

אליה וקוץ בה: הזכות להסברתיות של מערכות בינה מלאכותית

השימוש הגובר בבינה מלאכותית, בין אם כחלופה ובין אם ככלי מסייע לקבלת החלטות אנושית, מעלה חששות לגבי שקיפות, אחריות והוגנות. בתגובה, רגולטורים ברחבי העולם מקדמים את "הזכות להסברתיות", האמורה להקנות לנשוא ההחלטה הסבר להחלטת מכונה אוטונומית המשפיעה עליו. לשם מימוש הזכות להסברתיות הרגולטורים פונים לשלל פרקטיקות טכנולוגיות ליצירת "הסבר" כהגדרתם של מומחי למידת מכונה, והידועות יחד כתחום מחקר הקרוי eXplainability או XAI.

במאמר זה אבחן באופן ביקורתי את מידת התאמתן של פרקטיקות XAI למימוש "הזכות להסברתיות" הרגולטורית. לשם כך אנתח ראשית את מהות המושג "הסבר" – כחלק מ"חובת הנמקה" – בעולם המשפט. סקירה תמציתית של המופעים השונים של חובת הנמקה במשפט הציבורי, הפרטי ואף הבינלאומי מסייעת לחלץ את התכליות הנעוצות בליבו של מושג משפטי זה ואשר עומדות ביסוד השימוש התדיר בו. ניתוח זה מעלה כי התכליות המהותיות לחובת הנמקה המשפטית כוללות את הבטחת איכות ההחלטה, כיבוד האוטונומיה האנושית, מימוש הזכות להליך הוגן וחיזוק הסמכות של מערכת קבלת ההחלטות. לאחר מכן המאמר פונה לבחינת התחום הידוע כ-XAI, תוך עמידה על התמריץ לפיתוחו על ידי מומחים ללמידת מכונה – צורך פרקטי טכנולוגי לשפר את ביצועי המערכת – וסקירה של הפרקטיקות השונות המפותחות במסגרתו. העמדת XAI אל מול התכליות שמבוקש לממש

* הכותבת היא Ph.D Candidate ועמיתה במרכז צבי מיתר ללימודי משפט מתקדמים ומרצה במרכז לבינה מלאכותית ומדעי הנתונים של אוניברסיטת תל אביב TAD. הכותבת מבקשת להודות במיוחד למרכז מאיר שמגר למשפט דיגיטלי וחדשנות בפקולטה למשפטים של אוניברסיטת תל אביב על תמיכתו הנדיבה במחקר זה. המאמר מבוסס על עבודה לקבלת תואר L.L.M. באוניברסיטת תל אביב (Hofit Wasserman-Rozen, AI) תודתי שלוחה למנחי העבודה, פרופ' ניבה אלקין-קורן (הפקולטה למשפטים) ופרופ' רן גלעד-בכרך (הפקולטה להנדסה) על תרומתם שלא תסולא בפז למחקר, וכן לחברי מערכת "משפט, חברה ותרבות" על עבודת העריכה המסורה. תודה ליעל וואלי כרמון על תמיכתם לכל אורך הדרך ותודה מיוחדת לאייל, עלמה, לוטן ועילם על היותם הרוח מתחת לכנפיי, ולאמי סימה על הניגונים ששתלה בי.

באמצעות חובת ההנמקה מעלה שלל אתגרים ואף סיכונים פוטנציאליים הטמונים במימוש הזכות להסברות במצבה הנוכחי. המאמר נחתם בקריאה לאמץ פרשנות תכליתית למימוש הזכות להסברות, הן על מנת לכוון את הרגולטורים ומפתחי ההסברים הטכנולוגיים לייצור הסברים שישרתו את "חובת ההנמקה" והן על מנת לקדם קיומם של מוסדות ומתווכים שיהפכו את הזכות להסברות ליותר מ"אות מתה" בספר הרגולציה של מערכות בינה מלאכותית.

מבוא

אנו חיים בעולם שבו החלטות באמצעים אוטומטיים מסייעות להחלטות אנושיות, או מחליפות החלטות אלו, בקצב מסחרר. למהלך זה עשויות להיות השלכות מרחיקות לכת הכוללות פגיעה בזכויות אדם,¹ סכנה אשר לא חמקה מעינם של רגולטורים ברחבי העולם.² אחד הפתרונות המוצעים באופן תדיר להתמודדות עם סכנות הבינה המלאכותית הוא חובת מתן הסבר, או הנמקה, להחלטות אלגוריתמיות. "הסברות" למערכות בינה מלאכותית, המוגדרת לרוב כאפשרות להסביר לנמענים שונים – בהם משפטנים, מפתחים ומשתמשים – כיצד המערכת מקבלת החלטות,³ אמורה "[ל]אפשר לבקר את תהליך העיבוד של המערכת, ו[לה]בטיח יישום אחראי יותר של מסקנותיה".⁴ עקרון ההסברות תופס אחיזה בהליכי רגולציה של מערכות בינה מלאכותית בישראל⁵

- 1 ראו למשל Mathias Risse, *Human Rights and Artificial Intelligence: An Urgently Needed Agenda*, 41 HUM. RTS. (2019). ראו גם Rowena Rodrigues, *Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities*, 4 J. RESPONSIBLE TECH (2020).
- 2 ראו למשל Subcommittee on Human Rights, Artificial Intelligence and Human Rights: Using AI as a Weapon of Repression and its Impact on Human Rights, PE 754.450 (2024), [https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA\(2024\)754450_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf); *Regulation essential to curb AI for surveillance, disinformation: rights experts*, UN News, (June 2, 2023), <https://news.un.org/en/story/2023/06/1137302>.
- 3 ראו למשל המיזם הלאומי למערכות נבונות ועדת הבינה המלאכותית ועדת המשנה של המיזם הלאומי למערכות נבונות בנושא אתיקה ורגולציה של מערכות נבונות: דין וחשבון 13 (2019). הסברות מוגדרת כ"... יכולת להסביר את תהליך קבלת ההחלטה של המערכת" וכוללת הבהרה כי היא מתייחסת ליכולת הסבר זו "ברמת המשתמשים כפרטים, גם ברמת הכלל אם המערכת משפיעה על קבוצות, וגם עבור מפעילי המערכת עצמם".
- 4 הפורום לתשתיות לאומיות למחקר ולפיתוח ועדת בינה מלאכותית ומדע הנתונים 62 (2020).
- 5 ראו משרד החדשנות, המדע והטכנולוגיה מדיניות, רגולציה ואתיקה בתחום הבינה המלאכותית בישראל (2022) (להלן: *מדיניות רגולציה* 2022); ראו גם דוח משרד

ובעולם,⁶ ולשם מימושו גויסו שלל פרקטיקות טכנולוגיות לייצור מה שהקהילה הטכנולוגית מגדירה כ-"הסברים" למערכות מסוג זה. טכניקות אלו, השייכות לתחום מחקר הידוע בשם (XAI) eXplainability,⁷ והמבוססות לרוב על פיתוחים של תורת המשחקים,⁸ התהוו מתוך אתגרים טכנולוגיים של מפתחים הנדרשים להבין את דרך פעולתן של מכונות אוטונומיות סבוכות.

עקרון "ההסברות" יונק מחובת ההנמקה המשפטית, בהתבסס על השימוש הרגולטורי הנפוץ בשקיפות ככלי לקידום אחריותיות (Accountability).⁹ כפי שיפורט להלן, ביסוד השימוש הנפוץ בחובת הנמקה בעולם המשפט עומדות ארבע תכליות חשובות: א. שיפור איכות ההחלטה, ב. הכרה באוטונומיה האנושית, ג. מימוש הזכות להליך הוגן ו-ד. חיזוק האמון במערכת קבלת ההחלטות. ניכר כי המרכזיות היחסית של המכשיר המשפטי הקרוי "חובת הנמקה" (במשפט הציבורי, הפרטי ואף הבינלאומי) עומדת ביסוד אימוצו על ידי רגולטורים גם בהקשר של קבלת החלטות אוטומטיות. ברם, כפי שאטען במאמר זה, הפרקטיקות הטכנולוגיות לייצור הסברים באמצעות XAI מתקשות לממש את שלוש התכליות הראשונות, והן אף עלולות לחזק את הסיכון הטמון במימוש מופרז של התכלית הרביעית, סיכון העשוי להתבטא במתן אמון בלתי מוצדק במערכת קבלת ההחלטות. קושי זה נובע מהשונות בתפיסת המושג "הסבר" בין שתי הדיסציפלינות: המשפטית וההנדסית. בעוד שעולם המשפט פונה למכשיר הקרוי "הסבר" לשם הגשמת התכליות שפורטו לעיל, עולם למידת המכונה מתייחס להסבר ככלי אחד בארגז הכלים של בעלי מומחיות בתחום לשם טיוב המערכת.¹⁰ על מנת

המשפטים ומשרד החדשנות, המדע והטכנולוגיה **עקרונות מדיניות, רגולציה ואתיקה בתחום הבינה המלאכותית** (2023).

- 6 ראו *Blueprint for an AI Bill of Rights*, THE WHITE HOUSE, <https://www.whitehouse.gov/ostp/ai-bill-of-rights> גם *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 (Apr. 21 2021).
- 7 תחום המחקר של היכולת להסביר מערכות בינה מלאכותית נקרא לעיתים גם Interpretability או Intelligibility, אך לעיתים קיימת גם הפרדה בין מונחים אלו.
- 8 ראו למשל Emanuele Albini et al., *On the Connection Between Game-theoretic Feature Attributions And Counterfactual Explanations*, 2023 AAAI/ACM CONF. AI, ETHICS & SOC'Y 411 (2023).
- 9 "אחריותיות" – אחריות של בעלת תפקיד לתת דין וחשבון על פעולותיה ולקבל עליהן אחריות. ראו "אחריות דיווח" **האקדמיה ללשון עברית**, <https://www.nevo.co.il/shorturl/9B2F8E>.
- לנושא של אחריותיות בהקשר של בינה מלאכותית ראו למשל Madalina Busuioc, *Accountable Artificial Intelligence: Holding Algorithms To Account*, 81 PUB. ADMIN. REV. 825, 825–836 (2021).
- 10 SIMON CHESTERMAN, WE, THE ROBOTS?: REGULATING ARTIFICIAL INTELLIGENCE AND THE LIMITS OF THE LAW 152 (2021).

לגשר על פער זה המאמר מציע לאמץ גישה של יישום תכליתי של הזכות להסברתיות בנוגע להחלטות מכונה אוטומטיות. כך, טיב ההסבר להחלטת מכונה אשר יידרש, ודרישות המשנה הנלוות לו, יושפעו מהתכלית העיקרית שאותה נרצה לממש בכל מקרה נתון.

מהלך הדיון: **בחלק א** אתאר את עליית העיקרון של חובת מתן הסבר להחלטות אוטומטיות כחלק מהתפתחות הרגולציה בארה"ב, באירופה ובישראל ואת התקוות הנתלות במימושו של עיקרון זה. **בחלק ב** אבחן את משמעות מושג ההסבר בעולם המשפט, תוך הגדרת מושגים מהותיים כגון "הסבר", "הצדקה" ו"הנמקה", סקירה קצרה של השימוש בחובת הנמקה בעולם המשפט וניתוח התכליות העומדות ביסוד שימוש נפוץ זה. **חלק ג** יתאר את צמיחת תחום המחקר הטכנולוגי שאמון על יצירת הסברים למערכות בינה מלאכותית (XAI), ויבחן את מידת היכולת של הטכניקות המתפתחות בו לממש את התכליות העומדות ביסוד הדרישה הרגולטורית להטיל חובת הסברתיות על החלטות אוטומטיות. מאחר שבחינה זו מעלה מספר אתגרים למימוש התכליות, **חלק ד** יציע לאמץ פרספקטיבה של יישום תכליתי ולהתאים את הדרישות המשפטיות והרגולטיביות מהסברים להחלטות מכונה אל התכליות העיקריות שמבוקש לממש באמצעותם.

א. הזכות להסברתיות במערכות בינה מלאכותית

קבלת החלטות אנושית מאופיינת באלמנטים כגון חמלה, שיקול דעת, שימוש מותאם במידע ומומחיות וידע נצבר על העולם – תכונות ההולכות ונרחקות ככל שקבלת החלטות משתנה מאנושית לאוטומטית.¹¹ ברם, המעבר מקבלת החלטות אנושית

Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 1971 (2021). ביקורת המושמעת תדיר בהקשר של תכונת "הקופסה השחורה" של מערכות בינה מלאכותית היא היותו של המוח האנושי קופסה שחורה בפני עצמו. למעשה, קבלת החלטות אנושית היא דבר מורכב, וכך גם ההסברים הניתנים לה. ראו למשל Christine Jolls et al., *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471 (1998); Jennifer S. Lerner & Philip E. Tetlock, *Accounting for the Effects of Accountability*, 125 PSYCH. BULL. 255 (1999); Daniel Kahneman, *Maps of Bounded Rationality: Psychology for Behavioral Economics*, 93 AM. ECON. REV. 1449 (2003). אין בכוונת מאמר זה לטעון שקבלת החלטות אנושית וההסברים הניתנים לה על ידי בני אדם מאופיינים בשקיפות מוחלטת, ואפילו לא בשקיפות הרצויה (אף על פי שמתן הסבר אכן משפיע על מקבלי החלטה אנושיים, ראו Zhuang Liu, *Does Reason Writing Reduce Decision Bias? Experimental Evidence from Judges in China*, 47 J. LEGAL STUD. 83 (2018)). אסתפק רק בהבחנה כי בשל השוני המובנה בין קבלת החלטות אנושית לאוטומטית, הכלי שפותח להתמודד עם עמימות המאפיינת קבלת החלטות אנושית – "הסבר" – אינו בהכרח אפקטיבי להתמודדות עם עמימות המאפיינת קבלת החלטות באמצעות "קופסה שחורה" מלאכותית.

לקבלת החלטות אוטומטית באמצעות מכונות, בין אם באופן מלא ובין אם באופן חלקי (למשל כאמצעי מסייע), כרוכה בויתור על אלמנט משמעותי נוסף – שקיפות.¹² המורכבות ההולכת וגוברת של מכונות אלו מגבירה את בעיית "הקופסה השחורה", מונח המבטא את תכונת העמימות של מערכות כה סבוכות עד כי "לא ניתן להבין כיצד מערכת בינה מלאכותית מקבלת החלטות".¹³ במבחן המעשה, קיים קושי להציג הסברים לדרך פעולתן של מכונות אלו באמצעים שיובנו על ידי בני אדם. עמימות זו כרוכה בסכנה לפגיעה בחירויות הפרט כגון חופש הביטוי, הזכות לשוויון, הזכות להליך הוגן¹⁴ ועוד. מומחים אף הזהירו מסכנת הפיכתנו ל"עבדים לאלגוריתם",¹⁵ דהיינו שברייס בידי מכונות מסתוריות המכוונות את חיינו בדרכים שאינן מובנות לנו.¹⁶ על מנת להתגבר על תכונת העמימות הנ"ל, ולאור מרכזיותן של מכונות אלו בחיינו, הולך וגובר העניין ב"זכות להסברות".¹⁷ את הולדתה של זכות זו נהוג לשייך לחקיקה האירופית בנושא פרטיות מידע, ה-GDPR,¹⁸ והיא מהווה חלק מהגישה הנפוצה כיום שלפיה כדי להבין קבלת החלטות אלגוריתמית יש צורך בהכרח "להרים את מכסה המנוע" ולהסביר מה מתרחש בתוככי המכונה.¹⁹ כך, מתן הסברים להחלטות אוטומטיות מוזכר ככלי להגברת אחריותיות,²⁰ זיהוי וצמצום אפליה והטיות,²¹ שמירה

- 12 Heike Felzmann et al., *Transparency you can trust: Transparency Requirements for Artificial Intelligence Between Legal Norms and Contextual Concerns*, 6(1) BIG DATA & SOC'Y 1 (2019).
- 13 משרד המשפטים **בינה מלאכותית במגזר הפיננסי: שימושים נפוצים, אתגרים וסקירה השוואתית של התמודדות רגולטורית** 8 (2022).
- 14 מעיין פרל "פרטיות, שליטה ופיקוח בעידן של נתוני עתק: חובת הנמקה על החלטות אלגוריתמיות" **משפט, חברה ותרבות** ב 167, 173 (מיכאל בירנהק עורך 2019) (להלן: פרל "פרטיות, שליטה ופיקוח").
- 15 Lilian Edwards & Michael Veale, *Slave to the Algorithm: Why a Right to an Explanation is Probably Not the Remedy You Are Looking for*, 16 DUKE L. & TECH. REV. 18, 20 (2017).
- 16 ראו בכלליות, שם.
- 17 שם, בעמ' 20; ראו גם CHESTERMAN, לעיל ה"ש 10.
- 18 Bryce Council Regulation 2016/679, 2016 O.J. (L 119) 1 (להלן: GDPR); ראו גם Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"*, 38(3) AI MAG. 50 (2017).
- 19 Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1117 (2018).
- 20 Finale Doshi-Velez et al., *Accountability of AI Under the Law: The Role of Explanation*, 2017 BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y; Alison Smith-Renner et al., *No Explainability Without Accountability: An Empirical Study of Explanations and Feedback in Interactive ML*, 2020 CONF. HUM. FACTORS COMPUT. SYS., 1; Talia B. Gillis & Josh Simons, *Explanation < Justification: GDPR and the Perils of Privacy*, 2 J. L. & INNOVATION 71 (2019).

על עקרון ההוגנות,²² הבטחת הליך נאות וממשל תקין,²³ והוא אף נחשב הכרחי על מנת לממש את היכולת לערער על החלטות אוטומטיות.²⁴ מופעים של הזכות להסברות מצויים בצורה נרחבת ברגולציה הבינלאומית המתהווה בנושא בינה מלאכותית. בארה"ב, המתווה ל"מגילת זכויות" של העם האמריקאי בהתייחס לבינה מלאכותית שפורסם בסוף שנת 2022 כולל התייחסות מפורשת לקיומו של הסבר כאחד מחמשת עקרונות היסוד שאמורים להנחות את התכנון, השימוש וההטמעה של מערכות אוטונומיות, במטרה מוצהרת להגן על הציבור האמריקאי בעידן של בינה מלאכותית.²⁵ מוסדות ציבוריים אמריקאים מובילים בחקר ובפיתוח בינה מלאכותית עורכים או תומכים במחקר בנושא הסברות של מערכות בינה מלאכותית.²⁶ אחת היוזמות המשמעותיות ביותר לאסדרת טכנולוגיות בינה מלאכותית בממשל הפדרלי כוללת אף היא את עקרון ה"הסבר" כאחד מחמשת עמודי התווך ליצירת "SAFE Innovation for AI".²⁷ בשנים האחרונות שחקנים טכנולוגיים גלובליים פרטיים מצהירים גם כן על רצונם לעודד שקיפות והסברות בבניה המלאכותית שהם מפתחים ומטמיעים במוצריהם ובשירותיהם.²⁸ אסדרה-עצמית מסוג

- Maja Brkan, *Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond*, 27 INT'L J. L. & INFO. TECH. 91, 118 (2019); Gaspar Isaac Melsión et al., *Using Explainability to Help Children Understand Gender Bias in AI*, 2021 INTERACTION DESIGN & CHILD. 87
- Jonathan Dodge et al., *Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment*, 2019 INT'L CONF. INTELLIGENT USER INTERFACES 275, 277–278.
- Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 119 (2014)
- Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 204 (2019). ראו גם Kaminski & Urban, לעיל ה"ש 11, בעמ' 1980.
- ראו *Blueprint for an AI Bill of Rights*, לעיל ה"ש 6.
- Explainable Artificial Intelligence, Defense Advanced Research Projects Agency, <https://www.darpa.mil/program/explainable-artificial-intelligence> (last visited Aug. 7, 2024); AI Fundamental Research – Explainability, National Institute of Standards and Technology, <https://www.nist.gov/artificial-intelligence/ai-fundamental-research-explainability> (last visited Aug. 7, 2024); National Science Foundation, Computer and Information Science and Engineering (CISE): Research Initiation Initiative (CRII), NSF 21-585, at 1 (Aug. 25, 2021), <https://www.nsf.gov/funding/opportunities/nsf-program-fairness-artificial-intelligence-collaboration/505651/nsf21-585/solicitation>.
- Scott Wong, *'A moment of revolution': Schumer unveils strategy to regulate AI amid dire warnings*, NBC NEWS (June 21, 2023), <https://www.nbcnews.com/politics/congress/schumer-call-hands-deck-approach-regulating-ai-rcna90193>.
- IBM'S Principles for Data Trust and Transparency, IBM BLOG (May 30, 2018), ראו <https://www.ibm.com/policy/trust-principles>; *Microsoft Responsible AI Standard*,

זה, אף שאינה חפה מבעיות,²⁹ טומנת בחובה גם יתרונות כגון עידוד חדשנות³⁰ ואפשרות גמישות נדרשת.³¹ באירופה, הדיון המשפטי התמקד בראשיתו בעצם השאלה האם ה-GDPR, אשר נכנס לתוקף בשנת 2018, אכן כולל "זכות לקבלת הסבר", בהתבסס על סעיפים 13–15, סעיף 22 ורסיטל 71.³² פרסומן של ה-"Ethics Guidelines for Trustworthy AI"³³ בשנת 2019 הרחיב את גבולות הדיון כשהכיר במכשולים הטכנולוגיים העומדים בפני יישום של עקרון חובת ההסבר בפועל והציע שילוב של אמצעים רגולטוריים נוספים, כגון עריכת ביקורות (Audits) והקפדה על יכולת העקיבות (Traceability).³⁴ מאידך גיסא, החקיקה הנוכחית של האיחוד האירופי בנושא בינה מלאכותית (AI ACT), אשר

- V2, MICROSOFT BLOGS (June 2022), <https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf> (last visited Dec. 24, 2023).
- סביר כי התכליות ביסוד הצהרות מסוג זה קשורות לאינטרסים של גופים מסחריים ופרטיים, וביניהם קידום האמון במוצרים ובשירותים, אימוץ הטכנולוגיה, ואף מסר לרגולטורים בדבר היעדר הצורך בהסדרה מרכזית ממשלתית.
- Rodrigo Ochigame, *The invention of 'ethical AI': How big tech manipulates academia to avoid regulation*, THE INTERCEPT (Dec. 20, 2019, 1:19 PM), <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence>.
- Daniel E. Walters & Hannah Jacobs Wiseman, *Self-regulation in Emerging and Innovative Industries*, 22 L. & ECON. CTR. GEO. MASON U. SCALIA L. SCH. RSCH. PAPER SERIES, 14-27.
- Peng Hwa Ang, *The Role of Self-Regulation of Privacy and the Internet*, 1 J. INTERACTIVE ADVERT. 1, 5 (2001).
- Gianclaudio Malgieri & Giovanni Raou, *לעיל ה"ש 18*. Raou Goodman & Flaxman, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, 7 INT'L DATA PRIV. L. 243 (2017).
- Andrew D. Selbst & Julia Powles, *לעיל ה"ש 24*. Raou Kaminski, *Meaningful Information and the Right to Explanation*, 7 INT'L DATA PRIV. L. 233 (2017); Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIV. L. 76 (2017); Sandra Wachter et al., *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J. L. & TECH. 841 (2018).
- High-Level Expert Grp. on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (Apr. 8, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.
- Sebastian Bordt et al., *Post-Hoc Explanations Fail to Achieve Their Purpose in Adversarial Contexts*, ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 891 (2022).

אוסרה ביום 13.3.2024³⁵ כוללת את סעיף 86(c) שכותרתו "A Right to Explanation of Individual Decision-Making"³⁶ כפי שכותרת זו מעידה, הדיון המשפטי בדבר קיומה של זכות לקבלת הסבר מתגבש לכיוון של הסכמה כללית בדבר הצורך במימוש של עקרון ההסברות בהתייחס לפחות לחלק מטכנולוגיות הבינה המלאכותית. בישראל, שורה של דו"חות כללה התייחסות לשקיפות או הסברות כחלק מהעקרונות שצריכים להדריך את מקבלי ההחלטות בהסדרת תחום הבינה המלאכותית.³⁷ בהינתן שטרם הובהר באיזו גישת הסדרה תבחר ישראל,³⁸ כבאירופה או כבארה"ב, טיב ועוצמת האכיפה של זכות להסברות באופן מקומי נותרו בשלב זה לוטים בערפל.

כפי שסקירה זו, שאינה תמציתית, מלמדת, רגולטורים מייחסים חשיבות רבה לכלי של "חובת הסבר" בהתייחס לקבלת החלטות אוטומטית וככלי לקידום אחריותיות מצד השחקנים המרכזיים בתחום. השימוש בכלי הרגולטורי של הגברת שקיפות כדרך להטלת אחריותיות אינו חדש,³⁹ ושורשיו נעוצים בחשיבותו של המושג "הסבר" בעולם המשפט, שאותו אבחן להלן.

35 Artificial Intelligence Act, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024, Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 2024 O.J. (L 304) 1, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689

36 בהתאם למבנהו של ה-AI ACT, המחלק מערכות בינה מלאכותית לפי לרמת הסיכון הצפויה מהן לזכויות אדם, חובת מתן הסבר כאמור צפויה לחול בעיקר על מערכות המוגדרות "סיכון גבוה" (למעט חריגים המוגדרים בחוק). יחד עם זאת חשוב לזכור כי ה-AI ACT אינו מאיין את תחולתו של ה-GDPR ועל כן שימוש במידע הנופל בגדרו של ה-GDPR עשוי להיות כפוף ל"זכות להסברות" מכוחו, גם במקרים שבהם סעיף 86(c) של ה-AI ACT אינו חל.

37 ראו ועדת בינה מלאכותית ומדעי הנתונים, לעיל ה"ש 4; מדיניות רגולציה ואתיקה בתחום הבינה המלאכותית בישראל, לעיל ה"ש 5, בעמ' 84–85. לסקירה רחבה של התהוות הרגולציה המתגבשת בישראל בנושא בינה מלאכותית ראו שם, בעמ' 23–25.

38 סנונית ראשונה ניתן למצוא אולי במסמך עקרונות מדיניות, רגולציה ואתיקה בתחום הבינה המלאכותית שפורסם על ידי משרד החדשנות, המדע והטכנולוגיה ומשרד המשפטים ביום 14.12.2023, ואשר מנחה את הרגולטורים לפעול במסגרת של רגולציה ענפית ולא רגולציה רוחבית, שם.

39 David Weil et al., *The Effectiveness of Regulatory Disclosure Policies*, 25 J. POL'Y ANALYSIS & MGMT 155, 155 (2006).

ב. "הסבר" בעולם המשפט

קבלת החלטות היא פרקטיקה הנעוצה בלב-ליבו של העיסוק המשפטי.⁴⁰ בחברה דמוקרטית, החלטות מלוות לרוב בהסברים, ואזרחים מצופים להסביר זה לזה את הבסיס התומך בהחלטותיהם בדרך שתאפשר לאחרים להעריך עד כמה החלטות אלו עולות בקנה אחד עם ערכים של שוויון וחירות.⁴¹ השימוש בהסברים מסוג זה, אשר ניתן להגדירם כ-"Reason-Explanations",⁴² נעשה לרוב כאשר בני אדם מנסים להבין פעולות ולהתמודד עם חוסר הסכמות. בעולם המשפט נגדיר אותם כ-"הנמקות" ("Reason-Giving"),⁴³ מושג המשלב מתן סיבות⁴⁴ והצדקות.⁴⁵ בעוד שסיבות אמורות לסייע "להבין דבר שאינו ברור",⁴⁶ הצדקות מייחסות "סיבה צודקת להתנהגות".⁴⁷

Keith Hawkins, *On Legal Decision-Making*, 43 WASH. AND LEE L. REV. 1161, 1162 (1986).

JOHN RAWLS, *POLITICAL LIBERALISM* 217-218 (2005) ("Democracy involves, as I have said, a political relationship between citizens within the basic structure of the society into which they are born and within which they normally lead a complete life; it implies further an equal share in the coercive political power that citizens exercise over one another by voting and in other ways. As reasonable and rational, and knowing that they affirm a diversity of reasonable and philosophical doctrines, they should be ready to explain the basis of their actions to one another in terms each could reasonably expect that others might endorse as consistent with their freedom and equality. Trying to meet this condition is one of the tasks that this ideal (of democratic politics asks of us").

Kevin Baum et al., *From Responsibility to Reason-Giving Explainable Artificial Intelligence*, 35 PHIL. & TECH. 1, 17 (2022).

"הנמקה" אלישבע שקד ואביעד הכהן מילון משפטי: עברי-אנגלי אנגלי-עברי (1992).

"הנמקה" האקדמיה ללשון העברית, <https://hebrew-academy.org.il/keyword/%D7%94%D6%B7%D7%A0%D6%B0%D7%9E%D6%B8%D7%A7%D6%B8%D7%94>.

דפנה ברק-ארו "ההנמקה השיפוטית: דילמות ישנות וחדשות" משפטים נב 5, 9-11 (2022). יודגש כי במאמרה (ראו שם, ה"ש 1) המחברת מתייחסת באופן ייחודי להנמקה שיפוטית (בניגוד להנמקות מטעם גוף פוליטי) השונה לדבריה מהחלטות אחרות היכולות להיות מבוססות "על כוח ועל רצון". אך השוו להגדרת "הנמקה" אצל Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633, 634 (1995) ("[T]he practice of engaging in the linguistic act of providing a reason to justify what we do or what we (decide)".

"הסבר" האקדמיה ללשון העברית, <https://hebrew-academy.org.il/keyword/%D7%94%D6%B6%D7%A1%D6%B0%D7%91%D6%BC%D6%B5%D7%A8>.

Gianclaudio Malgieri, "Just" Algorithms: Justification (Beyond Explanation) of Automated Decisions Under the General Data Protection Regulation, 1 LAW &

במילים אחרות, הנמקה מגייסת סיבות, הסברים וצידיקים על מנת לתמוך בהחלטה ובפעולה ולשכנע כי היא נכונה, חוקית ומוסרית.⁴⁸

1. חובת הנמקה – היכן?

אף ש"חובת ההנמקה ממוקמת בספרה הציבורית",⁴⁹ היא הולכת וקונה לה אחיזה במשפט הפרטי ואף הבינלאומי. אסקור להלן בקצרה מופעים מספר של חובת הנמקה ואחלץ מתוכם את התכליות לקיומה. בהכללה, חובת ההנמקה במשפט הציבורי משקפת תכליות של שיפור ההחלטה וקיום הליך נאות, חובת הנמקה במשפט הפרטי משקפת תכלית של הכרה באוטונומיה האנושית וחובת הנמקה במשפט הבינלאומי משקפת תכלית של חיזוק הסמכות של מערכת קבלת ההחלטות.⁵⁰

חובת הנמקה במשפט הציבורי

גופים מנהליים

חובת הנמקה הפכה לעקרון יסוד במשפט הציבורי אשר מקורו בזכות להליך נאות.⁵¹ התעצמותה מאפיינת את התנועה ההדרגתית להאצלת סמכויות חקיקה לגופים

Bus. 16, 18 (2021) ("[A]n act of spotting the main reasons or factors that led to a (particular consequence, situation, or decision")

47 "הצדקה" האקדמיה ללשון העברית, <https://hebrew-academy.org.il/keyword/%D7%94%D6%B7%D7%A6%D6%B0%D7%93%D6%BC%D6%B8%D7%A7%D6%B8%D7%94>. ראו גם שם, בעמ' 20.

48 Andrew Stranieri et al., *A Hybrid Rule – Neural Approach for the Automation of Legal Reasoning in the Discretionary Domain of Family Law in Australia*, 7 A.I. & L. 153, 155 (1999). על מנת להמחיש את ההבדל בין הסבר/סיבה להצדקה נשתמש בדוגמה הבאה: נניח שבחברה (company) מסוימת הוחלט על הורדת שכר בצורה גורפת. הסבר להחלטה זו יהיה, למשל, מיעוט הזמנות בשוק שמיתרגם לקיטון בהכנסות החברה. צידוק לעומת זאת ימשיך וינמק את ההחלטה בכך שללא הורדת שכר יהיה צורך לפטר עובדים, החלטה אשר הנהלת החברה מבקשת להימנע ממנה כל עוד ניתן. ההנמקה תשלב הן את ההסבר (הסיבה, משמע ירידה בהכנסות החברה) והן את ההצדקה (בחירה להימנע מפיטורים תוך איזון עם פגיעה מידתית בשכר).

49 פרל, לעיל ה"ש 14, בעמ' 188.

50 האפיון של תכליות העולות מתחומי משפט מסוימים המובאת בפרק זה נעשתה על מנת להדגיש את המהלך התאורטי של חילוף התכליות מן המופעים. מובן כי התכליות באות לידי ביטוי במידות שונות בכל אחד מתחומי חובת ההנמקה שיתוארו להלן, ואין בהדגשת התכלית בתחום מסוים כדי לאיין את השפעתה בתחום אחר. כך למשל ניתן לטעון כמובן כי חובת הנמקה במשפט הציבורי משרתת בין היתר את ההכרה באוטונומיה האנושית.

51 Katherine J. Strandburg, *Rulemaking and Inscrutable Automated*, 119 COLUM. L. REV. 1851, 1865 (2019); דפנה ברק-ארוז **משפט מנהלי** כרך א 423 (2010).

מנהליים,⁵² והיא מקבלת ביטוי גובר והולך הן באיחוד האירופי, הן בארה"ב והן בישראל. באירופה, חובת הנמקה מעוגנת הן ברמה הלאומית והן ברמת האיחוד ומלווה בפסיקות של ה-ECJ.⁵³ בארה"ב, ה-APA⁵⁴ עיגן חובת הנמקה שמקורה בהחלטות שיפוטיות קודמות,⁵⁵ ואשר ממשיכה להתפתח בפסיקות עוקבות.⁵⁶ בישראל, עיקר עיגונה של חובת ההנמקה של גופים מנהליים מצויה בחוק לתיקון סדרי המנהל (החלטות והנמקות), התשי"ט-1958 (להלן: חוק ההנמקות),⁵⁷ והיא מהווה "עיקרון בסיסי בתורת המינהל".⁵⁸ על פי הסדר חוקי רחב זה, מוטלת חובת הנמקה על עובד ציבור שנתבקש בכתב להשתמש בסמכות שניתנה לו על פי דין וסירב לבקשה.⁵⁹ חוק ההנמקות קובע כי על ההנמקה להינתן בכתב,⁶⁰ במועד ההחלטה או במהירות הראויה

- 52 Martin Shapiro, *The Giving Reasons Requirement*, 1992 U. CHI. LEGAL F. 179, 180 (1992); William F. Pedersen Jr., *Formal Records and Informal Rulemaking*, 85 YALE L. J. 38 (1975). יודגש כי בד בבד עם האצלת סמכויות חקיקה לרשות המבצעת, נותרו בידיה גם סמכויות של קבלת החלטות ברמה הפרטית (טריבונלים, ועדות ערר ובתי דין העוסקים בשפיטה או גופים העוסקים באכיפה). לכל סוג החלטה תידרש מטבע הדברים הנמקה שונה. כך למשל החלטת מימוש מדיניות תצריך פירוט של התשתית המקצועית שהובילה לקבלת ההחלטה, בעוד שהחלטה פרטנית תצריך פירוט של עובדות המקרה וכיצד הן משליכות על המצב החוקי הרלוונטי.
- 53 Monica Delsignore & Margherita Ramajoli, *The 'Weakening' of the Duty to Give Reasons in Italy: An Isolated Case or a European Trend?*, 27 EUR. PUB. L. 23, 23 (2021). ראו גם Consolidated Version of the Treaty on the Functioning of the European Union art. 296, oct. 26, 2012, 2012 O.J. (C 326) 175–176; Charter of Fundamental Rights of the European Union, art. 41 2016 O.J. (C 202) 401–2. Ingrid Opdebeek & Stéphanie De Somer, *The Duty to Give Reasons in the European Legal Area: A Mechanism for Transparent and Accountable Administrative Decision-Making? A Comparison of Belgian, Dutch, French and EU Administrative Law*, 2 ROCZNIK ADMINISTRACJI PUBLICZNEJ 97, 102 (2016).
- 54 Administrative Procedure Act § 5 U.S.C § 551–559. זהו חוק בארצות הברית אשר מסדיר את ההליך המנהלי הפדרלי, לרבות את ההליך הרגולטורי, ו"יש אף הרואים בו 'חוקה' של המשפט המנהלי בארצות הברית". ראו משרד המשפטים **חקיקת מסגרת לרגולציה: סקירה משווה** 45 (2021).
- 55 Lisa Schultz Bressman, *Procedures as Politics in Administrative Law*, 107 COLUM. L. REV. 1749, 1777 (2007).
- 56 Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process*, 95 YALE L. J. 455, 468–469 (1986).
- 57 יצחק זמיר **הסמכות המינהלית** כרך ב 1273 (מהדורה שנייה מורחבת 2011).
- 58 רע"א 8996/04 **שכטר נ' נציגות הבית המשותף**, פ"ד נט(5) 17, 17 (2004); זמיר, לעיל ה"ש 57, בעמ' 1270.
- 59 ס'2 לחוק ההנמקות.
- 60 שם.

לאחר מתן ההחלטה,⁶¹ ברם החוק אינו מבאר את רמת פירוט הנימוקים נשוא ההחלטה, ובהתאם לפסיקה "ההיקף והעומק של ההנמקה תלויים בנסיבות המקרה".⁶² יודגש כי חובת ההגינות מחייבת את הרשות לחשוף את הנימוקים האמיתיים שהובילו לקבלת ההחלטה.⁶³ ודוק – חובת ההנמקה טומנת בחובה שתי חובות: ראשית החובה לקבל החלטה מנומקת, ושנית למסור את נימוקי ההחלטה.⁶⁴ במילים אחרות, עצם חובת ההנמקה מחייבת לייצר החלטה שטעמיה עימה. חובת ההנמקה עוברת כחוט השני אף מחוץ לגדרי חוק ההנמקה בהסדרים מיוחדים כדוגמת ס' 33(ו) לחוק הנכים (תגמולים ושיקום) הקובע כי "ועדת ערעור תיתן נימוקים להחלטתה",⁶⁵ ס' 75(ג) לחוק המקרקעין 1969 הקובע כי על מפקח אשר החליט לסטות מדיני הראיות ומסדרי הדין הנוהגים בבתי המשפט לרשום את "הטעמים שהניעוהו לכך"⁶⁶ וס' 40(א) לחוק בתי דין מנהליים הקובע כי "החלטת בית דין תהיה מנומקת, תינתן בכתב, ותיחתם בידי חברי בית הדין".⁶⁷

דוגמה עדכנית לנפקות של חובת הנמקה בתחום המשפט המנהלי, שאף נוגעת לקבלת החלטה אוטומטית, ניתנה לאחרונה בעניין ברגר.⁶⁸ עסקינן בהחלטה אלגוריתמית אשר הופקה בנתב"ג בדמות "רשימת הכללה" שעל פיה נערכים חיפושים בנוסעים החוזרים לארץ. במהלך הדיון בביהמ"ש נחשף כי – לדעת המשטרה – מדובר ב"קופסה שחורה" אשר אין נגישות לפרמטרים של קבלת ההחלטה שלה.⁶⁹ ביהמ"ש קבע כי לא ניתן לקבל קיומה של רשימה אשר היווצרותה אינה מעוגנת בחוק או בנוהל מסודר ואשר לא ניתן לוודא כי אינה מונעת משיקולים כגון שרירות-לב ואפליה, וכי בהעדר הבנה של המערכת נשללת למעשה "[ה]אפשרות לביקורת מנהלית ולביקורת שיפוטית יעילות".⁷⁰ כפועל יוצא מכך, בהינתן אותה עמימות סביב קבלת החלטה על ידי גוף מנהלי, קיימים קשיים במימוש זכותו של נאשם להליך הוגן.

- 61 שם, ס' 2(א) לחוק.
- 62 זמיר, לעיל ה"ש 57, בעמ' 1278.
- 63 ברק-ארז, לעיל ה"ש 51, בעמ' 425.
- 64 שם.
- 65 ס' 33(ו) לחוק הנכים (תגמולים ושיקום) [נוסח משולב], התשי"ט-1959.
- 66 ס' 75(ג) לחוק המקרקעין, התשכ"ט-1969.
- 67 ס' 40(א) לחוק בתי דין מנהליים, התשנ"ב-1992.
- 68 צ"א (מחוזי מר') 24474-01-22 ברגר נ' מדינת ישראל (נבו 11.9.2022).
- 69 שם, בעמ' 8.
- 70 שם. לאחרונה פורסם כי פעילות המערכת הופסקה בסמוך למועד מתן פסק הדין בעניין ברגר. ראו תומר גנון "הפרקליטות מאשרת לבג"ץ: 'הופסק השימוש במערכת הפרופיילינג בנתב"ג'" כלכליסט (15.8.2024), https://www.calcalist.co.il/local_news/article/rjawbis5a.

גופים שיפוטיים

על מרכזיותה של חובת הנמקה שיפוטית ניתן לעמוד מתוך תפיסה של שני התפקידים השונים של בית המשפט בהקשר החברתי: הראשון הוא **פתרון סכסוכים** והשני הוא **יצירת תקדים** הנוצר באמצעות פרסום פומבי של הסיבות שעמדו ביסוד פתרון הסכסוכים הנ"ל.⁷¹ תקדימים אלו (ונימוקיהם בצידם) מהותיים עבור שורה ארוכה של בעלי עניין כגון הציבור הרחב, יועצים, רגולטורים, ערכאות נמוכות וכיו"ב, העושים בהם שימוש לתכנון חייהם ופעולותיהם. חובת ההנמקה הוטמעה בשיטת המשפט הישראלית מיסודה, בניגוד להתרחבותה האיטית בשיטות משפט מערביות אחרות.⁷² למעשה, היעדר הנמקה המלווה החלטה שיפוטית נחשב "לדבר כפירה של ממש".⁷³ חובת הנמקה של החלטה שיפוטית כוללת הסבר של הוראת החוק שעליה היא מבוססת, ביאור מדוע היא חלה (או אינה חלה) על המקרה הספציפי הנידון, ובמקרה שבו אין הוראת חוק קיימת – על אילו עקרונות התבסס בית המשפט במתן החלטתו.⁷⁴ ברם, בתי משפט אינם רק יצרנים של הנמקות בעצמם, אלא גם מבקרים של הנמקות שניתנו על ידי גופים אחרים בשבתם של בתי המשפט כערכאות ערעור. ביקורת זו נעשית על נימוקים של ערכאות נמוכות,⁷⁵ דברי חקיקה,⁷⁶ והחלטות של רשויות מנהליות.⁷⁷ בשבתו כערכאת ערעור, בית המשפט אוסף את חובת ההנמקה המוטלת על גופים אלו הלכה למעשה. יפים לענייננו דבריה של כב' השופטת (בדימוס) עדנה ארבל בעניין **מקומן כל הצפון**: "ההנמקה הינה חלק בלתי נפרד מכל פסק דין. ההנמקה היא

71 William M. Landes & Richard A. Posner, *Adjudication as a Private Good*, 8 J. L. Stud. 235, 236 (1979).

72 ראו דורון מנשה "תכליתה ומהותה של חובת ההנמקה העובדתית במשפט הישראלי" **עלי משפט** יא 399, 404–406 (2014).

73 שם, בעמ' 403.

74 ראו רות גביון "בית המשפט וחובת ההנמקה" **משפטים** ב 89, 93 (התש"ל–התשל"א). השוו למיגל דויטש "התקנה התמוהה ברפורמת סדרי הדין: שופט לא חייב לנמק דחיית ערעור" **גלובס** (7.2.2019), <https://www.globes.co.il/news/article.aspx?did=1001272204>.

75 מנשה, לעיל ה"ש 72, בעמ' 403.

76 אכן, לא תמיד כפופים המחוקקים לחובת נימוק פוזיטיבית (ראו James J. Brudney, *Congressional Commentary on Judicial Interpretations of Statutes: Idle Chatter or Telling Response?*, 93 MICHIGAN L. REV. 1, 35 (1994)). ברם, ערכאת הערעור יכולה להשתמש באינדיקציות אחרות כדי ללמוד על שיקול הדעת של המחוקקים (ראו JOHN FINNIS, *REASON IN ACTION: COLLECTED ESSAYS VOLUME I*, 234 (2011)), או לבחון את החלטת הרשות המחוקקת בפרספקטיבה הידועה כ-"Excluded Reasons", דהיינו לבחון אילו נימוקים אינם מקובלים כתמיכה להחלטה (ראו Richard H. Pildes, *Avoiding Balancing: The Role of Exclusionary Reasons in Constitutional Law*, 45 HASTINGS L.J. 711, 712 (1994)).

77 יואב דותן "חובת הנמקה של רשויות מינהל וגופים נבחרים" **מחקרי משפט** יט 5, 17 (2002).

שמאפשרת לצדדים להבין את הבסיס לפסק הדין והיא שמאפשרת לבית משפט שלרעור לבקרו. יתרה מזאת, להנמקה חלק חשוב בתהליך של גיבוש ההחלטה ובלעדיה לא ניתן לומר שזכה להתברר בבית המשפט.⁷⁸

מן הסקירה הקצרה שלעיל עולות שתי תכליות חשובות לחובת הנמקה, כדבריו היפים של הנשיא (בדימוס) משה לנדוי: "הנמקה מלאה מחנכת את השופט למחשבה ברורה ולהעלאת נימוקיו – כולל תחושתו האינטואיטיבית, עליה דיבר פאונד – מעל לסף תודעתו אל אור היום, על מנת שיעמדו למבחן הביקורת של דרג הערעור, של אנשי המקצוע ושל הצבור כולו".⁷⁹ ללמדנו כי חובת ההנמקה במגזר הציבורי תכליתה כפולה היא – לחדד את שיקול הדעת ("מחשבה ברורה")⁸⁰ ולאפשר ביקורת על ההחלטה שניתנה.⁸¹ במילים אחרות, משכיחות השימוש בחובת ההנמקה במגזר הציבורי ניתן לחלץ שתי תכליות חשובות – שיפור איכות ההחלטה ומימוש הזכות להליך הוגן.

חובת הנמקה במשפט הפרטי

חובת הנמקה נתפסת כשייכת בעיקרה לממלכת המגזר הציבורי. ברם, כחלק מאשפת החיצים הרגולטורית הנקראת "שקיפות", היא הולכת וקונה לה אחיזה גם במגזר הפרטי. מעבר להחלת חובת ההנמקה במצבים כגון גופים דו-מהותיים⁸² וגופים שהוענקה להם סמכות על פי דין,⁸³ דיני הגנת הצרכן, דיני חוזים ודיני נזקין כוללים חובות הנמקה, גילוי ושיקיפות מסוגים שונים, בין אם מכוח דין או מכוח נוהג (בדומה להנמקה, "גילוי" מהווה את אחת מהדרכים למימוש ערך השקיפות).⁸⁴ לדוגמה, חקיקה

78 רע"א 1982/05 מקומון כל הצפון נ' פלוני, 4 (נבו 2006.1.23).

79 משה לנדוי "הלכה ושיקול דעת בעשיית משפט" ספר לנדוי כרך א 17, 31 (אהרן ברק ראינער מזוז עורכים 1995).

80 ראו בעניין זה ניסוי שבמסגרתו נתבקשו שופטים לנמק את החלטותיהם לפני מתן ההחלטה הסופית. התוצאות הראו השפעה חיובית על יכולת השופטים לנטרל משתנים שאינם רלוונטיים ואף מפלים שהוטמעו במכוון בתיאור המקרים שהובאו בפניהם. Liu, לעיל ה"ש 11.

81 ברק-ארז, לעיל ה"ש 51, בעמ' 435.

82 גדי פרל ותהילה שוורץ אלטשולר מודל ליצירת שקיפות אלגוריתמית 19 (2022).

83 זמיר, לעיל ה"ש 57, בעמ' 1274.

84 ראו למשל Bernard I. Finel & Kristin M. Lord, *The Surprising Logic of Transparency*, 43 INT'L. STUD. Q. 315, 316 (1999) ("we define transparency as legal, political, and institutional structures that make information about the internal characteristics of a government and society available to actors both inside and outside of the domestic political system [...] In general, transparency is increased by any mechanism that leads to the public disclosure of information such as a free press, open government hearings, and the existence of nongovernmental organizations with an incentive to release objective information about the government"; פרל ושוורץ אלטשולר, לעיל ה"ש 82. אולם ראו הבחנה בין גילוי וולונטרי לבין גילוי מנדטורי שיסודו בדרישה רגולטורית, David Weil et al., *The*

בתחום של הגנת הצרכן כוללת חובת גילוי לצרכן בנושאים שונים,⁸⁵ המכבדת את יכולתו של הצרכן לקבל החלטות מידעיות ומושכלות כיצור אוטונומי. באופן דומה, מערכות יחסים חוזיות כוללות חובות גילוי ושקיפות⁸⁶ העומדות ביסוד התחייבות הצדדים עצמה. בנוסף, דיני הגנת המידע באירופה דורשים אף הם הסכמת הפרט לעיבוד מידע על אודותיו שתהיה, בין היתר, מודעת.⁸⁷ אולם המופע המובהק ביותר של חובת הנמקה במשפט הפרטי הוא הדרישה למתן "הסכמה מדעת" להליך רפואי.⁸⁸ הסכמה זו מהווה עקרון יסודי בביר-אטיקה,⁸⁹ והיא בנויה על כך שהמטופל קיבל מידע ואינפורמציה אשר ביכולתו להבין.⁹⁰ מידע זה עומד במרכזה של יכולת המטופל לתת הסכמה להליך המוצע⁹¹ ומסירתו מבטאת את "הרעיון שזכותו של הפרט לאוטונומיה עשויה, בנסיבות מסוימות, להטיל על פרטים אחרים את החובה לנקוט פעולות לשם הבטחת קיומם של התנאים להיות החלטת הפרט אוטונומית".⁹² יודגש כי "על-מנת שהסכמתו של חולה לטיפול רפואי שעתיד להיעשות בגופו תהא 'הסכמה מדעת' אין די בהחתמת המטופל על טופס הסכמה, אלא יש לספק לו מידע הולם על מצבו, על מהות הטיפול המומלץ ומטרתו, על הסיכונים והסיכויים הטמונים בו ועל אלטרנטיבות טיפוליות סבירות לטיפול האמור".⁹³ בשאלה המורכבת איזה מידע יש למסור למטופל בהקשר זה נקבע כי "המבחן למידע שחובה למסור למטופל... [הינו] צרכיו של המטופל

Effectiveness of Regulatory Disclosure Policies, 25 J. POL'Y ANALYSIS & MGMT. 155 (2006).

85 ראו ס' 4 לחוק הגנת הצרכן, התשמ"א-1981. ראו גם Hardy Cross Dillard & Harris, *Product Liability: Directions for Use and the Duty to Warn*, 41 VA. L. REV. 145 (1955).

86 ראו הגדרת "הטעיה" בס' 15 לחוק החוזים (חלק כללי), התשל"ג-1973, הכוללת "אי-גילויין של עובדות אשר לפי דין, לפי נוהג או לפי הנסיבות היה על הצד השני לגלותן". ראו גם Melvin A. Eisenberg, *Disclosure in Contract Law*, 91 CAL. L. REV. 1645 (2003).

87 סעיף 4 ל-GDPR, לעיל ה"ש 18. לדיון בניתוח סוגיית ההסכמה בהקשר של פרטיות ראו למשל מיקי זר הפוליטיקה של הפרטיות: הדרמה הטכנולוגית בזירת המידע 89-93 (2022). ראו גם ביקורת על מידת ההצלחה של כלי רגולטורי זה לאפשר לאינדיבידואלים להגן על פרטיות המידע שלהם, Frauke Kreuter et al., *Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent*, 38 SOC. SCI. COMPUT. REV. 53 (2020).

88 פרק ד לחוק זכויות החולה, התשנ"ו-1996.

89 McLEAN SHEILA, *AUTONOMY, CONSENT AND THE LAW* 6 (2010).

90 שם, בעמ' 41.

91 ALASDAIR MACLEAN, *AUTONOMY, INFORMED CONSENT AND MEDICAL LAW: A RELATIONAL CHALLENGE* 134 (2009).

92 נילי קרקו-אייל דוקטרינת ההסכמה מדעת בחוק זכויות החולה 300 (2008).

93 ראו ע"א 2781/93 דעקה נ' בית החולים "כרמל" חיפה, פ"ד נג(4) 548, 526 (1999).

למידע כדי להחליט אם לקבל את הצעת רופאו.⁹⁴ דרישות שקיפות, גילוי, הסבר והנמקה בתחום האזרחי חושפות בפנינו אפוא תכלית נוספת העומדת ביסודה של חובת ההנמקה – הכרה באוטונומיה האנושית.

חובת הנמקה במשפט הבינלאומי

אחד הממדים המפתיעים שבהם חובת ההנמקה צוברת תאוצה בשנים האחרונות הוא יחסים בינלאומיים. בהדרגה, הסיבות לפעולתה של מדינה הופכות להיות חשובות באותה מידה כמו הפעולה עצמה.⁹⁵ ההתפשטות (לעיתים מנדטורית אך לרוב וולונטרית) של תופעת מתן הסברים ונימוקים על ידי מדינות מושפעת מ"תרבות ההצדקה"⁹⁶ ההולכת וקונה לה אחיזה בדינים בינלאומיים,⁹⁷ הן מנקודת מבט של *Opinio Juris*⁹⁸ (עמדה משפטית מקובלת) והן מתוך הסכמה ומחויבות לנוהג המקובל.⁹⁹ באותה מידה היא מגלמת את היותה של המדינה – כמקבלת החלטות – חלק ממארג ואקוסיסטם המקיף אותה ואשר עליו היא שואפת להשפיע כדי ליהנות מלגיטימציה ותמיכה רבות יותר ובכך לחזק את אפקט קבלת החלטותיה על ידי הקהילה הבינלאומית. בישראל, החלטת ממשלה מס' 1936 בעניין הקמת מערך הסברה לאומי היא דוגמה למגמה זו.¹⁰⁰ ההחלטה שמה דגש מיוחד על חשיבות הסברה פעולתיה של ישראל בעולם, בצינה "ברחל בתן הקטנה" כי "הסברה ישראל בעולם הינה מרכיב מרכזי בביטחון הלאומי של מדינת ישראל וכלי לחיזוק מעמדה, דמותה

94 ע"א 434/94 ברמן קטינה באמצעות הוריה נ' מור – המכון למידע רפואי בע"מ, פ"ד נא(4) 205, 206 (1997).

95 Chimène L. Keitner, *Explaining International Acts*, 63 McGill L. J. 649, 651 (2018) דוגמאות עדכניות לתהליך זה ניתן לראות בעלייתו של תחום Global Administrative Law (ראו למשל Benedict Kingsbury, *The Concept of 'Law' in Global Administrative Law*, 20 EUR. J. INT'L L. 23, 41 (2009) וכן בחוקים ואמנות המסדירים סחר בינלאומי, ראו למשל Maurizia de Bellis, *A Duty to Provide Reasons: Definitive Safeguards Measures on Imports of Certain Steel Products*, in GLOBAL ADMINISTRATIVE LAW: CASES, MATERIALS, ISSUES 81, 81–85 (Sabino Cassese et al., eds., 2008).

96 Keitner, שם בעמ' 651 (The 'culture of justification' that exists at the international level includes an expectation that states will articulate the legal and policy bases for their actions, particularly when such actions depart from accepted norms of state behavior).

97 Kingsbury, לעיל ה"ש 95, בעמ' 47.

98 Harold Hongju Koh, *The Legal Adviser's Duty to Explain*, 41 YALE J. INT'L L. 189, 190 (2016).

99 Jo Lynn Slama, *Opinio Juris in Customary International Law*, 15 OKLA. CITY U. L. REV. 603, 603–604 (1990).

100 החלטה 1936 של הממשלה "הקמת מערך הסברה לאומי" (8.7.2007).

ותדמיתה של ישראל בעולם...¹⁰¹ בכך נחשפת תכלית חשובה נוספת ביסוד השימוש בנימוק להחלטות ולמעשים – הרצון להשפיע על הקהילה (הבינלאומית וכן הקהילה מבית), לזכות בלגיטימציה להחלטות המנומקות ולחזק את סמכות מערכת קבלת ההחלטות עצמה.

2. התכליות ביסוד חובת ההנמקה

הסקירה הקצרה של מופעי חובת ההנמקה במשפט הציבורי, הפרטי והבינלאומי שהובאה לעיל מלמדת על התכליות העומדות בבסיסה של חובת ההנמקה אשר, על אף שורשיה האיתנים במשפט הציבורי, הולכת וקונה לה אחיזה בתחומי משפט נוספים, כגון האזרחי והבינלאומי. תכליות אלו כוללות: (1) שיפור איכות ההחלטה; (2) מימוש הזכות להליך נאות; (3) הכרה באוטונומיה האנושית ו־ (4) חיזוק מערכת קבלת ההחלטות. חלק זה ירחיב את היריעה בנוגע למשמעות תכליות אלו.

שיפור איכות ההחלטה: בלב השימוש בחובת ההנמקה נעוצה יכולתה לשפר את ההחלטה,¹⁰² והעובדה כי היא משקפת את התימוכין החוקיים והמוסריים העומדים ביסודה.¹⁰³ שיפור ההחלטה נובע משורה של תהליכים פנימיים וחיזוניים המופעלים עקב הצורך לנמק החלטות. ראשית, עצם ההכרח להציג נימוקים משכנעים מוביל להאטה אוטומטית של תהליך קבלת ההחלטה,¹⁰⁴ שהרי "הנמקה דורשת מחשבה סדורה והגיונית לפני קבלת ההחלטה".¹⁰⁵ כמנגנון בקרת איכות פנימי,¹⁰⁶ חובת נימוק דוחפת את מקבל ההחלטה לבחינה מדוקדקת של השיקולים שהובילו לקבלת החלטה מסוימת¹⁰⁷ ולזיהוי סיבות שאינן לגיטימיות כאשר הן מהוות את התימוכין היחידים להחלטה המסתמנת.¹⁰⁸ יש המרחיבים לטעון כי די באפשרות שמקבלי החלטות יצטרכו לנמק את החלטותיהם כדי להשיג את אפקט האטה ההחלטה וצמצום הסיכון להחלטות שרירותיות שאינן הוגנות,¹⁰⁹ כלומר, ייתכן שיש בעצם קיום חובת הנימוק העתידית –

101 שם, בס' 13.

102 Ashley S. Deeks, *Secret Reason-Giving*, 129 YALE L.J. 612, 627 (2019). ראו גם מנשה, לעיל ה"ש 72, בעמ' 408.

103 Katie Atkinson et al., *Explanation in AI and Law: Past, Present and Future*, 289 A.I. 1, 2 (2020); Lon L. Fuller, *The Forms and Limits of Adjudication*, 92 HARV. L. REV. 353, 388 (1978).

104 Schauer, לעיל ה"ש 45, בעמ' 657. ראו גם Deeks, לעיל ה"ש 102, בעמ' 667.

105 זמיר, לעיל ה"ש 57, בעמ' 1269.

106 Katherine J. Strandburg, *Rule Making and Inscrutable Automated Decision Tools*, 119 COLUM. L. REV. 1851, 1868 (2019).

107 Shapiro, לעיל ה"ש 52, בעמ' 180.

108 Schauer, לעיל ה"ש 45, בעמ' 658.

109 Shapiro, לעיל ה"ש 52, בעמ' 184.

אף אם לא תמומש בפועל ודרישת נימוק לא תוצב – כדי להפעיל לחץ פסיכולוגי על מקבלי החלטות לייצר החלטות טובות, חוקיות והוגנות יותר.¹¹⁰ יתרון נוסף הטמון בחובת ההנמקה הוא העובדה כי היא מעודדת דיון ציבורי בנוגע להחלטות, היכול בעצמו להשפיע על החלטות קיימות ועתידיות, להאיר זוויות ראייה חדשות ומגוונות ולהעשיר את מנגנון קבלת ההחלטה בפרט, ובפורום הציבורי בכלל.¹¹¹ בהינתן שמגוון דעות עשוי לעיתים להוביל לשינוי מגמה עתידי,¹¹² חובת הנמקה חשובה אפוא לדיון הציבורי גם – ואולי בפרט – כאשר רשות ערעור אינה נתונה (כגון החלטות ערכאה עליונה).

הכרה באוטונומיה האנושית: אחת מהתכליות המצויות ביסוד חובת ההנמקה היא הגושפנקא לאוטונומיה האנושית,¹¹³ הן של מקבל ההחלטה והן של נשוא ההחלטה. עבור נשוא ההחלטה, מתן נימוקים מכבד את היותו אדם אוטונומי. נימוקים משקפים את העובדה שמכלול השיקולים הרלוונטיים נלקח בחשבון, לרבות שיקולים המשפיעים על נשוא ההחלטה ושהוא אינו רק "כלי על לוח השחמט".¹¹⁴ מתן נימוקים אף משרת את יכולתו של נשוא ההחלטה לערער עליה או לכל הפחות לבקר אותה, ובכך הוא הופך לחלק בלתי נפרד מתהליך קבלת ההחלטה עצמו.¹¹⁵ מבחינת מקבל ההחלטה, קיומם של נימוקים התומכים בה מאשש את האוטונומיה העצמית שלו עצמו, בהיותו בעל שיקול דעת, יכולת להבחין בין טוב לרע וחוש מוסרי. תפיסה אריסטוטלית זו מתייחסת לשיקול הדעת כחלק מהותי מהיות בני האדם יצורים רציונליים,¹¹⁶ אשר שואפים מעצם טבעם לקיומו של רציונל משמעותי למעשיהם.¹¹⁷ יש בהנמקה כדי לבטא את ציוויו של עמנואל קאנט כי על האדם לראות הן את עצמו והן את סובביו כתכלית ולא כאמצעי,¹¹⁸

Mathilde Cohen, *The Rule of Law as the Rule of Reasons*, 96 ARCHIV FÜR RECHTS- UND SOZIALPHILOSOPHIE 1, 15 (2010) 110

Niva Elkin-Koren, *Contesting Algorithms: Restoring the Public Interest in Content Filtering by Artificial Intelligence*, 7(2) BIG DATA & SOC'Y 2, 4 (2020) 111
Elkin-Koren & Maayan Perel, *Speech Contestation by Design: Democratizing Speech Governance by AI*, 50 FLORIDA ST. U. L. REV. 611, 618–623 (2022)

עדות לכך ניתן למצוא למשל בעובדה שדעות מיעוט הופכות לעיתים, ברבות השנים, לעמדות רוב בפסיקה. 112

Gillis & Simons, לעיל ה"ש 20, בעמ' 75. 113

Jerry L. Mashaw, *Small Things like Reasons are Put in a Jar: Reason and Legitimacy in the Administrative State*, 70 FORDHAM L. REV. 17, 19 (2001) 114

Jerry L. Mashaw, *Reasoned Administration: The European Union, the United States, and the Project of Democratic Governance*, 76 GEO. WASH. L. REV. 99, 104–105 (2007); ראו גם CHARLES LARMORE, *THE MORALS OF MODERNITY* 137 (1996) 115

John Gardner, *The Mark of Responsibility*, 23 OXFORD J. L. STUD. 157, 158–159 (2003) 116

שם. 117

IMMANUEL KANT, *KANT: THE METAPHYSICS OF MORALS* XXVII (Lara Denis ed., R. George Wright, *Treating Persons as* Mary Gregor trans., 2nd ed. 2017) 118

ובמילים אחרות, החלטות מבוססות נימוקים הן תנאי בסיסי לאוטונומיה האנושית ומחזקות אותה בעצם קיומן.¹¹⁹

מימוש הזכות להליך נאות: חובת הנמקה נחשבת לרוב כאמצעי למימוש זכויות אחרות,¹²⁰ כגון הזכות להליך נאות,¹²¹ הטומנת בחובה את הזכות לדיון¹²² ולערעור על ההחלטה.¹²³ אין מחלוקת על חשיבות ידיעת הנימוקים כדי לתקוף החלטה בצורה מושכלת, אך יש לנימוקים תפקידים נוספים בהקשרי ערעור. ראשית, ידיעת הנימוקים מאפשרת לצד התופס עצמו כנפגע מההחלטה להעריך את סיכויי הצלחת הערעור בטרם ישקיע משאבי זמן וכסף בתקיפתה. שנית, בנסיבות מסוימות, ידיעת הנימוקים עשויה אף לשכנע את נשוא ההחלטה לראותה באור חדש וחיובי יותר. שלישית, חובת נימוק בפני עצמה מחייבת יצירת תיעוד מצד מקבל ההחלטה, תיעוד אשר עשוי לשמש לא רק את נשוא ההחלטה אלא גם את הרשות האמונה על שמיעת הערעור. אותו תיעוד מסייע בידי רשות הערעור לשחזר את שיקול הדעת שהוביל לקבלת ההחלטה ולגבש את הביקורת ביחס אליו.¹²⁴ רביעית, נימוקים מאפשרים למקבל ההחלטה עצמו לזווג הליך ביקורת פנימי שיכלול הן העמדת הגנה בפני הערעור והן בחינה מחודשת של ההחלטה והפגמים שהיא מציפה בנוגע להליך קבלת ההחלטות עצמו.¹²⁵

חיזוק הסמכות של מערכת קבלת ההחלטות: זו אחת מהתכליות החשובות ביותר בלב קיומה של חובת הנמקה.¹²⁶ חובת הנמקה תורמת לחיזוק הסמכות של מערכת קבלת ההחלטות באמצעות אישוש הלגיטימיות שלה, עיגון יכולת ההסתמכות על החלטותיה והכוונת התנהגות בהתאם להן, והעצמת תכונת האחריות של המערכת עצמה.

חיזוק הלגיטימיות של המערכת חשוב בהיותה של הלגיטימיות המפתח לקבלת החלטותיה על ידי הציבור הרחב.¹²⁷ חובת הנמקה מחזקת את הלגיטימיות והקבלה של

-
- Ends in Themselves: The Legal Implications of a Kantian Principle*, 36
U. RICH. L. REV. 271 (2002).
- Ruth Bader Ginsberg, *The Obligation to Reason Why*, 37 U. FLORIDA L. REV. 205, 119
222 (1985).
- Mathilde Cohen, *Reasons for Reasons*, in APPROACHES TO LEGAL RATIONALITY 119, 120
120 (Dov M. Gabbay et al. eds., 2010).
- Henry J. Friendly, *Some Kind of Hearing*, 123 U. PENNSYLVANIA L. REV. 1267, 121
1293–1294 (1975).
- שם. 122
- Kaminski & Urban, לעיל ה"ש 11, בעמ' 1974. 123
- Shapiro, לעיל ה"ש 52, בעמ' 183. 124
- Aaron L. Nielson & Christopher J. Walker, *The New Qualified Immunity*, 89 S. 125
CALIF. L. REV. 1, 59 (2015).
- Cohen, לעיל ה"ש 120, בעמ' 121. 126
- Tom R. Tyler & Jonathan Jackson, *Future Challenges in the Study of Legitimacy* 127
and Criminal Justice, in LEGITIMACY IN CRIMINAL JUSTICE: AN INTERNATIONAL
EXPLORATION 83, 88 (Justice Tankebe & Alison Liebling eds., 2013).
-

מערכת קבלת ההחלטות משום שהיא מקטינה את החשש מהחלטה שרירותית,¹²⁸ מחייבת את המערכת לבסס את ההחלטה על תהליך סדור ושקול,¹²⁹ מסייעת לציבור להבין את צדקתה של ההחלטה,¹³⁰ מעודדת דיון ציבורי הכולל מגוון רחב של השקפות ועמדות,¹³¹ ושומרת על ערך אמון הציבור.¹³²

יכולת ההסתמכות על החלטות המערכת חשובה אף היא, משום שהיא מאפשרת לכפופים לה לפעול לפי הנורמות החברתיות הגלומות בהחלטותיה ולתכנן את התנהגותם בהתאם.¹³³ במובן זה יש לנימוקים תפקיד חשוב אף יותר מההכרעה עצמה, שהינה מטבעה ספציפית וייחודית לנסיבות המקרה, בעוד שהנימוקים חושפים את שיקול הדעת, הנורמות והכללים שהובילו להכרעה הנ"ל.¹³⁴ יודגש כי לתכלית ההסתמכות הטמונה בנימוקים יש השפעה גם על מערכת קבלת ההחלטות עצמה, הן משום שהיא מכפיפה את מערכת קבלת ההחלטות לאותו רציונל הגלום בנימוקים שניתנו גם בהחלטות עתידיות¹³⁵ והן משום שרציונל זה מעודד אכיפה-עצמית של הנתינים וקונפורמיות ביחס לכללי המערכת.¹³⁶

לבסוף, נימוקים מחזקים את האחריות של המערכת, באמצעות הפעלת שורה של לחצים רגשיים-חברתיים על מקבל ההחלטה האנושי. פרופ' Jerry Mashaw מחיל על קבלת החלטות במערכת המשפטית תאוריה כללית בנוגע לנימוקים שפותחה על ידי

- 128 זמיר, לעיל ה"ש 57, בעמ' 1269. ראו גם גביוון, לעיל ה"ש 74, בעמ' 92.
- 129 ברק-ארז, לעיל ה"ש 51, בעמ' 423–424.
- 130 מנשה, לעיל ה"ש 72, בעמ' 407.
- 131 Glen Staszewski, *Reason-Giving and Accountability*, 93 MINNESOTA L. REV. 1253, 1278 (2009).
- 132 פרל "פרטיות, שליטה ופיקוח", לעיל ה"ש 14, בעמ' 189. על פי התאוריה של "צדק תהליכי" (Procedural Justice) די בכך שההליך עצמו נתפס כהוגן כדי לתרום לתפיסת הלגיטימיות של המערכת. ראו למשל Tom R. Tyler, *Psychological Perspectives on Legitimacy and Legitimation*, 57 ANN. REV. PSYCH. 375, 382 (2006). מחקרים מאוחרים יותר מפתחים תאוריה זו ומעמידים את הלגיטימיות של המערכת על שלוש רגליים: א. עוצמת החובה לציית למערכת; ב. רמת הביטחון והאמון במערכת; ו-ג. באיזו מידה החלטות המערכת תואמות את מערכת הערכים של הקהילה. ראו Tom R. Tyler & Jonathan Jackson, *Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation, and Engagement*, 20 PSYCH. PUB. POL'Y & L. 78 (2014).
- 133 גביוון, לעיל ה"ש 74, בעמ' 92.
- 134 Schauer, לעיל ה"ש 45, בעמ' 641.
- 135 שם, בעמ' 656.
- 136 Mark J. Richards & Herbert M. Kritzer, *Jurisprudential Regimes in Supreme Court Decision Making*, 96 AM. POL. SCI. REV. 305, 305 (2002); MARTIN M. SHAPIRO, *THE SUPREME COURT AND ADMINISTRATIVE AGENCIES* (1968).

פרופ' Charles Tilly.¹³⁷ תאוריה זו רואה מתן נימוקים כמיזם בינאיש במהותו,¹³⁸ ובהתאם מזהה סט לחצים המופעל על מקבל החלטה אנושי, לרבות רצון להימנע מביקורת פנים והסנקציות שבצידה, שאיפה לזכות בשיתוף פעולה ובהערכה מצד קולגות וחשש מביקורת ציבורית.¹³⁹ סט לחצים זה מקדם בעת ובעונה אחת תחושת אחריות פנימית של מקבל ההחלטה¹⁴⁰ וגם מחיל עליו סנקציות בצורה חיצונית במקרה של סטייה מאחריות זו. הנ"ל, בתורן, מביאות להנפקת החלטות איכותיות יותר באמצעות המערכת, ובכך תורמות למידת הסמכות הנתפסת שלה על ידי הציבור.¹⁴¹ סיכומי של דבר, חובת הנמקה הינה חובה משפטית החלה על החלטות אנושיות ונועדה למלא תכליות הכרוכות בשיקול הדעת האנושי. בחינת מהותו של מושג ה"הסבר" בעולם המשפט, או חובת ההנמקה, שהובאה לעיל, מדגישה הן את השימוש הנפוץ במכשיר זה והן את התכליות העומדות ביסודו. הפרק הבא יסקור את הכלי הטכנולוגי ליצירת הסברים להחלטות מכונה הידוע כ-XAI, ויבחן את יכולתו של הטכניקות ליצירת הסברים למערכות בינה מלאכותית הנכללות בו למלא את אותן תכליות משפטיות.

ג. XAI – טכנולוגיה למימוש הזכות להסברות?

חלקים א רב של מאמר זה תיארו עד כה את עליית הכלי הרגולטורי הידוע כ"זכות להסברות" ביחס למערכות בינה מלאכותית וכן את היסוד המשפטי בבסיסו של כלי זה – חובת ההנמקה. אולם, במקביל לתהליך זה, התחום הטכנולוגי האמון על פיתוח, הטמעה ושימוש במערכות אלו ("למידת מכונה") החל לפתח ביוזמתו שלו פרקטיקות להגברת רמת ההסברות שלהן. המטרה ביסוד פיתוח הטכנולוגיה של למידת מכונה היא קבלת החלטות אוטומטית ברמות דיוק גבוהות מאלו של בני אדם, אולם דיוק אינו חזות הכול ובהתאם קיימים משתנים חשובים נוספים כגון מידת האמינות של המערכת, היכולת להעריך נכונה את מידת הדיוק שלה ומידת הצלחתה להתמודד עם התקפות אדברסריות.¹⁴² שדה המחקר הידוע כ-XAI eXplainable AI¹⁴³ התפתח בשנים

CHARLES TILLY, WHY? WHAT HAPPENS WHEN PEOPLE GIVE REASONS... AND WHY 137 (2006). פרופ' טילי מפתח בספרו הבחנה בין סוגי נימוקים בהתאם לסוג מערכת היחסים שבמסגרתה הם נמסרים.

Mashaw, לעיל ה"ש 115, בעמ' 101.

Rebecca Ingber, *Bureaucratic Resistance and the* שם, בעמ' 101–103. ראו גם National Security State, 104 IOWA L. REV. 139, 164–165, 183–184 (2018).

ראו דותן, לעיל ה"ש 77, בעמ' 9.

אך השוו Lerner & Tetlock, לעיל ה"ש 11.

Wojciech Samek et al., *Explaining Deep Neural Networks and Beyond: A* ראו למשל Review of Methods and Applications, 109 PROC. IEEE 247 (2021).

Alejandro Barredo Arrieta et al., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI*, 58 INFO. FUSION 82, 84 (2020).

האחרונות ועוסק בפיתוח טכניקות מתקדמות מסוגים שונים שמטרתן המשותפת הינה להסביר החלטות שהתקבלו על ידי מערכות בינה מלאכותית.¹⁴⁴ שורשיו של XAI נעוצים בשנות ה-80 וה-90 של המאה הקודמת,¹⁴⁵ אך הפיתוח והשימוש בטכנולוגיה של רשתות עצביות מלאכותיות,¹⁴⁶ המורכבות לעיתים ממיליארדים של פרמטרים, הפכה את העיסוק בו לדחוף מתמיד.¹⁴⁷ זאת, משום שסיבוכיות זו מאלצת את האדם (מפתח/ת, משתמש/ת וכו') לקבל לכאורה גישה לכמות עצומה של מספרים כדי להסביר מערכת מסוג זה. בפועל לא קיימת כיום דרך להבין מודל כה מורכב בצורה שלמה.¹⁴⁸

XAI מורכב מקשת של טכניקות מסוגים שונים,¹⁴⁹ רבות מהן יוצרות אלגוריתם פשוט יותר "המשחזר" את קבלת ההחלטה של האלגוריתם המורכב, בצורה הניתנת להבנה אנושית.¹⁵⁰ קיומן של טכניקות שונות נובע מכך שקיימת שונות רחבה בין סוגי מודלים,¹⁵¹ ובהתאם התפתחו טכניקות ייצור הסברים השואפות לספק הסבר לכמה סוגי מודלים (Model Agnostic) בצד טכניקות ייצור הסברים המתמקדות במודלים מסוג מסוים (Model Specific). בין הבולטים ניתן למנות את תת-התחום הקרוי "הסברים

144 מדיניות רגולציה ואתיקה בתחום הבינה המלאכותית בישראל, לעיל ה"ש 5, בעמ' 62.

145 Tim Miller et al., *Explainable AI: Beware of Inmates Running the Asylum or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences*, EPRINT ARXIV (2 dec. 2017), <https://arxiv.org/abs/1712.00547>.

146 להסבר בנוגע לרשתות נוירונים מלאכותיות ראו למשל המיזם הלאומי למערכות נבונות בטוחות להעצמת הביטחון הלאומי והחוסן המדעי-טכנולוגי: אסטרטגיה לאומית לישראל דו"ח מיוחד לראש הממשלה חלק א' 46, נספח ב' (2020) ("אחד הכיוונים המתקדמים והמבטיחים ביותר בתחום הבינה המלאכותית הוא טכנולוגיית הלמידה העמוקה [Deep Learning] שבה במקום להתבסס על 'אלגוריתמים מוכתבים' מראש לביצוע פעולות מנסים לחקות תהליך מוחי אנושי ע"י אימון לפעולה, הזנה בנתונים וניתוח ביצועה. למידה עמוקה מתבצעת באמצעות שכבות רבות של רשתות נוירונים מלאכותיות שכל אחת מטפלת במשימה חיונית פשוטה בתהליך עיבוד המידע, במטרה לאפשר יחד ביצוע של פעולה מורכבת").

147 Mohammad Nagahisarchoghaei et al., *An Empirical Survey on Explainable AI Technologies: Recent Trends, Use-Cases, and Categories from Technical and Application Perspectives*, 12 ELEC. 1092, 1092 (2023); Meike Nauta et al., *From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI*, 55(13) ACM COMPUTING SURV. 295 (2023).

148 Rudresh Dwivedi et al., *Explainable AI (XAI): Core Ideas, Techniques, and Solutions*, 55(9) ACM COMPUTING SURV. 194 (2023).

149 שם.

150 Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACH. INTEL. 206 (2019).

151 Samuli Laato et al., *How to Explain AI Systems to End Users: A Systematic Literature Review and Research Agenda*, 32(7) INTERNET RSCH. 1, 70 (2022).

בדיעבד" (Post-Hoc Explanations), המייצר הסברים לאחר מתן ההחלטה האוטומטית וכולל למשל הסברים ויזואליים (כגון Correlation Plots), וכן הסברים המתמקדים במאפיינים החשובים ביותר (Feature Importance) שהובילו לקבלת ההחלטה האוטומטית (כגון SHAP ו-LIME).¹⁵² תת־תחום נוסף הינו הסברי "מה אם?" (Counterfactual Explanations) שבמסגרתו נבדק מהו השינוי הקטן ביותר שניתן לבצע בנתונים המוצגים למודל במקרה ספציפי ואשר יוביל לשינוי בהחלטה האוטומטית בנוגע לאותו מקרה.¹⁵³ הבחנה נוספת היא בין הסברים להחלטות ספציפיות של המערכת (Local Explanations) לבין הסברים המתייחסים למאפיינים הכוללים יותר של המערכת (Global Explanations), המספקים הבנה לגבי המידע שבו השתמשו לאימון המערכת, המתודולוגיה שמאחורי בחירת סוג מודל ספציפי, וכיו"ב.¹⁵⁴

קהילת ה-XAI עודנה שרויה בחוסר הסכמה לגבי המתודולוגיה הנכונה ליצירת הסברים, וחסרה כללים מוסכמים שיענו לשאלה כיצד יש לבחור בשיטת ייצור הסבר כזו או אחרת. בפועל, הטכניקות אשר פותחו על ידי הקהילה הטכנולוגית כדי לתקן ולשפר מערכות בינה מלאכותית מסיבות מקצועיות¹⁵⁵ טרם התגבשו לכדי פרקטיקה מקובלת למימוש ה"זכות להסברות",¹⁵⁶ עיקרון רגולטורי המהווה גם כיום אתגר טכנולוגי משמעותי.¹⁵⁷ אציע כי הסיבה לכך היא העובדה שקיים חוסר התאמה בסיסי בין כלי טכנולוגי שמטרתו איתור ותיקון תקלות במערכת (Debugging) והמיועד לקהל טכנולוגי מקצועי, לבין זכות משפטית המבוססת על התכליות ביסודה של חובת ההנמקה. בעוד שעבור יצרני מערכת בינה מלאכותית ומפתחיה המניע העיקרי לייצור הסברים להחלטותיהן הוא להגביר את מידת האמינות שלהן ולהימנע עד כמה שניתן

- 152 Nagahisarchoghaei et al., לעיל ה"ש 147, בעמ' 7–8.
- 153 Wachter et al., לעיל ה"ש 32, בעמ' 844–845. על מנת להבין את הרציונל שמאחורי הסברי "מה אם?" נשתמש בדוגמה של דחיית בקשה לקבלת הלוואה על ידי המודל. נבדוק מה יקרה אם נשנה את גובה ההכנסה החודשית של מבקשת הלוואה באמצעות תוספת של 1,000 ש"ח. האם המודל יחליט כעת כי היא ראויה לקבלת הלוואה? במידה שכן, מדובר בהסבר שמאפשר למבקשת הלוואה לבצע שינוי בחייה שיהפוך אותה בכל זאת לזכאית לקבלת הלוואה. מובן כי לטכניקת ייצור הסברים מסוג זה בעיות לא מעטות, ראו למשל, Dylan Slack et al., *Counterfactual Explanations Can Be Manipulated*, 35TH CONF. ON NEURAL INFO. PROCESSING SYS. 62 (2021).
- 154 Swetha Hariharan et al., *XAI for Intrusion Detection System: Comparing Explanations Based on Global and Local Scope*, 19 J. COMPUT. VIROLOGY AND HACKING TECHNIQUES 217 (2023).
- 155 Brent Mittelstadt et al., *Explaining Explanations in AI*, PROCS. CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 279, 283 (2019).
- 156 ELLEN P. GOODMAN & JULIA TRÉHU, AI AUDIT WASHING AND ACCOUNTABILITY 9 (2022).
- 157 Riccardo Guidotti et al., *A Survey of Methods for Explaining Black Box Models*, 51 ACM COMPUTING SURV. 1, 2 (2018).

מקיומן של תקלות וטעויות,¹⁵⁸ הקהילה המשפטית מחפשת למלא את התכליות העומדות ביסודו של המושג "הסבר" כפי שפורטו בחלק ב למאמר זה. בהתאם לאמור לעיל, אבחן להלן את יכולתן של פרקטיקות ייצור ההסברים הנכללות ב-XAI למלא תכליות משפטיות אלו בהצלחה.

XAI בשירות שיפור איכות ההחלטה: אחד התפקידים החשובים של חובת הנמקה הוא השפעתה על מקבל ההחלטה האנושי. השפעה זו באה לידי ביטוי באופנים שונים, כגון צמצום טווח שיקול הדעת האנושי וגידורו לטווח הנורמות החברתיות המקובלות. XAI מתקשה כיום לעמוד במשימה זו משלוש סיבות. ראשית, העובדה שמקבל החלטה אנושי מוחלף במכונה מאיינת את סט הלחצים החברתיים-רגשיים המצוי ביסוד השפעה זו. שנית, גם אם סט לחצים זה עשוי במידה מסוימת להשפיע על בני אדם הסובבים את המכונה (כגון מפתחים, משתמשים, מטמיעים וכו'), הרי שלא ברור מהי רמת השפעה זו.¹⁵⁹ שלישית, גם כשעסקינן בהחלטות אנושיות הנעזרות בהחלטות מכונה, הנטייה לסמוך בצורה עיוורת על החלטות מכונה¹⁶⁰ מקשה על מקבלי החלטה אנושיים להפעיל שיקול דעת עצמאי.¹⁶¹ בעוד שיש הקוראים להחלפת החלטות אנושיות בהחלטות אוטומטיות בדיוק בשל השאיפה להימנע מסט הלחצים החברתיים-רגשיים שהוזכר לעיל,¹⁶² ישנן סיטואציות שבהן קבלת החלטות המשקפת תכונות אנושיות עשויה דווקא להיות רצויה.¹⁶³ מקרה COMPAS עשוי לרמוז על כיוון לשימוש ב-XAI

158 Dwivedi et al., לעיל ה"ש 148.

159 ברי שעל מנת לבחון את מידת ההשפעה הזו ראוי לבצע בחינה אמפירית, שמאמר זה אינו מתיימר לבצע.

160 תופעה הידועה כהטיית האוטומציה (Automation Bias), שבעטייה בני אדם ייטו לקבל מסקנות שאליהן הגיעה מערכת קבלת החלטות אוטומטית בקלות יותר מכפי שיקבלו מסקנות אנושיות. ראו למשל- Linda J. Skitka et al., *Does Automation Bias Decision-making?*, 51 INT'L J. HUM. COMPUT. STUD. 991 (1999).

161 Malin Eiband et al., *The Impact of Placebic Explanations on Trust in Intelligent Systems*, EXTENDED ABSTRACTS 2019 CHI CONF. ON HUM. FACTORS COMPUTING SYS., David Lyell & Enrico ; 1 (2019) Kaminski & Urban, לעיל ה"ש 11, בעמ' 1961 ; Coiera, *Automation Bias and Verification Complexity: A Systematic Review*, 24 J. THE AM. MED. INFORMATICS ASS'N 423 (2017).

162 ראו למשל Daniel Kahneman et al., *Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making*, 94 HARV. BUS. REV. 38 (2016).

163 ההבנה כי קבלת החלטות אנושית טומנת בחובה יתרונות לחברה עומדת ביסוד דרישה רגולטורית נפוצה אחרת הידועה כ-Human in the Loop. דרישה זו מבקשת לשלב מקבלי החלטה אנושיים במערכות החלטה אוטומטיות מתוך כוונה להרוויח כמה אלמנטים חשובים: ראשית, עצם האנושיות עשויה להוות יתרון בהחלטות שכוללות שיקולים מוסריים ונורמות חברתיות. שנית, בני אדם מחזיקים בהקשרים וידע ביחס לעולם שמכונות כרגע מנועות מלהחזיק בו. שלישית, גמישות מחשבתית המלווה מקבלי החלטות אנושיים מהווה יתרון במצבי קיצון שבהם יש להפעיל שיקול דעת ולחרוג מתבנית קבלת החלטות מקובעת. רביעית, בני אדם הינם בעלי יכולת הכללה

לשיפור ההחלטה. בתחקיר שפורסם בשנת 2016 על ידי אתר התחקירים האמריקאי PROPUBLICA נטען כי COMPAS – אלגוריתם שנעשה בו שימוש במערכת המשפט ברחבי ארה"ב לחיזוי מידת הרצידיביזם (מועדות חשודים לשוב ולבצע עבירות) – מפלה על רקע גזעי.¹⁶⁴ מפתחת המערכת, חברת Northpointe Inc, פרסמה דו"ח המפריך לטענתה את תוצאות התחקיר.¹⁶⁵ ניכר כי המחלוקת נסבה סביב המודד הנכון לבדיקת "הוגנות" המערכת, בהינתן כמה מדדים אפשריים.¹⁶⁶ יחד עם זאת, יש מקום להנחה כי בעקבות טענותיה של PROPUBLICA ערכה Northpointe Inc בדיקה כלשהי (ולו בשל הצורך לייצר את הדו"ח המגן על המערכת מטעמה), וייתכן שהוסקו מסקנות לצמצום מידת האפליה בכל זאת. ניתן לדמיין השפעה דומה של הסברים המופקים באמצעות XAI על שיקול הדעת של מקבל ההחלטה האנושי המעורב בפיתוח, בהטמעה ובשימוש במערכת קבלת החלטות אוטומטית. כיוונים נוספים עשויים להיות

(Generalization), אנלוגיה והפשטה שמאפשרת ליישם כללים במצבים שונים. חמישית, אומנם נכון הדבר שהסברים אנושיים לא בהכרח משקפים את הליך קבלת ההחלטה האוטומטי, אבל במהלך השנים התפתחו טכניקות להתגברות על מכשול זה במידת מה, כגון מתן עדות וחקירה. ראו Rebecca Crotoft et al., *Humans in the Loop*, 76 VAND. L. REV. 429, 462–463 (2023). דוגמה למצבים שבהם שיקול דעת אנושי עדיף על זה של מכונה היא למשל תאונה קטלנית שבה נהרגה הולכת רגל המוליכה אופניים בשעות החשכה עקב טעות בזיהוי על ידי רכב אוטונומי במדינת אריזונה בשנת 2018. חקירה העלתה לכאורה כי המערכת אומנה על זיהוי הולכי רגל בסמוך למעברי חציה בלבד ועל כן לא זיהתה את הולכת הרגל שחצתה את הכביש באזור שלא כלל מעבר חציה. מובן כי טעות זיהוי שכזו לא הייתה מתבצעת בידי אדם, שלו מערך הקשרים רחב שאינו קושר בין הולכי רגל לבין מעברי חציה בהכרח, והיא מבטאת את השוני המובנה בין ראייה אנושית לבין "ראייה" של מערכות אוטומטיות. Phil McCausland, *Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk*, NBC NEWS (Nov. 9, 2019), <https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281>.

Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.¹⁶⁴

NORTHPOINTE INC., COMPAS RISK SCALES: DEMONSTRATING ACCURACY EQUITY AND PREDICTIVE PARITY (2016), *הגיבה לטענות Northpointe בהמשך*, Jeff Larson & Julia Angwin, *Machine Bias Technical Response to Northpointe*, PROPUBLICA (July 29, 2016), <https://www.propublica.org/article/technical-response-to-northpointe>.¹⁶⁵

Sam Corbett-Davies, Emma Pierson, Avi Feller & Sharad Goel, *A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear*, THE WASHINGTON POST (Oct. 17, 2016), <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas>.¹⁶⁶

פיתוח XAI לשם יצירת הצדקות ולא רק הסברים לקבלת החלטות אוטומטית,¹⁶⁷ שימוש בתוצרי XAI ללמידה של המודל,¹⁶⁸ ובמקביל עריכת ביקורת (Audit) שגרתית של מדדים של המערכת, כגון המידע שעליו אומנה ונבחנה, המשקלות השונים הקיימים בה וכיו"ב.¹⁶⁹

XAI בשירות ההכרה באוטונומיה האנושית: המעבר לקבלת החלטות אוטומטית כולל שקלול תמורות (Trade Off) בין דיוק ויעילות לבין שמירה על הגורם האנושי כפונקציה משמעותית בתהליך קבלת ההחלטה. במובלע, "המשמעות של קבלת החלטה שלא ניתן להבין או להסביר על ידי מכונה, עלולה לעורר חשש לפגיעה בכבוד האדם ובאוטונומיה שלו, בכך שהוא לכאורה נתון להחלטה שרירותית".¹⁷⁰ האם בכוחן של מכונות "לכבד" בני אדם ולהכיר באנושיות ובאוטונומיה של יצורים אנושיים? בעוד שישנם הקוראים להטמעת ערכים אנושיים, לרבות הכרה באוטונומיה האנושית, בפיתוח של מכונות,¹⁷¹ אחרים הכירו כבר בכך שמכונות אינן יכולות להוות סוכני מוסר, להיות בעלות חובות משל עצמן, או "לכבד".¹⁷² מנגד, ניתן לשאול האם קיומם של הסברים בדיעבד, או הסברים התורמים ליכולת נשוא החלטה לערער, בכל זאת ממלאים פונקציה של הכרה באוטונומיה האנושית של נשוא ההחלטה, ולו במידה חלקית. לדוגמה, אחד התחומים שבהם החשש לשחיקה באוטונומיה האנושית משמעותי הוא עלייתה של מדינת הרווחה הדיגיטלית, שבמסגרתה הולכים ומתרחבים השימושים האלגוריתמיים לחלוקת קצבאות, שירותים סוציאליים וכו'.¹⁷³ כניסתן של מערכות אוטומטיות להחלטות בתחומי הרווחה יוצרת סיכונים משמעותיים לזכויות אדם, בדגש על פגיעה בערכים כגון כבוד האדם, פרטיות והאוטונומיה האנושית.¹⁷⁴

- 167 אחד הכיוונים המעניינים יכול להיות למשל קידום יצירת הצדקות ולא דווקא הסברים. ראו למשל, בכלליות, Gillis & Simons, לעיל ה"ש 20.
- 168 ראו למשל: Roni Paiss, Hila Chefer & Lior Wolf, *No Token Left Behind: Explainability-aided Image Classification and Generation*, EUR. CONF. ON COMPUT. VISION 334 (2022).
- 169 לעניין ביקורות על מערכות בינה מלאכותית ראו למשל GOODMAN & TREHU, לעיל ה"ש 156.
- 170 משרד המשפטים ומשרד החדשנות, המדע והטכנולוגיה **מדיניות רגולציה** 2022, לעיל ה"ש 5, בעמ' 49.
- 171 Ibo Van de Poel, *Embedding Values in Artificial Intelligence (AI) Systems*, 30 MINDS AND MACHINES 385, 387 (2020). ראו גם Rafael A. Calvo, Dorian Peters, Karina Vold & Richard M. Ryan, *Supporting Human Autonomy in AI Systems: A Framework for Ethical Enquiry*, in ETHICS OF DIGITAL WELL-BEING: A MULTIDISCIPLINARY APPROACH 31 (Christopher Burr & Luciano Floridi eds., 2020).
- 172 Arto Laitinen & Otto Sahlgren, *AI Systems and Respect for Human Autonomy*, 4 FRONTIERS A.I. 1,2 (2021).
- 173 ראו למשל U.N. Secretary-General, *Extreme poverty and human rights*, U.N. Doc. A/66/265 (Aug. 4, 2011), <https://digitallibrary.un.org/record/709907?v=pdf#files>.
- 174 שם, בעמ' 19.

ברם, ייתכן שהתרופה נמצאת בקבלת החלטות משותפת לאדם ולמכונה. מחקר שפורסם בשנת 2022 בדק את השימוש במערכת אוטומטית לצורך החלטה על פתיחת חקירה של שירותי הרווחה במקרי חשד למצב של ילדים בסיכון במחוז אלגני במדינת מרילנד שבארה"ב.¹⁷⁵ המחקר העלה כי קבלת החלטות משותפת של אדם ומכונה (עובד/ת שירותי הרווחה בצירוף המערכת) הובילה לתוצאות מפלות פחות בהשוואה להחלטת מכונה באופן בלעדי. ככל שתוצרי XAI יתרמו לשיתוף פעולה זה הם עשויים בכך גם לעודד את ההכרה באוטונומיה האנושית במשתמע.

XAI בשירותי הזכות להליך נאות: חובת הנמקה נמצאת בבסיס קיומו של הליך נאות והיא מורכבת מהסברים, הצדקות – וביחד הנמקות – השופכות אור על שיקול הדעת והרציונל האנושי שהובילו להחלטה. אך כיצד ניתן לנהל הליך נאות כשעסקינן בהחלטה אלגוריתמית המהווה "קופסה שחורה"? בהליך שהתנהל לאחרונה בבית המשפט הגבוה לצדק של האיחוד האירופי (ECJ)¹⁷⁶ נידון עניינה של QO, מבקשת אשראי שניתן לה דירוג אשראי נמוך ע"י חברת SCHUFA Holding AG (SCHUFA), בהסתמך על מערכת החלטה סטטיסטית. QO טענה כי היא זכאית לדעת באיזה מידע ספציפי בעניינה נעשה שימוש, ואילו משקלות ניתנו לאותו מידע ספציפי במערכת בעת קביעת דירוג האשראי.¹⁷⁷ SCHUFA נעתרה לחשוף בפני QO, מבקשת האשראי, את דירוג האשראי שניתן לה בצירוף הסבר כללי על תהליך חישוב דירוגי אשראי, אך לא מעבר לכך. ביהמ"ש בגרמניה שדן בסמכותו המדינתית פנה אל ה-ECJ לשם קביעה האם עסקינן ביצירת פרופיל אישיות (פרופילינג) כהגדרתו ב-GDPR. בהחלטה תקדימית פסק ה-ECJ כי דירוג אשראי אכן מהווה פרופילינג לעניין ה-GDPR והשיב את התיק לביהמ"ש בגרמניה לפסוק איזו צורה תלבשנה ההגנות והזכויות העומדות ל-QO במצב זה מכוח ה-GDPR, לרבות "הזכות להסברות" הקבועה לכאורה בסעיף 22 וברסיטל 178.71 מהליך זה עולה לכאורה כי יכולתם של נשואי החלטה אלגוריתמית לממש את זכותם להליך נאות תלויה ביכולתם לקבל הסברים, הצדקות ונימוקים כדי

Hao-Fei Cheng et al., *How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions*, PROC. 2022 CHI CONF. ON HUM. FACTORS COMPUTING SYS. 1 (2022). תוצאות המחקר תוקפו באמצעות בחינת בסיס נתונים חדש ומעודכן בהמשך. ראו Logan Stapleton et al., *Extended Analysis of How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions*, ARXIV.ORG (2022) <https://arxiv.org/pdf/2204.13872>.

Case C-634/21, SCHUFA Holding AG v. Land Hesse, ECLI:EU:C:2023:220, (Oct. 15, 2021) Andreas Häuselmann, *The ECJ's First Landmark Case on Automated Decision-Making – a Report from the Oral Hearing before the First Chamber*, EUR. L. BLOG (Feb. 20, 2023), <https://www.europeanlawblog.eu/pub/the-ecjs-first-landmark-case-on-automated-decision-making-a-report-from-the-oral-hearing-before-the-first-chamber/release/1>.

שם. 177

Case C-634/21, SCHUFA Holding AG v. Land Hesse, ECLI:EU:C:2023:957, (Dec. 7, 2023). 178

לערער על החלטה אלגוריתמית (למשל דירוג אשראי), או לפחות להעריך נכונה את סיכויי של ערעור שכזה. יחד עם זאת, במצבה הנוכחי של הטכנולוגיה הסברית XAI מתקשים לספק את הנ"ל. בעוד שהטכניקות השונות עשויות לספק **רמזים** להסברים, הצדקות והנמקות עצמם, או את המשקלות של המרכיבים (Features) העומדים ביסוד ההחלטה,¹⁷⁹ הנ"ל יתורגמו ל"הסבר", "הצדקה" ו"הנמקה" בתנאי שעבודת חקר אנושית תעקוב אחר אותו "שביל פירורים" ותחשוף את ההנמקות בלב ההחלטה האוטומטית.¹⁸⁰ בעוד שתכונה זו מוכרת היטב לבעלי ידע בתחום למידת המכונה, היא אינה כה גלויה עבור בני אדם חסרי השכלה טכנולוגית. למעשה, כפי שמראה **מוצג א**, פרקטיקות הפקת הסבר XAI כיום רחוקות מיצירת הסברים נגישים לחסרי ידע טכנולוגי,¹⁸¹ עובדה המחייבת קיומם של "מתווכים" בין הטכנולוגיה לבין לקוחות ההסברים. בנוסף, אנו מצפים בצורה אינהרנטית שהסבר יינתן בהקשר רלוונטי ומבוסס על ידע של העולם שבו אנו חיים. ברם, לפחות עד שמערכות בינה מלאכותית כללית (Artificial General Intelligence) תהפוכנה מחזון למציאות, מערכות אוטונומיות "יודעות" רק מה שהן למדו או הוגדר ש"ידעו",¹⁸² דהיינו הן "לוח נקי" (Tabula rasa).¹⁸³ יכולתו של האדם מן היישוב להבין את ההסבר מהווה תנאי מקדמי ליכולתו להבין את ההחלטה, לגבש עמדה ביחס אליה, ולהעריך את סיכויי הערעור עליה. אשר על כן, היכולת להבין את ההסבר המופק להחלטת המערכת היא תנאי מהותי לקיום הליך נאות בהתייחס להחלטות מכונה, אולם על ההסבר להיות מותאם לאותו צורך.

179 Bordt et al., לעיל ה"ש 34, בעמ' 898.

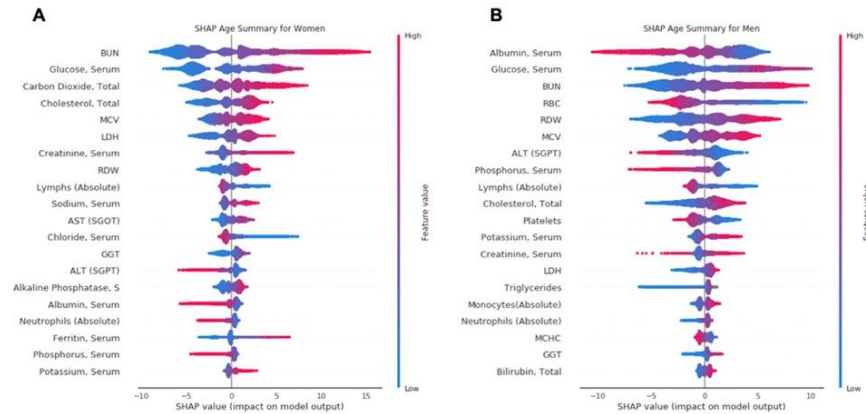
180 Mittelstadt et al., לעיל ה"ש 155, בעמ' 279.

181 Umang Bhatt et al., *Explainable Machine Learning in Deployment*, PROC. CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 648, 656 (2020).

182 Bordt et al., לעיל ה"ש 34, בעמ' 897. ראו גם Zachary C. Lipton, *The Mythos of Model Interpretability: In Machine Learning, the Concept of Interpretability is Both Important and Slippery*, 16 QUEUE 31, 33 (2018).

183 Jarek Gryz & Nima Shahbazi, *Futility of a Right to Explanation*, EDBT/ICDT WORKSHOPS 1, 4 (2020).

מוצג א



מוצג א, דוגמה של פלט XAI. במקרה המוצג כאן השתמשו החוקרים בשיטת הסבריות SHAP או SHAPLY כדי להבין את מידת ההשפעה של משתנים שונים על יכולתה של המכונה לזהות גיל של פציינטים. הפלט מתאר, הן עבור נשים (A) והן עבור גברים (B), את מידת ההשפעה של כל משתנה ומשתנה ביחס להחלטת המכונה, והאם מדובר בהשפעה שלילית או חיובית (כלומר תורמת להחלטת המכונה הסופית או גורעת ממנה). מוצג זה מעיד על רמת המורכבות של תוצרי XAI הדורשים מומחיות כדי לנתחם. נלקח מ־ Thomas R. Wood et al., *An Interpretable Machine Learning Model of Biological Age*, F1000RESEARCH 8, no. 17, 5 (2019).

XAI בשירות חיזוק הסמכות של מערכת קבלת ההחלטות: יצירת הסברים להחלטות מכונה אוטונומיות באמצעות XAI מזכרת תדיר בהקשר של בניית אמון של הציבור בכלל, ונשוא ההחלטה בפרט. היעדרו של אמון זה נתפס כמעכב את ההסתמכות האנושית על תחזיותיהן של המכונות ואת הטמעתן,¹⁸⁴ ואף מוגדר כאחד מהכשלים המרכזיים של מערכות בינה מלאכותית כיום.¹⁸⁵ ברם, מחקרים עדכניים מעוררים חשש בדבר התבונה שבעידוד מתן אמון במכונות אוטונומיות בהתבסס על הסברי XAI, כאשר הוא מנותק מבדיקת איכות ואמינות המערכת לגופה. ראשית, קיימות התנגדויות עקרוניות ליצירת הסברים למכונות אוטונומיות משיקולים כגון פגיעה פוטנציאלית בזכות לפרטיות, לקניין רוחני, לבטיחות ולשמירה על סודות

Marco Tulio Ribeiro et al., "Why Should I Trust You?" *Explaining the Predictions of Any Classifier*, 2016 PROC. 22ND ACM SIGKDD INT'L CONF. ON KNOWLEDGE DISCOVERY & DATA MINING 1135 (2016).

EXPLAINABLE AI: FOUNDATIONS, METHODOLOGIES AND APPLICATIONS v-vi (Mayuri Mehta, Vasile Palade & Indranath Chatterjee eds., 2022).

מסחריים.¹⁸⁶ שנית, מפתחי המערכות עצמם מסתייגים לעיתים מיצירת הסברי XAI ל"המונים" שיפרשו אותם בצורה שאינה מקצועית ואף ינצלו אותם לרעה לשינוי התנהגות שיוביל להערמה על המערכת.¹⁸⁷ התנגדויות אלו עשויות לקדם יצירת הסברים עמומים וחסרי משמעות מלכתחילה. שלישית, ניגוד העניינים מתעצם כאשר בעל המערכת המחליטה אמון גם על יצירת ההסבר להחלטותיה, שהרי הוא מצוי במערכת יחסים אדברסרית מול נשוא ההחלטה, המעוניין בהסבר לרוב כדי לשנות החלטה בלתי רצויה מבחינתו.¹⁸⁸ בהינתן שפרקטיקות XAI כיום מאפשרות לייצר הסברים מסוגים שונים לאותה החלטה, היכולת לסמוך עליהם כמעידים על אמינות המערכת עלולה להיות פגומה מיסודה.¹⁸⁹ אחת הדוגמאות לשימוש פוטנציאלי בהסברים לחיזוק אמון המשתמשים במערכת אוטונומית הוא סדרת ניסויים של חברת General Motors אשר במסגרתם נבדקה השפעת הסברים מסוגים שונים על נהגים בסימולטור נהיגה אוטונומי.¹⁹⁰ מטרת השימוש בהסברים הייתה לבחון האם ניתן להגביר את אמון הנהגים במערכת ולצמצם את התערבות הנהגים בנהיגה האוטונומית.

- Alison B. Powell, *Explanations as* Wachter et al. 186
 32, בעמ' 871. ראו גם *Governance? Investigating Practices of Explanation in Algorithmic System Design*, 36 EUR. J. CMMC'N 362, 365 (2021); Smitha Milli et al., *Model Reconstruction from Model Explanations*, 2019 PROCS. CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 1; Florian Tramèr et al., *Stealing Machine Learning Models via Prediction APIs*, 2016 PROC. 25TH USENIX CONF. ON SEC. SYMP. 601
- Yujia Zhang et al., *Why Should You Trust My* Powell 187
 שם, בעמ' 370. ראו גם *Explanation? Understanding Uncertainty in LIME Explanations*, ARXIV.ORG (2019)
- ראו גם Gryz & Shahbazi, לעיל ה"ש 183, בעמ' 4.
- Bordt et al. 188
 לעיל ה"ש 34, בעמ' 2.
- 189 להדגמה כיצד ניתן לבצע מניפולציה בהסברים המופקים למערכות בינה מלאכותית ובכך לשקף את הסכנה של שימוש בהסברים לביסוס אמון במכונות, ראו Boty Dimanov et al., *You Shouldn't Trust Me: Learning Models Which Conceal Unfairness From Multiple Explanation Methods*, 2560 CEUR WORKSHOP PROC. 63 (2020); להדגמה של שימוש בהסברים על מנת להסתיר קיומה של אפליה במערכת במסגרת ביקורת, ראו Joyce Zhou & Thorsten Joachims, *How to Explain and Justify Almost Any Decision: Potential Pitfalls for Accountability in AI Decision-Making*, 2023 ACM INT'L CONF. PROC. SERIES 12
- Ramaravind K. Mothilal et al., *Explaining Machine Learning Classifiers Through Diverse Counterfactual Explanations*, 2020 PROCS. 2020 CONF. ON FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY 607. לשיטת המחברת של מאמר זה, עצם היכולת לייצר הסברים מסוגים שונים עשויה להקנות לבעל העניין האחראי על פעילות המערכת את האפשרות לבחור הסבר המיטיב עימו.
- Claudia V. Goldman & Ronit Bustin, *Trusting Explainable Autonomous* 190
Driving: Simulated Studies, 2022 IEEE INTELLIGENT VEHICLES SYMP., 1255; Claudia V. Goldman et al, *Adaptive Driving Agent: From Driving a Machine to Riding with a Friend*, 2020 PROC. INT'L. CONF. ON HUM. AGENT INTERACTION 179

בעוד שמתן אמון במערכת המתפקדת היטב הינו תוצאה רצויה, אפקט זה עשוי להתעורר גם במצבים שבהם המערכת לא בהכרח ראויה לאמון האנושי.¹⁹¹ במצב כזה, בני אדם עלולים לשים מבטחם במערכות שמובילות לתוצאות שגויות מתוך הנחה מוטעית שההסבר מעיד בצורה אותנטית על אמינות המערכת. ניתן לדמיין למשל מצב של החלטה שגויה של המערכת האוטונומית לבצע מעבר לנתיב הנגדי בכביש, כפי שקרה לאחרונה בפיניקס שבארה"ב.¹⁹² אנו מצפים שאדם סביר הנמצא ברכב ייקח לידי שליטה ידנית ויחזיר את הרכב למסלול הנכון כך שלא יסכן את עצמו ואת העוברים והשבים. אולם מערכת שתיוצר הסבר משכנע לאותו אדם ברכב, למשל שהמערכת סבורה שזו החלטה נכונה בנסיבות העניין, או שהמעבר נעשה מכיוון שיש מכשול בנתיב הנסיעה המקורי, עלולה לעמעם את שיקול הדעת האנושי תוך עידוד הסתמכות מוטעית על המערכת גם במקום שבו היא טועה ומבצעת פעולה מסכנת חיים. על כן קידום גורף של יצירת הסברים להחלטות מכונה עלול לעיתים לגרום ליותר נזק מאשר תועלת.¹⁹³ בנוסף, ככל שהסברים יתגלו כמטעים, עמומים, או כמפעילי מניפולציה על לקוחותיהם האנושיים, כך עשוי לצנוח בהתאמה אמון הציבור הן בתוצרים ההסברתיים והן במערכות הבינה המלאכותית עצמן.

שלוש התפתחויות נוספות הראויות לציון בהקשר זה הינן: א. זיהוי הפוטנציאל של טכניקות XAI להפעיל מניפולציות על בני אדם, במכוון או מבלי דעת (תופעה הידועה כ-Explainability Pitfalls);¹⁹⁴ ב. התפתחות תחומי המחקר של XAI מתרחקת מהגברת שקיפות – המטרה שביסוד הזכות להסברות – לכיוון מטרות כגון הגברת אמון המשתמשים במערכת, שכנוע משתמשים, והגברת הסיפוק שלהם מהשימוש במערכת (ראו **מוצג ב ומוצג ג**); ג. כניסתם לתמונה של מודלי בסיס (Foundation Models), המאתגרת אף יותר את הדרישה לקבלת הסבר. לאור כמות הדאטה המסיבית שהם מאומנים לאורה, ובהינתן רמת הסיבוכיות העצומה שהם ניחנים בה, הדיסציפלינה הטכנולוגית עודה מתקשה בעצמה להבין את דרך פעולתם, את עצם היכולות שלהם

191 ראו Skitka et al., לעיל ה"ש 160.

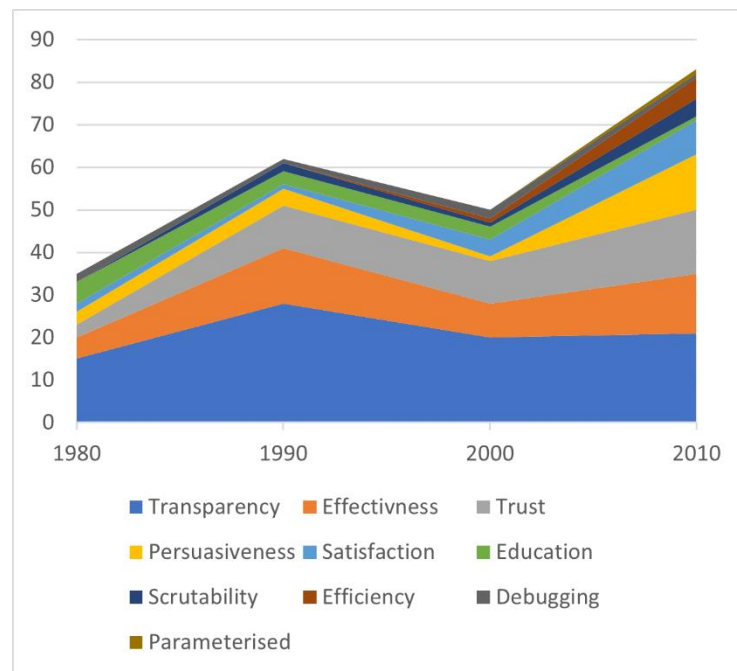
192 Wes Davis, *Police pulled over a Waymo car for driving in the oncoming lane*, THE VERGE (July 6, 2024), <https://www.theverge.com/2024/7/6/24193094/phoenix-waymo-car-pulled-over-traffic-stop-oncoming-lane-police>.

193 Ellen J. Langer et al., ראו *The Mindlessness of Ostensibly Thoughtful Action: The Role of "Placebic" Information in Interpersonal Interaction*, 36 J. PERSONALITY & SOC. PSYCH. 635 (1978).

194 למרות התקדמות משמעותית בתחום ה-XAI, מחקרים חדשים מציפים גם תופעות לוואי שליליות של הסברים אלו, כגון טקטיקות הטעיה לעיצוב "הסברי פלצבו" (ללא תוכן) ולייצור אמון מלאכותי במערכות בינה מלאכותית. תופעות אלו עשויות לקרות גם בתום לב במהלך מתן הסברים למשתמשי קצה עקב הרגישות הגבוהה של בני אדם להסברים. ראו למשל Upol Ehsan & Mark O. Riedl, *Explainability Pitfalls: Beyond Dark Patterns in Explainable AI*, PATTERNS 1, 1 (2021).

ואת הסיבה להתנהגויות מסוימות מצידם.¹⁹⁵ לסיום יוזכר כי מודלי שפה מתקדמים כגון ChatGPT ודומיו מפגינים יכולות שכנוע (הנמקה) גבוהות גם במקרים של טעויות ברורות,¹⁹⁶ תכונה אשר ניתן להניח כי רק תלך ותגבר תוך התאמה לזהות לקוח ההסבר בכל מקרה ומקרה.

מוצג ב

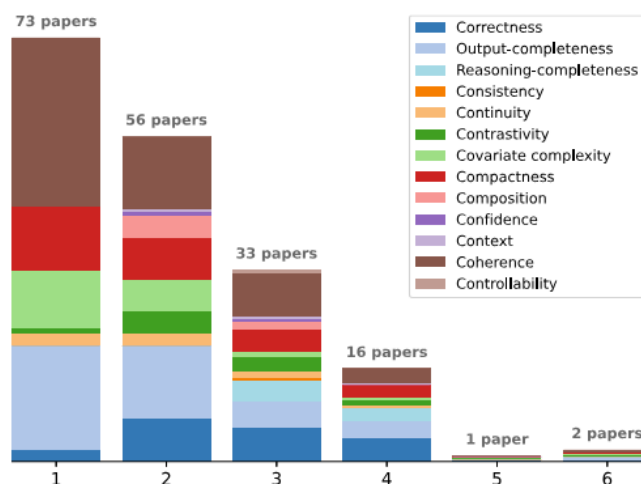


מוצג ב מעיד על ירידה עקבית בפיתוח טכניקות XAI (עד שנת 2010) למטרה של שקיפות (במובן של הסבר כיצד המערכת עובדת – כחול – Transparency) ועליית מטרות כגון עידוד האפקטיביות של הסברים (במובן של עזרה למשתמשים לקבל החלטות טובות – כתום – Effectiveness), עידוד אמון המשתמשים במערכת (אפור – Trust) ומידת השכנוע של ההסברים על משתמשים לנסות או לקנות (צהוב – Persuasiveness). שימוש במידע שהוצא מ־Ingrid Nunes & Dietmar Jannach, *A Systematic Review and Taxonomy of Explanations in Decision Support and Recommender Systems*, 27 USER MODELING AND USER-ADAPTED INTERACTION 393, 410 (2017).

¹⁹⁵ Rishi Bommasani et al., *On the Opportunities and Risks of Foundation Models* (2021).

¹⁹⁶ Sébastien Bubeck et al., *Sparks of Artificial General Intelligence: Early Experiments with GPT-4*, EPRINT ARXIV 60 (13 Apr. 2023), <https://arxiv.org/abs/2303.12712>

מוצג ג



(b) Total number of unique Co-12 properties quantitatively evaluated in a paper that introduces an XAI method.

מוצג ג – סקירה של 361 מאמרים המציגים פרקטיקות ל-XAI בשנים 2014–2020, מתוכם רק 58% כללו בדיקה כמותית של מאפיינים שונים בהסברים המונפקים על פי אותה שיטה. מבין המאפיינים הנ"ל רק מעטים בודקו בצורה כמותית את מידת הנכונות של ההסבר (כחול כהה – Correctness), כלומר עד כמה ההסבר נאמן למודל שאותו הוא מסביר. המוצג נלקח מ־Nauta et al., לעיל ה"ש 147, Fig. 6. (b).

מסקנת פרק זה, שבו הועמדו התכליות לחובת הסבר בעולם המשפט אל מול הפרקטיקות המפותחות בתחום למידת מכונה לייצור הסברים למערכות בינה מלאכותית (XAI), הינה כי קיים מרחב של אי התאמות בין השתיים. כך, XAI מתקשה לענות על התכלית של שיפור איכות ההחלטה, קיים ספק לגבי יכולתן של פרקטיקות אלו לכבד את האוטונומיה האנושית, הן מתקשות להתאים לתכלית של מימוש הליך נאות והן עלולות לעודד – במודע ולא במודע – הסתמכות שאינה מוצדקת על מערכת קבלת ההחלטות. חוסר התאמה בסיסי זה מעורר קושי בלתי מבוטל בנוגע למימוש הזכות להסבריות של מערכות בינה מלאכותית באמצעות XAI כיום, אולם הוא גם מצביע על הכיוון שאליו יש להתקדם על מנת להשיג את התכליות המבוקשות. בהסתכלות צופה פני עתיד, ובהנחה שטכנולוגיית הפקת הסברים למערכות בינה מלאכותית תתפתח בעתיד, ניתן לחלק את ההסברים למערכות בינה מלאכותית לשלושה סוגים (מוצג ד): א. הסברי XAI שאינם משתווים להסברים אנושיים (כגון הסברים שאינם מאפשרים זכות ערעור בשל היותם מסובכים מדי או עמומים, הסברים שלא ניתן לבצע באמצעותם בקרה או עדכון למערכת וכו'); ב. הסברים המשתווים להסברים אנושיים (כגון הסברים אשר באמצעות תיווך יוכלו לאפשר זכות ערעור,

יתרמו לאוטונומיה של נשוא ההחלטה כבן אנוש ויאפשרו ביקורת על פעולת המערכת); ג. הסברים העולים על הסברים אנושיים (כגון הסברים החושפים דרכי פעולה חדשות, תובנות שלא ניתן היה לקבל מהסבר אנושי, יכולת הבנה עמוקה יותר של מושגים עמומים כגון "הוגנות" ו"סבירות" ואינדיקציות מהימנות לכך שהמערכת הינה אמינה ונטולת פניות). מובן כי עשויות להתקיים וריאציות משולבות שבהן אלמנטים מסוימים של הסבר XAI יעלו על הסבר אנושי בעוד שאלמנטים אחרים יהיו נחותים או שווים לו.

מוצג ד



מוצג ד – אי ההתאמה בין מהות המושג "הסבר" בעולם המשפט לבין מושג זה בדיסציפלינה הטכנולוגית מתורגמת לשלושה סוגי תוצרים פוטנציאליים של XAI: 1. הסברים נחותים יחסית להסברים אנושיים, אשר מתקשים למלא את התכליות המשפטיות המבוקשות; 2. הסברים המשתווים להסברים אנושיים מבחינת מילוי התכליות המשפטיות המבוקשות; 3. הסברים העולים על הסברים אנושיים ביכולתם למלא את התכליות המשפטיות המבוקשות.

ד. סיכום ומסקנות

כאשר רגולטורים דורשים "חובת הסברתיות" להחלטות מערכות בינה מלאכותית – למה הם מתכוונים? מאמר זה זיקק את התכליות העומדות בבסיס חובת ההנמקה הנעוצה בלב הזכות להסברתיות של מערכות בינה מלאכותית: שיפור איכות ההחלטה, הכרה באוטונומיה האנושית, תמיכה במימוש הזכות להליך נאות וחיוק סמכותה של מערכת קבלת ההחלטות. העמדתן של תכליות אלו כנגד הטכניקות השונות ליצירת הסברים באמצעות XAI כיום מעלה כי לא רק ש-XAI מתקשה לענות על שלוש

התכליות הראשונות, אלא שהוא אף בעל פוטנציאל הצלחה מופרז ביחס לתכלית הרביעית והאחרונה. חוסר התאמה זה צריך להדריך רגולטורים לכיוון גישה זהירה בנוגע להחלת חובת הסברות כללית של מערכות בינה מלאכותית. במה דברים אמורים? יובהר מייד כי אין במאמר זה קריאה "לשפוך את התינוק יחד עם המים". הסברים אנושיים (וקבלת החלטות אנושיות בציודם) אינם חפים מבעיות וכשלים בעצמם, והם אינם מכשיר מושלם למימוש התכליות העומדות ביסודם.¹⁹⁷ בנוסף, קבלת החלטות אלגוריתמית עשויה בתנאים מסוימים לקדם את התכליות שבסודן של חובת ההנמקה במידה שתעלה על היכולת האנושית. כך למשל הקטנת העומס על בתי המשפט באמצעות מעבר להחלטות אוטונומיות צפויה להיטיב עם יכולת המימוש של הליך נאות וזיהוי דפוסים במידע אמור לקדם החלטות איכותיות יותר.¹⁹⁸ בנוסף, ל-XAI תפקיד חשוב ביצירת תובנות בתחום למידת המכונה ואף, בתנאים המתאימים, עבור בעלי עניין אחרים (מטמיעי מערכות, משתמשי המשיך בשרשרת הערך, רגולטורים, מתווכים, משתמשי קצה ועוד). לדוגמה, היכולת לאתר חוסר הוגנות ואפליה מתוך הסברים להחלטות אלגוריתמיות עשויה להיות גבוהה יותר מאשר איתור תופעות אלו בהתבסס על הסברים להחלטות אנושיות.¹⁹⁹

בעוד שהשימוש בזכות להסברות כדרך לייצור שקיפות הוא הגיוני בבסיסו, על רגולטורים להכיר בכך שהוא מביא תועלת מוגבלת ואף עלול להזיק אם אין ממקדים אותו בהקשר הנכון. היכולת להפיק תועלת ממכשיר רגולטורי כזה בהקשר של הגברת האחריות תלויה לא רק ביכולת הטכנולוגית לייצר הסברים אלא גם בציפיות של החברה מהם ובסט ההקשרים החברתיים, הקוגניטיביים והמבניים שבהם הזכות להסברות מיושמת. אין דין הסבר למשפטנים כדין הסבר למהנדסים, ואין דין הסבר המלווה החלטה הניתנת בתחום המשפט הפלילי כדין החלטה על תיעדוף ומזור בצומת. במובן זה, העובדה כי מדובר בטכנולוגיה צעירה יחסית ומתפתחת מהווה יתרון המאפשר לעצב אותה בצורה המיטיבית כדי למלא את התפקיד המיועד לה רגולטורית וחברתית.

197 Kahneman et al., לעיל ה"ש 162. ראו גם Andrea Bonezzi et al., *The Human Black-box: The Illusion of Understanding Human Better than Algorithmic Decision-making*, 151 J. EXPERIMENTAL PSYCH. GEN. 2250 (2022).

198 דוגמה נוספת הינה השוואה בין אלגוריתם לחיזוי אל מול שוטרת בהקשר של ביצוע חיפוש המבוסס על "חשד סביר". ניתן לשאול, בהנחה שאלגוריתם יהיה מדויק יותר, האם אין עדיפות לקבלת החלטה באמצעות אלגוריתם אף על פי שהחלטת השוטרת לכאורה מלווה בהסבר. ראו נתי פרל "החלטה על קיומו של 'חשד סביר' בעולם ה-Big Data: מהנמקה אנושית לאלגוריתמים לחיזוי פשיעה" *דין ודברים* יד 237, 283 (2017).

199 ראו למשל Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, 10 J. L. ANALYSIS 113, 152–153 (2018).

כדי להימלט ממלכודת ה־Technosolutionism²⁰⁰, חשוב למסגר את XAI כמרכיב הכרחי בתהליך למימוש זכות להסברתיות, אך כזה שאינו יכול לעמוד לבדו. כך, על מנת שתפקידים אלו יוכלו להתבצע בצורה מספקת יש לאמץ גישה הבוחנת את השימוש ב־XAI לאור תכליות מנגנון ההסבר בצורתו האנושית, ולבחון בצורה ביקורתית באיזו מידה יכול הסבר שמקורו במכונה למלא אותן. מיקוד האתגר בהשגת ארבע התכליות המשפטיות לקיומו של הסבר מאפשר להדגיש את שילוב הפרספקטיבות הנדרש מיצירת הסברים להחלטות אוטומטיות. מן הראוי אפוא לנקוט בפרשנות תכליתית בנוגע למימושה של הזכות להסברתיות באופן המקשר בין סוג וטיב ההסבר המבוקש לבין התכלית העיקרית העומדת ביסוד הדרישה לקבלו. כך למשל, ככל שאנו מעוניינים בהסבר בעיקר לשם שיפור איכות החלטות המערכת, פעולה אשר סביר שתבוצע על ידי מומחים בלמידת מכונה, נבקש לקבל הסברים טכנולוגיים יותר מטבעם (בין אם בשיטות XAI, גישה למידע שעליו אומנה המכונה, ביקורת וכיו"ב). לעומת זאת כאשר מטרת ההסבר הינה לאפשר זכות ערעור לנשוא החלטה, על ההסבר להיות נגיש להדיוט, או לפחות מתווך לחסרי השכלה טכנולוגית באמצעות מתווכי משנה (בדומה ליועצי ביטוח המשמשים מתווכי משנה לעולם הביטוח, עורכי דין המשמשים מתווכי משנה לעולם המשפט וכיו"ב). על ההסבר אף להיות כזה הנתון לביקורת שיפוטית אפקטיבית. הסבר אשר תכליתו העיקרית הינה הכרה באוטונומיה האנושית עשוי למשל להעניק למקבל ההסבר את האוטונומיה לבחור את רמת ההעמקה והמורכבות של ההסבר, ובכך יבטא את עקרון "האדם במרכז"²⁰¹. ולסיום, הסבר אשר תכליתו העיקרית הינה חיזוק סמכותה של מערכת קבלת ההחלטות יצטרך להעיד בצורה אותנטית על מהימנותה של המערכת ועל כך שהיא אכן ראויה לאמון אנושי.

הפרשנות התכליתית עשויה לשמש גם לצורך הערכת ההסברים המופקים ברמה החברתית. בעוד שחלק מהתכליות תוכלנה להימדד על ידי מומחים או מתווכים בעלי ידע (באיזו מידה הסברים משמשים לטיוב ההחלטה), לרבות באמצעות תקינה ייעודית ומתן זכות גישה²⁰² חלק אחר מהתכליות יוכל להימדד על ידי משתמשי קצה (אפקטיביות היכולת להליך הוגן) או על ידי ארגונים חברתיים ודעת קהל (באיזו מידה ההסברים מקדמים אמון לא מבוסס במערכת או מבטאים הכרה באוטונומיה האנושית).

200 הרעיון כי כל בעיה ניתנת לכאורה לפתרון טכנולוגי המוביל לאימוץ חסר ביקורת של פתרונות טכנולוגיים. ראו EVGENY MOROZOV, TO SAVE EVERYTHING, CLICK HERE: THE FOLLY OF TECHNOLOGICAL SOLUTIONISM (2013)

201 ראו למשל Jordan Richard Schoenherr et al., *Designing AI Using a Human-Centered Approach: Explainability and Accuracy Toward Trustworthiness*, 2023 4 IEEE TRANSACTIONS ON TECH. & SOC'Y. 9 (2023).

202 בדומה לסעיף 40 לחוק השירותים הדיגיטליים האירופי הידוע גם כ־"DSA", המעניק לחוקרים זכות גישה למידע בתנאים מסוימים, למטרות של בקרה: Council Regulation (EU) 2022/2065 on a single Market for Services and Amending Directive 2000/31/EC (Digital Services Act) (2022) O.J (L 277).

לשם מימוש הנ"ל יש להקים מוסדות מתווכים שישולבו בהם ידע וכישורים מדיספלינות מגוונות ולהשקיע בחינוך הציבור. בנוסף, יש להכווין את הפיתוח של XAI באופן אקטיבי למיקוד בתכליות הרלוונטיות לחברה האנושית ולצמצם סיכונים לשימוש בעייתי בהסברים למערכות בינה מלאכותית (למשל באמצעות דרישות תיעוד, הסברים למערכת בכללותה בצד הסברים להחלטה מסוימת, וכן דרישות להצגת כמה הסברים משיטות שונות לאותו מודל). סביר כי במקרים מסוימים תוצרי הסברתיות ישרתו בעלי מומחיות שונים, והאיזון בין התקדמות טכנולוגית ויעילות אל מול הסברתיות יכתיב ויתורים על הנמקה מלאה. אולם ראשית דבר בהכרה מוסדית בכך ש"הזכות להסברתיות" אינה קלה ליישום, והיא דורשת פרשנות תכליתית על מנת שלא תהפוך לאות מתה כזכות שאינה בת מימוש.