מוסד שמואל נאמן
למחקר מדיניות לאומית

מדע וטכנולוגיה

בינה מלאכותית, מדעי הנתונים ורובוטיקה חכמה

דו"ח בנושא אתיקה, משפט ופרטיות

ד"ר דפנה גץ
אושרת כץ שחם
רינת קליין
ד"ר רועי צזנה
שלמה רוזנברג
אבידע שהם
אלה ברזני
ד"ר ערן לק
סימה ציפרפל

דצמבר
2018

מוסד שמואל נאמן
למחקר מדיניות לאומית

בינה מלאכותית, מדעי הנתונים ורובוטיקה חכמה

דו"ח בנושא אתיקה, משפט ופרטיות

מוגש למועצה הלאומית למחקר ופיתוח במשרד המדע והטכנולוגיה

חוקרים:

ד"ר דפנה גץ

אושרת כץ שחם, רינת קליין, ד"ר רועי צזנה, שלמה רוזנברג, אבידע שהם, אלה ברזני, ד"ר ערן לק, סימה ציפרפל

דצמבר, 2018

קריית הטכניון, חיפה 3200003 טל. 04-8292329 פקס. 04-8231889 info@neaman.org.il

תודות:

תודתנו נתונה לד"ר תהילה שוורץ אלטשולר, פרופ' דב גרינבאום, פרופ' אור דונקלמן ועו"ד קרני שגל-פפרקורן על הסיוע ועל הרקע והתובנות שסיפקו לנו.

אין לשכפל כל חלק מפרסום זה ללא רשות מראש ובכתב ממשרד המדע והטכנולוגיה ו/או ממוסד שמואל נאמן מלבד לצורך ציטוט של קטעים קצרים במאמרי סקירה ופרסומים דומים תוך ציון מפורש של המקור.

הדעות והמסקנות המובאות בפרסום זה הן על דעת המחבר/ים ואינן משקפות בהכרח את דעת מוסד שמואל נאמן.

תוכן העניינים

1. הקדמה 4
2. אתגרים כלליים 4
3. אתגרים בהשפעה על המסחר והכלכלה 10
4. מדיניות של מדינות 12
5. מדיניות של חברות 18
6. סיכום 24
7. מקורות 26

1.הקדמה

בחלק זה של הדו"ח נסקור את האתגרים החדשים הגדולים שנפתחים בעקבות ההתפתחויות האחרונות בתחום הבינה המלאכותית, וננסה להבין כיצד מציעים ארגונים בינלאומיים, מדינות וחברות תעשייתיות להתמודד עמם.

2.אתגרים כלליים

◀ אתגר הסיבתיות (קאוזליות)

אתגר הסיבתיות מתאר את הקושי של יצרני בינות מלאכותיות, אלגוריתמים לומדים ורובוטים לחזות את הדרך בה יתנהלו המוצרים מהרגע בו ישתחררו לאוויר העולם. המקור לאתגר מגיע בעיקר מתפישת המחשוב הנוכחית הרווחת עבור בינה מלאכותית: זו של רשתות עצבים מלאכותיות. מהנדסי הבינה המלאכותית אינם טורחים לגבש כללים מדויקים עבור מערכות עצבים מלאכותיות, אלא מפתחים מודלים סטטיסטיים עבור תחום מסוים של בעיה, ואז מאמנים את המודלים הללו על דוגמאות רבות ושונות, כדי לעדן ולשפר את הדיוק שלהן.

מערכות העצבים המלאכותיות מוצלחות במיוחד בהבנת העולם שמסביבן, אך מוגבלות ביכולות הלוגיות שלהן – הן אינן נסמכות על כללים מדויקים, אלא "על מה שעובד מספיק טוב, מספיק מהפעמים". באמצעות אימון ניתן 'לאלף' אותן להפיק את התוצאות הקרובות ביותר להיות נכונות ומתאימות. עם זאת, המשמעות היא גם שמערכות אלו, מהרגע שהן משתחררות מהמפעל או מסביבת התכנות הבטוחה והמסוגרת, יכולות לזכות באימון שיערער את הידע שרכשו לפני כן.

הבעיה חמורה עוד יותר מאחר שמערכות העצבים המלאכותיות מהוות (לפחות נכון להיום) קופסא שחורה במובנים רבים. המהנדסים הפועלים עליהן יכולים להבין את הקלט שהן מקבלות ולנתח את הפלט שהן מספקות, אך אופן עיבוד המידע בתוך המערכת מהווה עדיין תעלומה עבורם. וכפי שהוסבר במגזין של MIT -

"אינך יכול פשוט להסתכל ברשת עצבית עמוקה ולראות איך היא עובדת. ההיגיון של הרשת מוטמע בפעולתם של אלפי נוירונים מדומים, המסודרים בעשרות או אפילו מאות שכבות מורכבות המחוברות זו לזו."[1]

דוגמה טובה לכך מגיעה מהמקרה בו שחררה חברת מיקרוסופט בוט חדש בשם "טאי" לרשת החברתית. טאי עוצבה בדמותה של צעירה אמריקנית בת 19, ונועדה להיות 'הפנים' של מיקרוסופט מול צעירי אמריקה. ברם, הציבור הרחב ראה בטאי סוג של אתגר, ורבים שלחו לה הודעות שנועדו לאמן אותה במילים ומושגים לא-ראויים. תוך פחות מ- 24 שעות החלה טאי לספק תשובות משלה שהיו מבוססות על מה שלמדה מהציבור, והודיעה שהנאצים צדקו אחרי הכול[2].

ניתן לראות מקרה זה כקוריוז משעשע ותו לא, אך ברור שמערכות לומדות מתחילות להשתלב בכל התחומים. מערכות לומדות נמצאות בבסיסם של רכבים אוטונומיים, של אלגוריתמים למסחר בבורסה, בייעוץ רפואי ועוד. כל ניסיון לעודד בתי-עסק וחברות לשלב בינות מלאכותיות במודל העסקי שלהם, לא ינחל הצלחה מבלי התמודדות עם הסוגיות המשפטיות של אתגר הקאוזליות. חברות קטנות יהסוס – ובצדק – לפתוח את עצמן לתביעה מצד הלקוחות או אפילו מצד הרגולטור, בשל פעולתן של מערכות שיצאו מהמפעל ברמה גבוהה של פעילות, אך תפקודן השתבש ברגע שהחלו לקבל מסרים סותרים ומערערים מהציבור ומהלקוחות.

המחוקק יצטרך למצוא פתרון שימזער את האחריות הנופלת על כתפי היצרנים והמשתמשים בבינה המלאכותית כתוצאה משימוש בבינה מלאכותית היוצאת מכלל שליטה, צד בצד עם הדרישה מצד היצרנים והמשתמשים לעמוד בסטנדרטים גבוהים-מספיק של בטיחות כדי לצמצם את הסיכוי לאסונות ותאונות חמורים. מבחינה זו, ברור למשל שאין דינו של בוט המתדיין ברשת החברתית – שניתן לכבות אותו בלחיצת כפתור – כדינו של רכב אוטונומי שעלול ללמוד להתנהג בכביש באופן קלוקל, ובכך לסכן חיים באופן ממשי.

◀ אתגר הטעות האינהרנטית

מערכות בינה מלאכותית צפויות להיות בעלות רמה גבוהה של דיוק בקבלת החלטות ובניתוח מצבים שונים בתחומים עבורן אומנו. ככל שמערכות אלו ישתכללו יותר, כך ישתפרו גם יכולותיהן עד שיגיעו לרמה על-אנושית – כפי שניתן לראות כבר היום במטלות שונות, כקריאת שפתיים[3], זיהוי שחפת[4], פענוח צילומי רנטגן[5] ועוד. עם זאת, גם

בינות מלאכותיות ברמת הדיוק הגבוהה ביותר צפויות לשגות במצבים מסוימים – בין שכתוצאה מחוסר יכולתן לעבד ולהבין את כל המורכבויות שבמצב מסוים, או כתוצאה מפעולתם של שחקנים אנושיים שאינה ניתנת לחיזוי ולצפייה מראש. אנו מכנים קושי זה "אתגר הטעות האינהרנטית", מאחר וכל ייעוץ מצד הבינה המלאכותית טומן בחובו סיכוי מסוים לטעות.

כדוגמה לאתגר הטעות האינהרנטית ניתן לחשוב על מצב בו בינה מלאכותית מספקת ייעוץ רפואי ברמה גבוהה, וממליצה שלא לכרות את רגלו של מטופל החולה בסוכרת, מכיוון שקיים סיכוי גבוה – 99% - שהפצעים לא יזדהמו. באחוז אחד מהמקרים, יזדהמו הפצעים שברגל אחרי הכול, והמטופלים ומשפחותיהם עלולים לתבוע את מפעילי הבינה המלאכותית על ייעוץ שגוי.

באופן דומה, רכבים אוטונומיים עשויים לקבל החלטות הגיוניות ומנומקות-היטב על הכביש, שאמורות להוביל לתוצאות הטובות ביותר לפי התכנות וההנחיות שקיבלו. עם זאת, נהגים אנושיים אינם מצייתים תמיד לאותם כללי היגיון, אתיקה ומוסר, ועלולים 'לחתוך' ולבצע עבירות תנועה שונות שהרכבים האוטונומיים יתקשו לקחת בחשבון בחישוביהם. המשמעות היא שאפילו הרכב האוטונומי המוצלח ביותר צפוי לשגות מדי פעם בחישוביו, גם אם כתוצאה מפעולתן של ישויות אנושיות, או אירועים אקראיים שלא ניתן לחזות מראש, כמהמורות הנפערות בכביש, או שכבת קרח דקה וחלקה שהתהוותה על האספלט.

◀ אתגר הנציגות ההוגנת

הספרות המשפטית מכירה בקיומה של "בעיית הנציג", המתארת מצב בו נציגו של לקוח מסוים פועל בדרכים שאינן מספקות מענה מיטבי לרצונו של הלקוח. הסיבה לכך היא שנציגים אנושיים מתקשים לייצג באופן מיטבי את לקוחותיהם משתי סיבות מרכזיות: ראשית, בשל הקושי ביכולתו של הנציג להבין את דרך חשיבתו של הלקוח, ושנית בשל מערכת התמריצים השונה במסגרתה פועל הנציג, ואשר מכווינה אותו לביצוע פעולות שאינן בהכרח לטובת הלקוח. כך, למשל, לעורכי דין ישנו אינטרס מובנה 'למשוך' את המשפט זמן רב יותר, מכיוון שבמקרים רבים הם זוכים לתשלום לפי שעות עבודה. מערכות תמריצים אלו אינן מצוינות בקול רם, אך ברור כי התמריצים הללו משוקללים באופן מודע או שאינו-מודע על-ידי הנציג.

בעיית הנציג קשה עוד יותר במקרים בהם הלקוח אינו מבין היטב את תחום המקצוע של הנציג (כפי שקורה בדרך כלל). קיימים מקרים רבים בהם סוכנים מכל סוג שהוא – בביטוח, בבורסה, בראיית חשבון ועוד – ניצלו לרעה את חוסר ההבנה או חוסר העניין של הלקוח בתחומם, על מנת לספק לו שירותים ברמה נמוכה, או אפילו לרמות אותו במובהק.

לכאורה, אלגוריתמים מסוגלים לספק מענה לבעיית הנציג, מכיוון שהם חפים משיקולים זרים, ומסוגלים לייצג באופן מיטבי את הלקוח. עם זאת, ראייה נאיבית מסוג זה אינה עומדת במבחן המציאות. מנועי בינה מלאכותית מיוצרים כיום על-ידי חברות, שבמקרים מסוימים שולטות ביד רמה על נישה מסוימת (כגון גוגל בתחום מנועי החיפוש, פייסבוק בתחום הרשתות החברתיות, איירבנב בתחום האירוח השיתופי, ועוד). מערכת השיקולים לפיה מקבלים המנועים הללו החלטה נותנת, באופן טבעי, משקל גדול יותר לצרכי החברה שפיתחה ומפעילה אותם, על חשבון צרכי הלקוחות. במקרים קיצוניים עלולים הלקוחות לגלות כי שירותי הנציגות שניתנו להם – בין שמדובר בשירותי ייעוץ ביטוחי, רפואי, או אפילו בתוצאות חיפוש ברשת – מוטים לטובת החברות באופן שפוגע בלקוחות.

דוגמה למקרה מסוג זה הופיעה בשנת 2017, כאשר האיחוד האירופי קבע כי גוגל העדיפה להמליץ על שירותיה שלה, במקום על שירותים מוצלחים יותר (לכאורה) של חברות מתחרות[6]. זוהי דוגמה ברורה לאתגר הנציג, מאחר וגוגל מספקת למעשה 'נציג דיגיטלי' – סוכן אלגוריתמי המנסה לפענח מהם רצונותיו של הלקוח המבצע פעולת חיפוש ברשת באמצעותו. הנציג הדיגיטלי של גוגל אמור לספק את התוצאות המתאימות ביותר ללקוח, אך כפי שניתן ללמוד מקביעת האיחוד האירופי, הנציג של גוגל מעדיף – באופן טבעי מבחינת גוגל – להמליץ על שירותיה של החברה שייצרה אותו.

טענות דומות כנגד מנוע החיפוש של גוגל נשמעו מצד חברת Yelp, המספקת בעצמה שירותי המלצה חכמים[7]. באופן דומה, ניתן לטעון כי גם החלטתה של חברת איירבנב למנוע מהנציג הדיגיטלי שלה להציג ללקוחות דירות להשכרה בהתנחלויות, מהווה דוגמה נוספת לבעיית הנציגות ההוגנת[8]. טוויטר זוכה גם היא להתמודד עם טענות כנגד הטיה של נציגה הדיגיטלי, עקב השתקת קולות שמרניים ברשת החברתית[9].

כפי שניתן להבין, אתגר הנציגות ההוגנת הופך להיות חמור יותר עקב תופעת "המנצח לוקח הכול" הרווחת בעולם הדיגיטלי. תופעה זו מביאה לכך שבעולם הדיגיטלי, שירותים מסוימים נוטים להשתלט לגמרי על נישות שונות, ולהעניק כוח גדול לחברות שמספקות אותם. מבלי תחרות הולמת במתן אותם שירותים, הלקוחות נאלצים להסתפק

בנציג דיגיטלי אחד בלבד, שמוטה באופן ברור לצרכי החברה המייצרת אותו. הם יכולים, במקרים מסוימים, לבחור בנציגי חברות קטנות יותר, אך אלו לרוב אינם ברמה גבוהה מספיק, או שאינם מותאמים למשתמש הממוצע[10].

גורם נוסף המקשה על הלקוחות ועל הרגולטור בהתמודדות עם הנציגים הדיגיטליים, הוא שקיים קושי ברור בפענוח מכלול השיקולים העומדים מאחורי ההחלטה הסופית שמקבלים הנציגים. נציגים דיגיטליים רבים מבוססים כיום על רשתות עצבים מלאכותיות, שלא ניתן עדיין להבין את המנגנונים המדויקים לפיהם הן פועלות. כתוצאה מכך, אפילו החברות המפעילות את הנציגים הללו אינן יכולות לעיתים להסביר מדוע בדיוק קיבלו אלו את ההחלטה אשר קיבלו.

אתגר הנציגות ההוגנת אינו צפוי לפגוע באופן ישיר בחברות – למעשה, החברות המייצרות את הנציגים הדיגיטליים צפויות דווקא להרוויח ממנו בטווח הקצר – אך הוא בעל השלכות הרסניות בטווח הארוך על החברה בכללותה. המשתמשים עלולים לאבד אמון בשירותים הדיגיטליים שהחברות מציעות להם ולבחור להחרים את הנציגים הדיגיטליים, כפי שנעשה לאחרונה במקרה של פייסבוק[11]. בראייה רחבה יותר, נציגים דיגיטליים מוטים יכולים להשפיע לרעה על כל תחומי החיים: מהטיית הראיה הפוליטית של האזרחים[12], [13], ועד למתן שירותי ייעוץ רפואי המתעדפים את שיקולי החברות הרפואיות על חשבון המטופלים[14]. ברור שהציבור הרחב יצפה מהמחוקק להתמודד באופן יעיל עם נציגים דיגיטליים המועלים באמון לקוחותיהם.

◀ אתגר הפיקוח על מערכות אוטונומיות

מערכות אוטונומיות מסוגלות ומחויבות מטבען לקבל החלטות בזמן-אמת. המערכות האוטונומיות המדוברות ביותר כיום הן הרכבים האוטונומיים, המתחילים להגיע לכבישי הערים. ניתן למצוא כיום רכבים אוטונומיים ביותר משבע-עשרה ערים בכל העולם, בהן הם עוברים ניסויים תוך כדי שהם מתמודדים עם התנועה בכבישים[15]. הרכבים האוטונומיים נדרשים לקבל החלטות בפרקי-זמן של שברירי-שניות: מתי לבלום ובאיזו עצמה, מתי לסטות ימינה או שמאלה, באיזו מהירות עליהם לנסוע, וכן הלאה.

על מנת לקבל החלטות באופן מיטבי, הרכבים האוטונומיים חייבים להיות מצוידים במערכת של היוריסטיקות (heuristics) – כללי אצבע שמכווינים את פעולתם. הם חייבים גם להיות מחוננים בסט של 'ערכים' לפיהם ייקבעו ההיוריסטיקות הללו. ערך כזה לדוגמה עשוי להיות ערך השוויון, הקובע כי הרכב האוטונומי צריך להתייחס לכל בני-האדם כשווים, ולא להפלות לפי גיל, מין, גזע או מעמד כלכלי או חברתי. כך, אם נאלץ הרכב האוטונומי להחליט בין שתי אפשרויות שיכולות לפגוע בבני-האדם, הוא אינו רשאי להתייחס לאחד מהקריטריונים הללו בניסיון לקבוע מה עליו לעשות.

ועדה שכונסה מטעם משרד התחבורה בגרמניה המליצה כבר על כללים אתיים שיחולו על מכוניות אוטונומיות. ההמלצות, שיכולות להיות תקפות ברובן גם עבור ישראל, הן כדלקמן[16] –

1. מטרתן המרכזית של מערכות תחבורה אוטונומיות הינה לשפר את בטיחות התחבורה לכל המשתמשים בדרכים. מטרה משנית היא להגדיל את אפשרויות התחבורה הפתוחות לאזרחים.
2. ההגנה על האינדיבידואל חייבת לקבל קדימות בהשוואה לכל שאר השיקולים התועלתניים. המטרה היא לצמצם את רמת הפגיעה עד שהיא נמנעת לחלוטין. מערכות אוטונומיות לא יורשו להתנייד בכבישים אלא אם יבטיחו להביא לפחות לירידה בפגיעה בהשוואה לנהיגה אנושית.
3. המגזר הציבורי אחראי להבטחת בטיחותן של המערכות האוטונומיות בכבישים הציבוריים. מערכות לנהיגה בכבישים עשויות להזדקק לפיכך לאישור רשמי ולניטור מתמיד.
4. מערכות אוטונומיות צריכות למנוע תאונות בכל מקרה בו ניתן לעשות זאת באופן פרקטי. עם זאת, יש לתכנן את הטכנולוגיה כך שמצבים קריטיים לא יצוצו מלכתחילה. הרכבים צריכים להיות מתוכננים ומתוכנתים כך שייסעו באופן מתגונן ומוכן-לכל ("defensive and anticipatory") כך שהסיכון שיהוו לאחרים יהיה מינימלי.
5. כניסתן של מערכות נהיגה אוטונומיות משוכללות יותר, במיוחד כאלו עם אפשרות למניעת התנגשות אוטומטית, יכולה להיות מקובלת חברתית ואתית אם קיים בה הפוטנציאל לצמצום נזקים.
6. במקרים קיצוניים שלא ניתן להימנע מהם, למרות כל ההכנות ואמצעי הזהירות, העדיפות העליונה היא הגנה על חיי אדם. המערכות צריכות להיות מתוכנתות כך שיסכינו עם נזק לבעלי-חיים או לרכוש, אם המשמעות היא שניתן למנוע פגיעה אישית.
7. לא ניתן לבצע סטנדרטיזציה מראש של אירועים יוצאי-דופן המחייבים קבלת החלטות מורכבת מצד הרכב – למשל, מקרה בו הרכב נאלץ לבחור בין דריסת אדם אחד על הכביש, או שניים על המדרכה. מסיבה זו, כל מקרה מסוג זה יעבור ניתוח בפני עצמו על-ידי רשות ציבורית במטרה להפיק לקחים לעתיד.

8. במקרה של תאונה בלתי-נמנעת, הרכב אינו רשאי להבדיל בין בני-אדם לפי מאפיינים אישיים כגיל, מגדר, או חוסן פיזי או נפשי. הוא גם אינו רשאי 'לאזן' קורבנות מצדדים שונים. ניתן להצדיק את תכנות הרכב במטרה לצמצם את מספר הפגיעות האישיות, אך הישויות המעורבות בחישוב הסיכונים אינן רשאיות להקריב ישויות שאינן-מעורבות.
9. במקרה של מערכות נהיגה אוטונומיות, האחריות על הנהיגה עוברת מהנהג האינדיבידואל ליצרני ותפעלי המערכות, ולגופים שאחראיים על קבלת ההחלטות התשתיתיות, המדיניות והחוקיות.
10. יצרני ומפעילי הרכבים האוטונומיים מחויבים לבצע אופטימיזציה של הטכנולוגיות באופן מתמיד, ולשפר גם את המערכות שכבר הגיעו לצרכנים, במידה והדבר אפשרי והגיוני מבחינה טכנולוגית.
11. הציבור זכאי לקבל ידע אודות טכנולוגיות חדשות ופריסתן בשטח. המידע צריך להגיע באופן שקוף ככל האפשר, לעבור לציבור ולהיסקר על-ידי גוף עצמאי ומקצועי.
12. לא ניתן עדיין לומר היום האם בעתיד ניתן יהיה לספק שליטה מרכזית וקישוריות לכל הרכבים בכבישים. מוסריותה של קישוריות מרחוק שכזו עומדת בשאלה במידה והיא תאפשר ניטור מוחלט אחר השימוש בדרכים ותפעול הרכבים מרחוק.
13. נהיגה אוטונומית מוצדקת רק במידה שבה מתקפות אפשריות, ובמיוחד תפעול מרחוק של הרכב האוטונומי או ניצול נקודת תורפה במערכת, אינה גורמת נזק שינתץ את אמונם של אנשים לאורך זמן בתחבורה באופן כללי.
14. הנוסעים ברכב הם היחידים שיכולים להחליט האם המידע המופק מהנסיעה ברכב יועבר לידי ישויות אחרות.
15. חייבת להיות אפשרות להבדיל בבירור בין מצב בו המערכת האוטונומית פועלת לבין מצב בו הנהג האנושי מקבל את האפשרות לאכוף את רצונו על המערכת, ולפיכך נושא באחריות לפעולות הרכב. הממשקים שמאפשרים לנהג האנושי לשלוט ברכב חייבים להיות ברורים והגיוניים.
16. התכנה והטכנולוגיה ברכבים ברמת אוטונומיות גבוהה חייבות להיות מתוכננות כך שהעברת השליטה לנהג האנושי במקרה חירום תתבצע בצורה מהירה ופשוטה. לשם כך צריכות המערכות להתאים עצמן להתנהגות ולרמת התפקוד האנושית הבסיסית, במקום לדרוש מבני-האדם לתפקד ברמה גבוהה.
17. ניתן להתיר מבחינה אתית את קיומן של מערכות המסוגלות ללמוד, בתנאי שהם משפרות את הבטיחות הכולל. אין לפרוס מערכות הלומדות בעצמן אלא אם הן עומדות בדרישות הבטיחות ואינן מפרות את הכללים שהוגדרו במסמך. הוועדה ממליצה לבחון תרחישים אפשריים בגוף ניטרלי, במטרה לפתח סטנדרטים אוניברסליים הולמים.
18. במצב חירום, הרכב חייב להיכנס ל- "מצב בטוח" באופן אוטונומי ומבלי סיוע אנושי. לשם כך יש להגדיר מהו המצב הבטוח, וכיצד נכנס אליו הרכב.
19. יש להטמיע שיעורים בנושא השימוש במערכות אוטונומיות כחלק מהחינוך הדיגיטלי הכללי. בשיעורי נהיגה יש להנחות את התלמידים גם בתפעול הולם של מערכות נהיגה אוטונומיות.

ועדת האתיקה של משרד התחבורה הגרמני שמה דגש על זכויות האינדיבידואל ועל הבטיחות, אך גישה זו אינה היחידה הנהוגה בעולם. חוקר הבינה המלאכותית הסיני קאי-פו לי, למשל, הסביר בספרו AI Superpowers כי סין מתייחסת לרכבים אוטונומיים בראייה תועלתנית יותר, ומתמקדת בתועלת הגדולה לציבור שרכבים אלו אמורים לספק בסופו של דבר. מכיוון שכך, קאי-פו מאמין שסין לא תגביל ניסויים ברכבים אוטונומיים מסיבות בטיחות באותה המידה שמדינות המערב עושות זאת. היא גם לא תגביל את גריפת המידע מהרכבים שתוביל לשיפור היעילות והבטיחות של השירות. סין, למעשה, תהיה מוכנה במוצהר לשלם בחיי אדם בתקופת הפיתוח והלמידה של הרכבים האוטונומיים, מתוך מטרה להביא את המערכות הללו לערים בהקדם האפשרי. הרציונל, לפי קאי-פו, הוא שמהרגע שהרכבים האוטונומיים יהיו בטוחים מספיק, הם יצילו מאות-אלפים בסין מתאונות מדי שנה[17].

קשה להאמין שמדינות אירופה יבחרו בגישה דומה לזו של ממשלת סין בנושא, ועל כן נראה סביר שמדינות שונות כמו סין וגרמניה ימצאו עצמן בשני קצוות מנוגדים של הספקטרום האתי – לפי פרשנותן – בתחום הרכבים האוטונומיים.

בינתיים, קיימות סוגיות אתיות שנוגעות גם באופן תפקודם של הרכבים האוטונומיים במקרי חירום בלתי-צפויים. הוגי דעות מנסים ליישם לעתים פילוסופיות ברורות עבור הרכבים האוטונומיים (כפי שעושות ממשלות גרמניה וסין), אך המלצות לכללים אתיים מגיעות גם מכיוונים אחרים, כגון באמצעות שיתוף הציבור. במחקרים עדכניים, למשל, מוזמן הציבור להצביע על דרך הפעולה האתית ביותר בה צריך הרכב לנקוט במגוון אינסופי של מצבים לא-צפויים[18]. באמצעות שקלול 1.3 מיליון הצבעות שונות, הצליחו עורכי המחקר לפתח מערכת מוסרית שאמורה

להיות מסוגלת להתמודד גם עם דילמות אתיות חדשות על הכביש. עם זאת, המחקר הראה כי המצביעים חלוקים בדעותיהם בנושאים רבים, וכי החלטה מוסרית ואתית בתרבות אחת יכולה להיחשב מגונה בתרבות אחרת[19].

אפילו בקרב הציבור בתרבות אחת ואחידה ניתן למצוא קונפליקטים: אנשים תומכים באופן כללי במוסר תועלתני, לפיו אמורות המכוניות האוטונומיות לחתור להצלת מספר רב ככל האפשר של חיים. בה באותה העת, הם אינם מעוניינים לנסוע ברכב אוטונומי שעלול להקריב את הנוסע על מנת להציל את חייהם של עוברי-אורח[19][20]. ניתן לסכם, לפיכך, שעדיין לא ברור מהם כללי המוסר האוניברסליים לפיהם צריכים הרכבים האוטונומיים לפעול, או אם כאלו קיימים בכלל. מציאתם עשויה להיות אחת מפריצות הדרך החשובות ביותר עבור חדירת הרכבים האוטונומיים לכבישים, אך לפחות בינתיים ברור שהרגולטור הישראלי צריך לשתף פעולה עם אנשי אתיקה, מוסר ומשפט כדי לזהות לפחות את הערכים שיהיו מקובלים על הציבור בישראל.

◀ אתגר הבינה המלאכותית הרמאית

מנועי בינה מלאכותית יכולים לשמש חברות כדרך לרמות את המחוקק ולעקוף את החוקים שנקבעו להגנה על הציבור. כדוגמה לכך, ב- 2015 חשפה הסוכנות להגנת הסביבה בארצות הברית כי חברת פולקסווגן הטמיעה ברכביה אלגוריתם פשוט שנכנס לפעולה ברגע בו נכנס הרכב למבחן שהיה אמור לכמת את רמת פליטות הגזים המזיקים שלו. האלגוריתם הפעיל מנגנונים למניעת זיהום שהיו כבויים באופן רגיל, וכך סייע לרכבים לעבור בהצלחה את מבחני הזיהום האמריקניים[21]. בפועל, בעת נסיעה שגרתית בכביש פלטו הרכבים תחמוצות חנקן וגזי חממה בשיעור הגבוה בעשרות מונים מהמותר בחוק[22]. פולקסווגן נאלצה לשלם מיליארדי דולרים בפיצויים למשלה ולרוכשי הרכבים[23], ומנכ"ל החברה בארה"ב הואשם ברמאות ובקשירת קשר[24]. באמצע 2018 התגלה כי יצרנית הרכבים דיימלר פעלה, לכאורה, באופן דומה בגרמניה, אך התחייבה להחזיר 774,000 כלי-רכב מרצונה החופשי ולתכנת אותם מחדש על מנת שיעמדו בתקני זיהום האוויר הגרמניים[25].

מקרים אלו חושפים סוג חדש של אלגוריתמים – כאלו שנועדו להתמודד עם רגולציות מחמירות באמצעות שינוי זמני של אופן פעולת המכשיר או ספק השירות. אלגוריתמים אלו אינם מוגבלים רק למכוניות. בשנת 2013 התגלה כי גם סמסונג הטמיעה בטלפונים החכמים שלה אלגוריתמים שזיהו אפליקציות שנועדו לבחון את יכולותיהם, ואלו גרמו לטלפון לעבוד ברמה גבוהה יותר למשך הבדיקה בלבד[26]. אלגוריתמים דומים יכולים להופיע גם בשירותים המספקים ייעוץ – בין שמדובר בייעוץ ביטוחי, רפואי, או מכל סוג אחר. מדובר, למעשה, בסוג מיוחד של "אתגר הנציג" שסוקר מוקדם יותר בדו"ח. במקרה זה, 'הנציג' הדיגיטלי מרמה את הרגולטור עצמו, ולא רק את המשתמשים בשירותיו.

מומחים המסקרים פרשיות אלו מאמינים כי הדרך הטובה ביותר – ואולי היחידה – לוודא שפרשיות מסוג זה לא יחזרו על עצמן היא באמצעות דרישה לשקיפות גבוהה מצד החברות בנוגע לאלגוריתמים שהן מטמיעות במוצריהן[27] [28]. מעבר לכך, המומחים ממליצים לאפשר גם לחברות פרטיות ואפילו לחובבנים מהציבור הרחב לנתח את האלגוריתמים הפתוחים המוטמעים במכונות ובשירותים, ואת המידע המגיע מהם, מאחר וממשלות אינן מצליחות לחשוף רמאויות שכאלו במהירות או ביעילות.

◀ אתגר השליטה מרחוק

האינטרנט של הדברים מאפשרת מודלים חדשים של תפעול מוצרים, ואלו מעלים אתגרים חדשים עמם צריכים המשתמשים והמחוקק להתמודד. אחד האתגרים המרכזיים הוא זה של הספק המרוחק. מכשירים חכמים רבים כיום מחוברים לאינטרנט ונעזרים במרכזי עיבוד המידע והבינה המלאכותית שבענן על מנת לבצע את פעולותיהם.

מצב זה אינו בעייתי בפני עצמו, אך הוא חושף את הלקוחות למספר אתגרים שונים. דווקא הצורך בחיבור לאינטרנט אינו מהווה אתגר משמעותי כיום, מכיוון שרוב בתי האב בישראל מחוברים כבר לאינטרנט, או שמסוגלים להתחבר לאינטרנט סלולרי באמצעות טלפונים חכמים. אתגר גדול יותר נובע מהעובדה שהמוצרים הנמכרים ללקוחות הופכים להיות תלויים באופן בלתי-נמנע בחברה המספקת את השירות דרכם. כלומר, במידה והחברה מחליטה לחדול מלספק שירות מסוים דרך מוצריה, הרי שמכשירי הקצה הופכים לחסרי-תועלת בין לילה.

שתי דוגמאות מהשנים האחרונות למקרים מסוג זה הגיעו משני מוצרים שונים. באמצע 2017 התעוררה חמתו של מנכ"ל ומייסד חברת SoftComplex, שגילה שאחד מלקוחות החברה דירג את המוצר האחרון שלה – מכשיר המסוגל לפתוח או לסגור את דלתות המוסך דרך האינטרנט – בכוכב אחד בלבד באמזון. המנכ"ל הזועם הודיע מיד ללקוח המאוכזב שמרגע זה ואילך החברה תמנע ממנו שירות, והמכשיר שרכש לא יורשה להתחבר לשרת המרכזי

של החברה[29]. למעשה, החלטתו של המנכ"ל הביאה לכך שהמוצר שבידי הלקוח הפך לחסר-ערך לחלוטין, ושווה ביכולותיו ללבנה פשוטה. אין פלא שבארה"ב מכונה התופעה בשם Bricking[30].

במקרה אחר מהשנים האחרונות, רכשה חברת Nest (בבעלות אלפאבית), חברה מתחרה שפיתחה ומכרה עוזר ביתי חכם משלה. לאחר הרכישה, הודיעה נסט כי המכשירים של המתחרה-לשעבר יפסיקו לעבוד ויינעלו לשימוש. תגובתו של אחד המשתמשים מבהירה את גודל התרעומת בקרב הציבור מהחלטה זו -

"ב- 15 למאי, ביתי יפסיק לעבוד. תאורת החצר שלי תפסיק להידלק ולהיכבות, אורות האבטחה שלי יפסיקו להגיב לתנועה, ו[מערכת] הרתעת הפורצים תוצרת-בית שלי תפסיק לעבוד. זוהי החלטה מודעת של גוגל/נסט. ... הם אינם פשוט מפסיקים לתמוך במוצר, אלא מייעצים ללקוחות שב- 15 למאי, פחית חומוס תהיה יעילה הרבה יותר..."[31]

בשני המקרים, החברות האמורות נרתעו מכוונותיהן עקב המחאה הציבורית הרחבה כנגד חסימת השירותים[32], אך ברור כי מדובר במקרים ראשונים מני רבים. לא ניתן לצפות מחברות בעלות כוונת-רווח להמשיך לתמוך במכשירים ובשירותים שלא הצליחו 'לספק את הסחורה', והרגולטור יתקשה מאד לחייב אותן לעשות זאת ולשמר אפליקציות או שירותים שאינן מעוניינות לספק עוד. אף על פי כן, על מנת לשמר רמה גבוהה של אימוץ המוצרים והשירותים החדשים, הרגולטור יצטרך למצוא דרך-ביניים שתספק הן את רצון הציבור לקבל תגמול הולם לכספו לאורך זמן, והן את צרכי החברות.

◀ אתגר הפשיעה המרוחקת

"כשהכול מחובר, הכול פגיע" כתב מארק גודמן בספרו "פשעי העתיד"[33]. ואכן, אנו עדים כיום לשפע של ניסיונות פשיעה המתבצעים דרך האינטרנט ונסמכים על הבינה המלאכותית המוגבלת של המכשירים המחוברים לרשת. הדוגמה המפורסמת ביותר היא כמובן תולעת סטוקסנט הידועה, ששיבשה את פעילותם של בקרים חכמים ששלטו על פעולת הצנטריפוגות באיראן[34]. מאז סטוקסנט התרבו מקרי הפשיעה המערבים השתלטות מרחוק על בקרים ומנועי בינה מלאכותית.

באחד המקרים בעלי הפוטנציאל הנפיץ ביותר, התבצעה ב- 2017 מתקפת סייבר על הבקרים החכמים שבמפעל פטרוכימי בערב הסעודית. המתקפה הייתה אמורה לפגוע פיזית במפעל ולגרום לפיצוץ מכוון. התוקפים שיבשו מרחוק את פעולת הבקרים ששלטו על הלחץ והטמפרטורה במתקנים שבמפעל. בקרים דומים פועלים בכמעט 20,000 מפעלים מסביב לעולם, כולל תחנות כוח גרעיניות והתקנים לטיהור מים[35]. ברור מדוע קיים חשש גדול לפשיעה ולתקיפות טרור מרוחקות כבר בעתיד הקרוב.

הבקרים שהותקפו בשני המקרים המתוארים לא השתייכו למערכות הבינה המלאכותית מהדור החדש. עם זאת, אנו רואים כי דווקא הקניית בינה מלאכותית למכשירים יכולה לפתוח אותם לפריצות בדרכים חדשות ויוצאות-דופן. כך, למשל, בשנת 2017 העלתה חברת המזון המהיר בורגר קינג פרסומת בה נשמע המשפט – "אוקי גוגל, מהו הווופר בורגר?"

משפט תמים-לכאורה זה גרם להפעלתו של העוזר הביתי של גוגל במיליוני בתים. העוזר הביתי בדק בוויקיפדיה כיצד מוגדר וופר בורגר, ואז הקריא את ההגדרה בקול רם לכל הדיירים. תרגיל שיווקי זה גרם לכך שבורגר קינג הואשמה בהפרת החוק, מאחר והיא 'השתלטה' על מכשירים פרטיים בבתיהם של מיליוני אמריקנים[36]. לא ברור האם בורגר קינג אכן הפרה את החוק, אך המקרה מדגים כיצד ניתן לתפעל את הבינות המלאכותיות בדרכים לא-שגרתיות למטרת הפקת רווחים.

עדיין לא ברור לגמרי מהו דינה של פשיעה מרוחקת כאשר היא מבוצעת בדרכים לא-קונבנציונליות. בורגר קינג לא הועמדו לדין על ה- 'פריצה' שביצעו, למשל, אך ברור שלא ניתן לעבור על סדר היום במקרים אחרים. כך, למשל, מתרבים כיום אביזרי המין החכמים המחוברים לרשת, ורבים מהם ניתנים לפריצה בקלות יחסית[37]. האם אדם המשתלט על אביזר מין מרחוק, לפיכך, יכול להיות מואשם באונס? לחלופין, האם אדם המשתלט על רובוט ביתי, יכול להיות מואשם בפריצה? אלו שאלות שהמחוקק יאלץ להתמודד עמן יותר ויותר, ככל שהמכשירים החכמים מתפשטים בחברה.

◀ אתגר התיאום-ללא-הכוונה

מערכות אוטונומיות הפועלות בסביבה אנושית ימצאו עצמן מחויבות לשתף פעולה – בין שעם בני-אדם, או עם מערכות אוטונומיות אחרות. בעוד שהתקשורת עם בני-אדם הינה פשוטה וצפויה יחסית (על אף שגם בה ניתן למצוא אתגרים עבור המכונות הלומדות, כפי שסקרנו ב- "אתגר הסיבתיות"), התקשורת עם אלגוריתמים אוטונומיים ולומדים אחרים יכולה להוביל לשגיאות גדולות.

דוגמה לטעות פשוטה, בה שני אלגוריתמים בסיסיים למדו אחד מהשני, הופיעה באמזון ב- 2011 וגרמה למחירו של ספר מדעי – "The Making of a Fly" – להאמיר ל- 23 מיליון דולרים ו- 93 סנט, ועוד הוצאות משלוח. הסיבה לעלייה הדרמטית במחיר הספר הייתה ששני אלגוריתמים נכנסו למלחמת-מחירים זה עם זה. האלגוריתמים עדכנו באופן אוטומטי את מחיר הספר מדי יום, כאשר כל אחד מהם הסתמך על השני: האחד קבע את מחיר הספר כ- 0.9983 מהמחיר שהשני הציע, והשני קבע את המחיר כ- 1.2705 מהצעתו של הראשון. מערכת אילוצים זו הביאה לכך שמחיר הספר זינק מעלה ב- 27 אחוזים בערך מדי יום, עד לתוצאה הלא-הגיונית בעליל[38].

למרות שמערכת האילוצים שתיארנו נשמעת מוזרה, חשוב להבין שהיא נראתה הגיונית לשני המוכרים בעת קביעתה. המוכר הראשון רצה להתחרות בהצלחה עם המוכר השני, ולכן דרש מהאלגוריתם להציע מחיר נמוך יותר מכל הצעה של השני. המוכר השני, לעומתו, הסתמך כנראה על המוניטין המרשים יותר שלו וקיווה שהרוכשים הפוטנציאליים יהיו מוכנים לשלם מחיר גבוה יותר בעטיו. לכן, הוא הרשה לעצמו לדרוש מהאלגוריתם לקבוע מחיר גבוה יותר מכל הצעה של המוכר הראשון. וכך, שני האלגוריתמים נכנסו ללולאה של העלאת מחירים, שרק התערבות אנושית הייתה יכולה לעצור.

דוגמה אחרת, שתפסה יותר את דמיון הציבור, הגיעה משני צ'אט-בוטים של פייסבוק שדיברו אחד עם השני במעבדות המחקר של החברה. האלגוריתמים – שהיו מתוחכמים בהרבה מאלגוריתמי קביעת המחירים של אמזון – היו אמורים להתאמן אחד מול השני במשא ומתן. האלגוריתמים ניהלו שיחות ארוכות אודות ערכם הסובייקטיבי של כמויות שונות של כדורים, ספרים וכובעים, וניסו למצוא את הדרך המיטבית לחלק את השלל ביניהם[39]. הם תוכנתו להגיב אחד לשני וללמוד זה מזה, אך החוקרים לא סיפקו להם תמריץ חיובי להמשיך להתדיין באנגלית, וכך קרה שהאלגוריתמים עברו לפתח 'שפה פרטית' משל עצמם. באמצעות שפה זו הם הצליחו להעביר אחד לשני את דרישותיהם באופן יעיל יותר מכפי שהשפה האנגלית אפשרה להם[40].

בסופו של דבר החליט צוות המחקר של פייסבוק לקטוע את השיחה בין שני הבוטים, מכיוון שהם היו אמורים ללמוד להתדיין עם בני-אדם באנגלית, ולא אחד עם השני באופן שאינו מובן לאנשים מן השורה. המדיה חגגה על הבשורה עם כותרות סנסציוניות אודות הבינה המלאכותית של פייסבוק שיצאה משליטה ותכבוש בקרוב את העולם, אך האמת פשוטה יותר: מדובר רק בשני אלגוריתמים שביצעו בדיוק את מה שהיו אמורים לעשות. הצרה החלה כאשר הם למדו אחד מהשני מבלי שיהיו 'תנאי סף' הולמים, או פיקוח אנושי מתמיד בכל שלב בתהליך הלמידה העצמית.

בעולם בו בינה מלאכותית תהיה נוכחת בכל חלק מחיינו, לא ניתן יהיה לספק פיקוח אנושי עבור כל אלגוריתם בפני עצמו – על אף שאלגוריתמים רבים ילמדו זה מזה, ויפיקו תובנות שאינן תמיד תואמות לכוונת המתכנתים המקורית. מכיוון שכך, המחוקק צריך לחייב את חברות הבינה המלאכותית להגביל את יכולות הלמידה העצמאית של האלגוריתמים שהן מפתחות, ליצור תנאי סף הולמים לתהליך הלמידה, ולפתח מנגנוני בטיחות שיוודאו שהבינות המלאכותיות נותרות בתלם – אולי באמצעות אלגוריתמים-חוקרים הפועלים במקביל.

3. אתגרים בהשפעה על המסחר והכלכלה

לאלגוריתמי בינה מלאכותית צפויה להיות השפעה גדולה על המסחר והכלכלה בישראל. בחלק זה נבחן שלוש דוגמאות מרכזיות להשתלבותם הצפויה של אלגוריתמים בכלכלה הישראלית: בהשקעות וניתוב כספים מיטבי, במתן הלוואות ובמסחר בבורסה.

◄ בינה מלאכותית בתחום ההשקעות וניתוב כספים

מאז ומתמיד ניסו ממשלות וגופים פיננסיים לפענח מהם התחומים המתאימים ביותר להשקיע בהם את כספם. הבינה המלאכותית מתחילה לסייע כיום בתחומים אלו. כך, למשל, חברת נטפליקס נעזרה בבינה מלאכותית לניתוח המידע הגדול שאספה מהצופים בנוגע להעדפותיהם. האלגוריתמים זיהו כי הצופים מעוניינים לצפות בסדרות פוליטיות עם שחקנים ומפיקים מסוימים – וכך החליטה נטפליקס להשקיע בהפקת סדרת הטלוויזיה "בית הקלפים" שמילאה אחר כל התנאים הללו וזכתה להצלחה גדולה[41].

חברות וגופים אחרים, כגון חברת טלפוניקה, WR Hambrecht Ventures, גוגל ונצ'רס[42] וקורליישן ונצ'רס [43], נעזרים באלגוריתמים לבחירת חברות הזנק בהם ישקיעו, במידה מרשימה של הצלחה, לפי דיווחי החברות עצמן[44]. כפי שניתן להבין, האלגוריתמים אינם שקופים ורב הנסתר על הגלוי, אך נראה שהם בוחנים פרמטרים

הקשורים לחברות הזנק מחד, ופרמטרים הקשורים לסביבה הכללית, כגון מצב השוק והמתחרים. סביר להניח ששילוב של קבלת החלטות אנושי עם בינה מלאכותית יוכל להפיק תשואות גבוהות יותר מהרגיל. במידה ואכן כך יקרה, שוק ההשקעות יאמץ במהירות את הטכנולוגיה החדשה לחיקו.

טכנולוגיות אלו פותחות מספר אתגרים חדשים שהמחוקק עשוי למצוא עצמו נאלץ להתמודד עמם. אלגוריתמים, למשל, עלולים לתת משקל לגורמים הסותרים את ערך השוויון, ולמזער את סיכוייהם של בני מיעוטים לקבלת השקעות. בדרך זו ינציחו האלגוריתמים פערים חברתיים ומגדריים גם בתעשייה[45]. אתגר אחר שהמחוקק עשוי להידרש אליו הוא אובדן ההטרוגניות בהשקעות. מדינות כיום מעודדות את קיומם של שווקי השקעות מאחר והם תומכים בחברות רבות ושונות. חלק מהחברות יצליחו, ואחרות ייכשלו, אבל בראייה כוללת יותר הכלכלה כולה מרוויחה. ברם, אם אלגוריתמים יוכלו להצביע ברמת דיוק גבוהה על נתיבי השקעה מועדפים, ברור ששוק ההשקעות יתמקד בהם, וכך לא יכוסו נתיבים אחרים עם סיכויי הצלחה נמוכים יותר – אך כאלו שיכולים לשנות את העולם לטובה במידה ויצליחו למרות הסיכוי הנמוך.

◀ בינה מלאכותית בשוק ההלוואות

ניתן למצוא בעולם חברות רבות המשתמשות בבינה מלאכותית כדרך לקביעת רמת הסיכון של לווים פוטנציאליים, ובהתאם לכך – מחליטות האם לתת להם הלוואה ומהם אחוזי הריבית המתאימים ביותר. אפשר לראות את הפוטנציאל של האלגוריתמים הללו ממומש בחברות כמו Smart Finance, שפיתחה אפליקציה דרכה מתבצעות מיליוני הלוואות קטנות באופן אוטומטי. האפליקציה מבקשת גישה לחלק מהמידע בסמארטפון של מבקש ההלוואה. האלגוריתמים מבוססי הלמידה העמוקה שלה בוחנים 1,200 נקודות מידע, כגון מהירות ההקלדה של תאריך הלידה באפליקציה, אחוזי הסוללה שנותרה בטלפון, או את היום בשבוע. כל נקודת מידע כזו בפני עצמה מספקת אינדיקטור חלש בנוגע לסיכוייו של הלווה להחזיר את הכסף בזמן. אישור ההלוואה מגיע דרך האפליקציה תוך שמונה שניות או פחות, באופן אוטומטי לגמרי[46]. ב- 2017 ביצעה החברה שתי מיליון הלוואות בחודש, עם אחוזי החזר הלוואות מהגבוהים ביותר בשוק, לפי דיווחה[17].

חברות אחרות מסתמכות על פרמטרים אחרים, אך כולן מסתמכות על "אינדיקטורים חלשים" שבני-אדם לא היו שמים אליהם לב באופן רגיל, אך ביחד מספקים כוח חיזוי גדול יותר לחברות. במאמר שפרסמה סיסקו הובאו מספר אינדיקטורים חלשים ומורכבים אחרים, לפיהם למשל, סוחרים ברשת שמוכנים לשלוח את מרכולתם ללקוחות מקוונים בקליפורניה, הינם בעלי סיכוי גדול יותר להחזיר הלוואות שיקבלו[47].

חברות ההלוואות מתחילות להסתמך גם על מידע המגיע ישירות מהרשת החברתית. פייסבוק, למשל, מסוגלת כיום להגדיר דירוג אשראי למשתמשים שונים, לפיו ניתן יהיה לקבוע את סיכויו של כל משתמש להחזיר הלוואות[48].

באמצעות השימוש בבינה מלאכותית כ- 'סוכן הלוואות', יכולות חברות מתקדמות לפתח מודלים עסקיים חדשים. כך, חברת Lending Club מאפשרת למיליוני אנשים להשקיע את כספם באמצעות מתן מיקרו-הלוואות זעירות, שעומדות על מאות או אפילו עשרות דולרים. מבקשי ההלוואות מאופיינים באמצעות אלגוריתמים, המייחסים לכל אחד מהם רמת סיכון (וריבית) שונה – ונותני ההלוואות יכולים לבחור את רמת הסיכון בה יבחרו להשקיע[49].

על אף שמדובר כיום בטכנולוגיה מפציעה ((Emerging technology, ברור שגם בנקים ייאלצו לעבור להסתמך על אלגוריתמים מתוחכמים במתן הלוואות. מכיוון שכך, המחוקק יידרש להתמודד עם האופן בו ינותחו מקבלי ההלוואות, ויצטרך להחליט אלו פרמטרים ראויים להיבחן, ואלו לא. האם, למשל, יכול אלגוריתם להתייחס למגדר של מקבל ההלוואה או לצבע עורו? לכאורה, התייחסות לאינדיקטורים אלו מעוררת חשש לקבלת החלטות מוטה, על סמך פרמטרים שאינם בשליטתו של מקבל ההלוואה.

אתגר גדול אחר בתחום זה הוא בפרטיות של מקבל ההלוואה. בסין, בה רווחות אפליקציות למתן מיקרו-הלוואות, מקובל כי האפליקציות מבקשות גישה למידע שבטלפון החכם על מנת לקבוע את רמת הסיכון של מבקש ההלוואה. הלווים מגלים בשנים האחרונות כי במידה ואינם מצליחים להחזיר את ההלוואה בזמן, החברות אינן מהססות להשתמש בפרטיהם האישיים – למשל, במספרי הטלפון של הוריהם וקרובי-משפחתם – כדי לדרוש בשלום הלווה. ברור כי חדירה לפרטיות מסוג זה אינה ראויה, ויש מקום למחוקק להתערב על מנת למנוע ממנה מלהופיע גם בישראל[46].

◀ בינה מלאכותית למסחר בבורסה

בעשור האחרון הפכו שווקי הבורסאות לנחלתם של אלגוריתמים מתוחכמים. קיימות הערכות לפיהן עד לשבעים אחוזים מכל רכישות ומכירות המניות בבורסה מתבצעות באמצעות אלגוריתמים[50]. האלגוריתמים מגיבים לכל

מאורע בבורסה בפרקי-זמן הנאמדים בננו-שניות – משך זמן שקצר פי יותר ממיליון מיכולתו של אדם לקרוא אפילו משפט אחד בדו"ח זה, על אחת כמה וכמה לקבל החלטה לפיו[51]. האלגוריתמים מתחרים באלגוריתמים אחרים בניסיון להגדיל רווחים, ואף מנסים לעתים לתעתע במתחריהם בשיטות שונות, למשל באמצעות שיגור הודעות כוזבות על רכישת מניות.

בימים הראשונים של השימוש באלגוריתמים, גרפו החברות שהפעילו אותם רווחים גדולים. כל עסקה הניבה אמנם רווח זעום, אך האלגוריתמים ביצעו יותר מאלף עסקאות בשנייה. עם זאת, ככל שמספר הסוחרים האנושיים בבורסה יורד, ומספר האלגוריתמים עולה, מתקשות יותר החברות להפיק רווחים מנתיב פעולה זה[52].

קיימים אתגרים רבים העומדים בפני המחוקק בתחום זה. האם, למשל, ראוי שאלגוריתמים המופעלים על-ידי חברות גדולות ינהלו את המסחר בבורסה, במקום אינדיבידואלים שרוצים להשקיע מכספם? התוצאה של מגמה זו ברורה: הסוחרים האנושיים עוברים להשקיע ולסחור באגרות חוב במקום במניות. המחוקק יצטרך להחליט האם זוהי תופעה רצויה.

אתגר אחר הוא החוסר בשקיפות של האלגוריתמים המופעלים בבורסה. אלגוריתמים אלו פועלים במהירות עצומה ויכולים ליצור תוך שניות בועות-ענק ולפוצצן, או לשתף פעולה זה עם זה בדרכים שלא אופיינו או תוכננו מראש. אירועים שכאלו יכולים להוביל להפסדים של מיליארדי דולרים בבורסה תוך דקות ספורות, כפי שאירע כאשר באג בפעולתו של האלגוריתם שהפעילה חברת Knight Capital גרם למנוע הבינה המלאכותית לרכוש בטעות מניות בשווי כולל של שבעה מיליארד דולרים תוך פחות משעה[53]. המחוקק יצטרך להחליט כיצד להתמודד לפיכך עם מנועי הבינה המלאכותית הפועלים בבורסה וכיצד לבקרם על מנת לשמר את אמון הציבור בבורסה ולצמצם את הסיכוי למקרים שיובילו להתמוטטותה.

אתגר שלישי הוא בהגדרה מחדש של "פעילות לא הוגנת" בבורסה. בארה"ב אסר המחוקק על חברות המסחר האלגוריתמי למקם את מחשביהן בבנייני הבורסות, מכיוון שהדבר היה מקצר את משך הזמן עד שהמחשבים היו מקבלים את המידע ופועלים לפיו – דבר שלפי הרגולטור תאם להגדרה של קבלת מידע מוקדם, האסורה לפי החוק[54]. אלגוריתמים מסוגלים גם להשפיע על מחיריהן של מניות מסוימות, בדומה לדרך בה עושים זאת במכוון סוחרי בורסה אנושיים. מדובר באקט לא-חוקי בבירור כאשר הוא מתבצע על-ידי בני-אדם, אך המחוקק יצטרך לקבוע מה דינם של אלגוריתמים הנוקטים בפעילות דומה.

אתגר רביעי ואחרון (לעת עתה) עוסק בשאלת ההוגנות של המסחר האלגוריתמי. המסחר האלגוריתמי מבוצע בעיקר על-ידי חברות גדולות, הדוחקות את רגליהם של הסוחרים האנושיים אל מחוץ לשוק. במצב עניינים זה עלולה להתעורר תרעומת הן מצד הציבור והן מצד הסוחרים האנושיים, על המסחר האלגוריתמי המאפשר לעשירים להתעשר יותר – בעוד שכל היתר אינם מרוויחים מהשינויים בבורסה.

למרות כל האתגרים הללו, יש להבהיר כי למסחר האלגוריתמי יש גם חשיבות (על אף שקשה לכמת אותה) בהגברת הנזילות בבורסה. כמו כן, הרגולטור יכול לעשות שימוש באלגוריתמים מסוג זה בעצמו לתועלת הציבור. בארה"ב, למשל, החליטה הוועדה לניירות ערך ולבורסות (SEC) לשתף פעולה עם חברת Tradeworx, ולנצל את יכולת איסוף הנתונים של החברה כדי לגלות מסחר אסור בבורסה ולפעול כנגדו[52].

4. מדיניות של מדינות

◄ GDPR

ה- GDPR – רגולציית הגנת המידע הכללית של האיחוד האירופי – נכנס לתוקף ב- 2018 וקבע מחדש את הדרך בה מתנהלות חברות תעשייתיות מול האזרחים מבחינת המידע שהן אוגרות אודותיהם[55]. לא ניתן לדבר על מדיניות השימוש במידע מבלי להתייחס ל- GDPR ולנסות ללמוד ממנו לקראת חקיקה דומה בישראל. על אף שה- GDPR נוגע רק במידע, חברות המפתחות בינה מלאכותית יושפעו באופן משמעותי מכל חקיקה בנושאי מידע, מאחר והבינות המלאכותיות החדשות לומדות באמצעות הרצה על מאגרי מידע.

חברות שעוברות על חוקי ה- GDPR באיחוד האירופי יספגו קנסות משמעותיים – עד לעשרים מיליון אירו או ארבעה אחוזים מהכנסותיהן העולמיות. אין פלא שחברות דיגיטליות מכל העולם שינו את התנהלותן כך שתתאים לכללים החדשים באירופה.

ה- GDPR מורכב ממספר סעיפים מרכזיים, שעל החשובים ביותר ביניהם למטרות העבודה הנוכחית נעבור בקצרה בחלק זה של הדו"ח.

סוגי מידע

מכיוון שמטרתו המרכזית של ה- GDPR היא לספק זכויות והגנה גדולה יותר לאינדיבידואלים, הוא מפרק את המידע הנאסף אודות אינדיבידואלים לשני סוגים מרכזיים. הראשון הוא מידע אישי – שיכול לשמש לזיהוי אינדיבידואלים. השני הוא מידע אישי רגיש כמידע גנטי, נטיות מיניות והשקפות פוליטיות של אינדיבידואלים.

בסיסים לעיבוד מידע

ארגונים מורשים לעבד מידע אישי בתנאי שקיבלו אישור מפורש לכך מהאינדיבידואל שהמידע מתייחס אליו, או כאשר יש צורך לעבד את המידע. הצורך יכול להיות חוקי (למשל, כדי לציית לבית המשפט המחייב את החברה לספק את המידע, או לרשות ממשלתית רשמית), הגנתי (למשל, כדי להגן על האינטרסים של האינדיבידואל כאשר הוא אינו מסוגל לספק אישור בעצמו), או חיוני להצעת ומימוש חוזה עם מושא המידע. במידה והארגון מעבד מידע על אינדיבידואלים לפי צורך אך מבלי אישורם המפורש, הוא מחויב ליידע אותם בנוגע לכך.

זכויות האינדיבידואלים

ארגונים חייבים לספק לאינדיבידואלים מידע אודות הדרך בה הנתונים שלהם יעובדו. הפירוט צריך להיות ברור, קצר ונהיר לכל – במיוחד כאשר הוא מופנה לילדים. כאשר הארגון מנסה לקבל מידע ישירות מהאינדיבידואל, הוא חייב לספק הסבר בנוגע לסיבה לבקשה למידע, ומה יהיו ההשלכות במידה והאינדיבידואל יסרב לספק את המידע.

ה- GDPR מאפשר לכל אדם להגיש בחינם בקשה מיוחדת – SAR (Subject Access Request) – לחברה או לארגון שאוספים מידע אודותיו. הארגונים המקבלים את הבקשה חייבים לספק את המידע שנאסף על האינדיבידואל תוך חודש אחד. תקנה זו מוודאת שחברות טכנולוגיה גדולות וקטנות ייאלצו לספק שליטה גדולה יותר למשתמשים על המידע שברשותן. במצבים מסוימים, יכולים האינדיבידואלים לדרוש אפילו שהמידע האישי שלהם יימחק ממאגרי המידע של החברות.

תקנה חשובה במיוחד היא שהחברות מחויבות לספק לכל אינדיבידואל את המידע – לפי דרישה – בפורמט נפוץ בתעשייה. בדרך זו, אינדיבידואלים יוכלו לעבור מחברה לחברה תוך שהם 'נושאים על עצמם' את המידע שלהם, ומבלי שיהיו כבולים לחברה זו או אחרת.

החלטה רלוונטית במיוחד של ה- GDPR לנושא הדו"ח הנוכחי היא שלאנשים תהיה הזכות לדרוש שלא יהיו כפופים להחלטות אוטומטיות של המכונות אודותיהם, במיוחד אם לאלו יש השפעה גדולה עליהם. המשתמשים חייבים לקבל הסבר בנוגע לכל החלטה שהתקבלה עבורם.

אחריות ארגונית

במסגרת ה- GDPR, חברות אחראיות לטיפול במידע האישי של המשתמשים. הן מחויבות להגן על המידע, להעריך את השפעת המידע על האינדיבידואל במידה וישתחרר, ולתעד את דרך עיבוד המידע. עליהן ליישם את העקרונות הבסיסיים של הגנת מידע, כאנונימיזציה במידת האפשר וצמצום המידע השמור. החברות חייבות לדווח לרשויות על כל מקרה של "הרס, אובדן, שינוי, גישה או פרסום לא-מאושרים" של המידע, במידה ויכולה להיות השפעה שלילית מכל סוג שהוא (פיננסית, פגיעה במוניטין ועוד) על האנשים המצוינים במידע.

בחברות עם יותר מ- 250 עובדים, יש חובה לשמור תיעוד של הסיבה לשמירה על המידע האישי, לתאר את המידע שנאגר, את משך הזמן בו הוא נאגר, ותיאור של דרכי ההגנה על המידע. בחברות בהן מתבצע ניטור תדיר וסיסטמטי של אינדיבידואלים בקנה מידה נרחב, או שמעבדות כמות גדולה של מידע אישי רגיש, יש לפתוח משרת "קצין הגנת מידע".

◀ בריטניה

באמצע שנת 2017 פרסמה הוועדה הנבחרת בנושא בינה מלאכותית דו"ח – "בינה מלאכותית בממלכה המאוחדת: מוכנה, נכונה ומסוגלת?"[56]. בדו"ח דנה הוועדה בבינה המלאכותית והשלכותיה על המדינה, והגיעה למספר מסקנות בנושאי אתיקה ופרטיות הרלוונטיים לפיתוח בינות מלאכותיות, כדלקמן –

א. הבדלה בין נתונים למידע אישי

הבדלה בין נתונים (data) למידע אישי (personal data). "נתונים" הם כמעט כל סוג מידע שנאגר במחשב או שמיועד להיאגר במחשב. "מידע אישי", לעומת זאת, מתייחס באופן כללי למידע שקשור לאינדיבידואלים. איסוף נתונים בקנה מידה נרחב יספק כוח גדול לחברות ולארגונים, אך שאלות של פרטיות ובעלות על מידע מתייחסות בדרך כלל למידע אישי.

ב. חשיבותה של יכולת העברת המידע (data portability)

הוועדה הסכימה בנוגע לחשיבותה של יכולת העברת המידע (data portability). תקנה זו, שנקבעה על-ידי הרגולציה הכללית להגנה על המידע (GDPR) באיחוד האירופי, מחייבת חברות וארגונים לספק למשתמש את המידע האישי שלו באופן מסודר וברור, ללא תשלום. המטרה היא שהצרכנים יוכלו לקחת 'על גופם' את המידע האישי שלהם ולהעבירו משירות אחד למשנהו. בדרך זו אמורה התקנה למנוע מצב בו הצרכנים יישארו כבולים לספק שירות אחד מסוים.

ג. הקושי באנונימיזציה

גופים רבים עורכים את מסדי הנתונים שברשותם ומסירים פרטים מזהים של אינדיבידואלים כחלק מתהליך "אנונימיזציה" של מסד הנתונים. בדרך זו, למשל, מסד נתונים המכיל אלפי צילומי רנטגן לא יכלול שמות, כתובות או מאפיינים אחרים שיכולים להוביל לזיהוי האינדיבידואלים המצולמים. עם זאת, בינה מלאכותית מתקדמת יכולה לשמש לזיהוי-מחדש של אנשים ביעילות גבוהה, ולסתור אפילו שיטות אנונימיזציה מתקדמות. במטרה להתמודד עם הבעיה, חוק הגנת המידע הבריטי מגדיר פעילות זיהוי-מחדש שכזו כפלילית. הוועדה מרוצה, באופן כללי, מתהליכי האנונימיזציה הקיימים, אף שהיא מכירה בכך שלעולם לא יהיו מושלמים.

ד. הקושי בשליטה במידע

הוועדה סקרה עדים רבים בנושא השליטה במידע, וקיבלה שפע של דעות סותרות ומנוגדות. חלקם, כנציגי החברות הגדולות, טענו שהמצב הקיים מספק ברובו, ושאינדיבידואלים מקבלים פיצוי על השימוש במידע – בין שבאופן ישיר, או באופן לא-ישיר, מכיוון שהמידע מאפשר לחברות לפתח המצאות חדשות שייטיבו בסופו של דבר עם האוכלוסייה כולה. הדוגלים בעמדה זו הדגישו כי נתונים "אינם משאב סופי", וכי למונח "מונופולי נתונים" אין משמעות של ממש. לטעמם, למידע האישי אין משמעות עבור האינדיבידואל האוסף אותו, אלא הוא זוכה למשמעות רק כאשר הוא מרוכז ומנותח ביחד עם מידע שנאסף ממקורות נוספים רבים. המידע, לפיהם, אינו הדבר החשוב ביותר, אלא דווקא המיומנות שחברות רוכשות בעיבודו וביישום התובנות המתקבלות ממנו.

הדעה השנייה תמכה בקידום מאגרי מידע הזמינים ופתוחים לציבור, במטרה להתמודד עם הסכנות הכרוכות בהופעתם של מונופולי מידע. לפי עמדה זו, כמות גדולה ככל האפשר של מידע צריכה להיות זמינה לכל לשימוש. חלק מהעדים הוסיפו כי כל המידע שנאסף במסגרת פרויקט מחקר ציבורי הינו "משאב ציבורי וצריך להיות זמין בהגבלות מינימליות ככל האפשר". מספר עדים חשו שבנושא זה במיוחד, הממשלה יכולה וצריכה להתערב על מנת לעודד פיתוח סטנדרטים הקשורים במידע פתוח, ולקדם פתיחה של מאגרי מידע ציבוריים לציבור הרחב.

הדעה השלישית קידמה שיתופי פעולה והעברת מידע מרשויות ממשלתיות לחברות בתעשייה, אך תחת כללי התנהלות ברורים ומסודרים שיקבעו 'מחיר' לכל נתון ולכל פיסת מידע המועברת לתעשייה. הציבור, במילים אחרות, רשאי לדרוש להרוויח על הנתונים שהוא מעביר לחברות, אך התגמול לא יגיע לאינדיבידואלים אלא למדינה בכללותה.

הדעה הרביעית והאחרונה קידמה בעלות פרטית על המידע. לפי התומכים בעמדה זו, לאינדיבידואלים צריכה להיות בעלות על המידע האישי שלהם, והם צריכים להיות אחראים בנוגע לשימוש בו ולאופן אגירתו. עמדה זו נתמכת ביוזמות כ- The Hub of All Things, שעוזרת לפתח מאגרי נתונים אישיים מבוזרים, הנשלטים על ידי אינדיבידואלים שיכולים לצפות בכל המידע האישי שלהם במקום אחד, להבין כיצד נעשה בו שימוש, ולמכור או לסחור בו בתמורה לכסף ולהטבות[57].

ברור שקיים קושי גדול במציאת האיזון בין מגוון הדעות המנוגדות והסותרות שהוצגו לוועדה. מסיבה זו הצהירה הוועדה כי "קיים צורך דחוף בבהירות קונספטואלית בנושא הנתונים, אם נרצה למנוע מבלבול בתחום זה לעכב את התפתחות הבינה המלאכותית."

ה. המלצות בתחום המידע

ניתן לסכם את המלצות הוועדה בתחום המידע בקצרה במספר נקודות מרכזיות –

- פיתוח נאמנויות מידע: הוועדה ממליצה לפתח "נאמנויות מידע" (Data trusts) שתפקידן יהיה לפקח על העברת מידע אתית בין ארגונים. הוועדה ממליצה כי נאמנויות המידע הללו יכללו אפשרות לייצוג האנשים שהמידע שלהם נשמר – בין שבאמצעות התייעצות עמם או עם נציגיהם, או בדרכים אחרות.

- פתיחת מידע לחוקרי ומפתחי בינה מלאכותית: הוועדה ממליצה כי בכל הזדמנות, וכאשר הדבר מתאים, חוקרי ומפתחי בינה מלאכותית יקבלו גישה למידע ציבורי, לאחר אנונימיזציה ראויה.
- קיים צורך במנגנונים חוקיים וטכניים חדשים: הוועדה מכירה בכך שלא ניתן להסתפק בפתיחת מידע לציבור כדרך להפיכת המידע לזמין ולשימושי, בעיקר כאשר מדובר במידע רגיש או יקר-ערך. משום כך יש צורך בפיתוח מנגנונים חוקיים וטכניים חדשים לחיזוק השליטה האישית במידע ולשמירה על הפרטיות. יוזמות מסוימות, כ- Open Banking, עשויות לספק מענה ראשוני למנגנונים שכאלו, וראוי לבחון אותן בקפידה.
- יש להבין את הערך שבמידע הציבורי: למידע ציבורי יש ערך רב, והוועדה ממליצה להבהיר לגופים הציבוריים את הערך שבמידע שברשותם. הוועדה מציעה לסייע לגופים הציבוריים להעריך את המידע שברשותם, על מנת שיוכלו להפיק ממנו את המיטב ולהיכנס למשא ומתן הוגן ומבוסס-ראיות עם גופים מהמגזר הפרטי.

ו. הצורך בבינה מלאכותית מובנת

הוועדה עמדה על הקושי בפיתוח בינה מלאכותית 'מובנת' – כלומר, כזו שניתן להבין מדוע וכיצד קיבלה החלטות מסוימות. במילות הוועדה –

"אנו מאמינים שקשה, ואולי אפילו בלתי אפשרי, להגיע לשקיפות טכנית מלאה במערכות בינה מלאכותית מסוגים מסוימים הנמצאות בשימוש כיום, ובכל מקרה שקיפות שכזו לא תסייע או תתאים במקרים רבים."

יחד עם זאת, הוועדה הכירה בקיומם של מצבים מסוימים בהם הבטיחות חשובה במיוחד, ושם יש צורך דחוף בשקיפות טכנית ואלגוריתמית. הוועדה ממליצה לרגולטורים בתחומים אלו לחייב את השימוש במערכות בינה מלאכותית שקופות יותר, אפילו אם אלו באות על חשבון יכולות ודיוק גבוהים יותר.

במידה ולא ניתן לפתח מערכות בינה מלאכותית שקופות, הוועדה מוכנה להסתפק גם במערכות "מסבירות", המסוגלות להסביר את המידע עליו הסתמכו לקבלת החלטה מסוימת, ואת ההיגיון שהנחה אותן בקבלת ההחלטה.

הוועדה מאמינה שחובה לפתח מערכות בינה מלאכותית מובנות או בעלות יכולת הסברה עצמית, אם ברצוננו להפוך את הבינה המלאכותית לכלי אינטגרלי ואמין בחברה. הוועדה מאמינה שבינה מלאכותית "מסבירה" תהיה נוחה ומתאימה יותר לשימוש בידי הצרכנים והאזרחים. במידה ומערכות בינה מלאכותית אינן יכולות לספק הסבר מלא ומקיף אודות ההחלטות שקיבלו, הוועדה סבורה שאין מקום לפרוס אותן במצבים בהם יכולה להיות להם השפעה משמעותית על חייו של אינדיבידואל. במקרה של רשתות עצבים מלאכותיות, שאינן מסוגלות עדיין לספק הסברים בנוגע להחלטות שקיבלו, ייתכן שהמשמעות היא שפריסתן תתעכב בתחומים מסוימים עד שיימצאו פתרונות טובים יותר.

הוועדה הנחתה את הגופים הציבוריים המתאימים בבריטניה להפיק הנחיות אודות הצורך במערכות בינה מלאכותית מובנות. מהחברות המפתחות בינה מלאכותית מצופה לאמץ את ההנחיות הללו ולקבוע סטנדרטים משותפים שיהיו רלוונטיים לתחומים בהן הן פועלות.

ז. התמודדות עם הטיות

הוועדה מוטרדת מכך שמאגרי מידע רבים המשמשים לאימון מערכות הבינה המלאכותית כיום, מספקים ייצוג חסר של האוכלוסייה. מערכות בינה מלאכותית שלומדות מהמאגרים הללו, עלולות לקבל החלטות קלוקלות המשקפות את חוסר-הצדק של חברות ותרבויות בעבר ובהווה. חוקרים, ארגונים וחברות המפתחים בינות מלאכותיות מודעים לבעיות הללו, ומתחילים להתמודד עמן, אך יש צורך לעשות יותר כדי לוודא שהמידע מייצג אוכלוסיות מגוונות ואינו מנציח אי-שוויון חברתי.

הוועדה ממליצה שחוקרים ומפתחים יפעלו כדי לוודא שהמידע שהם מקבלים עובד-מראש על מנת לספק ייצוג מאוזן בכל עת שהדבר אפשרי. בנוסף, החוקרים צריכים להתייחס לסוגיות של הטיות במידע, אך מעבר לכך – הם צריכים גם לשקול את ההטיות הטמונות באלגוריתמים עצמם. לכן צוותיהם של החוקרים צריכים לכלול אנשים ממגדרים, מקורות אתניים ורקעים סוציו-אקונומיים שונים.

הוועדה ממליצה שהממשלה הבריטית תעודד יצירת כלים ומערכות לבחינת ובדיקת מאגרי מידע, כדי לוודא שהם מייצגים אוכלוסיות מגוונות. לאחר מכן יש לעודד את התעשייה לעשות שימוש בכלים שיפותחו.

ח. המלצות בנוגע למונופולי מידע

הוועדה סבורה שהמונופוליזציה של המידע מדגימה את הצורך במערכות אתיות ברורות וחזקות להגנה על המידע ולעידוד התחרותיות בבריטניה, ולערנות מתמשכת מצד הרגולטורים. הוועדה קוראת לממשלה לבחון באופן פרו-אקטיבי את החשש למונופוליזציה אפשרית של המידע מצד חברות הטכנולוגיה הגדולות הפועלות בבריטניה.

◀ צרפת

ראש ממשלת צרפת הטיל על המתמטיקאי וחבר הפרלמנט הצרפתי סדריק וילאני את המטלה להציע אסטרטגיה לפיתוח "בינה מלאכותית בעלת משמעות" בצרפת ב- 2017. הדו"ח שהפיק וילאני מתמקד בעיקר בשתי סוגיות אתיות שמאחורי בינה מלאכותית, כדלקמן[58] –

א. בעיית הקופסה השחורה

הדו"ח עוסק בבעיית הקופסה השחורה: מערכות בינה מלאכותית אינן מובנות היטב כיום, כך שאיננו יכולים לפענח בקלות כיצד הגיעו להחלטות אליהן הגיעו. הדברים רלוונטיים בעיקר עבור מערכות עצבים מלאכותיות, שאינן נסמכות על מערכות חוקים שנקבעו מראש. הדו"ח מסכם בנושא זה כי –

"בטווח הארוך, אחד מהתנאים לקבלת הטכנולוגיה מצד החברה הוא האחריות (accountability) של הטכנולוגיה. בנושאים מסוימים מדובר אפילו בשאלה עקרונית: כחברה, איננו יכולים להתיר להחלטות חשובות מסוימות להתקבל מבלי הסבר. למעשה, מבלי שנוכל להסביר החלטות שקיבלו מערכות אוטונומיות, יהיה קשה להצדיקן: נראה שיהיה בלתי-אפשרי לקבל את מה שאיננו יכולים להצדיק בתחומים קריטיים לחיי האינדיבידואל כגישה לאשראי, תעסוקה, דיור, צדק ובריאות."

ב. שוויון, הטיות ואפליה

כותבי הדו"ח מבהירים כי בינה מלאכותית יכולה להוביל להטיות אינהרנטיות בקבלת החלטות, באופן שמפלה חלקים מהאוכלוסייה. נכון להיום, עדיין אין התמקדות מספקת בניסיון למנוע הטיות שכאלו, בין היתר בשל גילן הצעיר של מערכות הבינה המלאכותית החדשות.

כותבי הדו"ח מציעים מספר קווים מנחים לרגולטור, שאמורים לסייע בהתמודדות עם הדילמות האתיות האמורות.

המלצה #1: מתן שירותי בחינה רשמיים לאלגוריתמים

כותבי הדו"ח מציעים לקבוע גוף של מומחים עם מיומנויות מתאימות לבחינת האלגוריתמים ומאגרי המידע ולבדיקתם בכל דרך הולמת. כל גוף רגולטורי יקבע בפני עצמו את הצרכים המתאימים לתחומים בהם הוא עוסק. ההבנה היא שהאזרחים אינם מסוגלים בכוחות עצמם לקבוע האם הבינה המלאכותית מספקת שירותים הוגנים, ועל כן יש צורך בגוף מרכזי שיבצע את הבדיקה עבורם.

לא ברור עד כמה שימושי (או אפשרי) להסיק מסקנות מבחינת קוד המקור של הבינה המלאכותית. מסיבה זו, הבודקים יצטרכו לבחון גם את תיעוד הפיתוח. ייתכן שהם יחליטו להסתפק גם בבחינת ההוגנות והשוויון של התכנה, למשל באמצעות הזנת מגוון רחב של מידע כוזב, או באמצעות יצירת מספר רב של פרופילי משתמש.

המלצה #2: פיתוח יכולות הערכה ציבורית של בינה מלאכותית

כותבי הדו"ח ממליצים לספק גם לציבור הרחב יכולות בחינה והערכה של הבינה המלאכותית. לשם כך הכותבים ממליצים "לשמן את גלגלי התקשורת" בין הרשויות, בין מכוני המחקר ובין עמותות ציבוריות. בנוסף יש למצוא דרכים להעניק לעמותות ולציבור גישה למידע ששמור בדרך-כלל בידי החברות. ניתן לשקול גם מימון וסיוע לגופים עצמאיים בפרויקטים (בין שבכסף או בתמיכה מדעית, הנדסית, משפטית וכדומה). אחרון חביב, ניתן לשקול מתן סיוע בפיתוח פרוצדורות לבחינת האלגוריתמים ולהנדוס-לאחור (reverse engineering) שלהם.

המלצה #3: פיתוח בינות מלאכותיות "הניתנות להסברה"

כותבי הדו"ח ממליצים לפעול לפי תכנית המחקר של דרפ"א ששמה דגש על פיתוח בינה מלאכותית "ניתנת להסברה" – כלומר, כזו שיכולה להסביר את הבחירות שעשתה. כותבי הדו"ח ממליצים להתרכז בשלושה נתיבי מחקר: הפקת מודלים שניתנים להבנה יותר בקלות, יצירת ממשקי משתמש מובנים יותר, והבנת התהליכים הקוגניטיביים שכרוכים בהפקת הסברים הולמים. נתיבי מחקר אלו משלבים מספר רב של מיומנויות מתחומים שונים: מדעי המחשב, מתמטיקה, תכנון ופיתוח, נוירו-מדעים, פסיכולוגיה ועוד. לפיכך, הם מחייבים שיתופי פעולה בינתחומיים.

המלצה #4: חינוך לאתיקה בתהליך ההכשרה של מהנדסים וחוקרים בתחום הבינה המלאכותית

כותבי הדו"ח ממליצים להוסיף חינוך לאתיקה ולמדעי החברה בתהליך ההכשרה של המהנדסים והחוקרים בתחום הבינה המלאכותית. קורסים בסיסיים באתיקה ובמדעי החברה צריכים להיכלל בסילבוס הקורסים האקדמי של כל מהנדסי המחשבים ומדעני המחשב. המטרה הסופית היא להפיק בוגרים המסוגלים לפתח מערכות בינה מלאכותית יעילות, ובמקביל להבין את השלכותיהן והשפעתן על החברה ועל האזרחים.

המלצה #5: פיתוח מנגנון הערכה להשפעת אפליה (Discrimination Impact Assessment)

החוק האירופי קובע כי במקרים מסוימים, במידה ואנשי החוק רוצים לעבד מידע אישי מסוים, עליהם לנסות קודם להעריך עד כמה פעולותיהם ישפיעו על זכויותיו ורצונותיו של האינדיבידואל אותו הם חוקרים. הערכה זו מכונה באופן רשמי "הערכת השפעת פרטיות" (Privacy Impact Assessment). בדרך זו, הרשויות אינן חייבות לחכות לקבלת החלטה מגבוה, אלא יכולות לבצע את ההערכה בעצמן ולפעול בהתאם לפי התוצאות, באופן יעיל וזריז.

כותבי הדו"ח ממליצים לאמץ אופן פעולה דומה עבור מערכות בינה מלאכותית שעשויות להפלות אינדיבידואלים או קבוצות מסוימות. יוצרי המערכות הללו חייבים לשקול את ההשלכות החברתיות של האלגוריתמים שהם מפתחים.

המלצה #6: מציאת דרכים להתמודד עם זכויות הציבור למידע

כותבי הדו"ח מזהים את אחד האתגרים החשובים בהתמודדות עם השליטה על המידע: החקיקה כיום בנוגע להגנה על המידע חלה רק על מידע אישי, או על אלגוריתמים שהחלטותיהם משפיעות באופן ישיר על אינדיבידואלים. האלגוריתמים החדשים נופלים מחוץ לגבולות חקיקה זו, אך ברור שיש להם השפעות מרחיקות-לכת על החברה. כך, למשל, אלגוריתמים השולחים שוטרים לבצע פטרולים בתדירות גבוהה מהרגיל באזורים מסוימים, יכולים להפלות לרעה חלקים מהאוכלוסייה. באופן דומה, אלגוריתמים המתעדפים באופן שונה מעבר שליחים בשכונות פשע יכולים להנציח את מצבם הסוציו-אקונומי הירוד של התושבים באותם אזורים.

הכותבים מציעים לממשלות לסייע לאזרחים להגיש תובענות ייצוגיות, שהן סוג תביעה בה אדם בודד תובע בשם הפגיעה שחוותה קבוצה שלמה. הם מציעים גם שסוג התובענות הייצוגיות בתחום זה בצרפת ישופר כך שהתובעים יוכלו לזכות בפיצויים במידה ויזכו בתביעה (המצב הנוכחי הוא כזה בו התובענה הייצוגית יכולה רק להביא להפסקת הפעילות הפוגענית).

המלצה #7: התקדמות זהירה בתחום השיטור החיזויי (Predictive policing)

כותבי הדו"ח מייעצים להתקדם בזהירות בכל הנוגע ליוזמות שיטור חיזויי (predictive policing). מערכות בינה מלאכותית אלו מוגבלות מטבען, מסוגלות לשגות, ואנו עשויים לגלות כי הן מפרות חירויות בסיסיות מסוימות, כזכות לפרטיות והזכות למשפט הוגן. במידה ומערכות אלו ייכנסו לשימוש רווח, הן עלולות להשפיע לרעה על שיקול דעתם של שוטרים ושופטים כאחד. שופטים, במיוחד, עשויים למצוא עצמם במצב בו המערכות מספקות להם 'פסקי-דין מוכנים מראש', עליהם יצטרכו רק לחתום. באופן טבעי, שופטים אנושיים רבים ימצאו עצמם הופכים למעין חותמת גומי עבור האלגוריתמים, מבלי שיטרחו להתעמק בפסק הדין שיצרו עבורם, או לנסות לבקר את החלטותיהם הממוחשבות.

על מנת למנוע פגיעה בחברה, הכותבים ממליצים לקבוע כי יש ליידע את האזרחים בנוגע לזכויותיהם: הזכות לטיפול מהיר והזכות לקבל הסבר בנוגע לעיבוד המידע עליו מתבססות ההחלטות האלגוריתמיות. שנית, הכותבים קובעים כי בכל שלב בתהליך השיטור החיזויי, האחריות להליך מסודר תיפול על נציג אנושי. אחרון חביב, יש לזהות תחומים בהם השיפוט האנושי – פגום ככל שיהיה – צריך להישמר בחזקת בני-האדם, ולא לעבור לידי המכונות. בתחומים אלו ממליצים הכותבים שלא לערב את הבינה המלאכותית.

המלצה #8: יצירת ועדת אתיקה לאומית: הכותבים ממליצים לייסד ועדה מייעצת לאומית בתחום האתיקה של הבינה המלאכותית וטכנולוגיה דיגיטלית. הוועדה תהיה אחראית לתיאום הדיון הציבורי באופן נגיש ומוסדר בחוק. הוועדה תצטרך להסביר את ההיגיון שמאחורי החלטותיה בראייה קצרת-טווח המתמקדת בהשלכות התעשייתיות והכלכליות, אך באותה העת תצטרך גם לקחת בחשבון נקודות מבט ארוכות-טווח, עם הבנת ההשלכות החברתיות של כל החלטה. הוועדה תכלול חוקרים ברמת מיומנות גבוהה, ותהיה עצמאית ככל שהדבר יתאפשר.

המלצה #9: התמקדות בתקשורת עם הציבור: הכותבים ממליצים להתייעץ גם עם הציבור בנושאים אתיים הקשורים לבינה מלאכותית. עם זאת, יש להגדיר היטב את דרכי התקשורת עם הציבור. ייתכן, למשל, שהוועדה מהסעיף הקודם תכלול נציגי עמותות ללא מטרות רווח, ונציגי ציבור המסוגלים לקחת חלק בבחינת נושאים שונים.

הוועדה תהיה אחראית גם על ארגון אירועים ציבוריים להרחבת הידע של הציבור בתחום הבינה המלאכותית, לניהול דיונים ציבוריים ולפיתוח כלים דיגיטליים ואחרים שיאפשרו דיונים פוריים, כולל סקרים. הוועדה תהיה גם אחראית לאיסוף, עיבוד וניתוח התוצאות, למתן משוב על המצב הקיים ולמיפוי צרכים וחששות מצד הציבור.

המלצה #10: הצטרפות לדיון הבינלאומי בנושאי האתיקה: הכותבים ממליצים להצטרף לדיון הבינלאומי בנושאי האתיקה בתחום הבינה המלאכותית. הם מבחינים כי מתחילה להיבנות רשת של ועדות אתיקה לאומיות, בסגנון G29 Network (רשת של רשויות לאבטחת מידע). כמובן שגם ישראל צריכה להצטרף לכל רשת כזו שתקום.

◀ סין

למרות המוניטין הלא-רשמי שיצא לסין כמדינה בה אין משמעות לפרטיות, המעצמה פרסמה בתחילת ינואר סטנדרט חדש לפרטיות מידע שמתחרה ב- GDPR האירופאי.

בסין קיים מאבק מתמיד בין שתי סיעות: זו המקדמת הגנה מוגברת על המידע ועל הפרטיות, וזו המקדמת פיתוח בינות מלאכותיות המסתמכות על מידע גדול, ומוכנה לפגוע בפרטיות האזרחים לשם כך[59]. הסטנדרט החדש מסדיר את הדרך בה חברות רשאיות לאסוף, לאגור, לתחזק ולשתף מידע אישי. הוא מגדיר, למשל, את המקרים בהם החברות חייבות לבקש מהמשתמש רשות לאסוף את המידע האישי שלו, ומבהיר כי במידה והמידע משותף מבלי רשותם של המשתמשים, הרי שהוא חייב לעבור תהליך אנונימיזציה. בכל מקרה, חברות צריכות להגן היטב על המידע של המשתמשים.

המרכז למחקרים אסטרטגיים ובינלאומיים (CSIS) שפך אור על מספר קווי דמיון ושוני בין הסטנדרט הסיני החדש לבין ה- GDPR האירופאי, שכוללים את הבאים –

- הסטנדרט הסיני מקיף יותר סוגי מידע: בעוד ש- GDPR מגדיר "מידע אישי רגיש", הסטנדרט הסיני מתייחס לכל סוג מידע אישי שיגרום נזק מכל סוג שהוא אם ייעלם או אם ייעשה בו שימוש זדוני.
- הסטנדרט הסיני כולל דרישות מחמירות יותר בנוגע לסוג המידע שחייב להיכלל בבקשות להיתר שמירת מידע מצד האזרחים.

5. מדיניות של חברות

לא מפתיע לגלות שגם החברות העוסקות בבינה מלאכותית מתחילות לפתח כללי אתיקה שינחו אותן במלאכת הפיתוח, הפריסה והתחזוקה של מערכות הבינה המלאכותית ומאגרי המידע. בחלק זה בדו"ח נעבור על מערכת כללי האתיקה שגיבשו ופרסמו שלוש חברות בינלאומיות גדולות – אינטל, יבמ וגוגל – בשנים האחרונות.

◀ מדיניות אינטל

אינטל פרסמה ב- 2018 דו"ח המתאר את מדיניות החברה בסוגיות האתיות הסובבות את הבינה המלאכותית ופרטיות [60]. הדו"ח כולל חמש אבחנות לגבי בינה מלאכותית ופרטיות, ושש המלצות מדיניות.

אבחנה ראשונה: אוטומציה מוגברת אינה צריכה לפגוע בהגנה על הפרטיות

אינטל מודעת לכך שהשינוי הטכנולוגי המהיר אינו מקבל התייחסות ראויה בחוקי הפרטיות של האיחוד האירופי. עם זאת, היא מאמינה כי החוקים הללו חשובים לשמירה על הפרטיות, ויש לעדכן ולהתאים אותם לזמנים החדשים ולטכנולוגיות החדשות. אינטל ממליצה גם לשים דגש מיוחד על חוקי הפרטיות שעדיין רלוונטיים לעידן החדש, מתוך מטרה לוודא שהאוטומציה המוגברת לא תתנגש עם חוקי ההגנה על הפרטיות.

אבחנה שנייה: יכולת הסברה מחייבת אחריות גדולה

אינטל מאמינה שעקרון השקיפות יאותגר בשנים הקרובות עקב הקושי להבין את ההיגיון מאחורי ההחלטות שיקבלו מנועי בינה מלאכותית מורכבים. היא מקדמת את זכותם של ארגונים ליישם עקרונות בינה מלאכותית גם מבלי יכולת הסברה, ובתנאי שהוכיחו כי הם מקיימים תהליכים ומקדישים משאבים למזעור סיכונים לפרטיות ולחיי האינדיבידואל.

אינטל מקדמת דרכי עבודה אחראיות, ומתחייבת ליצור כלים ולספק אימון לעובדיה על מנת לשמור על הפרטיות ולקדם מערכות לביקורת פנימית וחיצונית על המתנהל בחברה. היא מעוניינת לקדם גישות של "פרטיות בתכנון מראש" (Privacy-by-design) בכל תהליכי התכנון והייצור של בינה מלאכותית.

אבחנה שלישית: עיבוד מידע אתי מבוסס על פרטיות

אינטל מאמינה שפרטיות ואבטחת מידע הינן הבסיסים עליהם מושתתים חיינו כיום, והן המאפשרות את חירות האינדיבידואל ויכולתו לקבל החלטות עבור עצמו. הגנה על אינדיבידואלים והמידע אודותיהם אינו רק עניין של ציות עיוור לחוק, אלא נכונות לאמץ את הערכים החברתיים ולפעול לכינון אמון בטכנולוגיות ובהשפעתן החיובית על העולם.

אבחנה רביעית: פרטיות מגנה על מי שאנחנו

פרטיות מחייבת שהמידע יהיה אמין ושלא ייעשה בו שימוש לפגיעה באינדיבידואלים. היא אמורה למנוע גישה לא-מאושרת למידע אישי, מחיקתו או אובדנו. היא מחייבת מתן כבוד לחיים הפרטיים ולחיי המשפחה, לביתו של האינדיבידואל ולסודיות התקשורות שהוא מנהל.

בינה מלאכותית יכולה להשפיע על הדרך בה אחרים רואים אותנו, באמצעות תפעול אינדיבידואלים והמציאות שהם רואים ויצירת מידע כוזב אודות אינדיבידואלים מסוימים, באופן שישפיע על הדרך בה אחרים רואים אותם. בינה מלאכותית יכולה להשפיע גם על הדרך בה אנו רואים את עצמנו: הן מאפשרות לארגונים להבין אינדיבידואלים או לחזות את התנהגותם טוב יותר מכפי שהם עצמם היו מסוגלים לכך. ממשלות וארגונים מסוימים יוכלו לנצל יכולת זו לרעה, להתמקד באזרחים מסוימים ולהשפיע על דרך ראיית המציאות שלהם, ובכך על יכולת קבלת ההחלטות שלהם.

אבחנה חמישית: הצפנה חזקה יותר ואנונימיזציה מלאה מסייעות לפרטיות

בעולם בו יותר ויותר מידע נאסף, מעובד ומנותח על-ידי מנועי בינה מלאכותית, יש צורך בטכניקות הצפנה ברמה גבוהה ובאנונימיזציה מלאה כדרך להגן על פרטיות האזרחים. אנונימיזציה מלאה תהיה קשה יותר לאורך זמן, מאחר ומנועי בינה מלאכותית יוכלו להפר אותה בבוא הזמן. עם זאת, טכניקות מודרניות מסוימות מסוגלות להוסיף 'רעש' למידע אישי וכך להקשות על זיהוי-מחדש של האינדיבידואל.

המלצה ראשונה: יוזמות חוקיות ורגולטוריות חדשות צריכות להיות מקיפות, ניטרליות מבחינה טכנולוגית ואמורות לאפשר זרימת מידע חופשית

יוזמות חוקיות חדשות בתחום הפרטיות צריכות להיות מקיפות על מנת להימנע מיצירת חורים שניתן יהיה לנצל, וראוי שיכסו גם את השימושים במידע ואת הטכנולוגיות שעדיין אינן קיימות, או שמתקיימות מחוץ למנגנונים החוקיים של ההווה. חוקי פרטיות שיוגדרו עבור טכנולוגיות בינה מלאכותית מסוימות עלולות שלא להיות גמישות מספיק כדי לשרוד לאורך זמן.

בנוסף, היכולת לעבד מידע ולשנע אותו בין מדינות הינה קריטית לפיתוח טכנולוגיות חדשות. ערכה הגלובלי של זרימת מידע דיגיטלי גדל עשרות מונים בעשור האחרון. יוזמות חקיקה, לפיכך, צריכות לקדם זרימת מידע בינלאומית עם ביקורת הולמת.

המלצה שנייה: ארגונים צריכים לאמץ גישות מבוססות-סיכונים לאחריות

ארגונים צריכים לקבל אחריות על המוצרים שהם משחררים לשוק, ולהטמיע וליישם תהליכי פיתוח מבוסס-פרטיות (privacy-by-design), הכוללים שימוש בטכניקות כאנונימיזציה מלאה וכהצפנה ברמה גבוהה. חברות אחראיות צריכות לפתח הערכות השפעה על הפרטיות עבור הטכנולוגיות שהן מייצרות, ולזהות את הסיכון לפרטיות שהטכנולוגיות טומנות בחובן. בכל שלב בתהליך הפיתוח, המהנדסים והאנליסטים צריכים לזהות ולהעריך את

ההשלכות הלא-מכוונות האפשריות על משתמשי הקצה, ולהציע פתרונות לצמצום הסיכונים. ארגונים אחראיים זקוקים גם למנגנוני פיקוח, ניהול ובקרה פנימיים עבור תהליכי הפיתוח הללו, עם ועדות אתיקה פנימיות.

המלצה שלישית: יש לטפח קבלת החלטות אוטומטית, אך גם לאבטח אותה על מנת להגן על אינדיבידואלים

חוקי הפרטיות הינם כלי רב-ערך עבור האזרחים, אך הם מוגבלים מטבעם עקב האחריות הכבדה הרבה שהם מטילים על האינדיבידואל, שצריך להבין מה השימוש שנעשה במידע האישי שלו, וכיצד יושפע מכך. הכותבים ממליצים לבחון את "האינטרס הלגיטימי" של הישות המעבדת את המידע, כנגד הציפיות הלגיטימיות של האינדיבידואלים. אינטרסים לגיטימיים מסוימים, כגון הצורך לספק ביטחון פיזי לאינדיבידואלים, או לאבטח את הרשת כולה, אמורים לספק הצדקה לעיבוד מידע אישי – גם כשלא ניתן להשיג אישור ברור מהאינדיבידואל.

משמעות ההמלצה היא שיש להיזהר מהטלת הגבלות חמורות מדי על מערכות לקבלת החלטות אוטונומיות. גישה מחמירה מדי עלולה לעצור את החדשנות בתחום ולמנוע מאזרחים ליהנות משירותי בינה מלאכותית שישפרו את חייהם בתחומים רבים.

הכותבים ממליצים לשלב רמות שונות של מעורבות ופיקוח אנושיים במערכות אוטונומיות שונות. מערכות בינה מלאכותית צריכות להיות מתוכננות, להיבנות ולהיפרס כך שיאפשרו שליטה ושיפוט אנושי במקרים בהם אינדיבידואלים יושפעו מההחלטות שיקבלו מערכות אלו. לשם כך יש צורך בבחינת הרמות השונות של פיקוח אנושי הנחוצות לכל מערכת, לפי ניתוח סיכונים.

המלצה רביעית: **ממשלות צריכות לקדם גישה למידע**

ממשלות צריכות לקדם גישה למאגרי מידע ציבורי מובנים וזמינים. עליהן לתמוך באופן פעיל ביצירת מאגרי מידע אמינים (שיכללו מידע אישי של אינדיבידואלים), בהם יוכלו מפתחי בינה מלאכותית להשתמש על מנת לבחון פתרונות אוטומטיים ולהבין את איכות האלגוריתמים שלהם. הממשלה צריכה לספק תמריצים לשיתוף מידע בין הציבור והמגזר הפרטי ובין השחקנים השונים בתעשייה, ולתמוך ביצירתם של ממשקים משותפים (APIs) שיספקו גישה מהירה ודינמית למידע. בנוסף, הממשלה צריכה לתרום לכינון סטנדרטים בינלאומיים שיקלו על שיתוף המידע.

אחרון חביב, הממשלה צריכה לקדם מגוון (Diversity) במאגרי המידע שברשותה וברשות אחרים. מגוון גדול יותר יצמצם את הסיכון להטיות לא-מכוונות.

המלצה חמישית: חשוב לממן מחקר בתחום האבטחה על מנת להגן על הפרטיות

קצב ההתקדמות של הבינה המלאכותית מעולם לא היה מהיר יותר, אך עדיין יש צורך בעבודה רבה על מנת לשפר את יעילות המחשוב, צריכת האנרגיה והקישוריות במרכזי מידע ובמכשירי הקצה. יש לתמוך גם במחקר בתחום האבטחה על מנת לאפשר קצירת נתונים שימושיים ומשמעותיים ממידע מוצפן, מבלי לקרוא את המידע האישי.

המלצה שישית: צריך מידע כדי להגן על מידע

הפוטנציאל העצום של בינה מלאכותית כולל גם את האפשרות להשתמש בניתוח מידע כדי לתמוך במדיניות ציבורית – למשל, להגנה על הפרטיות, לאבטחת מידע ולאבטחת סייבר. אלגוריתמים יכולים לעזור לחשוף מקרים של אפליה לא-מכוונת ושל הטיה, גניבת זהות או איומי סייבר. מנועי בינה מלאכותית יכולים לספק לארגונים את היכולת למנוע ולמתן סיכונים מסוג זה. בדרך זו, השימוש במידע (שכולל מידע אישי) יסייע בסך-הכול בשמירה על הפרטיות.

◀ מדיניות יבמ

יבמ שחררה בסוף 2018 "מדריך אתיקה יומיומית פרקטי למפתחי ולמתכנני בינה מלאכותית". המדריך נועד לשמש כבסיס לדיון שיגדיר את האתיקה היומיומית בתחום הבינה המלאכותית. יבמ מאמינה ש- "חובה להטמיע אתיקה בתהליכי הפיתוח והתכנון כבר מהרגע הראשון ביצירת בינה מלאכותית", וש- "כמפתחי ומתכנני בינה מלאכותית, יש בידינו חלק גדול מיכולת ההשפעה הקולקטיבית. אנו יוצרים מערכות שישפיעו על מיליוני אנשים."

יבם התוותה במדריך חמישה עקרונות מרכזיים, עם המלצות נקודתיות למפתחי בינה מלאכותית עבור כל עקרון. יש לציין שהעקרונות וההמלצות שמים דגש גדול – אולי גדול מדי – על האחריות האישית של המפתחים, גם במצבים בהם האחריות הסופית צריכה ליפול על החברה שמייצרת את מנוע הבינה המלאכותית, או על הלקוח / צרכן שמשתמש בו בדרכים שאינן ראויות.

עקרון ראשון: אחריות

מפתחי ומתכנני בינה מלאכותית אחראים לחשיבה אודות פיתוח הבינה המלאכותית, תהליכי קבלת ההחלטות של הבינה המלאכותית והתוצאות הסופיות. בני-האדם שכותבים את האלגוריתמים הם המגדירים הצלחה או כישלון, והם מקבלים את ההחלטות אודות שימושי המערכת ומי יושפע ממנה. לפיכך, כל אדם המעורב בשלב כלשהו בתהליך יצירת הבינה המלאכותית, אחראי וחייב לשקול את השפעת המערכת על העולם.

המלצות נקודתיות:

- מדיניות ברורה: מדיניות החברה צריכה להיות ברורה וזמינה לצוותי הפיתוח והתכנון כבר מהיום הראשון, כך שאיש לא יכול להתבלבל בנוגע לאחריות המשותפת.
- הבדלה בין אחריות החברה לאחריות המפתח: חשוב להבין היכן אחריות החברה / מוצר מסתיימת. ייתכן שלמפתח הבינה המלאכותית לא תהיה שליטה על הדרך בה משתמש, לקוח או מקור חיצוני אחר ישתמש במידע או בכלי מסוים שהמפתח שם בידיו.
- תיעוד: חשוב לשמור תיעוד מפורט של תהליכי הפיתוח וקבלת ההחלטות.
- להכיר את הרגולציה: המפתחים צריכים להיצמד להנחיות ההתנהלות העסקית של חברתם. בנוסף, הם צריכים להכיר ולהבין את החוקים הלאומיים והבינלאומיים, הרגולציות וההנחיות שהבינה המלאכותית שיפתחו צריכה לפעול במסגרתם.

עקרון שני: הסדרת ערכים

הבינה המלאכותית צריכה לתכנן כך שתפעל בהתאם לנורמות ולערכים של קבוצת המשתמשים בה. למערכות הבינה המלאכותית אין הבנה אינטואיטיבית של ערכים אנושיים, ולכן תפקידם של המפתחים הוא לוודא שמנועי הבינה המלאכותית יתחשבו בערכים האנושיים הקיימים. חשוב להתייחס ברגישות הראויה למגוון רחב של נורמות וערכים תרבותיים, ולהבין שבינה מלאכותית צריכה לפעול בשיתוף פעולה עם המשתמשים, ולא לאכוף עליהם ערכים שונים מאלו המקובלים עליהם.

המלצות נקודתיות:

- הבנת תרבות היעד: המפתחים צריכים להבין את התרבות עבורה הם מייצרים את הבינה המלאכותית, ולהזמין לחברה מומחי מדיניות ואקדמאים שיעזרו לצוות הפיתוח להבין את נקודות המבט של המשתמשים.
- מיפוי הערכים: יבם ממליצה למפות את ערכי המשתמשים ולהתאים את פעולותיה של הבינה המלאכותית לערכים המתוארים.

עקרון שלישי: הסברתיות

חשוב לפתח בינה מלאכותית כך שבני-אדם יוכלו לזהות, לפענח ולהבין בקלות את תהליכי קבלת ההחלטות שלה. יבם מכירה בכך שאיננו סומכים על בני-אדם – או על מערכות ממוחשבות – שאיננו מבינים את השיקולים מאחורי תהליך קבלת ההחלטות שלהם. ככל שהבינה המלאכותית תגדל ביכולותיה ותשפיע יותר על העולם, כך יהיה צורך גדול יותר להסביר את תהליכי קבלת ההחלטות שלה באופן שגם אנשים מן השורה יוכלו להבין.

המלצות נקודתיות:

- שקיפות: המשתמשים צריכים לדעת תמיד כאשר הם מתקשרים עם בינה מלאכותית. בינה מלאכותית שאינה מזדהה ככזאת, אינה בינה מלאכותית אתית.
- הרשות לשאלות: משתמשים צריכים להיות מסוגלים לשאול את הבינה המלאכותית מדוע היא עושה את מה שהיא עושה.

- פתיחות לסקירה: תהליכי קבלת ההחלטות חייבים להיות ניתנים לסקירה, במיוחד אם הבינה המלאכותית עובדת עם מידע אישי רגיש במיוחד, כגון מידע רפואי או ביומטרי.
- הסבר למשתמשים: כאשר בינה מלאכותית מסייעת למשתמשים בקבלת החלטות רגישות במיוחד, המנוע חייב להיות מסוגל לספק להם הסבר הולם אודות המלצותיו, המידע עליו הסתמך וההיגיון שמאחורי ההמלצות.
- אימות: לצוותי הפיתוח צריכה להיות גישה לתיעוד של החלטות הבינה המלאכותית, על מנת שיוכלו לוודא את נכונות תהליכי קבלת ההחלטות.

עקרון רביעי: הוגנות

בני-אדם נוטים במיוחד להטיות לא-מודעות, וכאשר הם יוצרים את הבינה המלאכותית, הם עשויים להקנות לה את אותן הטיות. לפיכך, הבינה המלאכותית חייבת להיות מתוכננת כך שתמזער הטיות ותקדם ייצוג כולל.

המלצות נקודתיות:

- ניתוח בזמן-אמת: יש לבצע ניתוח בזמן-אמת של פעולת הבינה המלאכותית, על מנת לחשוף הטיות מכוונות ולא-מכוונות. כאשר ברור שיש הטיה בתוצאות או במידע, הצוות חייב לחקור ולהבין את מקור ההטיה וכיצד ניתן למנוע אותה.
- הימנעות מהטיות: יש לפתח בינות מלאכותיות חפות מהטיות מכוונות, ולתאם סקירות מצד הצוות על מנת להימנע מהטיות לא-מכוונות. הטיות לא-מכוונות עשויות לכלול סטריאוטיפיזציה, הטיית אישוש והטיית העלות השקועה/אבודה (sunk cost bias).
- משוב ודיאלוג: יש לכונן מנגנון לקבלת משוב או לאפשר דיאלוג פתוח עם המשתמשים, על מנת לאפשר למשתמשים עצמם לזהות בעיות או הטיות.

עקרון חמישי: זכויות המידע של המשתמשים

בינה מלאכותית חייבת להיות מפותחת כך שתגן על המידע של המשתמשים ותשמר את יכולתם של המשתמשים לקבל החלטות בנוגע לגישה למידע ולשימושים בו. צוות הפיתוח אחראי על מתן כוח ושליטה למשתמשים.

המלצות נקודתיות:

- הכוח בידי המשתמש: המשתמשים צריכים לשמור תמיד על שליטה בנוגע לדרך השימוש במידע שלהם. הם יכולים לסרב לספק לאפליקציות גישה למידע אישי שאינם רוצים לשתף עמן.
- בקשת רשות: הבינה המלאכותית צריכה לבקש רשות מהמשתמשים כדי לקבל מהם מידע. הגדרות הפרטיות צריכות להיות ברורות, קלות לזיהוי וניתנות להתאמה.
- שקיפות מלאה: יש לספק הצהרה שקופה, ברורה ומלאה בנוגע לדרך בה המידע האישי ישמש את האלגוריתמים, או הדרך בה ישותף עם אחרים.
- הגנה על המידע: יש להגן על המידע של המשתמשים מגניבה, משימוש לרעה או מהשחתת המידע.
- איסור על שיתוף במידע: כאשר מפתחים שירות בינה מלאכותית חדש, יש לאסור על שימוש במידע המגיע מחברה אחרת מבלי רשות מפורשת.
- ציות לחוקים לאומיים ובינלאומיים: יש להכיר ולציית לתקנות וחוקים לאומיים ובינלאומיים בנוגע לשימוש במידע אישי.

◀ מדיניות גוגל

באמצע 2018 שחררה גוגל מסמך המתאר את שבעת עקרונות האתיקה עליהם היא מסתמכת בפיתוח ובשימוש בבינה מלאכותית[61]. לא מדובר בכללים רשמיים, אלא בעקרונות כלליים בלבד. עם זאת, גוגל רואה אותם כסטנדרטים שינחו את תהליכי המחקר והפיתוח בחברה וישפיעו על החלטותיה העסקיות.

עקרון ראשון: חשיבות חברתית

גוגל מתחייבת להתייחס למגוון רחב של גורמים חברתיים וכלכליים בעת פיתוח ושימוש בבינה מלאכותית. היא תתמקד בנתיבים בהם היא צופה שהיתרונות לחברה יאפילו על הסיכונים והקשיים. גוגל תחתור להפוך מידע מדויק ובאיכות-גבוהה זמין לכל באמצעות בינה מלאכותית, תוך שהיא מכבדת נורמות תרבותיות, חברתיות וחוקיות במדינות בהן היא פועלת. היא תמשיך להעריך בזהירות מתי ואיך לאפשר שימוש בטכנולוגיות שלה ללא-תשלום.

עקרון שני: הימנעות מיצירת או חיזוק הטיה שאינה הוגנת

בינות מלאכותיות ומאגרי מידע יכולים לשקף, לחזק או למזער הטיות לא-הוגנות. גוגל מבינה שלא קל להבדיל בין הטיה הוגנת וכזו שאינה הוגנת, על אחת כמה וכמה בתרבויות ובחברות שונות. גוגל תנסה להימנע מהשפעות לא-הוגנות על אנשים, במיוחד כאלו הקשורות לתכונות רגישות כגזע, מוצא, מגדר, לאום, רמת הכנסה, נטייה מינית, יכולות, ואמונה פוליטית או דתית.

עקרון שלישי: התמקדות בבטיחות

גוגל תמשיך לפתח וליישם פרקטיקות אבטחה ובטיחות על מנת להימנע מתוצאות לא-מכוונות שעלולות לגרום לנזק. גוגל תתכנן את מערכות הבינה המלאכותית שלה כך שיהיו זהירות באופן הולם למצב, ותשאף לפתח אותן לפי הפרקטיקות המתקדמות ביותר בחקר הבטיחות בבינה מלאכותית. במקרים מסוימים, גוגל תבחן טכנולוגיות בינה מלאכותית בסביבות מוגבלות, ותמשיך בניטור אחר פעולתן גם לאחר ההשקה.

עקרון רביעי: אחריות בפני המשתמשים

גוגל תתכנן מערכות בינה מלאכותית שיספקו למשתמשים הזדמנויות הולמות למתן משוב, לקבלת הסברים מתאימים – ולערעור עליהם. טכנולוגיות הבינה המלאכותית של גוגל יהיו כפופות להנחיה ולשליטה אנושית הומת.

עקרון חמישי: שילוב עקרונות פרטיות בפיתוח

גוגל תשלב את עקרונות הפרטיות שלה בפיתוח ובשימוש בטכנולוגיות בינה מלאכותית. היא תיידע את המשתמשים בנוגע לשימוש שנעשה במידע שלהם ותעודד יצירת ארכיטקטורות תוכנה עם התייחסות לפרטיות והצורך לשמור עליה.

עקרון שישי: שמירה על סטנדרטים גבוהים של מצוינות מדעית

החדשנות הטכנולוגית היא תוצר ברור של השיטה המדעית, ומחויבת לפיכך לחשיבה פתוחה וסקרנית, לחריצות אינטלקטואלית, ליושרה ולשיתוף פעולה. לכלי בינה מלאכותית יש את הפוטנציאל לפתוח ממלכות חדשות של מחקר וידע מדעי בתחומים קריטיים לאנושות כביולוגיה, כימיה, רפואה ומדעי הסביבה. גוגל תשאף לרמות גבוהות של מצוינות מדעית בפיתוח הבינה המלאכותית. לשם כך היא תפעל עם בעלי-עניין ברמה הגבוהה ביותר, ותשתף מידע באופן אחראי באמצעות פרסום תוצאות, פרקטיקות עבודה מוצלחות ומחקרים שיאפשרו ליותר אנשים לפתח יישומי בינה מלאכותית שימושיים.

עקרון שביעי: היצמדות לשימושים למטרות מסוימות

גוגל תפעל להגבלת השימוש בבינה המלאכותית שהיא תפתח, על מנת למנוע שימושים שעלולים לגרום לנזק נרחב. שימושים אפשריים יישקלו לפי הגורמים הבאים –

- מטרה ושימוש עיקריים: המטרה המרכזית והשימוש הסביר ביותר בטכנולוגיה, עם התייחסות לחשש שהפתרון יוכל להיות מותאם לגרימת נזק.
- טבע וייחודיות: גוגל רוצה לייצר טכנולוגיה שזמינה לכל, ולא לשימושים ייחודיים בלבד.
- היקף: גוגל רוצה ליצור טכנולוגיות בעלות השפעה נרחבת.
- טבע המעורבות של גוגל: גוגל עשויה לספק כלים כלליים, לשלב כלים לטובת הצרכנים, או לפתח פתרונות ייחודיים.

הימנעות משימוש בבינה מלאכותית

גוגל תימנע מפיתוח או פריסת בינה מלאכותית בתחומים הבאים –

- טכנולוגיות שגורמות או סביר שיגרמו נזק כללי. במקום בו יש חשש לנזק, גוגל תתקדם רק במידה והיא מאמינה שהיתרונות עולים בהרבה על הסיכונים, ותשלב מגבלות בטיחות הולמות במוצר הסופי.
- נשקים או טכנולוגיות אחרות שמטרתן המרכזית או יישומן מביאים לגרימת נזק לאנשים.
- טכנולוגיות שאוספות או משתמשות במידע למטרות ניטור, תוך שהן מפרות את הנורמות הבינלאומיות המקובלות בתחום.
- טכנולוגיות שמטרתן מתנגשת עם העקרונות המקובלים של המשפט הבינלאומי וזכויות אדם.

למרות שגוגל לא תפתח בינה מלאכותית לשימוש בנשקים, היא תמשיך לעבוד עם ממשלות ועם הצבא במגוון תחומים. אלו יכללו אבטחת-סייבר, אימון, גיוס צבאי, ביטוח בריאות לאנשי צבא לשעבר ומשימות חיפוש והצלה.

6. סיכום

ממשלות וחברות רבות מפרסמות בימים אלו המלצות לגיבוש עקרונות אתיים להתמודדות עם איסוף המידע על אינדיבידואלים ועם מנועי בינה מלאכותית שאמורים לעשות שימוש במידע. בחלק זה בדו"ח עברנו על האתגרים הגדולים בפניהם עומדת הבינה המלאכותית כיום. לאחר מכן כיסינו את ההמלצות להתנהלות אתית של מספר מדינות מייצגות, כבריטניה, צרפת וסין, כמו גם על ה- GDPR של האיחוד האירופי. אחרון חביב, פירטנו את ההמלצות לפיתוח ותכנון אתי של בינה מלאכותית המגיעות מצד מספר חברות בינלאומיות כיבמ, אינטל וגוגל.

בעקבות הסקירה, ניתן לזהות מספר עקרונות משותפים ונקודות חשובות שראוי שהרגולטור בישראל יתייחס אליהן בכל ניסיון לפתח מסגרות אתיות הנוגעות לאיסוף מידע ועיבודו. את אלו נמנה בקצרה כאן –

עקרון משותף ראשון: מניעת הטיה

מאגרי מידע צריכים לספק ייצוג הולם לכלל האוכלוסייה, על-מנת לוודא שהבינות המלאכותיות הלומדות מהם אינן מפתחות הטיות לא-מכוונות. מפתחי הבינות המלאכותיות צריכים לעשות כל מאמץ כדי למנוע הטיות מכוונות או לא-מכוונות מצד המנועים שהם מפתחים.

עקרון משותף שני: אחריות ציבורית

מנועי הבינה המלאכותית צריכים להיות מתוכננים כך שתהיה להם אחריות ציבורית: הם (או הארגונים שמאחוריהם) צריכים להיות מסוגלים להסביר לאנשים את הסיבה להחלטות שקיבלו הבינות המלאכותיות. ארגונים וממשלות צריכים לקדם את הבנת הציבור במדעי המידע ובבינה מלאכותית, לקבל משוב מהציבור ולשלב ככל האפשר את הציבור בתהליכי קבלת ההחלטות.

עקרון משותף שלישי: פרטיות מרבית

הארגונים האוספים מידע על האזרחים צריכים להעביר אותו אנונימיזציה ברמה גבוהה, על מנת להקשות עד כמה שאפשר (גם אם לעולם לא בוודאות) על זיהוי האינדיבידואל.

עקרון משותף רביעי: קידום הגישה למידע מצד הממשלות

הרשויות הממשלתיות האוספות מידע צריכות לפתוח אותו לארגונים ולציבור הרחב, ולסייע בכינון או בהשתתפות בסטנדרטים בינלאומיים שיקלו על שיתוף המידע. כל זאת, לאחר אנונימיזציה ראויה של המידע.

עקרון משותף חמישי: ניתוח סיכונים אחראי

ארגונים צריכים להעריך באופן רשמי ומסודר את ההשפעה שתהיה לטכנולוגיה שלהם על פרטיות המשתמשים, ולהציע פתרונות אפשריים לצמצום הסיכונים. כדי לעשות זאת, הארגונים יזדקקו לכונן ולקיים מנגנוני פיקוח, ניהול ובקרה פנימיים בתחומים אלו.

עקרון משותף שישי: השכלה בתחום האתיקה

כל האנשים המעורבים בתהליכי הפיתוח של בינות מלאכותיות ומאגרי מידע, צריכים לרכוש השכלה בסיסית בתחום האתיקה. רצוי שהשכלה שכזו תתחיל עוד בתואר הראשון באוניברסיטה.

עקרון שביעי משותף: הצטרפות לדיון הבינלאומי בנושאי האתיקה

הדיון בנושאי האתיקה הופך לתופעה גלובלית, ורשת של ועדות אתיקה לאומיות מתחילה להיבנות בימים אלו. חשוב שישראל לא תיוותר מחוץ לדיון הבינלאומי, אלא תיקח בו חלק מרכזי ומתמשך.

7. מקורות

[1] W. Knight, “The Dark Secret at the Heart of AI - MIT Technology Review,” *MIT Technology Review*, 04-Nov-2017. [Online]. Available: https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/. [Accessed: 12-Dec-2018].

[2] D. Alba, “It’s Your Fault Microsoft’s Teen AI Turned Into Such a Jerk | WIRED,” *WIRED*. [Online]. Available: https://www.wired.com/2016/03/fault-microsofts-teen-ai-turned-jerk/. [Accessed: 12-Dec-2018].

[3] J. Condliffe, “AI Has Beaten Humans at Lip-reading - MIT Technology Review,” *MIT Technology Review*. [Online]. Available: https://www.technologyreview.com/s/602949/ai-has-beaten-humans-at-lip-reading/. [Accessed: 12-Dec-2018].

[4] “Tuberculosis control, and the where and why of artificial intelligence.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5478795/. [Accessed: 12-Dec-2018].

[5] S. University, “Algorithm outperforms radiologists at diagnosing pneumonia,” *Stanford News*, 15-Nov-2017. [Online]. Available: https://news.stanford.edu/2017/11/15/algorithm-outperforms-radiologists-diagnosing-pneumonia/. [Accessed: 12-Dec-2018].

[6] M. Scott, “Google Fined Record $2.7 Billion in E.U. Antitrust Ruling,” *The New York Times*, 20-Jan-2018.

[7] “Yelp to revive antitrust complaint against Google over search results,” *VentureBeat*, 23-May-2018. .

[8] “Listings in Disputed Regions,” *Airbnb Press Room*, 19-Nov-2018. .

[9] E. Dwoskin and C. Timberg, “Google, Twitter face new lawsuits alleging discrimination against conservative voices,” *Washington Post*. [Online]. Available: https://www.washingtonpost.com/news/the-switch/wp/2018/01/08/google-faces-a-lawsuit-over-discriminating-against-white-men-and-conservatives/. [Accessed: 12-Dec-2018].

[10] P. Barwise, “Why Tech Markets Are Winner-Take-All,” *Media Policy Project*, 14-Jun-2018. .

[11] T. Hfu, “For Many Facebook Users, a ‘Last Straw’ That Led Them to Quit - The New York Times,” *The New York Times*, 21-Mar-2018. [Online]. Available: https://www.nytimes.com/2018/03/21/technology/users-abandon-facebook.html. [Accessed: 12-Dec-2018].

[12] Y. Abutaleb, “Facebook’s political influence under a microscope,” *Reuters*, 29-Jun-2016.

[13] A. C. Madrigal, “What Facebook Did to American Democracy,” *The Atlantic*, 12-Oct-2017. [Online]. Available: https://www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502/. [Accessed: 12-Dec-2018].

[14] H. Tang, “Why does artificial intelligence lie to human patients?,” *AI Med*, 08-Nov-2018. .

[15] Verger Rob, “Where to find self-driving cars on the road right now | Popular Science,” 12-Nov-2018. [Online]. Available: https://www.popsci.com/self-driving-cars-cities-usa#page-5. [Accessed: 12-Dec-2018].

[16] Federal Ministry of Transport and Digital Infrastructure - Ethics Commission, “Automated and Connected Driving,” Jun. 2017.

[17] K.-F. Lee, *AI Superpowers: China, Silicon Valley, and the New World Order*. Boston: Houghton Mifflin Harcourt, 2018.

[18] “Project Overview ‹ A Voting-Based System for Ethical Decision Making,” *MIT Media Lab*. [Online]. Available: https://www.media.mit.edu/projects/a-voting-based-system-for-ethical-decision-making/overview/. [Accessed: 12-Dec-2018].

[19] “Moral Machine,” *Moral Machine*. [Online]. Available: http://moralmachine.mit.edu. [Accessed: 12-Dec-2018].

[20] “Top 5 Safety Features in Autonomous Car | Intellias Blog,” *Intellias*, 19-Jun-2018. [Online]. Available: https://www.intellias.com/top-5-features-to-ensure-you-re-still-alive-after-riding-in-autonomous-car/. [Accessed: 12-Dec-2018].

[21] Chappell Bill, “‘It Was Installed For This Purpose,’ VW’s U.S. CEO Tells Congress About Defeat Device,” *NPR.org*, 10-Aug-2015. [Online]. Available: https://www.npr.org/sections/thetwo-way/2015/10/08/446861855/volkswagen-u-s-ceo-faces-questions-on-capitol-hill. [Accessed: 12-Dec-2018].

[22] C. Coleman, “VW could face long legal nightmare - BBC News,” *BBC News*, 24-Sep-2015. [Online]. Available: https://www.bbc.com/news/business-34352243. [Accessed: 12-Dec-2018].

[23] C. Rogers and M. Spector, “Judge Slaps VW With $2.8 Billion Criminal Fine in Emissions Fraud,” *Wall Street Journal*, 21-Apr-2017.

[24] J. Ewing, “Ex-Volkswagen C.E.O. Charged With Fraud Over Diesel Emissions,” *The New York Times*, 18-Jul-2018.

[25] R. Stumpf, “Daimler Does Dieselgate: 1 Million Vehicles Have ‘Defeat’ Device - The Drive,” *The Drive*, 06-Dec-2018. [Online]. Available: http://www.thedrive.com/news/21471/daimler-does-dieselgate-up-to-1-million-vehicles-have-defeat-device-says-regulators. [Accessed: 12-Dec-2018].

[26] C. Arthur, “How Samsung inflated its performance scores,” *The Guardian*, 13-Oct-2013.

[27] B. Schneier, “Volkswagen and Cheating Software - Schneier on Security,” *Schneier on Security*, 30-Sep-2015. [Online]. Available: https://www.schneier.com/blog/archives/2015/09/volkswagen_and_.html. [Accessed: 12-Dec-2018].

[28] K. Finley, “VW’s Cheating Proves We Must Open Up the Internet of Things,” *Wired*, 24-Sep-2015.

[29] S. Gallagher, “IoT garage door opener maker bricks customer’s product after bad review,” *Ars Technica*, 04-Apr-2017. [Online]. Available: https://arstechnica.com/information-technology/2017/04/iot-garage-door-opener-maker-bricks-customers-product-after-bad-review/. [Accessed: 12-Dec-2018].

[30] C. Hoffman, “What Does ‘Bricking’ a Device Mean?,” *How-To Geek*. [Online]. Available: https://www.howtogeek.com/126665/htg-explains-what-does-bricking-a-device-mean/. [Accessed: 12-Dec-2018].

[31] A. Gilbert, “The time that Tony Fadell sold me a container of hummus.,” *Look Ma, I’m Blogging!*, 04-Apr-2016. [Online]. Available: https://arlogilbert.com/the-time-that-tony-fadell-sold-me-a-container-of-hummus-cb0941c762c1. [Accessed: 12-Dec-2018].

[32] “Google’s Nest Reaching Out to Revolv Owners Amid Controversy | Time.” [Online]. Available: http://time.com/4283408/nest-google-shuts-down-revolv/. [Accessed: 12-Dec-2018].

[33] M. Goodman, *Future Crimes: Everything Is Connected, Everyone Is Vulnerable and What We Can Do About It*, Reprint edition. Anchor, 2015.

[34] J. Fruhlinger, “What is Stuxnet, who created it and how does it work? | CSO Online,” *CSO*, 22-Aug-2018. [Online]. Available: https://www.csoonline.com/article/3218104/malware/what-is-stuxnet-who-created-it-and-how-does-it-work.html. [Accessed: 13-Dec-2018].

[35] N. Perlroth and C. Krauss, “A Cyberattack in Saudi Arabia Had a Deadly Goal. Experts Fear Another Try.,” *The New York Times*, 17-Oct-2018.

[36] “Prosecute Burger King for Their Illegal Google Home Attacks in Their Ads – Lauren Weinstein’s Blog.” .

[37] A. Hern, “Someone made a smart vibrator, so of course it got hacked,” *The Guardian*, 10-Aug-2016.

[38] O. Solon, “How A Book About Flies Came To Be Priced $24 Million On Amazon,” *Wired*, 27-Apr-2011.

[39] “Deal or no deal? Training AI bots to negotiate,” *Facebook Code*, 14-Jun-2017. .

[40] M. Wilson and M. Wilson, “AI Is Inventing Languages Humans Can’t Understand. Should We Stop It?,” *Fast Company*, 14-Jul-2017. [Online]. Available: https://www.fastcompany.com/90132632/ai-is-inventing-its-own-perfect-languages-should-we-let-it. [Accessed: 13-Dec-2018].

[41] “How Big Data & Social Made House of Cards a Hit,” *Cision*, 28-Feb-2014. [Online]. Available: https://www.cision.com/us/2014/02/how-big-data-social-made-house-of-cards-a-hit/. [Accessed: 13-Dec-2018].

[42] “Correlation Ventures.” [Online]. Available: https://correlationvc.com/. [Accessed: 13-Dec-2018].

[43] “Could a ‘Moneyball’ approach help VCs improve their success rate?,” *Fortune*. [Online]. Available: http://fortune.com/2015/08/05/venture-capital-hits-average/. [Accessed: 13-Dec-2018].

[44] V. Savov, “Telefónica will let an algorithm decide which startups to invest in,” *The Verge*, 12-Aug-2015. [Online]. Available: https://www.theverge.com/2015/8/12/9136663/telefonica-open-future-startup-algorithm. [Accessed: 13-Dec-2018].

[45] “How algorithms reproduce social and racial inequality | Salon.com.” [Online]. Available: https://www.salon.com/2018/09/15/how-algorithms-reproduce-social-and-racial-inequality/. [Accessed: 13-Dec-2018].

[46] “China’s New Lenders Collect Invasive Data and Offer Billions. Beijing Is Worried. - The New York Times.” [Online]. Available: https://www.nytimes.com/2017/12/25/business/china-online-lending-debt.html. [Accessed: 13-Dec-2018].

[47] “Online lenders harvest big data to extend loans where banks cannot.” [Online]. Available: https://newsroom.cisco.com/feature-content?type=webcontent&articleId=1725302. [Accessed: 13-Dec-2018].

[48] “Facebook patents technology to help lenders discriminate against borrowers based on social connections,” *VentureBeat*, 04-Aug-2015. .

[49] “Rates & Fees | LendingClub.” [Online]. Available: https://www.lendingclub.com/public/rates-and-fees.action. [Accessed: 13-Dec-2018].

[50] M. Ford, *Rise of the Robots: Technology and the Threat of a Jobless Future*, Reprint edition. New York: Basic Books, 2016.

[51] B. Conway, “Wall Streets Need For Trading Speed: The Nanosecond Age,” *WSJ*, 14-Jun-2011. .

[52] “How the Robots Lost: High-Frequency Trading’s Rise and Fall,” 07-Jun-2013.

[53] B. P. Eha, “Is Knight’s $440 million glitch the costliest computer bug ever?,” *CNNMoney*, 09-Aug-2012. [Online]. Available: https://money.cnn.com/2012/08/09/technology/knight-expensive-computer-bug/index.html. [Accessed: 13-Dec-2018].

[54] *כלכליסט*,” עלייתן של מכונות הבסף“, ש. אספריל - *www.calcalist.co.il*, 14-Jan-2010. [Online]. Available: https://www.calcalist.co.il/local/articles/0,7340,L-3386208,00.html. [Accessed: 13-Dec-2018].

[55] “General Data Protection Regulation (GDPR) – Final text neatly arranged,” *General Data Protection Regulation (GDPR)*. [Online]. Available: https://gdpr-info.eu/. [Accessed: 16-Dec-2018].

[56] “AI in the UK: ready, willing and able,” p. 183.

[57] “Hub-of-All-Things,” *Hub-of-All-Things*. [Online]. Available: https://www.hubofallthings.com/. [Accessed: 13-Dec-2018].

[58] C. Villani, “For a Meaningful Artificial Intelligence,” 2018.

[59] “New China Data Privacy Standard Looks More Far-Reaching than GDPR.” [Online]. Available: https://www.csis.org/analysis/new-china-data-privacy-standard-looks-more-far-reaching-gdpr. [Accessed: 16-Dec-2018].

[60] D. Hoffman and R. Masucci, “Intel’s AI Privacy Policy White Paper,” 2018.

[61] R. Nieva, “Read Google’s AI ethics memo: ‘We are not developing AI for use in weapons,’” *CNET*. [Online]. Available: https://www.cnet.com/news/read-googles-ai-ethics-memo-we-are-not-developing-ai-for-use-in-weapons/. [Accessed: 17-Dec-2018].

מדע וטכנולוגיה

מוסד שמואל נאמן
למחקר מדיניות לאומית

טל. 04-8292329 | פקס. 04-8231889
קרית הטכניון, חיפה 3200003
www.neaman.org.il