

המאמר משווה בין הפליה אנושית להפליה אלגוריתמית במערכות קבלת החלטות מבוססות אלגוריתם (ADM – Automated Decision Making) בזירה השוקית, ובוחר את השלכותיהן בהקשר של דיני ההפליה במשפט הישראלי. במאמר אנו מתארות שלושה סוגי הפליה אנושית: הפליה טעמים, המבוססת על העדפות אישיות, הפליה סטראוטיפית, המבוססת ברובה על דעות קדומות, והפליה סטטיסטית, המבוססת על הכללות מדויקות סטטיסטית, אך לא רצויות נורמטיבית.

נראה כי בעוד שמערכות בינה מלאכותית עשויות להפחית הפליה הנובעת מטעמים אנושיים, כמו הפליית טעמים והפליה סטראוטיפית, הן לא תוכלנה להימנע מהפליה סטטיסטית, שעשויה להפלות חלק ממשתמשי המוצר או צדדים שלישיים. הבינה המלאכותית פועלת על בסיס מאגרי נתונים גדולים, ומבצעת הכללות סטטיסטיות באופן אינהרנטי. לכן ככל שפערים הם מדויקים סטטיסטית – הם יבואו לידי ביטוי בפלט האלגוריתמי. בנוסף, גם כאשר מנסים באופן אקטיבי להסיר משתנים מפלים כמו מגדר או גזע, המערכות האלגוריתמיות מסוגלות להסיק את המשתנים הללו ממשתנים אחרים, והתחזקות אחר המשתנה בהחלטה המפלה היא כמעט בלתי אפשרית. כך, במקרים מסוימים התוצאות שתתקבלנה על ידי המערכות האלגוריתמיות עלולות דווקא לשעתק את ההפליה ואת אי-השוויון הקיימים בחברה.

טענתנו במאמר היא שהפרויקט האלגוריתמי מציב אתגר משמעותי לדיני איסור ההפליה הקיימים. אנו מציעות לפצל את הטיפול המשפטי: להמשיך להחיל את דיני איסור ההפליה המסורתיים על החלטות אנושיות, אך לפתח מסגרת משפטית נפרדת להחלטות אלגוריתמיות שתתמקד בבחינת התוצאות וההשפעות החלוקתיות שלהן. גישה זו תדרוש הכרעות נורמטיביות מורכבות לגבי היקף הפערים המותרים בין קבוצות, ולגבי אופן חלוקת העלויות הכרוכות בצמצומם.

* תמר קריכלי כץ פרופסור למשפטים, הפקולטה למשפטים ע"ש בוכמן, אוניברסיטת תל אביב.

נויה מטלון בוגרת תואר ראשון, הפקולטה למשפטים, אוניברסיטת תל אביב. תודה רבה ליותם דואק, שקד שינדר ומאיה רייסנר, חברי מערכת כתב העת "משפט חברה ותרבות", לנדב זק העורך, ולשופטים החיצוניים על התרומה הגדולה לחידוד הטענות. האחריות לתוכן הדברים היא של הכותבות בלבד.

“הסכנות שמציבה הבינה המלאכותית הן לא רק שאנשים יופלו לרעה, אלא שלא יראו אותם כפי שהם באמת. בין אם מדובר בבנק שמשתמש בבינה מלאכותית כדי להחליט אם אני זכאית למשכנתא או לא, או שירותים חברתיים של הרשות המקומית – אנחנו נרצה לוודא שאנחנו לא מופלים לרעה בשל המגדר, צבע העור או המיקוד שלנו”

(מרגרט וסטגר, הנציבה האירופית לתחרות, בריאיון לרשת ה-BBC הבריטית, יוני 2023).¹

מבוא

האנושות נמצאת בעיצומה של מהפכה טכנולוגית. יותר ויותר ממשקים הופכים למבוססים על אלגוריתמים ובינה מלאכותית, בשורה שמביאה עימה הבטחות ליעילות, חיסכון בזמן, כסף ומשאבים, וכן מתן תוצאות מדויקות וברורות יותר – כל זה בשניות בודדות. למרות ההבטחות הרבות, רבים אינם מתרגשים מהחלפת הבינה האנושית בבינה מלאכותית, בפרט בהליכי קבלת ההחלטות, וכפי שיורחב בהמשך המאמר; הם חוששים מהטיות אלגוריתמיות שתשפיענה לשלילה על התוצאה.

מאמר זה יתמקד במערכות קבלת החלטות מבוססות אלגוריתם (ADM – Automated Decision Making Algorithm) והחשש להפליה שוקית בגין שימוש בהן.² המאפיין העיקרי של מערכות אלו הוא שהן מספקות תהליך קבלת החלטות אוטומטי, שסופו בקבלת החלטה, ללא מעורבות אנושית.³ החשש, כפי שיתואר במאמר זה, הוא מתוצאה אלגוריתמית מפלה. אלגוריתמים עלולים לגרום להפליה כאשר הם מתחשבים באופן ישיר או עקיף בפרטים כמו מגדר, גזע או נטייה מינית של אדם. עם הגידול בשימוש במערכות בינה מלאכותית בחיי היום-יום, ובייחוד בתהליכי קבלת החלטות

1 Dan Milmo & Alex Hern, *Discrimination is a Bigger AI Risk than Human Extinction – EU Commissioner*, GUARDIAN (Jun. 14, 2023, 11:19 BST), <https://www.theguardian.com/technology/2023/jun/14/ai-discrimination-is-a-bigger-risk-than-human-extinction-eu-chief>.

2 הפליה שוקית היא מצב שבו מאפייני זהות שאינם רלוונטיים משפיעים על החלטות שוקיות. הפליה זו היא הפליה אסורה אם היא מנוגדת לדברי החקיקה שכלולים בדיני ההפליה שנוגעים בעיקר לזירות של עבודה, מוצרים ושירותים [חוק שוויון הזדמנויות בעבודה, תשמ"ח-1988 (להלן: חוק שוויון הזדמנויות בעבודה); חוק איסור הפליה במוצרים, בשירותים ובכניסה למקומות בידור ולמקומות ציבוריים, תשס"א-2000 (להלן: חוק איסור הפליה)]. להרחבה ראו ה"ש 8, 9 להלן.

3 NOYB, *Profiling & Automated Decision Making*, <https://noyb.eu/en/project/profiling-automated-decision-making>; כמו כן, ראו Paul B. Laet, *Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?*, 31 PHIL & TECH 525, 525–530 (2018) (להלן: (Paul B. Laet, *Algorithmic Decision-Making*)).

שבהם כמעט אין התערבות אנושית, מתעוררות סוגיות בנוגע להשלכות האפשריות של שימושים אלו על תופעות של הפליה באיסוף ובעיבוד הנתונים, מה שמשפיע גם על התוצאה הסופית.⁴

ואכן, סוגיית ההפליה האלגוריתמית זוכה לתשומת לב הולכת וגוברת בעיתונות, בשיח הציבורי ובספרות האקדמית.⁵ את הטכנולוגיה סובבים חששות נרחבים בנוגע להטיה, בצורה זו או אחרת, כבר בחיי היום-יום שלנו: בחיפוש גוגל, בהמלצות הקנייה שלנו, בשיחה יומיומית עם Siri, בהמלצות מותאמות אישית בשירותי הסטרימינג שלנו, בחישוב פרמיית הביטוח שלנו, ובהחלטה האם על סמך קורות החיים שלנו נקבל ריאיון עבודה. האפשרות של הטיה אלגוריתמית מטרידה במיוחד כאשר מדובר במערכות אוטונומיות או חצי-אוטונומיות, מכיוון שמערכות אלה אינן חייבות לכלול אדם ב"לולאת המידע" שיכול לזהות ולפצות על הטיות באלגוריתם או במודל.⁶ לכן, בעוד שמערכות אוטונומיות עשויות להיתפס על ידי אחדים כניטרליות או כנטולות משוא פנים, הן עדיין עלולות להשתמש באלגוריתמים מוטים שגורמים נזק ניכר, סמוי מן העין, שנותר ללא תיקון, וכך גם להעמיק את אי-השוויון הקיים בחברה. במאמר זה נשווה בין המנגנונים שגורמים לבני האדם להפלות לבין המנגנונים שמביאים להפליה אלגוריתמית, ונדון בהשלכות המשפטיות של השוואה זו לעניין דיני

4 Simon Friis & James Riley, *Eliminating Algorithmic Bias is Just the Beginning of Equitable AI*, HARV. BUS. REV., (Sept. 29, 2023), <https://hbr.org/2023/09/eliminating-algorithmic-bias-is-just-the-beginning-of-equitable-ai>.

5 ראו יעל פרואקטור "סכנת האלגוריתם" העין השביעית (2.3.2024) Ziad Obermeyer, Brian Powers, Christine Vogeli & Sendhil Mullainathan, *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366(6464) SCIENCE 344, 447 (2019); Zachary Small, *Black Artists Say A.I. Shows Bias, With Algorithms Erasing Their History*, N.Y. TIMES (July 4, 2023), <https://www.nytimes.com/2023/07/04/arts/design/black-artists-bias-ai.html>; Cade Metz, *We Teach A.I. Systems Everything, Including Our Biases*, N.Y. TIMES (Nov. 11, 2019), <https://www.nytimes.com/2019/11/11/technology/artificial-intelligence-bias.html>.

6 כאשר אדם נמצא ב"לולאת המידע", המשמעות היא שהאלגוריתם מספק המלצה, אך ההחלטה מתקבלת בסופו של דבר על ידי האדם. לעומת זאת, כאשר אין אפשרות לאדם לשנות החלטות אלגוריתמיות באופן פרטני, נאמר שהאדם נמצא "מחוץ למעגל" (Out of the loop). נטען ששילוב בין אינטליגנציה אנושית למכנית יכול לעיתים קרובות להוביל לתוצאות טובות יותר. ראו Keith Dowding & Brad R. Taylor, *Algorithmic Decision-Making, Agency Costs, and Institution-Based Trust*, 37 PHIL. & TECH., 67 (2024) (להלן: Dowding & Taylor, *Algorithmic Decision-Making*); Reva Schwartz et al., NAT'L INST. OF STANDARDS AND TECH., *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, SPECIAL PUBLICATION (NIST SP), 10 (2022).

איסור הפליה במשפט הישראלי. לשיטתנו, הצורך בהשוואה נובע מהרצון לגבש מדיניות שתסדיר את הטיפול בהפליה אלגוריתמית, מתוך הבנה שגם בהיעדר אלגוריתמים בני האדם עלולים להפלות בקבלת ההחלטות שלהם. במילים אחרות, כדי לגבש מדיניות מושכלת, עלינו לקחת בחשבון גם את האופן שבו אנשים מתנהגים ולהשוות בין שתי החלופות: עלינו להשוות בין העולם של קבלת ההחלטות האלגוריתמית לעולם של קבלת ההחלטות האנושית ולהבין את החסרונות והיתרונות, וכן את הסכנות לאי-השוויון, הגלומים בכל אחד מהעולמות. בנוסף, ההשוואה בין המנגנונים חשובה גם כי הם קשורים זה בזה: המערכות האלגוריתמיות עלולות אף לשעתק הפליה אנושית שהוטמעה במידע שעליו התבססה המערכת האלגוריתמית.

נציין כי לא כל הבחנה בין קבוצות מהווה הפליה אסורה במובן של דיני איסור ההפליה.⁷ בישראל ובעולם, הפליה אסורה בחוק רק אם היא נוגעת לזירות חיים שדיני איסור הפליה עוסקים בהן (כמו, למשל, שוק העבודה או המוצרים) ומדובר בקטגוריה מוגנת, כמו גזע או מגדר.⁸ הקטגוריות המוגנות נבחרו על ידי המחוקק כזקוקות להגנה, בדרך כלל משום שהן קטגוריות שסבלו מהפליית עבר ונלוות אליהן אמונות סטראוטיפיות בדבר נחיתות. לבסוף, נציין כי גם לאיסורי ההפליה קיימים חריגים, אם למשל ההבחנה נגזרת ממהות השירות, המוצר או המשרה.⁹

7 חוק שוויון הזדמנויות בעבודה; חוק איסור הפליה, לעיל ה"ש 2.

8 ס' 2(א) לחוק שוויון הזדמנויות בעבודה, לעיל ה"ש 2: "לא יפלה מעסיק בין עובדיו או בין דורשי עבודה מחמת מינם, נטייתם המינית, מעמדם האישי, היריון, טיפולי פוריות, טיפולי הפריה חוץ-גופית, היותם הורים, גילם, גזעם, דתם, לאומיותם, ארץ מוצאם, מקום מגוריהם, השקפתם, מפלגתם או שירות במילואים שלהם, של בני זוגם או של הורה האחר של ילדם, ובכלל זה קריאתם לשירות מילואים או השירות הצפוי בשירות מילואים, לרבות מחמת תדירותו או משכו, בכל אחד מאלה: קבלה לעבודה, תנאי עבודה, קידום בעבודה, הכשרה או השתלמות מקצועית, פיטורים או פיצויי פיטורים, הטבות ותשלומים הניתנים לעובד בקשר לפרישה מעבודה"; וכן ס' 3(א) לחוק איסור הפליה, לעיל ה"ש 2: "מי שעיסוקו בהספקת מוצר או שירות ציבורי או בהפעלת מקום ציבורי, לא יפלה בהספקת המוצר או השירות הציבורי, במתן הכניסה למקום הציבורי או במתן שירות במקום הציבורי, מחמת גזע, דת או קבוצה דתית, לאום, ארץ מוצא, מין, נטייה מינית, השקפה, השתייכות מפלגתית, גיל, מעמד אישי, הורות או לכישת מדי כוחות הביטחון וההצלה או ענידת סמליהם".

9 ראו ס' 1(ג) לחוק איסור הפליה, ס' 2(א)(ג) לחוק שוויון הזדמנויות בעבודה, לעיל ה"ש 2. ראו גם דברי הסבר לס' 3 בהצעת חוק איסור הפליה במוצרים, בשירותים ובכניסה למקומות ציבוריים, התש"ס-2000, ה"ח 2871. לפי דברי ההסבר, מטרת איסור ההפליה היא מניעת פגיעה בכבודו של אדם ששייך לקבוצה שלגביה קיימת היסטוריה של קיפוח והפליה בעבר. ואכן, גורם מנחה לדיני איסור ההפליה בעולם הוא היסטוריה של הפליה שיטתית כנגד קבוצות מסוימות. נתערב גם אם הקבוצה אינה במיעוט מספרי, כמו נשים. גבולותיו של חוק שוויון הזדמנויות בעבודה אוסרים על הפליית עובדים או דורשי עבודה עקב השתייכותם לקבוצה מוגנת, כמו נטייה מינית, היריון, הורות ועוד.

במאמר נבחן את סוגיית ההפליה האלגוריתמית תחת המטרייה השוקית של מוצרים ציבוריים שעליהם חל חוק איסור הפליה (כמו מערכת אלגוריתמית להשמת עובדים, או מערכת אלגוריתמית לחישוב פרמיית ביטוח). כפי שנראה בהמשך המאמר, הפרויקט האלגוריתמי מציב אתגר של ממש לדיני איסור ההפליה בפרט, ולזירת המשפט בכלל. ההשוואה שנערוך תראה כי אף על פי שהשימוש במערכות מבוססות אלגוריתמים יכול למתן הפליה אנושית הנובעת מהטיות אנושיות, טעמים, סטראוטיפים ותסריטים חברתיים, מערכות בינה מלאכותית אינן יכולות להימנע מהפליה סטטיסטית, שעלולה להפלות לרעה חלק ממשתמשי המוצר. הסטטיסטיקה היא לחם חוקן של האלגוריתם, ומתחייבת ממהותו. לכן, במאמר זה הצעתנו היא גיבוש מדיניות משפטית נפרדת עבור הפרויקט האלגוריתמי בכל הנוגע לדיני איסור הפליה: נציע למקד את דיני איסור ההפליה כך שיעסקו רק במקרים של קבלת החלטות אנושית. לעומת זאת, במקרים של קבלת החלטות אלגוריתמית, במקום להסתכל על השאלה האם החלטת האלגוריתם הייתה מפלה או לא, תיבחן התוצאה וההשפעות החלוקתיות שלה. כך, כאשר בני אדם מקבלים החלטה, נתייחס אליה דרך דיני איסור ההפליה הקיימים ונבחן את היותה החלטה מפלה; כאשר אלגוריתם מקבל החלטה, אנו סבורות כי על המשפט לקבל הכרעה חלוקתית-תוצאתית, לכייל את תוצאות האלגוריתם, ולמעשה לפקח על התוצאה.

דרך הילוכנו במאמר זה תהיה כדלקמן: **בפרק א** נתאר את סוגי ההפליה האנושית. לאחר מכן, נעבור **בפרק ב** לתאר את מעגל החיים של ההחלטה האלגוריתמית, קרי, כיצד האלגוריתם מקבל את ההחלטה (שתוצאתה עלולה להיות מפלה). **בפרק ג** נעסוק בהשוואה בין ההפליה האנושית וההפליה האלגוריתמית. **בפרק ד** נתאר את המסגרת המשפטית הקיימת ונדון באתגרים המשפטיים הנובעים מהאתגר שהפרויקט האלגוריתמי מציב כלפי דיני איסור ההפליה. לבסוף, נסכם.

א. אדם מול מכונה – הגורמים להפליה האנושית

הפליה היא, כפשוטה, יחס שונה לשווים. יש לה משמעויות כלכליות, אישיות, חברתיות ואחרות. כאשר אנחנו מקבלות יחס פחות טוב מאדם שמאפייניו דומים לשלנו, נגרעים מאיתנו משאבים שמקבלים אחרים. אנחנו מרגישות פגועות ומאוכזבות. כפי שהגדיר זאת השופט בך, "ההפליה היא נגע היוצר תחושת קיפוח ותסכול; חברה אשר בה נוהגת הפליה אינה חברה בריאה, ומדינה שבה נוהגת הפליה לא תוכל להיקרא מדינה מתוקנת"¹⁰.

להפליה האנושית פנים רבות. נהוג לחלק את גורמי ההפליה לשלושה במספר: הפליית טעמים (הכוללת גם הפליה נורמטיבית), הפליה סטראוטיפית והפליה סטטיסטית. הגם שמופעים שונים של הפליה עשויים להיות מונעים על ידי יותר מגורם אחד, נשתמש בהבחנות אלה על מנת להשוות בין הפליה אנושית להפליה אלגוריתמית. נציין כבר כאן כי דיני איסור ההפליה הישראליים אינם משתמשים בהבחנות אלה והם

10 בג"ץ 104/87 נבו נ' בית הדין הארצי לעבודה, מד(4) 749 (1990).

מבססים את איסורי ההפליה על התכונה שעל בסיסה נעשית ההפליה (מוצא ומגדר בהשוואה לצבע שיער, למשל) וזירת החיים (עבודה ושוקי שירותים ומוצרים בהשוואה להיכרויות רומנטיות, למשל). השימוש שלנו בהבחנות אלה נועד לדיון בשאלות מדוע אנשים ואלגוריתמים מפלים.

1. הפליית טעמים

הפליית טעמים מקורה בטעמים הסובייקטיביים המפלים או הגזעניים של אנשים. היא אינה מונעת מסטראטיפים או מהבחנות סטטיסטיות, וגם לא מטעמים כלכליים. כך, למשל, אנו עשויים כמעסיקות להפלות במודע גברים, או עובדים ממוצא מסוים, משום שאיננו נהנות לעבוד עימם. עם הפליית טעמים, אנו מוכנות לוותר על רווחים או לשלם יותר כדי להיענות לטעמים המפלים שלנו. כך, למשל, אנו עשויים להיות מוכנות להפסיד עובדים מצוינים או לקוחות שמוכנים לשלם לנו מחירים גבוהים בעבור השירות שאנו נותנות, כדי להיענות לטעמים המפלים שלנו. הטעמים המפלים שלנו עשויים להיות גם טעמים אישיים-מוסריים באופיים, כלומר טעמים שמושפעים מהעדפה נורמטיבית אישית הנגזרת מנורמה מקובלת בחברה או בתרבות. מקרה של טעמים אישיים-מוסריים מפלים נדון למשל על ידי בית המשפט העליון בארצות הברית בשנת 2018, שם נפסק לטובת בעל מאפייה מקולורדו שסירב לייצר עוגת חתונה עבור זוג גברים בשל התנגדותו הדתית לנישואים גאים.¹¹ בסיווג עובדות פסק הדין לענייננו, התנגדות לתת שירות מטעמים דתיים מהווה הפליה על בסיס טעמים נורמטיביים מפלים. בית המשפט בקולורדו פסק אומנם כי מדובר בהפליה אסורה, אך בית המשפט העליון הפך את ההחלטה מטעמים ספציפיים לתיק עצמו.

בין אם הטעמים המפלים הם טעמים "סתמיים" ובין אם הם טעמים "נורמטיביים", חלק מתומכי הגישה הכלכלית למשפט אינם נבהלים מהפליה זו, וגורסים כי היא תיעלם לאורך הזמן. בהיבט השוקי, יש כלכלנים שטוענים כי הפליה שכזו לא תשרוד את התחרות בין בעלי העסקים, משום שהעסקים המפלים ייוותרו עם עובדים מוכשרים פחות או עם לקוחות המעוניינים לשלם להם פחות.¹² עם זאת, כדי שהפליית הטעמים באמת תיעלם ללא התערבות, נדרשת תחרות בין בעלי העסקים וגם שהלקוחות עצמם לא יפלו אותן קבוצות, שאז היא עשויה להפוך דווקא ליעילה כלכלית, כמו למשל במקרה של מסעדה הנמנעת מלהעסיק מלצרים ששייכים למגדר, לגזע או לנטייה מינית מסוימים, וזאת על מנת להיענות להעדפות הלקוחות שלה.

¹¹ ;*Masterpiece Cakeshop, Ltd. v. Colorado C.R. Comm'n*, 584 U.S. 617 (2018) להרחבה ראו רויטרס "העליון בארה"ב פסק לטובת אופה נוצרי שסירב לאפות עוגת חתונה לזוג הומואים" **הארץ** (4.6.2018), <https://www.haaretz.co.il/news/world/america/2018-06-04/ty-article/0000017f-e735-dc7e-adff-f7bd91550000>.

¹² .GARY S. BECKER, *THE ECONOMICS OF DISCRIMINATION* (1957).

לכן, כפי שנעשה בחוק איסור הפליה, לטעמנו נכון ורצוי לאסור בנסיבות מסוימות אפליית טעמים על בסיס הבחנות מסוימות (לאום, מגדר, נטייה מינית) – במטרה להרתיע מעסיקים או בעלי עסקים ולאזנת כי ההתנהגות אינה ראויה.

2. הפליה סטראוטיפית

סטראוטיפים הם "הכללות רווחות, פשטניות בדרך כלל, על מאפיינים של קבוצה או של סוג אנשים".¹³ מערכות הסטראוטיפים עשויות להוות את הבסיס למסורת האישית שלנו, ולשמש כמנגנון לשמירה על מעמדנו החברתי. הן מציירות תמונה מסודרת ואולי אף עקבית של העולם, שבה השתלבו ההרגלים, הטעמים, היכולות, הנוחות והשאיפות שלנו.¹⁴

לצורך המאמר, ולשם יצירת הבחנה בין הפליה המבוססת על הכללות ואמונות שהן מוטעות או שאינן מדויקות לבין כזו המבוססת על הכללות ואמונות שהן מדויקות במהותן, נתייחס לסוג הראשון כאל הפליה סטראוטיפית ולאחרון כאל הפליה סטטיסטית. הצורך בהבחנה זו חשוב הן לשם הדגשת ההבדלים שבין ההפליה האנושית להפליה האלגוריתמית, והן לשם הדיון הנורמטיבי בסוגי ההפליה.

כמו הפליית טעמים, גם הפליה סטראוטיפית נוטה להתרחש ביחסים בין אנשים. בהפליה סטראוטיפית אנחנו משתמשים בסטראוטיפים או באמונות תרבותיות¹⁵ לגבי קבוצות של אנשים ופועלים לפיהם. במצבים רבים, השימוש בסטראוטיפים או באמונות תרבותיות חוסך זמן ומאפשר קואורדינציה ושיתוף פעולה בין אנשים. למעשה, אלו מסייעים לנו להתנהל בעולם, ומהווים שבילים ב"מפת הדרכים" של האינטראקציה החברתית, ויוצרים תסריטים חברתיים שמובילים אותנו בכל אינטראקציה עם בני אדם: במכולת, באוניברסיטה, במעלית, בבנק. התסריטים הללו עוזרים לנו, בבסיסם, לתקשר עם אנשים. בלעדיהם, האינטראקציות שלנו היו ארוכות יותר, מסובכות יותר, והיו דורשות מאיתנו אנרגיה רבה של תכנון והכנה. התסריטים הללו מכתיבים התנהגות, ולכן הם גם יוצרים מציאות.¹⁶

13 "למד לשונך 50" האקדמיה ללשון העברית <https://hebrew-academy.org.il/2003/09/14/למד-לשונך-50>.

14 WALTER LIPPMANN, PUBLIC OPINION 94 (1922).

15 הפליה סטראוטיפית והפליית טעמים משיקות, אך אינן זהות. הפליית טעמים מבוססת על ההעדפות האישיות שלנו, והיא מעוגנת בצרכים שוקיים. מנגד, סטראוטיפים הם דעות קדומות, ונגזרים בין היתר מהבניות תרבותיות, חברתיות והיסטוריות.

16 בתחום הפסיכולוגיה החברתית, תסריטים חברתיים הם מבנים המייצגים דפוסי התנהגות, מחשבות, ורגשות המקובלים בחברה או בתרבות מסוימת. הם מעין "מדריכים פנימיים" שמשמשים לנו כמצפן חברתי, ומסייעים לנו להבין כיצד לפעול בסיטואציות שונות בחיים החברתיים. הם מתפתחים מהניסיון האישי שלנו ומהתבוננות בסביבה החברתית. לדוגמה, תסריט של הגעה לבית קפה יכול להיות התיישבות על כיסא, הזמנת משקה, תשלום ונתינת טיפ. התסריט החברתי משתנה בין תרבויות, ולכן בתרבויות שבהן השארת טיפ אינה נהוגה, התסריט החברתי יהיה בלעדיה, ועשויים להתווסף אליו שלבים נוספים.

לצד היתרונות והיעילות שבהישענות על אמונות חברתיות ועל תסריטים תרבותיים כדי לפענח את עולמנו החברתי, הסטראוטיפים שלנו לגבי קבוצות שייכות של אנשים נוטים להיות לא מדויקים, ולהטות את ההתנהגות שלנו ואת האופן שבו אנחנו מעריכות את כישוריהן והתנהגותן של אחרות. בכך, הסטראוטיפים מעמיקים את אי-השוויון בין קבוצות מסוימות בחברה, ואף מצדיקים אותו.¹⁷

כך, למשל, בסדרת ניסויים מפורסמת נמצא כי משתתפים שהתבקשו לקרוא מאמר שנאמר להם שנכתב בידי גבר העריכו את המאמר כאיכותי יותר בהשוואה למשתתפים שהתבקשו להעריך אותו מאמר בדיוק לאחר שנאמר להם שכתבה אותו אישה.¹⁸ מחקרים אחרים הראו, הן בתנאי מעבדה והן בתנאי שדה, כי משתתפים בניסוי (ומעסיקים) נוטים להעריך את קורות החיים של גבר שמועמד לעבודה כמוצלחים יותר בהשוואה למשתתפים בניסוי (ומעסיקים) שמתבקשים להעריך אותם קורות חיים בדיוק לאחר שנאמר להם כי הם קורות חייה של אישה.¹⁹ סדרת מחקרים נוספת בחנה הטיות גילניות בהליכי גיוס, והשוותה מועמדים צעירים ("מועמד א") ומבוגרים ("מועמד ב") בעלי כישורים דומים, כאשר לשתי הקבוצות לא היה ניסיון תעסוקתי קודם הרלוונטי למשרה. למשתתפים הוצג סט מיומנויות המאפיין כל מועמד: סט המיומנויות ה"מבוגר" כלל מיומנויות כמו יישוב ויכוחים, הבנת נקודות מבט של אחרים ויכולת להתמודד עם אנשים אחרים בנימוס. סט המיומנויות ה"צעיר" כלל מיומנויות כמו למידת כישורים חדשים, יצירתיות ושימוש ברשתות חברתיות. הגם שמידע לגבי גיל המועמדים לא נחשף למשתתפי המחקר, אלו הניחו שמועמד א צעיר יותר, וככל שהנחה זו הייתה חזקה יותר, הם העדיפו להעסיק אותו על פני המבוגר (באחד המחקרים, 80% ממשתתפי המחקר בחרו להעסיק את המועמד עם הפרופיל ה"צעיר" יותר, אף על פי ששני המועמדים תוארו באופן מפורש כבעלי כישורים דומים). המחקרים מצאו שקיימות הטיות סטראוטיפיות בעלות הנחות גילניות, המעדיפות מועמדים צעירים על פני מועמדים מבוגרים.²⁰

להרחבה על תסריטים חברתיים, ראו: ROGER C. SCHANK & ROBERT P. ABELSON, SCRIPTS, PLANS, GOALS AND UNDERSTANDING: AN INQUIRY INTO HUMAN KNOWLEDGE STRUCTURES (1977).

17 תמר קריכלי-כץ "אישוויון מגדרי והפליה סטריאוטיפית – מחשבות בעקבות פסקהדין בעניין מירב נ' אי.די.איי" **עיוני משפט** מב 271, 272 (2019) (להלן: תמר קריכלי-כץ "אי שוויון מגדרי והפליה סטריאוטיפית").

18 להרחבה על הניסויים, ראו בהתאמה: Claude M. Steele & Joshua Aronson, *Stereotype Threat and the Intellectual Test Performance of African Americans*, 69 J. PERSONALITY & SOC. PSYCHOL. 797 (1995); Steven J. Spencer et al., *Stereotype Threat and Women's Math Performance*, 35 J. EXP. SOC. PSYCHOL. 4 (1999) וכן Philip Goldberg, *Are Women Prejudiced Against Women?*, 5 SOC'Y 28 (1968).

19 תמר קריכלי-כץ "אי שוויון מגדרי והפליה סטריאוטיפית", לעיל ה"ש 17, בעמ' 275–276.

20 Dominic Abrams et al., *Old and Unemployable? How Age-Based Stereotypes Affect Willingness to Hire Job Candidates*, 72 J. SOC. ISSUES 105, 107–108, 111–119.

ישנם כלכלנים שטוענים כי כמו הפליית טעמים, גם ההפליה הסטראוטיפית תיעלם.²¹ הם טוענים כי מי שייקח נשים לעבודה יצליח, ומי שיפלה אותן בהפליה סטראוטיפית – יפסיד עובדות מוכשרות, וכנגזרת מכך יפסיד כסף. עם זאת, לסטראוטיפים יש נטייה להנציח את עצמם, ולכן לעיתים יהיה קשה מאוד להילחם בהם.²² בנוסף, הפליה סטראוטיפית עשויה להיות לא מודעת, וכשהיא כזו קשה מאוד לשנות אותה, ולשוק קשה לתקן אותה. כך, למשל, אם סטראוטיפ מסוים הוא תסריט חברתי שרבים מאמינים בו ומתנהגים לפיו (כולל מי שהסטראוטיפ מופנה אליהן) – יהיה קשה לראות שהוא מוטעה ויהיה קשה מאוד להקטין את ההפליה הסטראוטיפית. לכן, לטעמנו נכון ורצוי לאסור בנסיבות מסוימות הפליה סטראוטיפית על בסיס הבחנות מסוימות (לאום, מגדר, נטייה מינית), במטרה להרתיע מעסיקים או בעלי עסקים ולאותת כי ההתנהגות אינה ראויה.²³

3. הפליה סטטיסטית

בדומה להפליה סטראוטיפית, גם ההפליה הסטטיסטית נשענת על הכללות לגבי קבוצות של אנשים. אלא שבשונה מההפליה הסטראוטיפית, כאן מדובר בהכללות שהן מדויקות סטטיסטית.

המושג "הפליה סטטיסטית" הוצג על ידי הכלכלנים אדמונד פלפס וקנת' ארו בתחילת שנות השבעים של המאה הקודמת. השניים פיתחו תאוריה כלכלית שלפיה במציאות שבה חסר מידע ישיר לגבי עניין עובדתי או תכונה מסוימת, מקבלת החלטות רציונלית עשויה להשתמש בממוצע במקום בערך עצמו, שאינו ידוע לה (לדוגמה, הבדל בגובה הממוצע בין נשים וגברים). כך, למשל, הפליה בשוק העבודה עשויה להיגרם מכך שמעסיקים אינם יודעים בוודאות את תפוקת העבודה הפרטנית של כל עובד ועובדת, ולכן יעדיפו לבסס את החלטתם לגבי העסקת עובדים פוטנציאליים לפי תכונות זמינות, כמו מגדר או גיל. המעסיקים ישתמשו בכך שיש קורלציה מסוימת בין התכונות

21 ראו William A. Darity & Patrick L. Mason, *Evidence on Discrimination in Employment: Codes of Color, Codes of Gender*, 12 J. ECON. PERSP. 63, 82 (1998); Kerry Kawakami et al., *Just Say No (to Stereotyping): Effects of Training in the Negation of Stereotypic Associations on Stereotype Activation*, 78 J. PERSONALITY & SOC. PSYCH. 871 (2000) וכן Larry Alexander, *What Makes Wrongful Discrimination Wrong? Biases, Preferences, Stereotypes, and Proxies*, 141 U. PA. L. REV. 149, 165 (1992).

22 Lawrence Blum, *Stereotypes & Stereotyping: A Moral Analysis*, 33(3) PHILOSOPHICAL PAPERS 251, 270 (2004).

23 כך גם המצב הקיים לפי דיני איסור ההפליה. ראו סעיף 1 לדברי הסבר להצעת חוק איסור הפליה במוצרים, בשירותים ובכניסה למקומות בידור ולמקומות ציבוריים, התשס"א-2000, ה"ח 2871.

הרצויות למעסיק לבין שיוך לקבוצה, וזאת על מנת לקבל החלטה מושכלת בהיעדר מידע לגבי העובד הספציפי.²⁴

ההפליה הסטטיסטית נפוצה בשוק הפרטי בעולם הביטוח ובשוק העבודה. בעולם הביטוח, חברות שונות רשאיות להפלות בין מבוטחים שונים במחירי הפוליסה שהם נדרשים לשלם על סמך הערכת סיכונים סטטיסטית, אפילו אם המבוטחים משתייכים לקבוצה שהמשפט רגיל לחשוב עליה כעל קבוצה מוגנת; חברות הביטוח הסיעודי רשאיות לגבות פרמיה גבוהה יותר ממבוטחים מבוגרים בהשוואה למבוטחים צעירים.²⁵ בעולם, מדינות נבדלות בשאלה אילו מופעים של הפליה סטטיסטית בתחום הביטוח הן מאפשרות. כך, למשל, בשנת 2012 פסק בית הדין האירופי לצדק שהפליה במחירי פוליסה על רקע מגדרי היא אסורה.²⁶ בארצות הברית אסורה הפליה במחירי פוליסה על בסיס גזע, אף על פי שהוכח סטטיסטית שישנם הבדלים בתוחלת החיים בין אנשים לבנים לבין אנשים שחורים.²⁷

כאמור, הבחנה סטטיסטית נעשית לרוב משיקולי יעילות: הסתמכות על השתייכותם של אנשים לקבוצה מסוימת נעשית במטרה לחסוך את העלויות הכרוכות בבדיקה פרטנית של תכונותיהם האישיות של כל אחת ואחד מהם.²⁸ עם זאת, כאשר מסתכלים על המקרים הפרטניים, ההבחנה הסטטיסטית עשויה ליצור תוצאה מפלה. כך, למשל, נמצא שאנשים החיים בעוני משלמים יותר עבור מגוון מוצרים כמו אשראי, חשמל

- Edmund S. Phelps, *The Statistical Theory of Racism and Sexism*, 62 AM. ECON. REV. 659 (1972); Kenneth J. Arrow, *The Theory of Discrimination*, in DISCRIMINATION IN LABOR MARKETS 3 (Orly Ashenfelter & Albert Rees eds., 1973).²⁴
- Ronen Avraham, *Discrimination and Insurance*, in THE ROUTLEDGE HANDBOOK OF THE ETHICS OF DISCRIMINATION 335–347 (Kasper Lippert-Rasmussen ed., 2007).²⁵
- לעניין איסור על הבדלים מגדריים בביטוח באיחוד האירופי, ראו הודעה לתקשורת, European Commission Press Release IP/12/1430, EU Rules on Gender-Neutral Pricing in Insurance Industry Enter Into Force (Dec. 20, 2012).²⁶
- James B. Kirby & Toshiko Kaneda, *Unhealthy and Uninsured: Exploring Racial Differences in Health and Health Insurance Coverage Using a Life Table Approach*, 47 DEMOGRAPHY 1035, 1043 (2010) (“Comparing total and healthy life expectancies at birth by race shows that life expectancy is 5.3 years shorter for blacks than for whites”); כמו כן, ראו את התייחסות האגודה הלאומית של המפקחים לביטוח בארצות הברית בנוגע להפליה גזעית בביטוח, NAT’L ASSOC. OF INS. COMM’RS., MILESTONES IN RACIAL DISCRIMINATION WITHIN THE INSURANCE SECTOR (2020), <https://content.naic.org/sites/default/files/cipr-report-milestones-racial-discrimination.pdf>.²⁷
- מיטל פינטו “מדוע שרירותיות כשלעצמה מפלה, גם אם אינה קשורה לעילות חשודות” עיוני משפט 241, 264 (2021); וכן שרון רבין מרגליות “המקרה החמקמק של הפליה בעבודה – כיצד מוכיחים את קיומה?” הפרקליט 529 (1999) (להלן: רבין מרגליות, הפליה בעבודה).²⁸

וביטוח גם אם ההבחנה הממוצעת אינה משקפת את מאפייניהם האישיים כצרכנים.²⁹ ההפליה הסטטיסטית עשויה לנבוע לאו דווקא ממידע ישיר (שהשימוש בו אסור בחלק מהמדינות) אלא ממשתנים מנבאים, כמו מיקוד דואר שיכול להעיד על מצב סוציו-אקונומי. תופעה זו מכונה לעיתים קרובות במונח "תיחום אדום" (redlining), שנטבע בתחילת שנות השבעים של המאה הקודמת על ידי סוציולוגים בשיקגו. המונח נובע מהפרקטיקה של סירוב במתן שירותים או העלאת מחירים של שירותים כגון משכנתאות או ביטוח בריאות לתושבים באזורים מסוימים, לעיתים קרובות על בסיס גזעי.³⁰ באופן דומה, אם מעמד סוציו-אקונומי מגביל את גישתו של אדם לשכונות "טובות" או בטוחות יותר, הוא עשוי לשלם פרמיית ביטוח גבוהה יותר בשל הסיכון הגבוה הנתפס כקשור למיקומו הגאוגרפי, גם אם רמת הסיכון האישית שלו נמוכה יותר.³¹ בספרות נדונה בהרחבה הסוגיה של הפליה סטטיסטית בנוגע להעמדת אשראי צרכני בידי מלווים בנקאיים וחוז' בנקאיים בישראל, המכונה "חיתום אשראי". מרכיב מרכזי במודלים לחיתום אשראי הוא דירוג אשראי שמופק על ידי לשכות אשראי חיצוניות, והוא מבוסס על מודל סטטיסטי שמאפשר להעריך את כושרם של לקוחות לעמוד בהתחייבויותיהם ולשלם את חובותיהם במועד. בעבר, הערכת אמינותו ויכולתו של מבקש אשראי לעמוד בהתחייבויות התבססה על שיקול דעת אישי וניסיון של נותני האשראי. כיום, חלק ניכר מהערכות אלו מתבצע באמצעות מודלים אלגוריתמיים. המודלים משתמשים בטכניקות סטטיסטיות כדי למיין ולדרג לקוחות על פי רמת סיכון האשראי שלהם, תוך הסתמכות על היסטוריית האשראי של הלקוח ונתונים של לקוחות אחרים בעלי מאפיינים דומים. כפי שנראה בהמשך המאמר,³² מודלים אלגוריתמיים המשמשים לחיתום אשראי מנבאים את סיכון האשראי על בסיס נתונים קיימים על לקוחות, שעשויים לשקף פרקטיקות מפלות.³³ ההפליה הסטטיסטית נדונה בהקשר אחר, גם בעניין אבו חנא, שם דן בית המשפט העליון בחישוב על סמך נתונים סטטיסטיים לצורך קביעת כושר השתכרות עתידי של ילדה בדואית. בפסק הדין, השופט ריבלין העביר ביקורת על שימוש זה, וקבע כי הוא

INSTITUTE AND FACULTY OF ACTUARIES, THE HIDDEN RISK OF BEING POOR: THE POVERTY PREMIUM IN INSURANCE (2021) 29

INST. FOR POL'Y. RSCH., IPR AT 40 (2009), ראו: 30
<https://www.ipr.northwestern.edu/documents/year-in-review/iprat40.pdf>
 Redlining in Boston, NORTHERSTERN UNIVERSITY (Mar. 22, 2016),
<https://servicelearningexpo.sites.northeastern.edu/2020/12/22/the-lasting-effects-of-redlining-in-boston>; ARTHUR CHAPMAN, INSURANCE, BIASES, DISCRIMINATION AND FAIRNESS 1, 3 (2024)

Faisal Kamiran & Toon Calders, *Classifying Without Discriminating*, 2ND INT'L. 31
 CONF. ON COMPUT., CONTROL & COMM'C'N (2009); ARHUR CHARPENTIER,
 INSURANCE, BIASES, DISCRIMINATION AND FAIRNESS 260 (2024)

ראו הערות שוליים 48 ו-57. 32

רות פלאטו-שנער ומעיין פרל "חיתום אשראי צרכני על סמך עיבוד מידע אלגוריתמי – 33
 היש סיבה לחשוש מהפליה אסורה?" **עיוני משפט** מה 553, 556, 570, 609 (2022).

נותן תוקף לחלוקת המשאבים הקיימת בחברה ואינו ראוי נורמטיבית.³⁴ כך, ניתן לומר כי הבחנות סטטיסטיות פוגעות באוטונומיה וביכולת של אנשים לכתוב את סיפור חייהם. אנחנו מגשימות את האוטונומיה שלנו בצורה אידיאלית כאשר יש לנו החירות והיכולת המלאה לבצע החלטות שיקדמו את סיפור החיים שאנו רוצות בו.³⁵ למעשה, ההבחנה הסטטיסטית עשויה לנתק אותנו מסיפורנו האישי ולהדביק אליו משתנים שאומנם זרים לנו כאינדיבידואליות, אך מאפיינים בממוצע את הקבוצה שאליה אנו שייכות. כך, החלטה לגבינו שבבדיקה פרטנית יכלה להיות הוגנת וצודקת, עשויה להיות מפלה רק בשל השייך הקבוצתי שלנו.

לכן, ואף על פי שתהליך קבלת החלטות באופן סטטיסטי עשוי להוביל את המעסיקים או את נותני השירותים לקבלת החלטות יעילות או מדויקות יותר בממוצע, הוא עשוי להיחשב כהפליה אסורה על פי דיני איסור ההפליה, שכן נוצרת פגיעה באנשים המשוויכים לקבוצה המובחנת. חשוב לציין שבמקרים שבהם תוכר התנהגות מסוימת כהפליה סטטיסטית, משמעות הדבר תהיה כי המעסיקים או נותני השירותים יאלצו לוותר על השימוש בהבחנה, ולכן יתבקשו לבחון את המאפיינים הספציפיים של מועמדים לעבודה או מקבלי שירותים, ולא יוכלו להסתמך על הבחנות סטטיסטיות לגביהם, גם אם הבחנות אלה מדויקות. כך, למשל, יידרשו מעסיקים לבחון כל מועמד או מועמדת מבוגרים לעבודה מאתגרת מבחינה פיזית באופן אינדיבידואלי, ולא יוכלו להסתמך על הבחנות קבוצתיות לגבי המגדר או הגיל שלהם, גם אם הבחנות אלה הן מדויקות מבחינה סטטיסטית בנוגע לעבודה מאומצת.

עניין **רקנט** עסק בסוגיה זו.³⁶ העניין נסוב על הסכם העבודה הקיבוצי של חברת התעופה אל-על שקבע כי אנשי צוות דיילי אוויר חייבים לפרוש מעבודתם באל-על בגיל 60, בעוד שיתר העובדים באל-על חייבים לפרוש בגיל 65. המחלוקת הגיעה לבית הדין האזורי לעבודה, ולאחר כמה גלגולים הגיעה אל שולחן בג"ץ. שם, חברת אל-על טענה כי קיים שוני רלוונטי: בגיל מסוים אנשים אינם יכולים לעבוד כדיילי אוויר שכן באירוע חירום דיילים מעל גיל 60 יתקשו בחילוץ הנוסעים. כמו כן טענה חברת אל-על

34 ע"א 10064/02 **"מגדל" חברה לביטוח בע"מ נ' רים אבו חנא**, פ"ד ס(3) 13, פס' 37 לפסק הדין של השופט ריבלין (2005). פסק הדין, מהזירה המשפטית של דיני נזיקין, עסק בשאלת הפיצויים בגין אובדן יכולת השתכרות כאשר הניזוק (או הניזוקה) הוא ילד הממשיך למגור או מגדר מסוים. זאת מכיוון שבשונה ממבוגרים, אצל קטינים קיימת עמימות לגבי עיסוקם ושכרם העתידי. בפסק הדין דובר על הילדה הבדואית רים אבו חנא, שנפצעה בעודה בת 5 חודשים בתאונת דרכים. חברת הביטוח טענה כי הסכום שנקבע בבית המשפט המחוזי, על סך 500,000 ש"ח, הוא גבוה מדי, שכן הסיכוי **שהיא** תשתכר כך בעתיד הוא אפסי, מכיוון שרוב הנשים ממגזר וממקום מגוריה של אבו חנא אינן משתלבות בשוק העבודה. הערעור לעליון נסב גם על שאלות של צדק חלוקתי והפליה; הפיצוי שקבע המחוזי (שבוסס על השכר הממוצע במשק) נשאר על כנו. ראו גם דפנה ברק ארז **תקדימיות – נשים קובעות תקדים במשפט הישראלי** 207 (2024).

35 JOSEPH RAZ, THE MORTALITY OF FREEDOM 369–370 (1986).

36 דנג"ץ 4191/97 **רקנט נ' בית-הדין הארצי לעבודה**, פ"ד נד(5) 330 (2000) (להלן: **עניין רקנט**).

כי אחד השיקולים שהנחו אותה בחיוב דיילי צוות אוויר לפרוש לגמלאות בגיל 60, כחריג לגיל הפרישה המקובל בחברה שעמד על 65, הוא הצורך שדייליה יהיו בעלי הופעה צעירה, כמצופה על ידי נוסעיה.³⁷ בית המשפט קבע שמדובר בהפליה פסולה, האסורה לפי חוק שוויון הזדמנויות בעבודה. נקבע כי שיקול הגיל אינו יכול להצדיק את הפיטורין, וכי ניתן לפטר עובדים רק בתלות במבחני כושר וכדומה.³⁸ במילים אחרות, נקבע שההפליה הסטטיסטית מחמת גיל הינה הפליה אסורה, וכי על מנת לפטר עובדים יש לבחון באופן אינדיבידואלי את הכושר והמסוגלות הפיזית שלהם.

ישנם כלכלנים אשר טוענים שבניגוד למנגנוני ההפליה האחרים שתיארנו, כמו הפליית טעמים או הפליה סטראוטיפית, ההפליה הסטטיסטית היא יעילה ומדויקת במהותה, ולכן לא מוצדק ולא יעיל לאסור על קיומה.³⁹ כאמור, חוק איסור הפליה בשירותים ומוצרים וחוק שוויון הזדמנויות בעבודה אוסרים להפלות כנגד קבוצות מוגנות, והפליה סטטיסטית נכללת בסל משפטי זה.⁴⁰

לשיטתנו, האתגר המשפטי בעניין דיני איסור הפליה מתרחב כאשר מדובר בהפליה אלגוריתמית, שכפי שנראה בפרק ג למאמר זה חבויים בחיקה כלל סוגי ההפליה שמנינו. בפרק הבא נתייחס למעגל ההחלטות האלגוריתמי ונראה כי כמעט בלתי אפשרי להתחקות אחר הגורמים להחלטה שקיבלה הבינה המלאכותית. נראה כי כל קבלת החלטות אלגוריתמית מבוססת במהותה על הבחנות סטטיסטיות (שאינן מופנות בהכרח כנגד קבוצות המוגנות על ידי דיני איסור ההפליה). לעיתים, ההבחנות הסטטיסטיות תתבססנה על תכונות חשודות (כמו לאום או מגדר), ואומנם תהיינה מדויקות סטטיסטית, אך לא רצויות נורמטיבית. לכן, אנו סבורות שהפרויקט האלגוריתמי מציב אתגר של ממש לדיני איסור הפליה. נדון בכך כעת.

ב. מעגל החיים של ההחלטה האלגוריתמית

הניסיון לענות על השאלה מהי מערכת בינה מלאכותית הוא מורכב, ובישראל לא נקבעה לה הגדרה מחייבת בדין.⁴¹ בדו"ח הצוות הבין-משרדי "בינה מלאכותית בסקטור הפיננסי" הוצע להגדיר את מערכת הבינה המלאכותית כ"מערכת מבוססת מחשב הפועלת בדרגות שונות של אוטונומיה, במטרה להפיק תוצרים כגון תוכן, תחזיות, המלצות או החלטות, אשר עשויה להיות לה השפעה על משקיעים, לקוחות או על פעילותו של הגוף המפוקח". אנו מתמקדות במערכות בינה מלאכותית (Algorithm)

37 רבין מרגליות, **הפליה בעבודה**, לעיל ה"ש 28, בעמ' 353.

38 נציין כי בפסיקה נקבע כי אם אל-על תבחר לקיים מבחני כושר, אלו יהיו תקופתיים ויחולו על כל הדיילים באשר הם, ולא רק על המבוגרים (**עניין רקנט**, לעיל ה"ש 36, בפס' 26 לפסק הדין של הנשיא ברק).

39 Gary S. Becker, *The Economics of Discrimination*, לעיל ה"ש 12, בפרק 2.

40 ראו סעיף 2 לחוק שוויון הזדמנויות בעבודה וסעיף 3 לחוק איסור הפליה.

41 רשות ניירות ערך **בינה מלאכותית בסקטור הפיננסי** 8 (2024). הצוות בראשות עו"ד אמיר וסרמן הורכב מנציגים מהרשות לניירות ערך, משרד המשפטים, בנק ישראל, משרד האוצר, רשות שוק ההון ורשות התחרות.

(Decision Making) שלתוצריהן תהיה השפעה על החלטות שוקיות. כך, למשל, באמצעות שימוש במערכות כאלה יוכלו מעסיקים לחסוך בעלויות המיון של מתמודדים לעבודה, ובמקום להשתמש בגורמים אנושיים ישתמשו במערכות בינה מלאכותית אשר תסרוקנה את נתוני המועמדים ותקבלנה החלטות בדבר המשך תהליכי המועמדות שלהם.

אותן מערכות קבלת החלטות מבוססות אלגוריתמים נועדו לסייע או להחליף את תהליך קבלת ההחלטות האנושי. מערכות אלו יכולות לקבל החלטות מהירות יותר, ובשאיפה גם טובות יותר מאלה האנושיות, ובו בזמן לאפשר לבני אדם להקדיש את זמנם ואת תשומת ליבם לפעילויות אחרות.⁴² למרות היתרונות של מערכות אלה, ויכולתן לצמצן הטיות אנושיות מסוימות, עולה החשש שחלק מההחלטות תהיינה נגועות בסוגים מסוימים של הפליה.⁴³ כפי שנראה בהמשך פרק זה, הגם שהבינה המלאכותית עשויה לצמצם פערים שמקורם בטעמים מפלים ביחסים בין בני אדם, ישנם פערים אחרים שהיא עשויה להעצים, לדוגמה כאשר תשחזר הטיה קיימת (שמוטמעת בנתונים שעליהם המערכת מבוססת), או תפיק תוצאה נכונה מבחינה סטטיסטית, אך לא רצויה נורמטיבית.⁴⁴

חרף הערפל הטכנולוגי, נתחום לצורך דיון זה את תהליך קבלת ההחלטות הכללי של המערכת האלגוריתמית ונתאר כיצד הוא עשוי להפלות. הליך זה מתאפיין בשלושה שלבים: איסוף (Data Collection), אימון ובניית המודל (Training & Model Construction), וניבוי (Model Use):⁴⁵

1. בשלב האיסוף עלינו לאסוף מערכי נתונים שישמשו כקלט (הנתונים המגיעים אל המערכת האלגוריתמית) לתהליך הלמידה עצמו. אלו עשויים להיות יצירת מאגר נתונים חדש, או שימוש במאגר נתונים פתוח. איכות הנתונים היא קריטית, שכן כל פגם בה עלול להשתקף ישירות במודל שיפותח לאחר מכן. נתונים פגומים, לענייננו, מתאפיינים בין היתר בכך שאינם מייצגים קבוצות מסוימות באוכלוסייה, חסרים במידע,

42 ראו Dowding & Taylor, *Algorithmic Decision-Making*, לעיל ה"ש 6, בעמ' 67; Bruno Lepri et. al, *Fair, Transparent, and Accountable Algorithmic Decision-Making Process*, 31 PHIL & TECH, 611–627 (2018).

43 Dowding & Taylor, *Algorithmic Decision-Making*, לעיל ה"ש 6, בעמ' 67.

44 ראו הדיון בעניין **אבו חנא**, לעיל ה"ש 34.

45 Laa, *Algorithmic Decision-Making*, לעיל ה"ש 3, בעמ' 525–530.

ומוטים היסטורית או חברתית.⁴⁶ נתונים פגומים כאלה עלולים לגרום לאלגוריתמים לייצר תוצאות לא הוגנות ולהעצים את ההטיות הקיימות בהם.⁴⁷

2. בהמשך, הנתונים הקיימים משמשים כחומר גלם לתהליך ההכשרה והאימון בלמידת מכונה. בשלב זה מיושמות טכניקות עיבוד מידע שונות,⁴⁸ ונבנה מודל שמתאים בצורה הטובה ביותר לנתונים שלו. מודל כזה מתפתח שלב אחר שלב, תוך הפחתת שיעור השגיאה שלו באופן מתמשך. המערכת לומדת את המידע ומפיקה ממנו מידע חדש, או לחלופין יוצרת ייצוג חדש של המידע הקיים.⁴⁹ כאמור, מודלים אלה נבנים למטרות ניבוי; לדוגמה, חיזוי מי זכאי להלוואה, איזו פרמיית ביטוח יש לקבוע, את מי לבדוק על העלמת מס או על פעילויות חשודות בשדה התעופה וכדומה.⁵⁰ חשוב להדגיש כי מטרת האימון היא להגיע לניבוי הדומה למופעים במאגר הנתונים עצמו. לכן, ככל שהוא נשען על החלטות עבר מוטות, יהפוך גם האלגוריתם המנבא למוטה. במילים אחרות, כאשר משתמשים במידע מוטה לאימון מודל, קיימת סבירות גבוהה שההטיות הללו תוטמענה בתוך המודל עצמו.⁵¹

נדגים זאת באמצעות מודל המיועד להערכת מועמדים לאוניברסיטה על בסיס נתוני עבר בדבר הצלחתם של תלמידים ותלמידות בלימודים. מעבר לעובדה שהמאפיינים של המדד הנבחן עשויים להיות בעצמם שנויים במחלוקת (למשל ציונים כמדד להצלחה אקדמית), המידע המשמש לאימון המודל עלול לסבול מייצוג חסר של אוכלוסיות מסוימות בשל החלטות עבר מוטות בדבר קבלה לאוניברסיטה. כך, למשל, בהקשר של מוצא אתני או זהות מגדרית, הרשומות עשויות להיות חלקיות, לכלול שגיאות רבות יותר, או להופיע בתדירות נמוכה יותר במדגם. כך, ייתכן שתלמידה מוכשרת ביותר ממגזר שסובל מתת-ייצוג בלימודים גבוהים לא תתקבל ללימודים אם למשל יינתן

46 הטיות היסטוריות עשויות להיות כנגד קבוצות שסבלו מהפליה בהיסטוריה, ואלו לרוב גם תיחשבה לקבוצות מוגנות (כמו מגדר, גזע, נטייה מינית ועוד). אם האלגוריתם יסתמך על נתונים מהשנים הללו, הוא עשוי לשעתק את ההטיות הללו ולגזור מהן החלטה מפלה היוצרת תמונת מציאות אחרת מזו הרצויה בימינו אנו.

47 Alexandra Honker & Julie Rogers, *What is Algorithmic Bias?*, IBM (Sept. 20, 2024), <https://www.ibm.com/think/topics/algorithmic-bias> (להלן: Honker & Rogers, *What is Algorithmic Bias?*).

48 עצי סיווג והחלטה, מכונות וקטור תמיכה (SVMs), שיטות הרכבה (ensemble methods), רשתות נוירונים וכדומה. ראו Paul B. Laet, *Algorithmic Decision-Making*, לעיל ה"ש 3, בעמ' 531.

49 לדוגמה, מערכת אלגוריתמית להערכת הלוואות מתפתחת מניתוח בסיסי של הכנסה ודירוג אשראי. היא מתפתחת לפי טכניקות העיבוד שלה ולומדת לזהות קשרים מורכבים כמו דפוסי החזר הלוואות ויחסי הוצאות-הכנסות, עד ליצירת תובנות חדשות כמו אינדיקטורים ייחודיים לאמינות פיננסית ופרופילי סיכון מדויקים, המאפשרים קבלת החלטות מותאמות אישית.

50 Paul B. de Laet, *Algorithmic Decision Making*, לעיל ה"ש 3, בעמ' 531; ראו ע"פ 8660-22 מדינת ישראל נ' עומר ברגר (נבו 9.2023), בו נדון בהמשך המאמר.

51 שם, בעמ' 532.

משקל-יתר לנתונים השליליים המייצגים את ההיסטוריה של אותו מגזר בקבלה ללימודים גבוהים, מה שיוריד את ממוצע סיכוייה להתקבל, שכן נתוניה ייבחנו על ידי המכונה באופן סטטיסטי ולא פרטני.⁵²

במילים אחרות, מודל הבינה המלאכותית יכול להיות מוטה שכן הוא מאגד בתוכו את כלל ההטיות האנושיות – טעמים, סטראוטיפים והטיה סטטיסטית. הוא תוצר של מפתחיו ושל הידע האנושי (כמו למשל בלוגים, טוקבקים, מאמרים שעשויים להיות "נגועים" בדעות קדומות וכן כל מידע שנאגר בשלב האיסוף), ואלו, כדרכו של עולם, נגועים בדעות קדומות ובסטראוטיפים. המודל יכול להיות מוטה אם הנתונים סובלים מאחד מליקויים אלו: הם משקפים הטיות אנושיות או תרבותיות קיימות בהחלטות עבר (הפליית טעמים או הפליה סטראוטיפית), הם חסרים לגבי קבוצות מסוימות או שהם כוללים חלוקה לא מאוזנת לקבוצות ומעמדות.⁵³ למעשה, האלגוריתם הוא כמו מראה,

52 בשנת 2020 תיאר הארגון Student Borrower Protection Center פערים ניכרים בריביות ובעמלות של הלוואות סטודנטים, שהיו בעלי נתונים זהים, על רקע מוסד ההשכלה הגבוהה שבו הם למדו, שמצוי במתאם עם קבוצה גזעית. ראו דוח הצוות הבין-משרדי **בינה מלאכותית בסקטור הפיננסי**, לעיל ה"ש 41, בעמ' 95; STUDENT BORROWER PROT. CTR., EDUCATIONAL REDLINING (Feb. 5, 2020), <https://protectborrowers.org/wp-content/uploads/2020/02/Education-Redlining-Report.pdf>

53 כך, לדוגמה, חברה המשתמשת באלגוריתם קבלת החלטות לסינון קורות חיים ודירוג מועמדים לראיונות עשויה להפלות מועמדים אם האלגוריתם מאומן על נתוני גיוס היסטוריים המשקפים את נוהלי הגיוס של החברה בעבר. כך, למשל, האלגוריתם עשוי לתעדף מועמדים מאזורים מבוססים כלכלית או בוגרי אוניברסיטאות יוקרתיות על פני מועמדים בעלי כישורים שווים הבאים מרקע פחות מבוסס, וזאת בשל הערכת יתר של האלגוריתם לסמנים הקשורים למעמד סוציו-אקונומי גבוה. מקרה שקרה בבריטניה בשנת 2019 עסק בסוגיה דומה. לאחר שהבחירות בוטלו עקב מגפת הקורונה, נעשה שימוש באלגוריתם לחיזוי ציוני תלמידים לצורך קבלה לאוניברסיטאות. תלמידים ממוסדות חינוך בעלי חוסן כלכלי זכו להערכות גבוהות יותר מעמיתיהם מבתי ספר במעמד סוציו-אקונומי נמוך יותר, וזאת גם בניגוד גמור לנתוני העבר שהצביעו על מגמה הפוכה בחלק מהמקומות. התברר כי המערכת האלגוריתמית העניקה עדיפות מופרזת לדירוג המוסד החינוכי, במקום להתמקד בהישגיהם הפרטניים של התלמידים. כתוצאה מכך, מועמדים רבים משכבות מוחלשות נדחו מהלימודים, מה שהוביל להתנגדות ציבורית ניכרת ולרפורמות בקריטריוני הקבלה. ראו Karla Adam, *The U.K. Used an Algorithm to Estimate Exam Results The Calculations Favored Elites*, THE WASH. POST (Aug. 17, 2020), https://www.washingtonpost.com/world/europe/the-uk-used-an-algorithm-to-estimate-exam-results-the-calculations-favored-elites/2020/08/17/2b116d48-e091-11ea-82d8-5e55d47e90ca_story.html.

מקרה מפורסם אחר קרה בארצות הברית, שם השתמשו במערכת COMPAS, כלי להערכת סיכונים המשמש בארצות הברית לחיזוי הסיכוי לחזרה לעבירות פליליות. מחקרים גילו כי נאשמים שחורים סווגו באופן לא פרופורציונלי כבעלי סיכון גבוה יותר מנאשמים לבנים, גם כאשר היו להם רישומים פליליים דומים. האלגוריתם הסתמך על משתנים עקיפים (proxy variables) כמו רקע סוציו-אקונומי, שהשפיעו באופן לא

והוא בסך הכול משקף את הנתונים שהוא מאומן עליהם. לכן, ניתן בעצם לומר שמערכת הבינה המלאכותית "טובה כמו הדאטה שלה".⁵⁴ עם השלמת תהליך פיתוח האלגוריתם המודל מוכן לשלב הניבוי, שבו מתקבלת ההחלטה האלגוריתמית. בשלב זה המודל מפיק תחזית על בסיס המידע החדש שהוזן, ובכך מספק תובנות חדשות – וגם תוצאה חדשה.⁵⁵ כך, הטיה באימון עשויה ליצור הטיה בעיבוד. בשלב זה גם עשויה להתעורר הטיה בפרשנות, כאשר השוני בין ייעוד האלגוריתם לבין שימוש הקצה יוצר עיוות של הפלט (התוצאה) וגרימה לפעולות מוטות. מכיוון שאלגוריתם נבנה למטרה ספציפית ומוכחת, שימוש בו למטרה אחרת, ללא ביצוע ההתאמות הנדרשות, עלול להוביל לתוצאות לא מדויקות שעשויות להיות גם מפלות.⁵⁶ כך, למשל, שימוש באלגוריתם שיוצר בסין על אוכלוסייה בארצות הברית עלול להיות מוטעה, בשל הבדלים תרבותיים וחברתיים מסיביים בין שתי התרבויות. לצורך הדוגמה ובהכללה, דמיינו מקרה שבו עיריית תל אביב יוצרת מערכת אלגוריתמית על בסיס הנתונים מבתי הספר בעיר כדי להחליט מי מהילדים יהיה זכאי למלגה עקב הישגיו הלימודיים המצוינים. גם אם המערכת מנבאת בצורה מצוינת מי זכאי למלגה בתל אביב, אם לפתע מערכת החינוך תשתמש במערכת אלגוריתמית זו באופן רוחבי, ההחלטות שתקבל עשויות להפלות ילדים בערים אחרות ומאשכולות סוציו-אקונומיים נמוכים יותר.

המערכת האלגוריתמית המשתמשת בתוצאות מוטות כבסיס לקבלת החלטות יוצרת לולאת משוב המחזקת את ההטיה לאורך זמן. תהליך זה, שבו האלגוריתם ממשיך ללמוד ולשחזר דפוסים מוטעים, מוביל לתוצאות שעשויות להפוך ליותר ויותר בלתי מאוזנות.⁵⁷ חששות בנוגע לסוגיית ההטיה האלגוריתמית הובילו למאמצים לפיתוח מודלים אלגוריתמיים "נטולי הפליה" או "מודעים להפליה". לכאורה, נראה כי הסרה של כל משתנה רגיש ממערך הנתונים תוכל להיות פתרון מתאים. עם זאת, מכיוון שהמערכת מלמדת את עצמה על בסיס הנתונים שאליהם היא חשופה (כאמור בתחילת הפרק, שלב האיסוף) ועל פיהם יוצרת מידע חדש, ייתכן שחלק מהמשתנים שנותרו

פרופורציונלי על קבוצות גזעיות מסוימות. ראו Ellora Thadaneey Israni, *When an Algorithm Helps Send You to Prison*, N.Y. TIMES (Oct. 26, 2017), <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html>.

Katherine Miller, *Data-Centric AI: AI Models Are Only as Good as Their Data Pipeline*, STANFORD UNIVERSITY HUM.-CTR. A.I. (Jan. 25, 2022), <https://hai.stanford.edu/news/data-centric-ai-ai-models-are-only-good-their-data-pipeline>; Xavier Ferrer et al., *Bias and Discrimination in AI: a Cross-Disciplinary Perspective*, 40(2) IEEE TECH & SOC. MAG. 72 (2021).

Paul B. de Laat, *Algorithmic Decision Making*, 356 בעמ' 356.

56 כך למשל בעניין ההפליה בביטוח הבריאות שבה דנו קודם; & Tamar Kricheli-Katz & Ronen Avraham, *Battling Discrimination in Artificial Intelligence* 1, 3 (2023) (עותק).

שמור בידי המחברים) (להלן: *Battling Discrimination*); & Kricheli-Katz & Avraham.

57 *Honker & Rogers, What is Algorithmic Bias?*, לעיל ה"ש 47.

במודל עדיין יהיו מתואמים עם המשתנים הרגשיים וכך ישעתקו את ההפליה. ככל שקיימת הטיה כלפי קבוצה באוכלוסייה, למשל בקבלה לעבודה, מערכת הבינה המלאכותית המאומנת על נתונים מוטים אלה עלולה לשעתק הטיית חברתיות קיימות, ובהמשך לשחזר ולהנציח את תוצאותיה המפלות שלה עצמה.⁵⁸ לכאורה, ניתן למחוק גם את המשתנים הנגועים החדשים. עם זאת, לעיתים קרובות לא נוכל לדעת מהו הפלח מהמידע שנגוע בהטיה. בנוסף, משמעות מחיקת מידע נוסף עשויה להיות מחיקת מידע בעל ערך עבור ניבוי התוצאה.⁵⁹

וכך, המערכות האלגוריתמיות מציבות אתגר במענה על השאלה "כיצד, ומדוע, התקבלה ההחלטה המפלה", מה שבעגה המקצועית ידוע גם כמונח "הסברתיות".⁶⁰ למודלים רבים של בינה מלאכותית יש מנגנוני עיבוד הנחשבים ל"קופסאות שחורות". המשמעות היא שלא תמיד ניתן לעקוב אחר מחזור החיים של האלגוריתם, מה שמקשה על הבנה ותחקור של תהליך קבלת ההחלטות שלו: בעוד שניתן לראות את הקלט והפלט, התהליכים שבאמצעותם הראשון הופך לשני נותרים בגדר תעלומה.⁶¹ לכן, למשל, אף על פי שגופים ציבוריים חבים בחובת שקיפות,⁶² ולפרט יש זכות לגשת למידע על אודותיו תחת חוק הגנת הפרטיות,⁶³ היעדר השקיפות מקורו לעיתים בבעיה שהיא טכנולוגית במהותה, ונוצרת בשל מורכבותן או גודלן של המערכות האלגוריתמיות.⁶⁴ באופן טבעי, במקרים רבים לא קיימת שקיפות בנוגע לדרך שבה קיבלה המערכת האלגוריתמית את החלטתה, וכך קשה לאתר את הנתון המסוים שבגיניו התרחשה ההפליה – באם זה נתוני הקלט, אימון המידע, או עיבוד המידע על ידי המודל שגרמו לתוצאה המפלה.⁶⁵

חוסר השקיפות עשוי דווקא לגרום למשתמשים ולקובעי המדיניות לסמוך על תוצאות האלגוריתמים כמובנות מאליהן, וככאלו שאינן דורשות בדיקה נוספת.⁶⁶

58 דוח הצוות הבין-משרדי **בינה מלאכותית בסקטור הפיננסי**, לעיל ה"ש 41, בעמ' 94; וכן Michael A. Griffith, *AI Lending and ECOA: Avoiding Accidental Discrimination*, 27 N.C. BANKING INST. 349 (2023).

59 Talia B. ; 352 בעמ' 3, לעיל ה"ש 3, Paul B. de Laat, *Algorithmic Decision Making*, 106 MINN. L. REV. 1175 (2022). Gillis, *The Input Fallacy*, 106 MINN. L. REV. 1175 (2022).

60 דוח הצוות הבין-משרדי **בינה מלאכותית בסקטור הפיננסי**, לעיל ה"ש 41, בעמ' 55.

61 FRANK PASQUALE, *THE BLACK BOX SOCIETY* 3, 4 (2016). לדוגמה, אדם עם דירוג אשראי נמוך עשוי להידחות לקבלת הלוואה מבלי להבין אי פעם כיצד הדירוג שלו חושב. גם המחירים שמוצעים ללקוחות שונים על ידי קמעונאים מקוונים נקבעים לעיתים קרובות על ידי אלגוריתמים סודיים.

62 גדי פרל ותהילה שוורץ אלטשולר **מודל ליצירת שקיפות אלגוריתמית** (2022).

63 ס' 13(א) לחוק הגנת הפרטיות, תשמ"א-1981.

64 דוח הצוות הבין-משרדי **בינה מלאכותית בסקטור הפיננסי**, לעיל ה"ש 41, בעמ' 55.

65 Kricheli-Katz & Avraham, *Battling Discrimination*, לעיל ה"ש 56, בעמ' 2.

66 מחקר גילה כי שופטים נוטים להסתמך על תוצאות האלגוריתם כדי להצדיק החלטות שכבר רצו לקבל, בייחוד במקרים בעלי פרופיל גבוה, ובכך מחזקים את ההטיות האישיות שלהם תחת מעטה של אובייקטיביות אלגוריתמית. ראו Lauren Feiner, *Judges Are*

כתוצאה מכך, החלטות אלגוריתמיות עשויות לעבור ביקורת פחותה מזו של החלטות אנושיות, גם אם התוצאות של ההחלטות האלגוריתמיות מראות בבירור בעיות של הטיה והפליה.⁶⁷ קל לדמיין נסיבות שבהן פיקוח לקוי בעיצוב של מערכת בינה מלאכותית עלול להשפיע על המשתמש שבקצה. כך, למשל, הטיה אלגוריתמית יכולה להוביל לשימור אי-השוויון החברתי דרך מתן תוצאה מפלה כמו אי-קבלה לעבודה, דחיית בקשות הלוואה, או תשלום נוסף עבור פוליסת ביטוח, כל זאת ללא סיבה מספקת או שקיפות בנוגע לדרך קבלת ההחלטה. הטיות כאלו עשויות גם להביא לשלילת חירותו של אדם.⁶⁸ בישראל קרה דבר דומה בשנת 2022, בעניין עומר ברגר נ' מדינת ישראל, ועליו נרחיב כעת.⁶⁹

עומר ברגר, אזרח ישראל ללא עבר פלילי, חזר ארצה בטיסה מרומניה. כאשר נחת, עבר חיפוש משטרי ומצא את עצמו נעצר בגין אישום של ייבוא סם מסוכן. התברר כי מעצרו היה בשל הופעת שמו ב"רשימת הכללה" – מערכת אלגוריתמית שחווה בצורה סטטיסטית את מעורבותם של נוסעים הנכנסים לישראל בעבירות סמים. ברגר ביקש להסיר את החיסיון על המידע שהוביל למעצרו, כדי לגלות את השיטות והאמצעים ששימשו את משטרת ישראל בהפעלת מערכת ההתראות בנתב"ג במהלך החקירה בעניינו. בדיון נמצא כי האלגוריתם המשטרי נבנה בדרך של היזון חוזר, כך שהמערכת מלמדת את עצמה את משקלם של נתונים ואת המשמעות הסטטיסטית שלהם. כמו כן, המשטרה הבהירה בבית המשפט כי היא אינה יכולה לדעת אילו נתונים נלקחו בחשבון של האלגוריתם, ומהו המשקל שניתן להם.⁷⁰

למערכת היה "שלב שני": המשטרה בחרה בצורה אנושית את המועמד החשוד מתוך רשימת הכללה שיצר האלגוריתם. למרות הגורם המפקח, המערכת לא חפה מסיכון להפליות. לכך התייחס השופט דרואן-גמליאל, באומרו שקיומו של שלב זה

Using Algorithms to Justify Doing What They Already Want, THE VERGE (Oct. 30, 2024), https://www.theverge.com/2024/10/30/24281924/pretrial-risk-assessment-algorithms-research?utm_source=chatgpt.com.

Minesh Tanna & William Dunning, *Bias and Discrimination*, in ARTIFICIAL INTELLIGENCE 422, 428 (Charles Kerrigan ed, 2022).

בארצות הברית נעצר אדם בגלל זיהוי שגוי של מערכת בינה מלאכותית שהייתה בשימוש משטרת מדינת מישיגן, כתוצאה מהטיה במערכת זיהוי-פנים אלגוריתמית. כאשר גילו את הטעות במשטרה, אמר הבלש לשותף שלו: "אני מניח שהמחשב טעה". להרחבה, ראו Kashmir Hill, *Wrongfully Accused by an Algorithm*, NYT (Aug. 3, 2020), <https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>.

צ"א (מחוזי מרכז) 24474-01-22 עומר ברגר נ' מדינת ישראל (נבו 11.9.2022).

שם, בעמ' 8 לפסק הדין. מחקרים מראים כי כאשר המשטרה עושה שימוש בנתוני פשיעה היסטוריים המוטעים לשם קבלת החלטות מבצעיות, העלויות מוטלות בעיקר על המושפעים מההחלטות ולא על הגורם המשתמש באלגוריתם. נזק מסוג זה מכונה "השפעות חיצוניות אלגוריתמיות": נזקים הנגרמים כתוצאה משימוש באלגוריתמים, אשר מוטלים על פרטים שאינם מעורבים בהחלטה על עצם השימוש באלגוריתם או על אופן הפעלתו. ראו Dowding & Taylor, *Algorithmic Decision-Making*, לעיל ה"ש 6.

אינו מקטין את הפגיעה בערכי יסוד, שכן ייתכן שעל בחירת האלגוריתם בחשודים פוטנציאליים תתווספה הטיות אנושיות.⁷¹ לצורך ההשוואה, חוק הבינה המלאכותית של האיחוד האירופי אוסר על שימושים מסוימים בבינה המלאכותית, שכוללים בין היתר מערכות בינה מלאכותית אשר חוזות את הסיכון של אדם לבצע פשע (במנותק מהערכה אנושית קודמת ביחס למעורבותו בפעילות פלילית), וכן מערכות בינה מלאכותית שמטרתן יצירה של מאגרי מידע לזיהוי פנים באמצעות קצירת מידע (scraping) של תמונות פנים מהאינטרנט או ממצלמות אבטחה.⁷² למרות הסיכונים, פעולות אלו טרם נעשו בזירת החקיקה הישראלית.

בפרק הבא נעבור להשוואת ההפליה האנושית להפליה האלגוריתמית. נראה כי ההשוואה מעלה תמונה מורכבת: לעיתים האלגוריתם יוביל לתוצאות מפלות יותר מבני האדם (למשל במקרים של הפליה סטטיסטית או מופעים של נתונים מוטעים), ולעיתים לתוצאות מפלות פחות (הפליית טעמים או הפליה סטראוטיפית). כפי שנראה מיד, מאחר שמערכת הבינה המלאכותית עובדת על מאגרי נתונים ענקיים (Big Data), ברצוננו לטעון כי בפועל כל קבלת החלטות אלגוריתמית מבוססת במהותה על הבחנות סטטיסטיות, וכך כל הבחנה אלגוריתמית היא במהותה הפליה סטטיסטית (שאינה מופנית בהכרח כנגד קבוצה המוגנת על ידי דיני איסור ההפליה). מכאן, טענתנו היא שגם אם האלגוריתם יהיה מדויק, הוא עדיין עלול להוביל להפליה סטטיסטית כנגד קבוצות מוגנות האסורה לפי דיני איסור ההפליה. הפליה כזו למעשה תנבע מכך שהאלגוריתם עשוי לשעתק אי-שוויון ופערים המתקיימים בעולם. הפליה סטטיסטית

71 שם, בעמ' 10. במחוזי התקבלה עתירתו של ברגר, ונפסק כי יצומצם החיסיון באופן משמעותי. עם זאת, המדינה ערערה לעליין וזה הפך את החלטת המחוזי. פסק הדין לא עסק בחוקיותה של שיטה ומערכת זו, אלא בחוקיות הראיה.

72 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 2024/1689), art. 5; וכן עמרי רחום-טוויג "המשפט לומד את המכונה: ה-AI Act האירופי כחקיקה טכנולוגית בזמן אמת" פורום עיוני משפט (תגובות משפט) מח (13.3.2024), <https://www.taulawreview.sites.tau.ac.il/post/rahum-twaig2>. החוק נכנס לתוקף באוגוסט 2024, אך בשלב זה לא חלו דרישותיו והגבלותיו של החוק, ונקבע כי תהליך זה יקרה בהדרגה. תחילת הטמעת איסורים והגבלות אלו קרתה בפברואר 2025 (פרק 1 ו-2 לחוק). למעקב אחר תהליך הטמעת החוק ראו Eu Artificial Intelligence Act, *Implementation Time, FUTURE OF LIFE INST*, <https://artificialintelligenceact.eu/implementation-timeline>.

כזו עשויה לשקף הבדלים בין קבוצות – אמיתיים, נכונים ומדויקים סטטיסטית – אך לא רצויים נורמטיבית.

ג. הפליה אנושית מול הפליה אלגוריתמית

נתחיל בהפליית טעמים. כאמור, הפליה זו מתרחשת במערכות יחסים אנושיות והיא קשורה בטעמים המפלים שלנו. ככזו, היא לא רלוונטית במישרין לאלגוריתמים ולבינה המלאכותית משום שלא לגוריתם אין שנאה או העדפה. עם זאת, כאמור, היא עשויה להיות רלוונטית בעקיפין אם למתכנתת שתכנתה או אימנה את האלגוריתם יש טעמים מפלים, והם משפיעים במודע או לא במודע על האופן שבו היא מתכנתת את האלגוריתם.⁷³ הפליית טעמים תשפיע באופן עקיף על האלגוריתמים גם במקרה שבו נתוני העבר שהאלגוריתם מתאמן עליהם מבוססים על החלטות שהתקבלו על רקע הטעמים המפלים של בני האדם שקיבלו את ההחלטות האלו בעבר. כמו שנבחר בהמשך, המלצתנו היא לעסוק בהחלטות אנושיות (וביניהן גם אלה שקיבלה המתכנתת) באמצעות דיני איסור הפליה (ככל שהם רלוונטיים) אך באלגוריתם באמצעות בדיקה תוצאתית.

כך או כך, אפליית טעמים מתרחשת במערכות יחסים בין אנשים (שהם בעלי הטעמים המפלים), ולכן היא לא רלוונטית במישרין לאלגוריתמים. למעשה, אלו עשויים להיות מצוינים בהשוואה לבני האדם דווקא ביכולת שלהם שלא לפעול על פיה. לעניין אפליית טעמים, בינה מלאכותית עדיפה על בינה אנושית, משום שהראשונה אינה מושפעת מטעמים מפלים. לכן, כפי שנטען בהמשך, נמליץ לטפל בה במסגרת דיני

⁷³ המפתחים גם הם מהווים שחקנים פוליטיים בזירה של הפליה אלגוריתמית. דוגמאות להטיה והפליה בעולם העבודה קיימות בכל אחת מהחברות הטכנולוגיות שמובילות את פיתוח טכנולוגיות הבינה המלאכותית. כך, למשל, עובדי חברת "מיקרוסופט" תבעו את המעסיקה בתביעה ייצוגית בטענה שהחברה כשלה באופן שיטתי בהתייחסות רצינית למאות טענות על הטרדה והפליה. ראו Dan Levine, *Microsoft Women Filed 238 Discrimination and Harassment Complaints*, REUTERS (Mar. 13, 2018), <https://www.reuters.com/article/idUSKCN1GP079>. בעניין אחר, חברת "אפל" כינתה הצעות ליצירת נוהלי גיוס מגוונים כ"מעמסה גדולה מדי", ראו Alexandra Sims, *Apple's Board Rejects Diversity Proposal as 'Unduly Burdensome and not Necessary'*, INDEPENDENT (Jan. 17, 2016), <https://www.nevo.co.il/shorturl/812700>. דוגמה נוספת היא ביקורת שזר שבוטלה בחברת "גוגל" על ידי מחלקת העבודה של ארצות הברית. זו מצאה פער של שש עד שבע סטיות תקן בין השכר לגברים ולנשים כמעט בכל קטגוריות התפקידים. בנוסף, תביעה שהוגשה נגד חברת "טסלה" מאשימה את החברה בהפליה מגדרית, נקמה, וסביבת עבודה עוינת. אחת העובדות טענה שבצוות שלה היו יותר גברים בשם "מאט" מאשר נשים. להרחבה, ראו: Sheelah Kolhatkar, *The Tech Industry's Gender-Discrimination Problem*, THE NEW YORKER (Nov. 13, 2017), <https://www.newyorker.com/magazine/2017/11/20/the-tech-industrys-gender-discrimination-problem>.

איסור ההפליה המסורתיים שאוסרים עליה אם היא מבוססת על הבחנות מסוימות (לאום, מגדר, נטייה מינית) ופועלת בזירות חיים מסוימות (למשל עבודה או שוקי מוצרים ושירותים). נציין לבסוף כי הפליית טעמים עשויה להשפיע **בעקיפין** על קבלת ההחלטות האלגוריתמית אם נתוני העבר שעליהם יתאמן האלגוריתם יהיו מבוססים על החלטות שהתקבלו על רקע הטעמים המפלים של מי שקיבלו את ההחלטות הללו בעבר.⁷⁴

בדומה להפליית הטעמים, גם הפליה שמבוססת על סטראוטיפים ואמונות תרבותיות מתרחשת במערכות יחסים אנושיות. לכן, אם הנתונים המקוריים שעליהם התאמן האלגוריתם מבוססים על החלטות עבר שקיבלו אנשים על בסיס סטראוטיפים, גם ההחלטות שיקבל האלגוריתם תהיינה מוטות, שכן הוא ילמד את הסטראוטיפים של מפתחיו ושל המידע שהוא שואב בהליך למידת המכונה.⁷⁵ החשש מתעצם כאשר מבינים את סדר הגודל של תוצאות קבלת ההחלטות על ידי האלגוריתם, שכן לא מדובר בהחלטה אנושית פרטנית, אלא במערכת ממוחשבת שמתעסקת עם כמות דאטה עצומה וכמות החלטות בסדר גודל אדיר.

אולם, בעוד שלבני האדם קשה יחסית להימנע מהפליה מסוג זה והיא במקרים רבים תהיה לא מודעת, יש דרכים שבהן ניתן לכייל את האלגוריתם כך שיסייע להימנעות מהפליה סטראוטיפית. הדרך העיקרית היא שימוש במאגרי נתונים מדויקים שאינם משקפים הטיות אנושיות. לעיתים נדרש להסרה מכוונת של משתנים נגועים בהפליה. למעשה, בינה מלאכותית ואלגוריתמים, כשייעזרו במאגרי נתונים מדויקים, יצליחו לתקן אי-דיוקים בקבלת ההחלטות האנושית ככל שהם מבוססים על סטראוטיפים או אמונות מוטעות.⁷⁶ דיוק הנתונים יכול להיעשות גם דרך "עיסוי הדאטה", קרי, עיבוד מקדים של נתונים.⁷⁷

- 74 Ronald Honker & Rogers, *What is Algorithmic Bias?*, לעיל ה"ש 47. להרחבה ראו: Ronald Rathelot & Mirna Safi, *Taste-Based Discrimination*, in HANDBOOK ON ECONOMICS OF DISCRIMINATION AND AFFIRMATIVE ACTION (Ashwini Deshpande ed., 2023).
- 75 Aylin Caliskan et al., *Semantics Derived Automatically from Language Corpora Contain Human-Like Biases*, 356(6334) SCIENCE 183 (2017); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674, 683 (2016).
- 76 Stephanie Bornstein, *ANTIDISCRIMINATORY ALGORITHMS*, 70 ALA. L. REV. 519, 534 (2018).
- 77 עיבוד זה קורה באמצעות מניפולציה של נתונים, כמו עיבוד מקומי של הנתונים כך שמקרים גבוליים יתויגו מחדש, שימוש ב"דגימה מועדפת" מקומית, הכוללת מחיקה או שכפול של מקרים באימון, וכך לוותר על נתונים שהם מוטים בעצמם או נתונים מיותרים שעלולים להיות בעלי השפעות שליליות על מערך הנתונים מנקודת מבט של הפחתת ההפליה. הליך זה צריך להיעשות בזהירות, שכן גם המפתחים עשויים להשליך, הגם שלא בזדון, את הטיותיהם האנושיות על האלגוריתם, וכך להנציח אותן דרך המכונה (פעולה שנקראת Designer Bias). ראו גם Zhisheng Chen, *Ethics and Discrimination in*

הליך בקרה נוסף יכול לכלול בדיקה הכוללת שינוי אקראי של התכונות הרגישות במערך הנתונים המשמש לאימון (למשל, שינוי מגדר מנקבה לזכר ולהפך). במסגרת הליך בקרה זה נעשה שימוש במדדי "השפעת קלט כמותית", המאפשרים להעריך האם מאפייני קבוצות (כמו מגדר או גזע) משפיעים באופן בלתי ראוי על התוצאות.⁷⁸ חשוב להדגיש ששילבים אלה כרוכים בהפעלת שיקול דעת מטעם המפתחים, שיקול שעשוי להתגלות בדיעבד כמוטה. חוקרים ממליצים שהצוות שמפתח ומנהל את האלגוריתם יהיה בעצמו מגוון מבחינה אתנית ומגדרית, שכן צוותים מגוונים קשובים יותר להטיות ולאפשרות של הפליה אלגוריתמית.⁷⁹

בהקשר זה, מעבר לסוגיות האתיות הברורות, ההפליה האלגוריתמית (ככל שאינה מדויקת ומבוססת על מאגרי נתונים מוטעים כתוצאה מהחלטות עבר) עלולה להביא לסיכונים ולתוצאות בעייתיות עבור החברות המשתמשות בבינה המלאכותית וכן עבור החברות המפתחות עצמן. מערכת בינה מלאכותית אשר מנתחת נתונים מוטעים עלולה להוביל לתוצאות שליליות כמו אובדן לקוחות פוטנציאליים, אספקת מוצרים לא מתאימים ללקוחות והחלטות גיוס פגומות.

כאמור, בעינינו ההפליה הסטטיסטית היא האתגר הגדול ביותר לפרויקט הבינה המלאכותית. כזכור, במקרים אלה ההפליה אומנם תהיה מדויקת מבחינה סטטיסטית, אך לא רצויה נורמטיבית. כמובן גם בני האדם עצמם נוטים להפלות הפליה סטטיסטית ביחסים שבין אנשים, בין אם הפרטיים ובין אם העסקיים, אך בשונה מהם, אלגוריתמים נוטים לעשות זאת באופן מדויק יותר, רחב יותר, ומבלי שניתן להסיר באופן אפקטיבי משתנים מוגנים כמו גזע או מגדר ממאגר הנתונים.

מערכות אלגוריתמיות הן מתוחכמות, ולכן עם נתונים מדויקים הן תגענה (או לפחות תתקרבנה) אל התוצאה המדויקת סטטיסטית. ככל שמאגרי הנתונים הם גדולים ומדויקים יותר – כך יהיה הניבוי הסטטיסטי מדויק יותר. כשפער בין קבוצות מתקיים בעולם, האלגוריתם "יתפוס אותו" טוב יותר מבני האדם. גם אם נקפיד לבצע את כל אמצעי הזהירות: פיקוח על הנתונים, הסרת התכונות "החשודות" (והמוגנות על ידי דיני איסור הפליה) של האנשים שלגביהם הוא מתבקש לקבל את ההחלטה, והסרה ממאגר המשתנים של תכונות "מנבאות" (proxy) כמו גזע, גיל, מגדר ולאום – האלגוריתם ידע להסיק אותן "בממוצע" מתוך שאר הנתונים הקיימים.⁸⁰ הוא יוכל להגיע לאותה תוצאה

Artificial Intelligence-Enabled Recruitment Practices, 10 HUMAN & SOC. SCI. COMMUN., 5-6 (2023).

78 Paul B. Laat, *Algorithmic Decision-Making*, לעיל ה"ש 3, בעמ' 533.

79 Marco Marabelli et al., *The Lifecycle of Algorithmic Decision-Making Systems: Organizational Choices and Ethical Challenges*, 30(3) J. STRATEGIC INF. SYSTEMS

Barocas & Selbst, *Big Data's Disparate Impact*; 11 (2021), לעיל ה"ש 75.

80 להרחבה על תכונות מנבאות ראו הכתוב ליד הערת שוליים 68, וכן את הדיון בפרק ג.

לא-שוויונית גם דרך הבחנות שגרתיות פחות, כמו תחביבים, סוג כתיבה, מוזיקה מועדפת ועוד, המסתמכות על קריטריונים סטטיסטיים מנבאים.⁸¹ למעשה, אלגוריתמים יוכלו לקבל החלטה מפלה סטטיסטית על רקע גזע או מגדר גם מבלי לדעת כלל את הגזע או המגדר של האנשים הרלוונטיים. זאת משום שגזע ומגדר נוטים להיות במתאם עם תכונות אחרות, כמו התחביבים, מקום המגורים, המצב הבריאותי, המשקל, המקצוע, שעות העבודה והמנוחה, ואפילו המילים והשפה שבהן נכתבו קורות החיים.⁸² בשונה מהפליה אנושית, הפליה אלגוריתמית עשויה להיות לא-אינטואיטיבית. הדבר מציב אתגר משמעותי למערכות משפטיות המבוססות על שיקולים המוכרים לנו דרך תסריטים חברתיים. סוג מסוים של תכונה מנבאת עשוי לא לעזור "דגלים אדומים", כפי שנתונים אחרים, בולטים יותר, היו מעוררים. לדוגמה, ידוע שהכנסה יכולה לשמש כתכונה מנבאת למגדר, אך כיצד נתונים כמו דפוסי גלישה, לייקים ברשתות החברתיות או כניסה לאתרים מתקשרים למאפיינים מוגנים? בזירה האלגוריתמית, איננו יכולות להסתמך על תחושות בטן אינטואיטיביות כדי להבין מתי וכיצד מתרחשת הפליה.⁸³ לכן, אנו סבורות שהפרויקט האלגוריתמי מציב אתגר של ממש לדיני איסור הפליה כשמדובר בהבחנות סטטיסטיות מדויקות על בסיס תכונות חשודות (כמו למשל לאום או מגדר). למעשה, לגבי הפליה אלגוריתמית ניצבים דיני איסור ההפליה בפני שוקת שבורה. ככל שהמערכת האלגוריתמית תהפוך לחכמה ומתוחכמת יותר, וככל שנתוני העתק (Big Data) יגדלו, הפליה עקיפה דרך תכונות מנבאות תהווה אתגר הולך וגובר עבור דיני איסור ההפליה, המבקשים לאסור הפליה על בסיס מאפיינים שיש להם קשר ישיר לניבוי התוצאה האלגוריתמית.⁸⁴ נדון בכך כעת.

ד. דיון: אתגרים משפטיים – שאלות ומחשבות להמשך

במשפט הישראלי, אסורה במפורש הפליה בשתי זירות חיים, שהן גם רלוונטיות מאוד לקבלת החלטות אלגוריתמית: שוק העבודה (חוק שוויון הזדמנויות בעבודה)⁸⁵ ושוק

81 Lori Andrews & Hannah Baker, *Automating Discrimination: AI Hiring Practices* 81 Anya E.R. Prince & וכן, *and Gender Inequality*, 44 CARDOZO L. REV. 145 (2022) Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1263 (2020) (להלן: Prince & Schwarcz, *Proxy Discrimination*).

82 שם.

83 Sandra Wachter et al., *Why Fairness Cannot be Automated: Bridging The Gap Between EU Non-Discrimination Law And AI*, 41 COMPUT. L. & SEC. REV. 1, 67 (2021) (להלן: Wachter et al., *Why Fairness Cannot be Automated*).

84 Prince & Schwarcz, *Proxy Discrimination*, לעיל ה"ש 81, בעמ' 1257, 1264.

85 חוק שוויון הזדמנויות בעבודה, לעיל ה"ש 2.

המוצרים והשירותים (חוק איסור הפליה במוצרים, בשירותים ובכניסה למקומות בידור ולמקומות ציבוריים).⁸⁶

דיני איסור הפליה מונים רשימה של קטגוריות – נגזרות בהפליה היסטורית – שעל בסיסן אסור להפלות באינטראקציות השוקיות שלנו, אלא בהתקיים החריגים המנויים בחוק (ההפליה נגזרת מאופיו של השירות או העיסוק).⁸⁷ כך, למשל, קובעים שני החוקים איסור הפליה על בסיס גזע, מין, נטייה מינית, לאום ועוד.⁸⁸ חשוב לציין, לענייננו, כי דיני איסור הפליה נוטים שלא להבחין אלא במקרים חריגים (לצורך האיסור) בין המנגנונים השונים שמובילים להפליה, אלא אוסרים בדרך כלל באופן גורף סוגים מסוימים של הפליה, בהתאם לקטגוריה שעומדת בבסיסם. יתרה מזאת, בעניין **רקנט** נקבע כי ההפליה הסטטיסטית מחמת גיל מהווה הפליה אסורה לצורך דיני איסור הפליה.⁸⁹

התמקדותו של המשפט הישראלי בקטגוריות המפללות, ומתוך כך האיסור הספציפי על הפליה סטטיסטית, נתקלת באתגר של ממש כשמדובר בהפליה **אלגוריתמית** סטטיסטית. כאמור, הפרויקט האלגוריתמי הוא במהותו פרויקט סטטיסטי. למעשה, עצם השימוש באלגוריתמים יוביל להפליה סטטיסטית כל עוד קיימים באוכלוסייה הבדלים סטטיסטיים בין קבוצות של אנשים אשר חלקם משתייכים לקבוצות שעליהן המשפט מבקש להגן מפני הפליה.⁹⁰ במילים אחרות, האיסור על הפליה גם אם היא הפליה סטטיסטית מעורר אתגר של ממש ככל שהדבר נוגע לפרויקט האלגוריתמי.

86 חוק איסור הפליה, לעיל ה"ש 2.

87 ס' 2(ג) לחוק שוויון הזדמנויות בעבודה; ס' 13(1) לחוק איסור הפליה, לעיל ה"ש 2.

88 נציין כי החוק אינו מגדיר במפורש מהי הפליה, ולא מדובר ברשימה סגורה. כפי שנפסק בפסק הדין של הנשיא ברק בעניין **רקנט**, "חוק שוויון ההזדמנויות אינו מגדיר 'הפליה' מהי. על-כן פירושו של ביטוי זה ייעשה על-פי התכלית המונחת ביסוד החוק. בהיעדר נתונים על תפיסה מיוחדת של המחוקק בעניין זה יש לצאת מתוך ההנחה כי הגישה המקובלת של המשפט הישראלי בעניין שוויון והפליה – ואלה אינם אלא שני צדדיו של אותו מטבע – חלה גם בגדריו של חוק שוויון ההזדמנויות". ראו **רקנט**, לעיל ה"ש 36, בעמ' 344. בדיון המקורי, בג"ץ 6051/95 **רקנט נ' בית הדין לעבודה פ"ד נא(3) 289** (1997), ציין השופט חשין כי "חזקה עליו [על המחוקק – א"ב] כי אימץ לעצמו את מושג ההפליה כפי שנתקבל בהלכה".

89 **רקנט**, לעיל ה"ש 36.

90 Wachter et al., *Why Fairness Cannot be Automated*, לעיל ה"ש 83, בעמ' 66. טענות נגד האלגוריתמים יכולות להיחשף רק לאחר מעשה, כלומר אם קורבן ההפליה מרגישה שנפגעה מההחלטה האלגוריתמית. לדוגמה, מועמדת לעבודה עשויה לעולם לא לדעת באיזה שלב, או האם בכלל, קורות החיים שלה סוננו בתהליך הגיוס. היא אינה יכולה לדעת מדוע נדחתה – האם זה משום שהיא אינה מתאימה לעבודה או כי האלגוריתם מוטה נגדה משום שהיא אישה? צרכן עשוי לא לדעת שהוא אינו נחשף לפרסומות מסוימות או שהוא מקבל הצעות למוצרים בתנאים פחות טובים באופן יחסי משכנו. מאחר שהמערכת האלגוריתמית פועלת על נתוני עתק, ומאחר שהיא מלמדת את עצמה באופן תמידי על בסיס נתונים אלו, ההפליה בהכרח תהיה משועתקת בתוכה על בסיס

נראה לנו כי מאופיו של תהליך קבלת ההחלטות האלגוריתמי נובע שהדרך קדימה לרגולציה של ההפליה האלגוריתמית הסטטיסטית כרוכה בשאלות נורמטיביות-תוצאתיות. כלומר, משום שאלגוריתמים (כשהם מבוססים על נתונים מדויקים) עשויים להנפיק תוצאות לא-שוויונות – הגם שהן מדויקות סטטיסטית – נשאלת השאלה הנורמטיבית מתי נרצה להתערב באי-השוויון הסטטיסטי שהאלגוריתמים מייצרים, ומה יהיו הקריטריונים להתערבות ולמידת ההתערבות.⁹¹

הכרעות אלה – לגבי פלטים מדויקים אך לא-שוויוניים שינביעו האלגוריתמים – עשויות להשתנות מזירה מזירה ומקבוצה לקבוצה. נידרש להכריע גם לגבי שיקוליהם וכישוריהם של מי שנרצה שיפקחו על אי-השוויון שייווצר ויכילו את האלגוריתם. בנוסף, תוצאה שוויונית תטיל עלויות על מעסיקים ונותני שירותים, מה שמעלה את השאלה על מי נטיל את העלויות שתהיינה כרוכות באכיפת תוצאה זו. בנוסף, נרצה לשאול האם לא ראוי שהציבור בכללותו יישא בעלויות שתהיינה כרוכות באימוץ החלטות שאינן מדויקות סטטיסטית אך רצויות נורמטיבית.

חשוב להדגיש שמשום שהפרויקט האלגוריתמי הוא סטטיסטי במהותו, ההכרעות על בסיסו תשקפנה פערים סטטיסטיים בין קבוצות, ככל שאלה מתקיימים גם במציאות (ובהנחה שמאגרי הנתונים הם מדויקים מספיק). לכן, לא יהיה מנוס מקבלת הכרעות נורמטיביות בדבר אותם פערים שהם מדויקים מבחינה סטטיסטית, אך עשויים להיות לא רצויים מבחינה נורמטיבית.

כדי להבין את טיבן של ההכרעות שנידרש לקבל נתמקד בדוגמה של מעסיקה המשתמשת באלגוריתם של השמת עובדות ועובדים לעבודה מאתגרת פיזית במהותה. האלגוריתם בוחר לעבודה רק מועמדים צעירים לאור ההבחנה הסטטיסטית שהם מתאימים יותר – בממוצע – לעבודות מסוג זה. בהינתן התוצאות החלוקתיות של האלגוריתם, נשאל את עצמנו האם הבחנה שכזו היא ראויה מבחינה נורמטיבית. לצורך העניין, נניח כי ההבחנות שעשה האלגוריתם הן מדויקות לגמרי ועל כן בממוצע בחירת המועמדים שנעשתה על ידו היא היעילה ביותר (מבחינת עלויות הבדיקה והמיון שהיא מטילה על המעסיקים). ייתכנו עדיין כמובן מועמדים לעבודה שיתאימו לדרישותיה אף על פי שהם מבוגרים יחסית, אלא שעלות איתורם (למשל בדרך של מבחני מסוגלות פיזית) גבוהה יותר מהתועלת שתתקבל מן האיתור.

נתונים מנבאים, חלקם אינטואיטיביים יותר וחלקם פחות, ולאנושות לא תהיה אפשרות להתחקות אחריה. לכן, כל שימוש יוביל להפליה סטטיסטית, חלקה אסורה בחוק וחלקה לא. ראו Dowding & Taylor, *Algorithmic Decision-Making*, לעיל ה"ש 6, בעמ' 68.

⁹¹ לדוגמה, שימוש באלגוריתם לחיזוי פשיעה כנראה יוביל לאכיפת יתר בקרב אוכלוסיות מוחלשות. סטטיסטית, אלגוריתמים כאלה עשויים להיות מדויקים: הם מזהים אזורים שבהם שיעור הפשיעה בעבר היה גבוה, ומפנים לשם כוחות שיטור באופן שיקטין את הסיכוי לאירועים נוספים. עם זאת, הדבר עשוי לשמור על ההטיה המבנית, לגרום לאכיפת יתר ולשחוק את אמון הציבור. להרחבה ראו את המאמר Sandra G. Mayson, *Bias in, Bias Out*, 128(8) YALE L.J. 2218 (2019), שסוקר בהרחבה עניין זה.

כעת, נידרש לשאול את עצמנו האם התוצאה המדויקת בממוצע מצדיקה את הפגיעה באוטונומיה ובהזדמנויות של אותם מועמדים מבוגרים אשר היו עשויים להתקבל לעבודה, לו רק הייתה נבחנת מועמדותם ללא קריטריון הגיל.

חשוב להדגיש כי גם ללא קבלת ההחלטות האלגוריתמית, ובהינתן האיסור על הפליה סטטיסטית הקיים במשפט הישראלי, רבים מהמועמדים המבוגרים לא היו מתקבלים לעבודה, משום שלא היו עומדים במיוני ההתאמה הספציפיים. עם זאת, חלקם של המועמדים, אלה השונים במאפייניהם ממוצע העובדים בגילם, היו עשויים להתקבל לעבודה. נזכיר בהקשר זה כי משמעותו של איסור ההפליה הסטטיסטית היא שמעסיקים אינם רשאים להסתמך על קריטריון הגיל כמשתנה בעת קבלת ההחלטות בדבר הקבלה לעבודה. עם זאת, מעסיקים רשאים עדיין שלא לקבל מועמדים מבוגרים לעבודה אם נכשלו במבחני ההתאמה האישיים שהם נדרשו לעבור.

ייתכן כמובן שגם מעסיקות המקבלות החלטות באופן עצמאי וללא מעורבותם של אלגוריתמים תפעלנה בניגוד לחוק ותקבלנה החלטות על בסיס הפליה סטטיסטית מחמת גיל. במילים אחרות, גם בהיעדר אלגוריתמים, מעסיקות עשויות להפלות הפליה סטטיסטית מחמת גיל. עם זאת, כפי שטענו בהרחבה לעיל, יכולתם של מעסיקים אינדיבידואליים לעסוק בהפליה סטטיסטית בהיעדר אלגוריתמים היא מצומצמת יותר מזו של אלגוריתמים, שהם מתוחכמים יותר, פועלים בסדרי גודל אחרים, ומסתמכים בהחלטותיהם על מאגרי נתונים גדולים. בנוסף, טענתנו היא שקבלת ההחלטות האלגוריתמית מערבת במהותה הפליה סטטיסטית.

בחזרה לקבלת ההחלטות האלגוריתמית, נציין בהקשר זה כי הסרתם של נתוני הגיל ממאגר הנתונים לא הייתה מסייעת להתגבר על התוצאה של האלגוריתם, משום שגיל נמצא במתאם עם הרבה מאוד מאפייני חיים אחרים – המערכת האלגוריתמית תוכל להסיק באופן סטטיסטי את גיל המועמד לפי היסטוריית הגלישה שלו, נתוני בריאות, מיקוד, הדברים שכתובים בקורות החיים והשפה שבה הם נכתבו. ובכן, השאלה הנורמטיבית הראשונה שניאלץ להכריע לגביה תהיה בדבר תמהיל הגילים של המועמדים שנרצה להפנות למעסיקה להמשך תהליך הבירור. ככל שנבקש להגדיל את חלקם של המועמדים המבוגרים (או את גילם) כך נטיל עלויות גבוהות יותר על המעסיקה שתיאלץ להשקיע משאבים ועלויות בסינון של יותר מועמדים שסיכויי ההתאמה שלהם לעבודה נמוכים יותר. חשוב להדגיש שהעלויות שבהן תתבקש המעסיקה לשאת הן עלויות הבירור והבדיקה (להבדיל למשל מעלויות העסקתם של עובדים לא מתאימים, שבהן איננו מבקשים את המעסיקה לשאת). עלויות אלה הן עלויות של ממש שאותן ביקש האלגוריתם לחסוך.

ההכרעה בדבר התמהיל תהפוך מורכבת אף יותר⁹² אם ניקח בחשבון שהמערכת האלגוריתמית תוכל להוביל לפערים בתוצאות גם לגבי קריטריונים מוגנים אחרים – כמו למשל המגדר של הפונים והפונות לעבודה. בדומה לסוגיית גיל המועמדים, שבה

92 למורכבות הגלומה ברגולציה של בינה מלאכותית וסוג השיקולים שנידרש לשקול, ראו: Yonatan Arbel et al., *Systematic Regulation of Artificial Intelligence*, 54 ARIZ. ST. L. J. 545, 551, 558–559, 585 (2023).

דנו קודם, גם המגדר שלהם עשוי להיות במתאם סטטיסטי עם התאמתם לדרישות הפיזיות של העבודה. במילים אחרות, נאלץ לשאול את עצמנו האם ואיך נרצה לכייל את האלגוריתם כך שהפלט יהיה מאוזן מבחינה מגדרית. כמו לגבי גיל כך גם לגבי מגדר, לא ניתן יהיה פשוט להסיר את נתוני המגדר ממאגר הנתונים משום שגם המגדר נמצא במתאם עם מאפייני חיים אחרים. בנוסף, גם לגבי מגדר, ככל שנבקש לכלול עוד נשים בקבוצת המועמדים שימשיכו את תהליך הבירור עם המעסיקה, כך תעלה עבודה עלות הבירור והמיון. משום שכל המועמדים הינם גם בעלי גיל וגם בעלי מגדר, נידרש לשאול את עצמנו כיצד לחשוב על ההצטלבות הראויה של תכונות אלה בקבוצה שאותה אנחנו מבקשים להעביר להמשך הבירור אצל המעסיקה. נשאל את עצמנו למשל האם ראוי שבקבוצה יימצא אותו שיעור של נשים מבוגרות וגברים מבוגרים, גם אם סיכויי ההתאמה של אחת הקבוצות לעבודה נמוכים מזו של האחרת, ועל כן עלות הבירור שכרוכה בהכללתה גבוהה יותר.

ייתכן שנניח מראש כי קבוצת המועמדים שנעביר להמשך הבירור צריכה לתאם את ההתפלגות של האוכלוסייה בכללותה; המשמעות של החלטה כזו עלולה לגרור עלויות בירור אדירות במקרים של עבודות מסוימות, אך נמוכות בהרבה במקרים של עבודות אחרות. לכן, נראה שההחלטה על תמהיל הפלט צריכה לקחת בחשבון את מאגר המועמדים, אופי העבודה ועלויות הבירור והמיון. ככל שנבחר בתוצאות אחדות יותר בנסיבות שונות, כך תהיה הטלת העלויות על המעסיקים שוויונית פחות (ומי שיישאו בנטל הגדול יותר יהיו מי שדרישות התפקיד בעבודות שהם מציעים נמצאות במתאם גבוה יותר עם קריטריונים מוגנים).

נציין שוב כי ההכרעה הנורמטיבית בדבר היקפם הראוי של פערים – שהם תוצר של פלטם מדויקים אך לא שוויוניים שינביעו האלגוריתמים – תושפע גם מזירת החיים שבה אנו פועלים. הדוגמה שבה עסקנו לעיל התמקדה בהחלטות הקשורות בשוק העבודה – זירת חיים שהיא חיונית לפרנסתם של אנשים, לחיי היום-יום שלהם ולקריירות שלהם. זירות אחרות עשויות לערב שיקולים אחרים כמו בריאות וקניין. ייתכן שפערים מגדריים בפרמיית ביטוחי החיים, למשל, ייתפסו על ידנו כראויים יותר או פחות מפערים מגדריים בשוק העבודה, גם בהינתן עלויות בירור יחסיות דומות. באופן דומה, ההכרעה הנורמטיבית בדבר היקפם הראוי של פערים, כאמור, תושפע גם מן השאלה מי היא הקבוצה הנפגעת. כך, ייתכן שנתפוס פערים דומים בהיקפם בשוק העבודה כראויים יותר או פחות, כשהקטגוריה של השונות היא גיל בהשוואה למגדר, למשל.

לסיום, נדגיש כי נראה שכל התערבות בתוצאותיו של אלגוריתם שהוא מדויק מבחינה סטטיסטית תחייב אותנו לקבל שורה של הכרעות נורמטיביות לגבי היקפם הראוי של הפערים בהינתן הפגיעה הכרוכה בהפליה הסטטיסטית והעלויות הכרוכות בבדיקה האינדיבידואלית.⁹³

93 פטר סאליב מציע במאמרו גישה חדשנית בשם "ביטול באמצעות אלגוריתם". גישה זו טוענת כי ניתן להשתמש באלגוריתמים מתקדמים של למידת מכונה על מנת לצמצם במידה רבה את התלות בשיטות המסורתיות של אכיפת חוק וכליאה, תוך שמירה על רמות

סיכום

עם התקדמות הטכנולוגיה, עולה באופן בלתי נמנע נושא ההטיה וההפליה בתחום הבינה המלאכותית. במאמר זה ביצענו השוואה בין ההפליה האנושית, שטבועה בכולנו, ובין ההפליה האלגוריתמית. הראינו כי אף שהשימוש במערכות מבוססות אלגוריתמים יכול למתן הפליה אנושית הנובעת מהטיות אנושיות, טעמים, סטראוטיפים ותסריטים חברתיים, מערכות בינה מלאכותית אינן יכולות להימנע מהפליה סטטיסטית, שעלולה להפלות לרעה חלק ממשתמשי המוצר או צדדים שלישיים.

לפרויקט האלגוריתמי יש כמובן גם השלכות חיוביות רבות, בין היתר לעניין אי-השוויון וההפליה. למשל, ניתן להשתמש באלגוריתמים על מנת לאתר הטיות בהחלטות שהתקבלו על ידי אלגוריתמים אחרים. לצד היתרונות הבולטים, לשיטתנו יש לתת את הדעת על כך שהפרויקט מציב אתגר של ממש לדיני איסור הפליה כשמדובר בהבחנות סטטיסטיות מדויקות על בסיס תכונות חשודות. כפי שתואר במאמר זה, ההפליה הסטטיסטית האלגוריתמית הינה אינהרנטית לשימוש באלגוריתמים. היא מורכבת, בהיותה מדויקת סטטיסטית, ועדיין עלולה לשעתק ולהנציח את אי-השוויון בחברה. צמצומה של הפליה מהסוג הזה אפשרי כנראה רק לאחר שהתקבלה התוצאה הלא-שוויונית בדרך של כיול מחדש של הפלט האלגוריתמי כך שיהפוך שוויוני יותר, אך כאמור יהיה כרוך בעלויות מיון ובדיקה.

הגם שבקרה על התוצאה של החלטת האלגוריתם עשויה להוות פתרון להפליה, בעיות אלו לרוב אינן ניתנות לפתרון בצורה טכנולוגית גרידא, שכן הן כרוכות בשאלות עמוסות ערכים, כגון מה היא החלוקה של הזדמנויות התעסוקה הראויה, אילו עלויות ראוי להטיל על מעסיקים ואילו גורמים צריכים להשפיע על סיכויי התעסוקה של אנשים. שאלות אלה הן שאלות נורמטיביות כבדות משקל שניאלץ להידרש להן בכל פעם שנבקש לתקן את תוצאותיה של ההפליה האלגוריתמית הסטטיסטית. משום

פשיעה נמוכות. ההצעה מתמקדת בהפחתת הפשיעה באמצעות שימוש אלגוריתמי חכם ומכוון כדי למקסם את היעילות של אכיפת החוק, לצמצם התערבויות מיותרות, ולהפחית את התוצאות השליליות של שיטור וכליאה. בין היתר סאליב טוען שהגישה של "ביטול באמצעות אלגוריתם" יכולה לחול לא רק במערכת המשפט הפלילית, אלא גם בתחומים אחרים כגון בריאות, פיננסים ותעסוקה. הרעיון המרכזי הוא שאלגוריתמים יכולים לתרום להפחתת הפליה מערכתית תוך התמקדות רק במקומות שבהם אכיפת החוק נחוצה. סאליב מדגיש כי המסגרות המסורתיות להבנת הפליה אלגוריתמית ומדידתה אינן מתאימות בהכרח בהקשר של ביטול באמצעות אלגוריתם. מסגרות מסורתיות אלו עשויות להעריך את המדיניות של ביטול באמצעות אלגוריתם כבלתי הוגנת, גם כאשר מדיניות כזו מפחיתה במידה רבה את מספר האנשים שנפגעים בגלל גזעם. מצד שני, הן עשויות להעריך מדיניות אחרת כהוגנת, אף על פי שהיא גורמת לפגיעה אפלייתית חמורה יותר. תובנה זו תצריך הערכה מחודשת של האופן שבו אנו תופסים היעדר הפליה או הוגנות על מנת לתפוס בצורה טובה יותר את המורכבויות שמביאות איתן ההתערבויות האלגוריתמיות. ראו Peter Salib, *Abolition by Algorithm*, MICH. L. REV. (forthcoming 2025).

שהפרויקט האלגוריתמי הינו במהותו פרויקט של הבחנה סטטיסטית, ומשום שניתן לטפל באי-השוויון שהוא מייצר אך באמצעות תיקון בדיעבד של תוצאותיו, נידרש בכל מקרה ומקרה להגדיר מה הן תוצאותיה של חלוקה צודקת (שאינה מדויקת סטטיסטית). בכל פעם שנימנע מלעשות כך, תהיינה ההחלטות מבוססות על הפליה סטטיסטית, שלגביה הכריע המשפט הישראלי שהיא בנסיבות של תכונות וזירות מסוימות הפליה אסורה.