miriam Lips, Corien Prins & Maurice Schellekens, eds., 2006)

Christopher Jensen, *50 Years Ago, "Unsafe at Any Speed" Shook the Auto World*, THE NEW YORK TIMES (Nov. 26, 2015) 91

Sam Peltzman, *The Effects of Automobile Safety Regulation*, 83 (4) JOURNAL OF POLITICAL ECONOMY, 677 (1975) 92

אלגוריתמית, יהיה צורך להתאים את מנגנוני הדיווח כך שיאפשרו שיח בין גורמי הטכנולוגיה לבין הרגולטורים על בסיס יכולות חישוביות ודיווחים אוטומטיים. הדבר גם ידרוש אוריינות דיגיטלית ממערך האכיפה והשופטים שיעסקו בנושא. עם זאת, לתפיסתנו, היציבות הצפויה בעקבות יישום המודל המוצע והבהירות שהמודל מציע באשר לחובות שיוטלו על השחקנים השונים – יביאו לתועלות חברתיות הגבוהות בהרבה מעלויות הפיקוח הנדרשות לשם יישום המודל.

## פרק 8

### המודל המוצע ויישומו

לאחר שהוצגו עקרונות המודל, נציג שני מקרי בוחן: האחד של שימוש מדינתי באלגוריתם, והשני של ממשק בין חברות פרטיות לבין המדינות בכל הנוגע לפיקוח על פיתוח טכנולוגיות של נהיגה אוטונומית. באמצעות מקרי בוחן אלו

ניתן לראות הן את הנזקים הנוצרים בשל היעדר שקיפות אלגוריתמית בפועל והן את הקושי בפרסום מדד ברור מראש ליצירת שקיפות. ננסה להראות גם כיצד המודל המוצע לאיתנות תהליכית יכול היה לפתור קשיים אלו.

#### 1. איתור הונאות במערכת הרווחה בהולנד

המחשה של הצורך בהוראות מפורשות לגבי שקיפות אלגוריתמית ושל ההשלכות השליליות של היעדרה ניתן למצוא בעניין מערכת "סירי" (Systeme Risico Indicatie – SyRI) של ממשלת הולנד לאיתור הונאות רווחה.<sup>93</sup>

סירי היא מערכת משותפת של כמה גופים ממשלתיים בהולנד שמטרתה איסוף מידע והעברתו למשרד המידע ההולנדי. שמות אנשים ואמצעים מזהים אחרים הוחלפו במערכת במספרי קוד (המפתח לקוד נשמר במיקום אחר). לאחר הצלבת המידע באמצעות אלגוריתמים, שדרך פעולתם לא פורטה לציבור, לפרלמנט או לבתי המשפט, המערכת הפיקה פלט אשר קבע כי עלה בידיה לאתר חשד להונאה אצל אדם כלשהו. ההתראה של המערכת שימשה בסיס לפתיחת חקירה מצד משרד הרווחה ההולנדי.

Max Vetzo, *The Netherlands – Algorithmic Fraud Detection System Violates Human Rights – The Case of SyRI*, 3 PUBLIC LAW, 650 (2021)

ברקע להטמעת מערכת זו בהולנד עמדו פעולותיהן של ממשלות במדינות כגון בריטניה ואוסטרליה אשר הטמיעו מערכות אלגוריתמיות שמטרתן איתור אוטומטי של הונאות בתשלומי הרווחה הניתנים לאזרחים. על מערכות אלו נמתחה ביקורת, בעיקר מכיוונם של ארגוני זכויות אדם, ובפרט מצד דווח האו"ם של מועצת זכויות האדם, אשר כתב על כך דוח מיוחד והעלה את הבעייתיות שבאיסוף הנרחב של המידע, ובשימת הדגש של מערכות אלו על איתור הונאות חלף סיוע לאזרחים במימוש זכויות הרווחה שלהם.

עניין פעולתה של המערכת ההולנדית הגיע לבית המשפט העליון בהולנד, אשר קבע כי השימוש במערכת אלגוריתמית זו הוא משום הפרה של הזכות לפרטיות המעוגנת באמנה האירופית לזכויות אדם.<sup>94</sup> כפי שמעירים ון־בקום וזיודרווין בורגזיוס (van Bekkum & Zuiderveen Borgesius)<sup>95</sup> בפסק הדין מוזכרים שלושה מאפיינים של היעדר שקיפות בדרך פעולתה של התוכנה. ראשית, במודל להערכת סיכון ההונאה שהמערכת מתבססת עליו; שנית, בדרך פעולת האלגוריתם בפועל; ושלישית, בסוגי ההצלבות המבוצעות בין מאגרי הנתונים השונים שתוכנן הועבר לידי משרד המידע.

אף שהשימוש בתוכנת סירי העלה חשד, עמדת בית המשפט ההולנדי הייתה כי די בהמלצה זו של המערכת כדי להוביל לשינוי במעמד המשפטי של אדם. לכן, סבר בית המשפט שעל המערכת צריכות לחול חובות מינהליות וחובת כיבוד זכויות אדם. בהמשך, בית המשפט קבע כאמור כי המערכת מפירה את הוראות הגנת הפרטיות של האמנה האירופית ושם דגש על שאלת הנחיצות והמידתיות בפעולתה. טיעון היעדר השקיפות עמד בבסיס קביעת בית המשפט בדבר היעדר המידתיות: "החקיקה של SyRI אינה שקופה דיה וניתנת לאימות כך שמתאפשר להסיק שההפרה של הזכות לפרטיות שהשימוש ב־SyRI עשוי לגרום היא הכרחית,

NJCM et al. v. The Dutch State (2020). The Hague District Court ECLI: 94  
NL: RBDHA:2020:1878 (SyRI)

Marvin van Bekkum & Frederik Zuiderveen Borgesius, *Digital Welfare* 95  
*Fraud Detection and the Dutch SyRI Judgment*, 23 (4) EUROPEAN JOURNAL OF  
SOCIAL SECURITY, 323 (2021)

מידתית ומידתית ביחס למטרות שהחקיקה חותרת אליה".<sup>96</sup> ובהתאם לכך קבע כי: "חקיקת SyRI אינה מספקת בשום אופן מידע על [...] אילו נתונים עובדתיים אובייקטיביים יכולים להוביל בצדק למסקנה שקיים סיכון מוגבר".

היעדר השקיפות הוביל את בית המשפט לקבוע גם שיש חשש של ממש כי המערכת הפלגתה אנשים על בסיס מעמד חברתי-כלכלי. בהיעדר מידע עובדתי בדבר דרך פעולת המכונה, בית המשפט הסתמך על נתונים בפועל שהיא הופעלה באזורים ממעמד חברתי-כלכלי נמוך כאינדיקציה מספקת לחשש מאפליה בהפעלת התוכנה.

אף שהיעדרה של שקיפות כשלעצמו לא היה עילה לפסילת המערכת, התעקשותה של ממשלת הולנד לשמור בסודיות את כל הפרטים של פעולת התוכנה לא רק שפגעה באמון הציבור, אלא אף בעצם החוקיות של התוכנה. בהקשר שלנו, ממשלת הולנד לא פעלה להוכיח את תקינות האלגוריתם ולא את דרך פעולתו.

לפני הממשלה ההולנדית, אשר הפעילה טכנולוגיה שלא הייתה מסוגלת לספק הסבר על פעילותה, לא עמדה חלופה חוקית ליצירת מסגרת לשקיפות באופן שהיה נותן מענה לדרישות הבקרה. ניסיונותיה להסתיר את דרכי הפעולה של התוכנה הסתיימו בפסילתה. אימוץ המודל המוצע על ידינו היה מאפשר לה לפעול על פי החלופה של פרסום הדרך של פעולת האלגוריתם; ולחלופין, אם חששה מפגיעה במערכת, קבלת אישור של מומחה צד ג' באשר לפעולתו. גם ללא פרסום הנתונים עצמם, ניתן להניח כי פרסום רשמי של מומחה, שהיה מאשר כי המאגר המזין את המערכת אמין ואינו מוטה וכי המערכת לניהול הסיכונים עומדת במבחנים מדעיים מחמירים, היה מניח את דעתו של בית המשפט כי המערכת נחוצה ומידתית לפגיעה שהיא יוצרת.



## 2. היעדר שקיפות בקביעת רף הבטיחות הנדרש ברכבים אוטונומיים

הטיפול של מדינות המערב בדרכי פעולתן של מערכות לנהיגה אוטונומיות למניעת תאונות דרכים אינו מתנהל על פי תוכנית סדורה וברורה בכל הקשור למידע הדרוש לשם קביעת רף הבטיחות לרכבים מסוג זה.<sup>97</sup> היעדר נורמות בנוגע לדרך שבה ישוקף מידע זה לציבור מעכב את אישור השימוש בטכנולוגיה וגם מביא לאובדן אמון הציבור בעתיד הנהיגה האוטונומית.

יצירת מארג רגולטורי שיאפשר כניסה מהירה של טכנולוגיות לנהיגה אוטונומית נתפס כמשאב מדינתי הן מאחר שהמדינה תיהנה מהרווחים הישירים של נהיגה אוטונומית, כגון בטיחות מוגברת וחיסכון במשאבים, והן מאחר שפיתוח כזה מעודד חברות טכנולוגיה להקים מפעלים ומרכזי פיתוח במדינה אחת ולא באחרת. תהליך כזה מתקיים בבריטניה, בסינגפור, בארצות הברית וגם בישראל – ונסקר בהרחבה בדוח השוואתי שפרסמה רשות החדשנות בעניין.<sup>98</sup>

הקושי העיקרי ביצירת רגולציה לרכבים אוטונומיים הוא היעדר שקיפות בעניין הפעולות שחברות הטכנולוגיה נוקטות למניעת תאונות דרכים. קשה להבין מהו מערך ניהול הסיכונים למניעת תאונות, מהן הפעולות המבוצעות למניעת קריסה של המערכות, ובעיקר, לאחר שהתרחשה תאונה – מהו מערך התחקור שנועד למנוע הישנות של מקרים מסוג זה. כפי שמתברר, גם מדינות שמבקשות לממש הליכים רגולטוריים אינן יודעות כיצד לבסס דרישות שקיפות בעניין זה. הסודיות סביב פיתוח הטכנולוגיות והצורך לשמר יתרון תחרותי, הם משום תמריץ שלילי

Bryant Walker Smith, *Automated Driving and Product Liability*, Mich. 97 St. L. Rev., 1 (2017)

World Economic Forum: *Autonomous Vehicle Policy Framework: Selected National and Jurisdictional Policy Efforts to Guide Safe AV Development* (2020) 98

לפרסום גלוי מצד היצרניות. לעומת זאת, הנזקים מסודיות זו, ובפרט מהיעדר השקיפות בדבר הקריטריונים לבטיחות, פוגעים באמון הציבור בטכנולוגיה.

דוגמה לכך היא אירוע דריסתה למוות של איליין הרצברגר בטמפה, אריזונה, בשנת 2018, על ידי רכב אוטונומי בשלבי הניסוי.<sup>99</sup> עד מועד התאונה הסתכמה הרגולציה על ניסוי של רכבים אוטונומיים במדינת אריזונה במסמך קצר, חתום בידי מושל המדינה, המאשר ניסויים אלו וקובע הוראות של חובות תיעוד וביטוח בביצוע הניסויים. לאחר התאונה התבררו הכישלון של מבנה רגולטורי זה והיעדר המוכנות לבקרת הבטיחות הנדרשת גם בשלבי הניסוי. כן התברר כי חברת אובר, שהניסוי היה מטעמה, לא פעלה מספיק למנוע תאונות כאלה. רק בדיעבד התכנסו הרשויות במדינה לבחינת נורמות הבטיחות בחברת אובר, והכשלים שאותרו הביאו, בין השאר, להפסקת פעילות הניסוי במדינה לחודשים ארוכים. העיסוק הציבורי הנרחב בתאונה פגע באמינות הטכנולוגיה בעיני הציבור, אף שנתוני הבדיקות הראו כי הטכנולוגיה בטוחה יותר מנהיגה של נהגים אנושיים, וכי המעורבות של רכבים אוטונומיים בתאונות נמוכה מאוד ביחס לכמות השעות שרכבים אלה היו בניסוי.<sup>100</sup>

פעולה מבעוד מועד על פי מודל האיתנות המדעית-תהליכית הייתה מאפשרת למדינת אריזונה להכניס לשימוש את הטכנולוגיה, ובד בבד לאפשר בחינה מדוקדקת של רמת הבטיחות שלה. הציבור היה יכול להיות חשוף להיקף הנסיעות של רכבים אלה, לאירועי הבטיחות שנבדקו ולאישור מצד אנשי מדע מוסמכים באשר לתקינות האלגוריתמים הנמצאים בתהליכי בחינה. בהתרחשות תאונה היה גם ניתן לפרסם מדדים שקופים לציבור.

Jack Stilgoe, *Who Killed Elaine Herzberg?*, WHO'S DRIVING INNOVATION? 99  
1-6 (2020)

John M. Scanlon, Kristofer D. Kusano, Tom Daniel, Christopher Alderson, Alexander Ogle, & Trent Victor, *Waymo Simulated Driving Behavior in Reconstructed Fatal Crashes within an Autonomous Vehicle Operating Domain*, 163 ACCIDENT ANALYSIS & PREVENTION (2021)

החשש שאנו מבטאים לגבי המצב הקיים הוא שכאשר מערכות נהיגה אוטונומיות יגיעו לשימוש מסחרי, היצרניות לא יסכימו לשתף את פרטי התוכנות, וגם אם יפרסמו חלקים נרחבים מהן, המידע שיתפרסם לא יוכל לספק למשתמשים הסבר קונקרטי מדוע הרכב פעל כפי שפעל. חוקיות הרכבים תעמוד למבחן, וללא הסבר – ייתכן שתתקיים חזקת אי־חוקיות למערכת קבלת ההחלטות של הרכבים.

היות שהטכנולוגיה נמצאת בהליכי ניסוי, נראה שהיעילות של מודל האיתנות מדעית־תהליכית המוצע כאן יכולה להיות גבוהה במקרה זה. פרסום המדדים שלפיהם תיבחן האיתנות של טכנולוגיות הבטיחות ברכב, ובפרט האלגוריתמים העוסקים בניהול הסיכונים – יגביר את אמון הציבור ואף יאפשר התקדמות מהירה משלבי הניסוי ליישום בפועל.

הדוגמאות שהובאו לעיל מחזקות את הצורך בחקיקת חובות של שקיפות אלגוריתמית. הן מאירות את הצורך בחקיקה שתאמץ את המודל המוצע ותיצור מארג חובות ברור שיאפשר ודאות לשחקנים בשוק – יצרנים ומשתמשים כאחד. היה וייושם המודל המוצע, יצרני התוכנה יוכלו לצפות מראש את הדרישות מהם, ובכך לכתחילה לייצר מוצר שיעמוד בסטנדרטים החברתיים. לאחר מכן, במקרה של כשל אפשרי, יהיו בידי מערכת המשפט הכלים להעריך את עמידת היצרנים בחובות שהוטלו עליהם.

## פרק 9

### סיכום

שנת 2021 החלה בסערת האחריות של הרשתות החברתיות לאלגוריתמים שגרמו להפצה של קונספירציית גניבת הבחירות בארצות הברית ולפרסומות לתכשירי הרזיה לנערות מתבגרות חסרות ביטחון. היא המשיכה בדיון נוקב

לגבי מעצרי שווא של אפרו־אמריקאים בגלל טעויות של מערכות לזיהוי פנים, והסתיימה בקריסתה של חברת Zillow, שהאלגוריתם שלה יצר טעויות חישוב בנוגע לשווי נדל"ן למגורים.

אלגוריתמים משמשים אותנו כיום בכל מקום, החל בסרטונים שאנחנו צופים בהם בטיקטוק, המשך בעדכון המסלול בווייז וכלה בשואב האבק הרובוטי. אבל נראה שבשנת 2022 מחלחלת ההבנה בנוגע להשפעתם העמוקה על חיינו, למגבלות שלהם, לנזקים שהם גורמים, ובעיקר לפער בין השימושים הנרחבים שחברות וממשלות עושות בהם לבין היעדר הפיקוח הציבורי והרגולטורי עליהם.

כמוזכר לעיל, לפי מכון ברקמן־קליין לחקר האינטרנט באוניברסיטת הרוורד, 92% מהמסמכים העוסקים ברגולציה ואתיקה של בינה מלאכותית מזכירים את המושג "שקיפות" כרכיב מכריע במארג הפתרונות המיועדים להתגבר על הסכנות הפוטנציאליות במערכות לומדות. המעבר מהצהרות למעשים רגולטוריים מחייב להפוך את המושג "שקיפות אלגוריתמית" לבריישום, ואולם הבעיה היא שהדרישה לשקיפות בכל המסמכים שניתחנו אינה כוללת הגדרה מדויקת של עצם המושג. כן היא אינה מציינת באיזה חלק של תהליך הפעולה של המערכות הלומדות נדרשת שקיפות; איזו רמה של שקיפות נדרשת לגבי שימושים שונים במערכות כאלה; על מי בשרשרת הערך חלה חובת השקיפות; ומה האיזון הרצוי בין התביעה לשקיפות לבין ערכים אחרים כמו סודיות מסחרית וחדשנות. לפיכך, בפועל, המושג "שקיפות אלגוריתמית" נשאר עכור מאוד.

הצלחנו לחלץ מספרות המחקר 13 משמעויות שונות של המושג "שקיפות אלגוריתמית" ובהן: החובה להפוך את החלטות התוכנה לבנות גילוי; יכולת מעקב אחר הליך יישום התכנון ההנדסי של תוכנה; עיצוב הגון של ממשק המשתמש

באופן שיאפשר לו להבין את פעילות המערכת; בהירות שתבטיח הבנה אנושית מה אירע ומדוע, והסברתיות שמשמעותה שמידע על פעילות התוכנה יובן על ידי המשתמש; יכולת לבדוק ולשחזר את דרך פעולתן של המערכות או לאמת את המידע שנעשה בו שימוש; וגם החובה לסמן במפורש שמי שמנהל את התקשורת עם המשתמשים איננו אדם אלא מכונה. בקצרה, כשמושג הופך להיות רב־משמעי כל כך, יש חשש לזיהום שלו. יש גם חשש מפני ethics washing – כלומר, הצהרת הצהרות אתיות חלולות וללא ביצוע אמיתי שלהן.

יתרה מזו, כשמדובר במערכות לומדות לעיתים גם מי שבנו אותן אינם יודעים כיצד הן פועלות. העניין הוא שאלגוריתמים – כסדרה של הוראות למערכת ממוחשבת – יכולים להיות בעלי כל מיני צורות: יש שהם מקודדים באופן צר, עם הוראות ברורות למחשב (למשל, סֵדֶר סדרה של שמות בסדר אלפביתי), ויש שהם מתוחכמים הרבה יותר ומבוססים על למידה עצמית של המכונה. באלה משתמשות פייסבוק, אינסטגרם וטוויטר כדי לבצע את הפרסונליזציה של התכנים שאנחנו מקבלים. לכן, התביעה מהן לחשוף את "המתכון הסודי" היא בעיקרה מופרכת.

לפיכך במחקר זה הצענו שני מסלולים: האחד יחייב את יצרן האלגוריתם לנמק מדוע התקבלה החלטה מסוימת ולא החלטה אחרת. מסלול זה תואם את המוכר לנו מהעולם החוץ־דיגיטלי. הבעיה היא שלא בכל מערכת אלגוריתמית אפשר לעשות זאת. במקרה הזה, אפשרות אחת היא לא לפקח על המערכת כלל, ואפשרות אחרת היא לאסור על השימוש בה. שתי האפשרויות האלה רעות. לכן הצענו לפתח במקביל הכרה במסלול חלופי שלא תידרש בו שקיפות אלא הוכחה של "איתנות התהליך". במסלול הזה יופעל מבחן משפטי־סטטיסטי משולב שיראה שאף שאי־אפשר להסביר מדוע המערכת קיבלה את ההחלטה שקיבלה, מהלך הייצור התבצע באופן מדעי מוכח ומתוך הקפדה על מימוש של ערכים שייקבעו מראש – בעיקר שוויון ומניעת נזק. התהליך שאת איתנותו יהיה צריך להוכיח יעסוק בבחינת סף משפטית של הגדרת מטרות התוכנה; במידע שמוכנס למערכת – כלומר בחינה סטטיסטית של איכות הנתונים שהוזנו למאגר הנתונים והתאמתם ליעדי התוכנה; ולבסוף בהוכחת התוקף המתמטי של עיבוד המידע לשם ייצור הפלט.

למרות התקווה והכמיהה (המיסטית למחצה) ברוב המקרים לא נוכל להשיג שקיפות אלגוריתמית כפי שאנחנו מדמיינים אותה מהעולם הישן. עם הבשלת ההבנה בעניין הנחיצות של שקיפות, נדרש להרחיב את גבולות הדמיון ולהבין מה יוכל להחליף אותה במקום שבו אי־אפשר להסביר מדוע התקבלו החלטות כאלה או אחרות.

**עו"ד גדי פרל** הוא תלמיד לתואר שלישי במשפטים באוניברסיטה העברית בירושלים; עמית מחקר במרכז מישאל חשין ללימודי משפט מתקדמים ועמית מחקר במרכז פדרמן לחקר הסייבר באוניברסיטה. חוקר רגולציה של מערכות מבוססות בינה מלאכותית.

**ד"ר תהילה שוורץ אלטשולר** היא עמיתה בכירה במכון הישראלי לדמוקרטיה וראשת התוכנית "דמוקרטיה בעידן המידע". חברת מועצת הארכיונים העליונה ולשעבר חברת נשיאות מועצת העיתונות. מומחית לאסדרת תקשורת ואתיקה עיתונאית ולמשק שבין טכנולוגיה, משפט ומדיניות.



המכון הישראלי  
לדמוקרטיה

[www.idi.org.il](http://www.idi.org.il)

מסת"ב:  
978-965-519-397-8