בפני יו"ר ועדת הבחירות המרכזית לכנסת ה- 26 **תב"כ 15/26**

כב' השופט נעם סולברג, המשנה לנשיא

1. **פרופ' ניבה אלקין קורן, ראשת מרכז הנשיא מאיר שמגר למשפט דיגיטלי וחדשנות, הפקולטה למשפטים אוניברסיטת תל אביב**
2. **עו"ד עמית אשכנזי, עמית מחקר בכיר, מרכז הנשיא מאיר שמגר למשפט דיגיטלי וחדשנות, הפקולטה למשפטים אוניברסיטת תל אביב**
3. **פרופ' יששכר (איסי) רוזן צבי, ראש מרכז אדמונד ולילי ספרא לאתיקה באוניברסיטת תל אביב**

שמענם לצורך המצאת כתבי בית דין:

הפקולטה למשפטים, אוניברסיטת תל-אביב

רמת אביב, תל-אביב 6997801

טל': 03-6407719

shamgarlaw@tauex.tau.ac.il

<u>המבקשים להצטרף</u>

בעניין:

1. **מפלגת "בנט 2026"**
2. **נפתלי בנט**

ע"י ב"כ עוה"ד נדב ויסמן ו/או עדן יגזאו ו/או עמית מולכו ו/או אח' ממשרד מיתר

דרך אבא הלל סילבר, 16 רמת גן 5250608

טל' 03-6103100 ; פקס' 03-6103111

meitar@meitar.com

<u>העותרים</u>

-נגד-

הליכוד, תנועה לאומית ליברלית

על ידי ב"כ עוה"ד אבי הלוי

מכנרת (מושבה) ד.נ עמק הירדן 15105

טל': 052-4326312 פקס': 04-6709443

avi.halevy@gmail.com

<u>המשיבה</u>

<u>בקשה מטעם המבקשים לצירופם כ"ידידי בית המשפט"</u>

בהתאם לסעיפים 5(ב) ו-33 להוראות הבחירות (דרכי תעמולה) (סדרי הדין בבקשות וערריים), התשע"ה-2015, מתבקש כב' יושב ראש ועדת הבחירות המרכזית, לצרף את המבקשים להליך כמשיבים, במעמד של "ידידי בית המשפט".

עניינה של העתירה הוא בשאלת הלגיטימיות של שימוש בטכנולוגית בינה מלאכותית יוצרת להפקת תכנים של דיפ-פייק (Deep-Fake או זיוף עמוק) כחלק מתעמולת הבחירות שעורכות המפלגות ורשימות המועמדים. ההלכה שתיקבע בעתירה תנחה את המפלגות והמועמדים בבחירות לכנסת ה-26, תכוון את פעילות התעמולה שלהם במהלך מערכת הבחירות ואת אופן השימוש בתקציבי התעמולה. אולם השפעתה של ההלכה שתיקבע אינה רק על המפלגות והמועמדים אלא על כלל ציבור הבוחרים ועל אינטרסים ציבוריים חשובים. שאלת הלגיטימיות לשימוש בבינה מלאכותית יוצרת להפקת דיפ-פייק משליכה על עקרונות המוגנים על ידי חוק דרכי תעמולה, ועל שורה של סיכונים לטוהר הבחירות, הכוללים חשש של ממש להתערבות של מדינות זרות בבחירות, פגיעה באמון הציבורי ביחס למידע המופץ בזירה הציבורית, וזכותם של הבוחרים לבחור ללא השפעה בלתי הוגנת וגניבת דעת.

כמפורט בהרחבה בעמדה זו, העמדה המובאת בה אינה בעד או נגד הצדדים הקונקרטיים להליך. בוחרים פוטנציאליים של מפלגות מכל הקשת הפוליטית עלולים ליפול קורבן לקמפייני דיפ-פייק תוצרת הארץ או תוצרת חוץ, כפי שנפרט בהמשך. מטרת העמדה להציג בפני יושב ראש הוועדה הנכבד את התשתית המשפטית, והמידע המדעי העדכני על אודות הסיכונים שבשימוש בדיפ-פייק, ולהצביע כיצד בנתיבי הפרשנות בהם פסעו יושבי ראש הוועדה בעבר בנושאים אלה, ניתן לצמצם את הסיכונים תוך מתן חופש ביטוי מירבי למפלגות, ומתן אפשרות מלאה ליהנות מהתועלות משימוש בבינה מלאכותית יוצרת.

על כן ההכרעה בהליך חשובה ועקרונית באופן שחורג מהמחלוקת הפרטנית שבין הצדדים, ובפרט בעת הזו. במועד זה עדיין ניתן להבהיר את כללי המשחק, באופן שיאפשר שימוש אחראי בטכנולוגיית הבינה המלאכותית היוצרת, ולהפיק ממנה את תועלותיה, ללא הסיכונים שיתוארו להלן.

מוסד ידיד בית המשפט נועד לאפשר לבית המשפט הדן בסכסוך לצרף להליך גורמים שיש בידם מומחיות ויכולת לסייע לבית המשפט להכריע בסכסוך שבפניו, בהתאם לתבחינים שנקבעו בעניין מ"ח 7929/96 **קוזלי נ' מדינת ישראל**, פ"ד נג(1) 529 553 (1999)). בהתאם לכך, ערכו המבקשים להצטרף חוות דעת מקיפה בנושא, ומבקשים שהיא תונח בפני יושב הראש הנכבד בעת הכרעתו בנושא.

המבקשים להצטרף הם בעלי מומחיות מובהקת בנושא זה, כחוקרים אקדמיים בתחומים הרלבנטיים, ומכוח הניסיון המקצועי בסוגיות אלה. המבקשים הם חוקרים מומחים מובילים בישראל ובעולם למשפט וטכנולוגיה, ובין היתר חוקרים את הסוגייה של שימוש באמצעים טכנולוגיים חדשים, ובמיוחד בבינה מלאכותית, לצורך יצירת תכנים סינתטיים לצרכי השפעה על דעת הקהל. על כן הידע של המבקשים עשוי לתרום תרומה ממשית להכרעה בעניין, מתוך ראייה רחבה.

פרופ' ניבה אלקין-קורן היא פרופסור מן המניין בפקולטה למשפטים באוניברסיטת תל אביב, מופקדת הקתדרה לחדשנות ותיאוריה משפטית ע"ש סטיוארט וג'ודי קולטון, ועמיתת מחקר בכירה במרכז ברקמן-קליין לאינטרנט וחברה באוניברסיטת הרווארד. היא לימדה כפרופסור אורחת בבתי הספר למשפטים באוניברסיטאות מובילות בעולם, ובין היתר הרווארד, אוניברסיטת ניו יורק, אוניברסיטת קולומביה, ואוניברסיטת קליפורניה. היא כיהנה בתפקידי מפתח בארגוני מחקר מובילים בעולם בתחום של משפט וטכנולוגיה, בין היתר יושבת ראש המועצה המדעית המייעצת של מכון אלכסנדר וון הומבולדט לחקר האינטרנט והחברה בברלין (HIIG) גרמניה.

היא עומדת בראש **מרכז הנשיא מאיר שמגר למשפט דיגיטלי וחדשנות** בפקולטה למשפטים אוניברסיטת תל-אביב וחברת הנהלת **המרכז האוניברסיטאי לבינה מלאכותית ומדעי הנתונים**. בטרם הצטרפה לאוניברסיטת תל אביב הייתה פרופ' מן המנין באוניברסיטת חיפה, שם ייסדה את המרכז למשפט

וטכנולוגיה, ואת המרכז לחקר סייבר, משפט ומדיניות. בשנים 2009-2012 כיהנה כדיקן הפקולטה למשפטים.

במסגרת תפקידה כראשת מרכז שמגר, היא מובילה מחקר אקדמי בנושא האתגרים המשפטיים ואתגרי המדיניות שמעורר העידן הדיגיטלי, בדגש על אסדרה של השימוש בבינה מלאכותית, אסדרה באמצעים טכנולוגיים, ואסדרת פלטפורמות דיגיטליות בדגש על זכויות יסוד וחופש הביטוי בעידן הדיגיטלי. מרכז שמגר הוא מרכז מחקר אקדמי המהווה פלטפורמה ללימוד וליבון סוגיות במשפט וטכנולוגיה ולדיאלוג מתמשך בין מעצבי מדיניות, יזמים טכנולוגיים, אנשי תעשייה, חוקרים, סטודנטים והציבור הרחב.

עו"ד עמית אשכנזי הינו עמית מחקר בכיר במרכז הנשיא מאיר שמגר למשפט דיגיטלי וחדשנות, ושימש כיועץ המשפטי של מערך הסייבר הלאומי במשרד ראש הממשלה (2014-2022) ושל הרשות להגנת הפרטיות במשרד המשפטים (2008-2013). בשנת 2018-2019 היה חבר עו"ד אשכנזי בתת ועדה לעניני אתיקה בבינה מלאכותית, שמונתה כחלק מהמיזם הלאומי למערכות נבונות, וחיבר את פרק האסדרה בדוח. בשנת 2019 ייצג עו"ד אשכנזי את ישראל בגיבוש המלצות ה- OECD בעניין בינה מלאכותית. בשנת 2022 ייעץ עו"ד אשכנזי למשרד החדשנות, המדע והטכנולוגיה בעניין גיבוש עקרונות מדיניות רגולציה ואתיקה בתחום הבינה המלאכותית. עו"ד אשכנזי חוקר את תחום ההתערבות הזרה במסגרת פרויקט מחקר במכון למחקרי ביטחון לאומי INSS משנת 2023. עו"ד אשכנזי חבר בוועדת המומחים החיצונית של ה- OECD לבינה מלאכותית מהימנה, וברשת מומחי הגנת הסייבר האירופית EU CyberNet. עו"ד אשכנזי מרצה בקורסים באוניברסיטאות בישראל בנושאים הקשורים למשפט וטכנולוגיה, ניהול סיכונים, הגנת הסייבר, הגנת הפרטיות, בינה מלאכותית מהימנה, וביטחון לאומי.

פרופ' ישששכר (איסי) רוזן-צבי הוא פרופסור מן המניין בפקולטה למשפטים באוניברסיטת תל אביב ומנהל מרכז אדמונד ולילי ספרא לאתיקה. תחומי מומחיותו המרכזיים הם משפט ציבורי, משפט השלטון המקומי, משפט מנהלי, סדר דין אזרחי ובתי המשפט, לצד עיסוק מתמשך בשאלות יסוד של דמוקרטיה ומנגנוני פיקוח ובקרה על מערכות שלטוניות. מחקריו התפרסמו בכתבי העת המשפטיים המובילים בעולם, ובהם *Stanford Law Review, University of Pennsylvania Law Review, UCLA Law Review, Cornell Law Review, Law & Society Review* ו *Journal of Empirical Legal Studies*. הוא הוזמן לשמש מרצה אורח באוניברסיטאות מובילות בעולם, ובהן Northwestern ו- Cornell בארה"ב, SciencesPo, Goethe-Universität ו- Oñati International Institute for the Sociology of Law.

במסגרת תפקידו כמנהל מרכז אדמונד ולילי ספרא לאתיקה באוניברסיטת תל אביב, מוביל פרופ' רוזן-צבי צוות מחקר אינטר-דיסציפלינרי הכולל משפטנים, חוקרי מדע המדינה, פילוסופים, כלכלנים וסוציולוגים, ועוסק בפיתוח ידע תיאורטי ויישומי בנושאי ליבה של המשפט הציבורי ואתגרי הדמוקרטיה. פעילות המרכז משלבת מחקר אקדמי, גיבוש ניירות עמדה, כנסים מקצועיים ויוזמות ציבוריות, במטרה להעמיק את השיח הנורמטיבי והמעשי על אודות מוסדות השלטון ומערכת היחסים ביניהם, שלטון החוק, ומנגנוני פיקוח ובקרה במשטרים דמוקרטיים (ובהם הפרדת רשויות אופקית ואנכית, מנגנוני בחירות, משאלי עם ועוד). במסגרת זו נבחנים תפקודם ועיצובם של מוסדות השלטון וכן שאלות הנוגעות לאיזונים בין רשויות, לאחריותיות, לשקיפות ולשמירה על זכויות אדם ואזרח. עבודתו במרכז מתמקדת גם בזיהוי ובניתוח האתגרים המוסדיים והמשפטיים הניצבים בפני דמוקרטיות בישראל ובעולם, בעידן של נסיגה דמוקרטית, ומקומם של מוסדות דמוקרטיים בעיצוב מרחב ציבורי יציב ומבוסס נורמות.

יודגש כי העמדות המובעות כאן הן עמדותיהן המקצועיות האישיות של המבקשים להצטרף מכוח מומחיותם ואינה מייצגת את המוסדות או הארגונים בהם הם פועלים.

נוכח האמור, ומתוך ראייה רחבה ורצון כי בפני יושב הראש הנכבד תעמוד חוות דעת מקצועית, בלתי תלויה, שמנתחת את הסוגייה, המבקשים סבורים כי צירוף חוות דעתם להליך תסייע להכריע במחלוקת הפרטנית. המבקשים להצטרף מסתפקים בהגשת חוות דעתם וצירופה לתיק. לכן אין בקבלתם כצד בהליך כדי להכביד על הדיון או לסרבל אותו.

יש לציין כי אף ששאלת צירופם של צדדים שלישיים במעמד של ידיד בית המשפט, בהליכים בפני יו"ר ועדת הבחירות, טרם נדונה לעומק, הרי שיושבי ראש ועדת הבחירות בעבר אכן אפשרו לגורמים מומחים לצרף חוות דעת בתיקים שהתנהלו בפניהם, ונעזרו בהם. למשל: תב"כ 8/21 **עו"ד שחר בן מאיר נ' מפלגת הליכוד** (27.2.2019); ער"מ 2/25 **עיריית טירת הכרמל אריה טל נ' סיעת הילה שלם והתושבים** (3.10.2022).

זהו הפתיח. דרך הילוכנו יהא כדלהלן: ראשית נביא את עקרי העמדה המקצועית של המבקשים, נסקור את הסיכונים מהפקת תוכן סינתטי בכלל ובמערכת בחירות בפרט, ובכלל זה נעמוד על הסיכונים מהשפעה זרה. לאחר מכן נדון בתחולתו של סעיף 13 לחוק הבחירות (דרכי תעמולה), תשי"ט-1959 (להלן – **חוק דרכי תעמולה**) והסעדים הראויים.

תוכן העניינים

עיקרי העמדה

1. לחוק דרכי תעמולה תכלית משולשת. ראשית, להבטיח שוויון בין המתמודדים השונים[1] ולמנוע מצב שבו בעלי ממון, השפעה או המפלגות שנמצאות בשלטון ישתלטו על המרחב הציבורי וישתיקו את יריביהן[2]; שנית, למנוע השפעה לא עניינית על הבוחר באמצעים לא הוגנים, כמו למשל בידור, מתנות וכד'; שלישית, להגן על חופש הביטוי הפוליטי בכך שהוא קובע כללי משחק אחידים, ומאזן בין התכליות השונות להגשמת הזכות לבחור ולהיבחר.[3] תכליות אלה מגולמות כולן בסעיף 13 לחוק.

2. אנו מגישים עמדה זו על מנת לסייע בבחינת היישום של חוק דרכי תעמולה על שימוש בכלים טכנולוגיים, ובכלל זה כלי "בינה מלאכותית",[4] להפקת "מודעת בחירות" כהגדרתה בחוק שהינה דיפ-פייק.[5]

3. דיני התעמולה, ככל שאינם מגבילים עצמם למדיום מסוים, חלים על פרסומים שעניינם השפעה על הבוחר בכל מדיום שהוא, בכפוף לתחולתם בזמן (סעיף 2 לחוק; תב"כ 8/21 **בן מאיר נ' הליכוד**; תב"כ 27/21 **ישראל ביתנו נ' שמיר מערכות**). עמדתנו, בתור מומחים למשפט וטכנולוגיה היא כי הדין והאינטרס הציבורי, מחייבים כי תובהר אופן תחולת החוק על שימוש של "מתמודד בבחירות", כהגדרתו בחוק, בכלים טכנולוגים, על מנת למנוע תקלות, אי הבנות, וסיכונים של ממש לתכליות והאינטרסים הציבורים המוגנים באמצעות החוק, ובפרט "הפרעה בלתי הוגנת".

4. כידוע, החוק מסדיר את דרכי התעמולה לצורך "הבטחת חופש הבחירה וטוהר הבחירות והבטחת שוויון בין המפלגות בניסיונן להגיע אל הבוחר".[6] אלה נועדו להבטיח את מימוש הזכות החוקתית המעוגנת בחוק יסוד: הכנסת לבחור ולהיבחר באופן שוויוני, וכל זאת בהתחשב בזכות החוקתית לחופש הביטוי הפוליטי, לצד השמירה על השוויון ועל דעתו הנקייה של הבוחר.[7] החוק נועד להבטיח מסירת "מידע

---

[1] בג"צ 524/83 **האגודה למען החייל בישראל נ' המנהל הכללי של רשות השידור** פ"ד לז (4) 85, 95 (1983).
[2] ראו דבריו של חה"כ ורהפטיג, בהציגו את הצעת החוק לכנסת; ד"כ 27 (התשי"ט), עמ' 2,251.
[3] בג"ץ 869/92 **זוילי נ' יו"ר ועדת הבחירות המרכזית לכנסת ה-13**, פ"ד מו(2) 692 (1992).
[4] הכוונה בעמדה זו הינה בעיקר ל"בינה מלאכותית יוצרת" (Generative AI) המייצרת תוכן סינתטי, המתוארת במסמך "עקרונות מדיניות רגולציה ואתיקה בתחום הבינה המלאכותית" שפרסמו משרד החדשנות, המדע והטכנולוגיה ומשרד המשפטים בשנת 2023 (להלן – **מסמך העקרונות**) כך: **מערכות "בינה מלאכותית יוצרת" מסוגלות ליצור תוכן מסוגים שונים, לרבות טקסט, וידיאו, ושמע, באיכות גבוהה המאתגרת את היכולת האנושית להבחין בין תוכן אמיתי לבין כזה אשר נוצר על ידי מכונה.**" (שם, בעמוד 8). בדוח מאוחר יותר, "בינה מלאכותית במגזר הפיננסי" (להלן – **דוח המגזר הפיננסי**), הדוח מגדיר מערכת בינה מלאכותית יוצר כ"סוג של למידת מכונה אשר מסוגל לייצר תוכן במדיות שונות: כתיבה, תמונות, מוסיקה ועוד. בינה מלאכותית יוצרת היא תוצר של אלגוריתם המתאמן על כמויות עצומות של נתונים לא מעובדים [...] ולומד דפוסים שמאפשרים למודל לייצר תכנים מתאימים בהתאם לשאלות או הוראות שמספקים לו. הבינה המלאכותית היוצרת היא אחת הטכנולוגיות המרכזיות שתורמות לפעולות דיסאינפורמציה". בדוח המגזר הפיננסי בחרו עורכי הדוח להגדיר "בינה מלאכותית" באופן שמותאם להגדרה של ארגון ה- OECD, כדלקמן: "ההגדרה המוצעת על-ידי הצוות לבינה מלאכותית בסקטור הפיננסי מבוססת בין היתר על הגדרות הOECD -והאיחוד האירופאי, בהתאמות המתבקשות לפעילות בסקטור זה: "מערכת בינה מלאכותית" – 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". יצוין כי הגדרת ארגון ה- OECD אשר עודכנה בשנת 2024 הינה: "AI system: An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment." https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#adherents. באופן כללי, בינה מלאכותית יוצרת המכונה גם בינה מלאכותית מחוללת היא תחום מתפתח בתוך הבינה המלאכותית, אשר עוסק ביצירת תוכן חדש כמו טקסט, תמונה, קול, וידאו או קוד על בסיס דפוסים שנלמדו ממאגרי נתונים רחבים. בניגוד לאלגוריתמים מסורתיים של חיזוי או סיווג, אשר מנתחים מידע קיים, מערכות גנרטיביות פועלות באופן יצירתי ומפיקות מידע חדש, לכאורה "יש מאין". מודלים אלה מבוססים לרוב על רשתות נוירונים עמוקות (Deep Neural Networks) ובפרט על מודלים גדולים של שפה (LLMs) כמו GPT המסוגלים להבין שפה טבעית ולהגיב בהתאם. Cole Stryker and Mark Scapicchio, What is generative AI? IBM Think, March 22, 2024, https://www.ibm.com/think/topics/generative-ai. במדינת קליפורניה מוגדרת מערכת בינה מלאכותית יוצרת (Generative) כ"בינה מלאכותית אשר יכולה לחולל תוכן סינתטי, לרבות טקסט, תמונה, וידאו ואודיו אשר מחקה מבנה או תכונות של המידע עליו אומנה המערכת." CAL. CIV.CODE § 3110 (3110(a), (c) § 2024).
[5] הביטוי "Deepfake" מחבר בין המילה "deep" המלמדת על אופן הפקת התוכן הסינטטי, באמצעות רשתות למידה עמוקה, לבין המילה "fake" העוסקת בהיותו של הפלט הסינתטי מזוייף או כוזב. לסקירה מקיפה ראו: Robert Chesney & Danielle K. Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 California Law Review 1753 (2019). Available at: https://scholarship.law.bu.edu/faculty_scholarship/640, בעמוד 1758-1763.
[6] תב"כ 8/21 **שחר בן מאיר נגד מפלגת הליכוד** (27.2.2019)(להלן – **תב"כ 8/21**), בפסקה 68.
[7] דנג"צ 1526/15 **טיבי נ' ישראל ביתנו** (23.08.2017); תב"כ 8/21, בפסקה 68.

מלא ולעודד שקיפות על מנת שהבוחרים יצביעו על פי צו מצפונם ולא על סמך אינטואיציות והנחות לא מבוססות".[8] תכליות אלה באות לידי ביטוי גם ביחס לשימוש במדיום הדיגיטלי להפקה והפצה של תכנים לצרכי תעמולה, בין היתר גם מסעיף 2א1 לחוק, שעניינו "שקיפות בתעמולת בחירות", והחלטות יושבי הראש של ועדת הבחירות המרכזית ביחס לסעיף 13 לחוק.

5. השימוש הגובר של מפלגות בכלים טכנולוגיים ובבינה מלאכותית יוצרת להפקת דיפ-פייק מסוגים שונים, כלומר תכנים שאינם אותנטיים, או תכנים שנראים מאוד מציאותיים, אך המתארים באופן כוזב אירועים שהתרחשו במציאות, ללא סימון מתאים, מעלה חשש כי שחקנים פוליטיים מניחים בטעות כי קיומה של טכנולוגיה המאפשרת בקלות הפקת תוכן כוזב, מטעה או מניפולטיבי, הופכת שימושים אלה למותרים.[9] מדובר בגישה שגויה, ההופכת את היוצרות – הטכנולוגיה היא אמצעי ולא מטרה, וקיומן של יכולות טכנולוגיות זולות ונגישות להפקת תוכן סינתטי אינו מתיר מעשים אסורים.

6. בעמדה זו, אנו מבקשים לעמוד על כך, שהמשך הלגיטימציה השגויה להפקה והפצה של דיפ-פייק באמצעות כלי בינה מלאכותית יוצרת, עלולה להוביל לשורה של סיכונים, שהחוק מבקש להגן מפניהם, החל מהטעיית בוחרים, וכלה בפגיעה בטוהר הבחירות, ובכלל זה התערבות זרה:

א. **הטעיית בוחרים** - ניתן להעריך כי חלק משמעותי מנמעני המסר, הבוחרים, יניחו כי תוכן סינתטי מציאותי מסוים [כגון התמונה נושא עתירה זו] הינו אותנטי, ויתחשבו בו כחלק מגיבוש עמדתם. מחקרים עדכניים מלמדים על קושי של ממש של בני אדם לזהות תוכן סינתטי ככזה ולהבחין כי אינו אמיתי.

ב. הכבדת הנטל על הבוחר – הציבור המבקש לגבש את עמדותיו בשוק הרעיונות התוסס, נדרש עתה לעסוק לא רק במטלה (הנכבדה) של איסוף ועיבוד המידע הנדרש לגיבוש עמדתו, אלא להשקיע בבירור בסיסי האם המידע המוצג אותנטי או מזויף. ודוק, השאלה אינה האם הבוחר נדרש לחשוב האם המסר הוא אמת אם לאו, אלא קודם כל האם מה שרואות עיניו ושומעות אזניו – הוא מידע סינתטי או אותנטי, ובמילים אחרות, האם איננו מזויף. אין מדובר בחשש מטעות, אלא מהטעיה וגניבת דעת.

ג. **תחרות לא הוגנת** - שימוש באמצעים טכנולוגיים לשם הטעיה והפרעה בידי מפלגות יוצר יתרון בלתי הוגן ביחס למפלגות המקפידות על עקרונות ההוגנות ופוגע בשוויון בתחרות בין המפלגות השונות. זאת, גם כאשר מדובר בשימוש לצורך תעמולה עצמית, שכן כל תעמולה עצמית נועדה לקדם את מועמדותך, נגד מועמדותו של היריב.

ד. חשש ל "מירוץ לתחתית" בין המפלגות בהפקת מידע כוזב - מתן לגיטימציה להפקת תוכן סינתטי המתחזה לאותנטי בידי מפלגות העומדות לבחירת הציבור, כחלק ממהלכי השכנוע שלהן, עלול לחולל "מירוץ לתחתית" בין המפלגות לשימוש בטכנולוגיה ליצירת תכנים מניפולטיביים ככל הניתן, המערערים את היכולת להבחין בין מה שאותנטי למה שכוזב, ומהווים גניבת דעת. כמפורט להלן, הדבר מהווה גם כר פורה להתערבות זרה בבחירות.

---

[8] תב"כ 8/21, בפסקה 70.

[9] ניתן להצביע על שימושים בטכנולוגיה זו לאחרונה בידי מפלגת עוצמה יהודית, מפלגת ישר, מפלגת כחול לבן ומפלגת ישראל ביתנו. יצוין כי אין כיום מאגר מידע שקוף הכולל את כלל התעמולה שנעשה בה שימוש בדיפ-פייק, וגם בכך יש אתגר משמעותי כפי שיפורט בהרחבה להלן.

ה. **ערעור האמון הציבורי במידע רלוונטי למערכת הבחירות וזיהום השיח הציבורי** – שכיחות גבוהה של תכנים סינתטיים המתחזים להיות אותנטיים תביא לפגיעה בנכונות של הציבור להאמין גם למידע אותנטי המופץ במסגרת תעמולה בידי המפלגות. זאת בפרט, ביחס למפלגות המקדמות דעות השונות מזה של הבוחר, ובכך לחתור עוד יותר תחת הרעיון של שוק רעיונות חופשי ותוסס.

ו. **פגיעה באמון הציבור באופן כללי במוסדות ובתהליכים הדמוקרטים** – מתן לגיטימציה להפצה והפקה של דיפ-פייק בידי השחקנים המרכזיים של הזירה הציבורית, המפלגות הפוליטיות ונבחרי הציבור, עלול לערער את האמון הציבורי במוסדות ובתהליכים הדמוקרטיים בכלל, לרבות הטלת ספק במהימנות של פרסומים מטעם גורמים רשמיים, כמו ועדת הבחירות המרכזית.

ז. **חשיפה להתערבות של גורמים זרים** - גורמים זרים עוינים שמטרתם המרכזית הגברת הקיטוב וחוסר האמון במוסדות (ללא קשר לתוכן בהכרח), מנצלים כלים טכנולוגיים אלו, כדי להציף את המרשתת בתכנים סינתטיים, כדי להתחזות לשחקנים פוליטיים ודוברים פוליטיים, מחמירים את ה- "ההפרעה הבלתי הוגנת" והכל באופן שמביא לפגיעה מערכתית נרחבת יותר ובכלל זה סיכון חמור לאיתנות המערכת הדמוקרטית ובפרט הליך הבחירות. בהקשר זה, אוי סימון תכנים סינתטיים, מקשה על איתור מעורבות זרה.

7. **יודגש כי אין מדובר בחששות תיאורטיים כי אם בסיכונים שהמחקר העדכני מצביע עליהם, ושבאו לביטוי בשימוש בכלים טכנולוגיים, בזירה הציבורית ברחבי העולם ובפרט סביב מערכות בחירות, וביחס לפעילות גורמים עוינים בשנים האחרונות במרחב הדיגיטלי בישראל**, כפי שנפרט בפרק ב' להלן. לא זו אף זו, הרגישות של מערכת הבחירות ורגעי השיא שלה (כגון סמוך ליום הבחירות) לתרחישי הטיה שונים, מחמירים מאוד את הסיכונים.

8. לצד זאת, אין חולק כי השימוש בכלים טכנולוגים להפקה של תעמולת בחירות מאפשר למפלגות העברת מסרים בדרך יצירתית וחדשנית ובאופן מוחשי. יכולות אלה מסייעות למפלגות, ובפרט למפלגות קטנות בעלות משאבים נמוכים יותר, בפעילותן בזירה הציבורית וב-"שוק הרעיונות". לפיכך, לשימוש **אחראי** בכלים טכנולוגים במסגרת גבולות מוגדרים עשוי להיות פוטנציאל בקידום "שוק רעיונות" תוסס שהינו חלק חיוני משיח דמוקרטי ובפרט בתקופת בחירות בה הבוחר אוסף ומעבד מידע לצורך גיבוש דעותיו ומימוש זכות הבחירה שלו.

9. על רקע האמור, כפי שנקבע בפרשת **טיבי**, פרשנו המוסמך של הדין בענייני בחירות נדרש לאזן בין חופש הביטוי הפוליטי של המפלגות והיכולת שלהן להשתמש בטכנולוגיה חדשנית במסגרת הדין, לבין החשש מהטיה אישית ומערכתית כוללת שיש בה גניבת דעת הבוחר, פגיעה "מובנית" ואינהרנטית בתחרות ההוגנת על קולו, פגיעה באמון וחשיפה להתערבות זרה.

10. הליכה בנתיבי פרשנות שנסללו עד ידי יושבי ראש ועדת הבחירות ביחס למשפט וטכנולוגיה, תוך יישום עקרונות קונקרטיים שאומצו בישראל בתחום ה-"בינה המלאכותית המהימנה"[10], מאפשר לבחון את

[10] הביטוי "בינה מלאכותית מהימנה" ("Trustworthy AI") מבוסס על המלצות ארגון המדינות המפותחות (ה- OECD) בנושא בינה מלאכותית, שהינו המסמך הבינלאומי הראשון הכולל עקרונות אתיים ומשפטים לבינה מלאכותית. ישראל לקחה חלק בניסוח ההמלצות והצטרפה אליהם. ראו: OECD, Recommendation of the Council on Artificial Intelligence, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

הסיכונים ולקבוע אמצעים לצמצומם באופן שמגשים את התכליות של החוק. כפי שנציג להלן, בחינה זו מובילה למסקנה כי המאפיינים **הטכניים והצורניים** של שימוש בבינה מלאכותית יוצרת הם אלה שיוצרים את הסיכונים לעיוות מערכתי של זירת השיח הציבורי, וכפועל יוצא מכך לסיכונים שתוארו לעיל להתערבות זרה, הטעיה, פגיעה באמון במידע בזירה הציבורית, ופגיעה בשוויון ההזדמנויות בבחירות. בהתאמה לכך, תוך שימוש באותם כלים ועקרונות, **התרופה לסיכונים הנובעים ממאפיינים אלה אף היא צורנית**, **בדרך של הטלת חובת סימון על התוכן הסינתטי**, כלומר – שהתוכן המדובר אינו אותנטי אלא נוצר באמצעים טכנולוגיים.

11. יודגש כי מענה טכנולוגי צורני זה, חובת סימון, משקף את הכללים המתפתחים במדינות המפותחות (ובכלל זה ישראל), הן במדיניות הפנים והן בהתחייבויותיה הבינלאומיות. לא פחות חשוב מכך, כלל זה מהווה נורמה וולונטרית מתפתחת בשדה הטכנולוגיה הגלובלי, כפי שיוצג להלן. נורמה זו באה לידי ביטוי בפרקטיקות המתגבשות בקרב השחקנים הטכנולוגים האחראים, וכל זאת במטרה לאפשר מיקסום התועלות מטכנולוגיה חדשנית זו, וצמצום הסיכונים ממנה. יחד עם זאת, אופן היישום אינו בהכרח זהה, פרט לוולונטריות של חלק מהפלטפורמות – ישנן פלטפורמות רבות, וכלים נוספים שאינם קובעים כללים אלה, ולכן יש צורך להחיל אותם גם על המתמודדים עצמם.

12. לבסוף, כפי שיפורט להלן, ידוע שהאתגר של התמודדות עם דיפ-פייק הינו אתגר מורכב, ואין לו "פתרונות קסם". לצורך התמודדות עם הסיכונים של הפקה והפצה של דיפ פייק נדרשת גישה רב שכבתית. לצד חובת הסימון שתפורט להלן, השחקנים השונים במערכת (מנהלי הקמפיין, המפלגות, ערוצי השידור והפלטפורמות) נדרשים להציג את הסימון באופן ברור גם בהמשך הפצתו. שנית, על מנת למנוע הסתננות של גורמים עוינים רצוי שיהיה מאגר פומבי מזוהה של כלל הפרסומות שהופקו בידי המפלגות כמקור אמין של התוכן שהפיקו. שלישית, על הגורמים המוסמכים ובהם וועדת הבחירות להיערך לתרחישי הפצה זדוניים של מידע סינתטי בידי גורמים עוינים ברגעים קריטיים של מערכת הבחירות. עם זאת, הטלת חובת הסימון הינה **נדבך ראשון**, המסייע בהגנה על זירת המידע ומאפשר מיקוד מאמצי הגנה נוספים ומשלימים.

13. בהתאם לכך, מהנימוקים שיוצגו להלן, מוצע כי שימוש של מפלגות או מי מטעמן בבינה מלאכותית יוצרת או אמצעים טכנולוגיים דומים להפקת תכנים סינתטיים המתחזים להיות אותנטיים יבוצע בהתאם לכללים הבאים:

א. תוכן סינתטי יסומן באופן בולט כיציר בינה מלאכותית/אמצעים טכנולוגיים, ויכלול גם סימן מים טכני, כלומר סימון טכנולוגי קריא מחשב, על גבי התוכן בדבר היותו יציר בינה מלאכותית;[11]

ב. במידה שהתוכן הסינתטי מתיימר להציג **תיאור עובדתי** של אירועים שהתרחשו טרם הפקת התוכן הסינתטי, ייכלל בתמונה, בקול או בסרטון בכל זמן הצגתם, כיתוב בולט כי: "התוכן אינו מהווה תיעוד של אירועים אמיתיים אלא נוצר באמצעים טכנולוגים לצרכי תעמולה".

ג. המפלגה שיצרה או מטעמה פורסם התוכן הסינתטי תכלול את שמה באופן בולט על גבי התוכן (כנדרש לפי סעיף 2א1 לחוק).

[11] סימן מים טכנולוגי מאפשר שמירה על היכולת לבדוק את האותנטיות של הקובץ, וכן מאפשר לפלטפורמות הצגת התוכן להציג בעצמן כי התוכן הינו יציר בינה מלאכותית, ובכך לסייע בהבאת מידע זה לידיעת הציבור והגנה טובה יותר מפני שימוש לרעה של גורמים זדונים המבקשים להסיר את הסימן. ראו עוד להלן.

ד. חובת הסימון בתעמולת בחירות שכוללת תוכן סינתטי, תהיה אחידה ושוויונית כדי שהסימון יהיה משמעותי ובולט ויגשים את התכלית של מניעת הטעיה וגניבת דעת, ותכלול גם סימון מים טכני, בהתאם לכללים המקובלים, כדי שהסימון יהיה משמעותי ובולט ויגשים את התכלית של מניעת הטעיה וגניבת דעת, ויקל על בעלי העניין השונים, גורמי הביטחון, המפלגות, כלי התקשורת, גורמי חברה אזרחית והבוחרים לזהות את אותו תוכן.

ה. תוכן סינתטי שלא ניתן לסמן– ייאסר.

14. יישום כללים אלה יביא להפחתה משמעותית של הסיכונים שתוארו לעיל, כמפורט להלן:

סיכון	אמצעי לצמצום הסיכון	השפעה
הטעיית בוחרים כי מדובר בתוכן אמיתי	סימון התוכן באופן בולט כיציר בינה מלאכותית, וכי מקורו במפלגה	מאפשר למפלגות לעשות שימוש בטכנולוגיה זו מבלי חשש להטעיה; מאפשר לפלטפורמות מפיצות תוכן לזהות תוכן זה כתוכן פוליטי לגיטימי ולפעול בהתאם לכללי ההפצה הקשורים אליו;
	שימוש בסימונים טכניים;	מאפשר למפלגות לעשות שימוש בטכנולוגיה זו; מאפשר להגן על יכולת הזיהוי בהמשך שרשרת הערך, בידי הפלטפורמות ובידי גורמים אחרים המבקשים להתחקות אחר מקורו, ולפעול בהתאם לכללי ההפצה הקשורים אליו; מאפשר לצמצם סיכונים של מחיקת סימנים מזהים;
	איסור על תיאור אירועים עובדתיים שהתרחשו בעבר ללא סימון מתאים;	מאפשר שימוש יצירתי "מוקומנטרי" תוך הבחנה בין תיאוריים עובדתיים לבין הבעת דעות או ביקורת;
פגיעה בשוויון ההזדמנויות בבחירות	סימון התוכן באופן בולט כיציר בינה מלאכותית, וכי מקורו במפלגה;	יצירת מרחב השפעה הוגן ושוויוני לכלל השחקנים הפוליטיים לשימוש בטכנולוגיה זו;
	איסור על תיאור אירועים עובדתיים שהתרחשו בעבר ללא סימון מתאים;	יצירת מרחב השפעה הוגן ושוויוני לכל השחקנים הפוליטיים;
ערעור האמון במפלגות ובמערכת	סימון התוכן באופן בולט כיציר בינה מלאכותית, וכי מקורו במפלגה;	מאפשר הבחנה ברורה בין תוכן סינתטי לתוכן אותנטי

הפוליטית, בתהליכים דמוקרטים בכלל	איסור על תיאור אירועים עובדתיים שהתרחשו בעבר ללא סימון מתאים;	מאפשר הבחנה ברורה בין תוכן סינתטי לתוכן אותנטי
חשיפה להתערבות זרה	סימון התוכן באופן בולט כיציר בינה מלאכותית, וכי מקורו במפלגה;	מאפשר לפלטפורמות לזהות תוכן זה כתוכן פוליטי לגיטימי שאין להתערב בו; מאותת לגופי הביטחון כי מדובר בשיח פוליטי פנימי שאין להתערב בו;

15. למען שלמות התמונה יודגש כי העמדה המובאת כאן אינה מתיימרת לכסות את כלל ההיבטים הקשורים בשימושים טכנולוגים ויישומי בינה מלאכותית בבחירות, (כפי שנעשה לאחרונה גם בהקשרי יישומי בינה מלאכותית במגזר הפיננסי.[12]) נושא זה ראוי לטיפול הוליסטי בידי הגורמים המוסמכים, לאחר בחינת מכלול ההיבטים הקשורים בכך. בהמשך לכך נציין כי לאחר מערכת הבחירות האחרונה, התייחס יו"ר ועדת הבחירות המרכזית, השופט עמית, לנושא זה:

> **"סעיף 2א1 לחוק דרכי תעמולה, שהתווסף בתיקון תשפ"ב- 2022, וחייב שקיפות במסרי תעמולה ממומנים או מטעמו של מתמודד בבחירות, היה אך הצעד הראשון בהתמודדות עם דרכים בלתי לגיטימיות להטיית דעת הבוחר ולהפצה של מסרים תעמולתיים או מטעים תחת כסות.**
>
> **לדעתי, יש צורך בצעדים נוספים על מנת להרחיב את חובת השקיפות ולהסדיר במרחב הדיגיטלי תופעות שנועדו לבצע מניפולציות על דעת הקהל, כגון: שימוש בטכנולוגיות של "דיפ פייק"; יצירת רשתות השפעה להדהוד מסרים והעצמתם בקרב הבוחרים על ידי שימוש בחשבונות פיקטיביים המופעלים על ידי בוטים, בין היתר, לשיתופים וחיבובים (לייקים); פרופילים מזוייפים תחת שמות בדויים או אף שמות של אחרים (פייקים); קבוצות "בתחפושת" שנועדו להריץ מפלגה או פוליטיקאים; הקמת רשת של חשבונות מזוייפים (אוואטרים) וכיו"ב תופעות של התנהגויות לא אותנטיות ברשתות החברתיות.**
>
> **להרחבה ראו: תהילה שוורץ אלטשולר וגיא לוריא, "תעמולה דיגיטלית והאיום על הבחירות", המכון הישראלי לדמוקרטיה דצמבר 2020")."[13]**

16. להשלמה יש לציין את ההסדרה המקיפה, הרב שכבתית, של תחום זה באיחוד האירופי הכולל כללי התנהגות לשחקנים השונים בשרשרת הערך של הפרסום הפוליטי בהוראות קונקרטיות, לצד הסדרים החלים בתחום הפצת תכנים ובתחום הבינה המלאכותית. תיאור מפורט של הסדרה זו לא נכלל בעמדה זו (מפאת קוצר הזמן) ומיקוד הדברים בדין הישראלי ופרשנותו.

17. עוד יצוין כי לדעתנו אין המדובר בפרשנות העולה בגדר הסדר ראשוני, בעל השלכה חדשה על זכויות יסוד ואינטרסים ציבורים, אשר מחייבת הסדרה בחקיקה ראשית. יישום החוק בדרך של הצורך **בסימון צורני טכני** ביחס לשימוש בתוצרי בינה מלאכותית יוצרת הינו פרשנות של סעיף 13 לחוק, על אמצעי

[12] ראו דוח המגזר הפיננסי, הערת שוליים 4 לעיל.
[13] ראו: ועדת הבחירות המרכזית, לקחי יו"ר ועדת הבחירות המרכזית מהבחירות לכנסת ה- 25, https://www.gov.il/BlobFolder/guide/knesset-elections-law/he/laws_old_decision_books_Lessons_from_cec_chairman_25.pdf

התעמולה העדכניים. גישה זו היא גם בהלימה לאופן המדוד והזהיר שהדין בישראל מתפתח ביחס ליישומי בינה מלאכותית באופן כללי, כפי שנציג להלן.

18. לבסוף יודגש כי העמדה המובאת כאן היא עמדתנו כמומחים בלתי תלויים. אנו מבקשים שבבוא יושב הראש להכריע בסוגייה, הוא יבחן גם את החשש כי מתן הכשר להפצת מידע סינתטי, ללא סימון, יוביל ל"מירוץ לתחתית" בין השחקנים הפוליטיים באופן שיחתור תחת התכליות היסודיות של כללי הבחירות בכלל ועל תעמולת בחירות בפרט, יפגע קשות בציבור הבוחרים, ביכולתם לגבש עמדה מושכלת, ויפגע באמון הציבורי בבחירות. לעומת זאת, קביעת כללים מנחים כמוצע בעמדה זו, יקבעו כללי משחק הוגנים על מנת למצות את התועלות מכלים טכנולוגים לצרכי תעמולה והעברת מסרים, תוך צמצום החשש מהסיכונים הנובעים מכלים אלה.

<u>הסיכונים מהפקת תוכן סינתטי</u>

19. על מנת להסביר מדוע שימוש בבינה מלאכותית יוצרת להפקת "דיפ פייק" מייצר סיכון ממשי להפרעה בלתי הוגנת בבחירות, נסקור בקצרה את הידע הקיים על אודות טכנולוגיה זו, את המחקר הקיים אודות השלכותיה החברתיות ובפרט המחקר אודות השימוש לרעה שנעשה בה במסגרת מערכות בחירות, וכחלק ממבצעי השפעה איראניים בישראל. יובהר כי הדברים מוצגים בקצירת האומר לנוכח לוחות הזמנים להגשת עמדה זו.

<u>תוכן סינתטי, יכולתו להטעות, ושימושיו במניפולציות מסוגים שונים</u>

20. במישור הטכני, ניתן לעשות שימוש בבינה מלאכותית יוצרת להפקת דיפ-פייק, תוכן סינתטי המתחזה לאותנטי, מסוגים שונים, הכוללים תמונה, קול, סרטונים וטקסט.[14] דוח ועדת המומחים הבינלאומי על "בטיחות AI", שפורסם בפברואר 2026 מדגיש את הסיכון **לייצור תוכן ריאליסטי מזויף** כסיכון המרכזי מבינה מלאכותית יוצרת. הדוח נערך בידי שורה של מומחים בינלאומיים, לבקשת מוסדות בינלאומיים, על מנת להניח בפני קובעי מדיניות ברחבי העולם תשתית עובדתית הנדרשת לפעילותם. הדוח, בהובלת חתן פרס טיורינג, יהושע בנג'יו, חובר בהשתתפות למעלה מ-100 מומחי בינה מלאכותית, ובשיתוף למעלה מ-30 מדינות וארגונים בינלאומיים. הדוח כולל סקירה מקיפה ועדכנית של המחקר המדעי העדכני ביותר בדבר היכולות והסיכונים של מערכות בינה מלאכותית.[15] הדוח מציין כי השיפור בכלי ה-AI בשנת 2026 הביא לכך שקשה הרבה יותר להבחין בין תוכן אמיתי לבין תוכן יציר AI. בנוסף, הדוח מציין כי אף שקיימים כלים אלגוריתמיים לזיהוי תוכן סינתטי, הם לא זמינים לציבור הרחב באותה מידה שהתוכן זמין.

21. מחקרים רבים מצביעים על ההטעיה המוחשית שנגרמת כתוצאה מתוכן סינתטי הנחזה לייצג נאמנה את המציגות. במחקר רוחב (meta analyses) שנערך בשנת 2024 (ומצוטט בדוח ועדת המומחים הבינלאומית), נבחנו באופן מקיף היכולות של אנשים לזהות תוכן סינתטי ככזה. המחקר מסיק כי היכולת לזהות תוכן סינתטי היא **אקראית**, וכי גם כאשר ננקטים אמצעים לשיפור יכולות הזיהוי השפעתם איננה עקבית.[16] ממצאים אלה מדגישים את עוצמת הסיכון של הטעיה ומניפולציה שיש

---

[14] הטכניקות ליצירת deepfake כוללות החלפת גוף או פרצוף ביחס לדמות, החלפת קול, שימוש בטכניקות הממירות טקסט לקול כדי להחליף קולות, שילוב פרצופים ("face morphing"), התחזות לדיבוב ("lip-syncing") וייצור טקסט. ראו למשל את הפרסום מטעם הארגון הממשלתי העוסק בהגנת הסייבר בגרמניה: Federal Office for Information Security. Deep Fakes – Threats and Countermeasures, https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html

[15] International AI Safety Report, p. 45, 47 https://internationalaisafetyreport.org

[16] A. Diel, et al. Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers p. 8: "Deepfake content is becoming increasingly widespread, realistic, and hard to detect. This meta-analysis, the

בשימוש בבינה מלאכותית יוצרת להפקת תכנים סינתטיים. ממצאים דומים נמצאו בניסוי ביחס ליכולת לזהות קול מלאכותי.[17] ממצאים אלה מאששים ממצאים ממחקרים מוקדמים יותר.[18]

22. סיכון מרכזי שמציין הדוח הבינלאומי הוא הכמות והקצב: הכלים **לזיהוי תוכן** יציר AI אינם עומדים בקצב השיפורים באיכות הפקת התכנים הסינתטיים, וקיים קושי לזהות את מי שיוצר את התוכן. עקב כך, טוענים המומחים, נדרשת תפיסה "רב שכבתית" (multilayered) על מנת להתמודד עם תופעה זו.[19]

23. לצד הסכנה להטעיה באופן כללי, הדו"ח הבינלאומי מצביע על סיכון משמעותי נוסף, **להשפעה ומניפולציה** ("influence and manipulation") באמצעות בינה מלאכותית יוצרת. מחברי הדוח מציינים כי מערכות בינה מלאכותית יכולות לייצר תוכן משכנע, שעלול להיות מניפולטיבי, ולשמש להנעת אנשים באופן שאינו לגיטימי. בין היישומים המתוארים גם שימוש ביישומי AI להפצת דעות קיצוניות.[20] מחברי הדוח מזהירים כי הפצת תוכן מניפולטיבי יציר AI יכול להחליש את אמון הציבור. הם מצטטים שורה ארוכה של מחקרים המעידים על האפקטיביות של תוכן יציר בינה מלאכותית **בהשפעה על אמונות של אנשים, ולפחות באותה אפקטיביות של בני אדם שאינם מומחים בנושא**.[21] כלים לצמצום הסיכון, כגון הגברת אוריינות דיגיטלית, וסימון תכנים יצירי AI כאמצעים להתמודדות עם מניפולציה, טרם הוכחו כאפקטיביים דיים.[22] הדבר מחייב, לפיכך, גישה רב שכבתית להגברת האפקטיביות של אמצעי צמצום הסיכון.

---

first on human deepfake detection, found synthesized human performance to be at chance. Across all stimulus types crossed chance levels. Thus, deepfake detection (audio, image, text, and video), 95% confidence intervals performance is not significantly above chance for any modality or overall. Performance remains relatively and video), with lower accuracy for detecting deepfakes than ,consistent across stimulus types (audio, image, text real stimuli. Finally, strategies to improve human deepfake detection generally succeed, though not consistently". https://www.sciencedirect.com/science/article/pii/S2451958824001714?via%3Dihub

[17] Barrington, S., Cooper, E.A. & Farid, H. People are poorly equipped to detect AI-powered voice clones. *Sci Rep* **15**, 11004 (2025). https://doi.org/10.1038/s41598-025-94170-3. "The quality and realism of AI-generated media is rapidly improving. Given the results reported here, there is good reason to believe that AI-generated voices will soon be indistinguishable from real ones both in terms of naturalness and identity. While this should be considered a triumph for those on the generative side, it raises real concerns for public safety. Our results highlight that relying on human perception to detect AI-generated voice clones is no longer consistently reliable. Thus, improved technologies that can detect AI-generated voices while protecting the user's privacy will become essential tools for preventing phone-based – and eventually, video-based – frauds."

[18] Köbis NC, Doležalová B, Soraperra I. Fooled twice: People cannot detect deepfakes but think they can. **Iscience.** 2021; 24(11).

[19] International AI Safety Report, p. 49. Detection and watermarking techniques have improved but remain inconsistent and face technical challenges (333, 335). Technical developments in AI content generation can also undermine their effectiveness. For example, a study found that deepfakedetection benchmarks – curated examples of AI-generated and real media designed to test the performance of deepfake detection tools – are outdated and perform about 50% worse on real world deepfakes than on the benchmarks usually used to evaluate them (317). **These limitations mean that multiple layers of techniques are likely needed to detect AI-generated content with a high degree of robustness."** (emphasis added)

[20] International AI Safety Report, p. 51.

[21] International AI Safety Report, p. 51: " Several studies have found that, in experimental settings, AI-generated content can influence people's beliefs at least as effectively as non-expert humans can. These studies generally measure people's self-reported agreement with a statement before and after exposure to AI-generated content: either static text or a multi-turn conversation (361, 365, 366). A large number of studies have found that exposure to AI-generated content can significantly change people's opinions and behaviour (367, 368, 369, 370, 371, 372, 373, 374, 375). Persuasivenessalso increases with the scale of the model used (Figure 2.4). Some of these studies have compared AI systems to humans and found that AI systems are as or more convincing than non-expert humans (see Table 2.2) (376, 377, 378, 379*, 380, 381, 382, 383), and can match the convincingness of human experts in writing static text (384, 385, 386)".

[22] International AI Safety Report, p. 55: Alternative mitigations, which focus on protecting users, provide some value but may not be sufficient on their own. Some researchers have suggested that improved education or AI literacy could mitigate manipulative effects (436, 437), but there is limited evidence for these claims (438). Labelling content as AI-generated has not proven effective at reducing manipulation (439, 440, 441), and users who are knowledgeable about AI or interact with it frequently are just as likely to be deceived (381).

24. במסגרת פרויקט מקיף של מכון התקנים האמריקני היישומי NIST (National Institute of Standards and Technology) בנושא התמודדות עם סיכונים כתוצאה מתוכן סינתטי, מוזכרים סיכונים שונים ממידע סינתטי כוזב וכן מיפוי של "שרשרת הערך" של הפצת תוכן סינתטי.[23] שרשרת ערך זו כוללת את שלב יצירת התוכן ("Creation"),[24] "פרסום" ("Publication")[25] ושלב ה-"צריכה" ("Consumption")[26] שבו הוא מגיע לנמעניו השונים. שלבי הפרסום וההפצה נעשים פעמים רבות באמצעות פלטפורמות הפצת התוכן, הרשתות החברתיות, ולכן הסיכונים מועצמים עקב מאפייני הפצת תכנים במרחב הדיגיטלי.[27]

25. באחד המחקרים המקיפים המוקדמים על סיכוני דיפ-פייק מציינים המחברים Chesney ו – Citron את שורת הנזקים שדיפ-פייק יכול לגרום הן ליחידים ולארגונים (בדרך של ניצול או פגיעה אחרת), והן לאינטרסים ציבורים.[28] המחקר מבהיר כיצד הטכנולוגיה מגדילה באופן משמעותי את היכולת "לעוות מציאות", משום שהזיופים נראים אמינים יותר, נפוצים מהר יותר, ומופצים בסביבה שמועצמת ע"י רשתות חברתיות והטיות קוגניטיביות, ובכך מעצימים את "שחיקת האמת" במרחב הציבורי. המחברים מדגישים לא רק את הנזק ליחיד, אלא גם את הנזק הכולל לחברה: עיוות השיח הדמוקרטי ופגיעה באמון במוסדות. בהקשר בחירות הם מצביעים על פגיעות מיוחדת כשמפיצים זיוף בעיתוי דוחק, בהינתן חלון זמן קצר לפני החלטה בלתי-הפיכה, כך שהשקר מספיק להתפשט אך אין זמן יעיל להפרכה. בנוסף, הם מציגים את "דיבידנד השקרן": ככל שהציבור מודע לכך שאפשר לזייף וידאו/אודיו בצורה משכנעת, כך קל יותר לבעלי עניין להכחיש ראיות אמיתיות ולטעון שהן דיפ-פייק, ובכך לחמוק מאחריות.

26. במחקרה המקיף על תוכן סינתטי פרופסור אלן גודמן, מדגישה את עוצמת הסיכון הנובע מיכולות אלה כתוצאה מכך שהשימוש בהן **הינו נגיש, זול קל להפעלה ומאפשר לא רק הפקת תוכן סינתטי באיכות גבוהה, אלא אף התאמה אישית שלו לנמען ספציפי.**[29] קפיצת מדרגה משמעותית זו ביכולות הטכנולוגיות מציבה רמת סיכון גבוהה וחדשה. פרופסור גודמן ממפה את הנזקים האפיסטמיים כתוצאה משימוש לרעה בתוכן סינתטי, תוך הבחנה בין שני סוגי נזק: נזק מסוג אחד הוא **הטעיה**, כשהתוכן הסינתטי גורם לאמונה שגויה, דהיינו גורם לאנשים להאמין שדבר מה הוא אותנטי בעוד שהוא מזויף. למשל, דימוי שמזייף אירוע (שלא התקיים) באופן משכנע, או בוטים המייצרים רושם

---

[23] National Institute of Standards and Technology, NIST Trustworthy and Responsible AI, NIST AI 100-4, Reducing Risks Posed by Synthetic Content, An Overview of Technical Approaches to Digital Content Transparency, November 2024, https://doi.org/10.6028/NIST.AI.100-4 ("NIST AI 100-4"), p.2-3.

[24] "Creation" - "Synthetic content is produced (generated from scratch or modified from other content) by AI tools, either from generative model developers or from downstream tool developers, and by users using these systems. Content may be edited and modified by users after initial creation."

[25] "Publication" - Synthetic content is uploaded and disseminated by users and publishers across platforms, websites, and other digital channels, or published or broadcast via non-digital channels such as print or radio.

[26] "Consumption" – "Audiences interact and engage with the synthetic content. This can include accessing and interpreting the content, reacting to or acting on it, and interpreting or acting on any accompanying disclosures or labels."

[27] על שיעור השימוש ברשתות חברתיות ושירותים מקוונים בישראל, ראו: איגוד האינטרנט הישראלי, **שימוש ברשתות חברתיות ושירותים מקוונים בישראל**: נתוני 2025 עם פילוחים דמוגרפים, המלמד על שיעור השימוש הגבוה ברשתות חברתיות והשפעתן המשמעותית על זירת המידע. על הקשר בין שימוש בטכנולוגית בינה מלאכותית יוצרת לבין הפלטפורמות, כולל ניצול כלי פרסום של הפלטפורמות לצורך הונאות פיננסיות, ראו: איגוד האינטרנט הישראלי, העוקץ האלגוריתמי: **רשתות חברתיות ובינה מלאכותית כאמצעי להונאות פיננסיות, סקירת איומי סייבר, קו הסיוע לאינטרנט בטוח**, מאי 2025, https://www.isoc.org.il/wp-content/uploads/2025/05/the-algorithmic-sting.pdf. בנושא תפקיד הפלטפורמות המקוונות בהפצת תכנים ראו: בג"צ 7846/19 **עדאלה נגד פרקליטות המדינה- יחידת הסייבר** (12.4.2021).

[28] Chesney& Citron, p. 1771-1785, הערת שוליים 27 לעיל.

[29] Goodman, Ellen P., Synthetic Content: Default to Distrust (December 09, 2025). **Case Western Reserve Law Review**, Vol. 75, Iss. 3, pp. 1- 47., Rutgers Law School Research Paper, Available at SSRN: https://ssrn.com/abstract=5236368 or http://dx.doi.org/10.2139/ssrn.5236368

שיקרי של תמיכה ציבורית רחבה. הנזק האפיסטמי מהסוג השני הוא מערכתי: שחיקת המחויבות לידיעה ולאמת, ולחשיבות הצורך להבחין בין עובדות לבדיה.

27. סיכונים אלה באים לידי ביטוי מוגבר כאשר מדובר על שימוש בתוכן סינתטי במערכת בחירות, שבה יש מצד אחד חשיבות עצומה לשיח פלורליסטי לצורך שכנוע הבוחרים, במסגרת הזמנים שהוקצבה לתהליך הבחירות, ומצד שני סיכון עצום הנובע מההשפעה האפשרית של הטכנולוגיה.

<u>הסיכונים מהפקת תוכן סינתטי במערכת בחירות</u>

28. מובן כי כלל הסיכונים האמורים לעיל עלולים לבוא לידי ביטוי בתעמולה פוליטית בתקופת בחירות. בהמשך לכך, הסיכונים המוגברים בתקופת בחירות נחלקים לשניים: לצד פוטנציאל ההטעיה והמניפולציה על בוחרים בתקופה הקצובה והרגישה של בחירות, מצטרפים גם סיכונים נוספים לפגיעה בשוויון בין מועמדים בשל היעדר כללים ברורים וכן הסכנות הנובעות מהתערבות זרה. שילוב היבטים אלה מביא לכך שיש בכלי טכנולוגי זה משום סיכון תשתיתי לשוק רעיונות, שמתבסס על שיח ציבורי מושכל ולא על הונאה, זיוף וגניבת דעת ולקיומן של בחירות הוגנות ושוויוניות, עקב עוצמת האיום שהוא משקף להפצת מידע מזויף שהינו משכנע במיוחד.

29. בדוח הסיכונים השנתי של הפורום הכלכלי העולמי (World Economic Forum) לשנת 2026 ניתן דגש מיוחד לסיכונים הנובעים מדיסאינפורמציה ושימוש לרעה בבינה מלאכותית, כולל התייחסות להתערבות בבחירות.[30] **הדוח מציין, כי בכל מערכות הבחירות שהתרחשו בשנתיים האחרונות ניתן היה לראות שימוש בבינה מלאכותית ליצירת תכנים מזויפים, וכן ניצול לרעה של תופעה זו בידי גורמים זרים לצורך קמפיין השפעה**. הדוח מציין כי ניתן להעריך שתופעה זו תחמיר ככל שהיכולות של בינה מלאכותית לייצר תוכן ישתפרו ויהיו משכנעות יותר. חשש נוסף הוא מפגיעה ביכולת של בוחרים לזהות תוכן פיקטיבי. בנוסף, אחד המחקרים הראה כי גם כאשר אנשים מתייחסים למידע פוליטי סינתטי כפחות מהימן, החשיפה אליו עדיין הביאה להשפעה שלילית על הדמות הפוליטית שתוארה בו.[31]

30. **במערכות בחירות קיימת רגישות מיוחדת להטעיה ומניפולציה במועדים קריטיים, המכונים בספרות "decisional chokepoints"[32], שהם מועדים שבהן מתקבלות החלטות בלתי הפיכות, ויש חלון זמן מוגבל להתמודד עם מידע מזויף ולהתמודד איתו ("debunk"), כגון מועדים הקרובים לקיום ההצבעה.**[33]

---

[30] Insight Report, World Economic Forum, The Global Risks Report 2026, 21st Edition, https://reports.weforum.org/docs/WEF_Global_Risks_Report_2026.pdf, p. 36: "Over the past five years, deepfake creation has become easier, cheaper, and more convincing. While the use of deepfakes during the 2024 "super election year" was still a relatively new phenomenon, they have started to proliferate and have a greater influence on politics and electoral processes. The weaponization of deepfakes can undermine trust in democratic institutions, contributing to more political polarization, and can lead to the incitement of political violence or social upheaval. **Recent elections in the United States, Ireland, the Netherlands, Pakistan, Japan, India and Argentina have all had to contend with such fabricated content on social media, depicting fictional events or discrediting political candidates, blurring the line between fact and fiction. As AI is used to make such content more personalized and persuasive, there is a risk of greater impact on elections.** For example, research has found that 87% of people in the United Kingdom are concerned about deepfakes affecting election results. **But while awareness is high, many lack confidence in their abilities to identify when content is manipulated."** (emphasis added)

[31] Hameleers M, van der Meer T, Dobber T. Distorting the truth versus blatant lies: The effects of different degrees of deception in domestic and foreign political deepfakes. **Computers in Human Behavior 152** (2024),

[32] ראו Chesney&Citron, הערת שוליים 27 לעיל, בעמוד 1778.

[33] Narrow Margins and Misinformation: The Impact of Sharing Fake News in Close Contests https://www.mdpi.com/2076-0760/13/11/571: "While it is not possible to ascertain the motives and knowledge of every candidate sharing political misinformation, there are proxy measures of related behavior available for analysis. Recent work by Kornberg et al. (2023) shows patterns of 2022 candidates sharing posts related to the "Big Lie"; on Twitter, Facebook, and Instagram, there was a much greater volume of shared bogus claims leading up

31. כאשר מופץ לציבור מידע סינתטי שנועד לייצר השפעה לא הוגנת בסמיכות לבחירות, גם אם לאחר מכן תהיה טענה שמדובר במידע מזויף, לא תמיד המידע המזים יגיע לציבור הרחב שנחשף למידע המזויף. זאת משום שהשקר נפוץ מהר יותר מאשר האמת.[34] יתרה על כך, לעיתים, כאשר הבוחר נחשף לפרסום הסינתטי בסמוך ליום הבחירות – ייתכן שהוא כבר יממש את זכותו להצביע, ויעשה כן נוכח ההטעיה, מבלי שיש בידיו לתקן. כך למשל, אירע בפרשה בה הופצה ידיעה שקרית על פרישתה של מפלגת גשר, בראשותה של אורלי לוי, בבחירות לכנסת ה-21 (תב"כ 61/21 **גשר בראשות אורלי לוי-אבקסיס נ' Facebook** (08.04.2019).

32. כך למשל, בבחירות בסלובקיה פורסם בסמוך מאוד לבחירות הקלטת קול (אודיו) מזויפת של מועמד מרכזי ולפיה הוא מתכנן לזייף את הבחירות.[35] בטורקיה פורסם סרטון פורנוגרפי מזויף של אחד המועמדים שהוביל לפרישתו ושינוי אפשרי בחלוקת ההצבעות.[36]

33. בטבלה שלהלן תיאור של אירועי שימוש בבינה מלאכותית יוצרת להפקת תוכן סינתטי לצרכי השפעה במערכות הבחירות באירופה בשנים 2024-2025.[37]

מדינה	פירוט	סיכונים והשפעה
אירלנד	**סרטון Deep Fake שבו מועמדת מודיעה על "פרישה"**[38] סרטון באיכות גבוהה שנוצר באמצעות AI הופץ במהלך העימות הטלוויזיוני האחרון לקראת הבחירות והציג כאילו המועמדת לנשיאות קתרין קונולי פרשה מהמירוץ. ועדת הבחירות הגיבה מהר ו-Meta הסירה את הסרטון, אך הוא המשיך להתפשט ברשת X.	• פגיעה בשוויון ההזדמנויות בבחירות • התערבות "בדקה ה-90"
רומניה	**ביטול תוצאות הבחירות לנשיאות בעקבות קמפיין השפעה רוסי נרחב**[39] מעל 85 אלף מתקפות סייבר וקמפיין דיסאינפורמציה במהלך הבחירות לנשיאות ב-2024 שתמכו במועמד פרו-רוסי, בין השאר באמצעות Deep Fake, קמפיין בוטים בטיקטוק, הפצת מידע כוזב באמצעות Troll Factory ועוד. בעקבות חקירה של רשויות אכיפת החוק ביהמ"ש החוקתי ברומניה ביטל את תוצאות הבחירות לנשיאות בשל התערבות רוסית.	• בית המשפט קבע כי היקף ההשפעה הזרה שלל את הוגנות הבחירות • התערבות "בדקה ה-90" • השפעה זרה

---

to, and eventually peaking on, Election Day. After this period, there was a substantial decline in the post-election period (Kornberg et al. 2023, p. 15)."

[34] Soroush Vosoughi *et al.,* The spread of true and false news online. *Science* 359, 1146-1151 (2018).10.1126/science.aap9559

[35] https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/

[36] Łabuz, M., Nehring, C. On the way to deep fake democracy? Deep fakes in election campaigns in 2023. ***Eur Polit Sci*** 23, 454–473 (2024). https://doi.org/10.1057/s41304-024-00482-9

[37] הסקירה מבוססת בין היתר על הפרסום מטעם המכון למחקר ביטחון לאומי, "מעגלי השפעה – ניוזלטר תוכנית תודעה והשפעה זרה" , https://www.inss.org.il/he/circles

[38] **Necva Tastan Sevinc, Anadolu Ajansı, 26.10.2025.** *Deepfake video targeting Ireland's president-elect sparks election integrity concerns.* Anadolu Agency. https://www.aa.com.tr/en/europe/deepfake-video-targeting-ireland-s-president-elect-sparks-election-integrity-concerns/3727224 (Anadolu Ajansı)

[39] **Rodica State (writing) / Oana Ghita (reporting), Agerpres, 16.09.2025.** *President Dan: Romania will not tolerate any foreign interference in its democratic process.* Agerpres. https://agerpres.ro/english/2025/09/16/president-dan-romania-will-not-tolerate-any-foreign-interference-in-its-democratic-process--1484695 (AGERPRES)

טאיוואן	**התערבות סינית בטאיוואן: קמפיין השפעה לקידום מועמדת בבחירות**[40] מבצע המיוחס למפלגה הקומוניסטית הסינית במימון של מעל 3 מיליון דולר השתמש ברשתות טיקטוק ויוטיוב כדי לקדם את המועמדת Li-wen בבחירות למפלגת ה-KMT ולהכפיש את יריבה Hau באמצעות כלי AI ו-Deep Fake, כולל סרטון בו האחרון נראה מנשק את ראש עיריית טאיפיי.	• פגיעה בשוויון ההזדמנויות בבחירות • השפעה זרה
יפן	**הפצת תעמולה ביפן על ידי רוסיה והדבקת כלי AI בנרטיבים כוזבים**[41] רוסיה הפעילה רשתות טרולים ובוטים שמציפות את הרשת, בכ-3.5 מיליון פריטי תעמולה בשנה, כולל עלייה משמעותית בתקופת הבחירות ביפן. עומס התוכן גורם לצ'אט בוטים של AI דוגמת Grok ו-Chat GPT, אשר סורקים את הרשתות, לשחזר מידע כוזב ולהפוך נרטיבים מעוותים לעובדות מקובלות.	• פגיעה בהוגנות של הבחירות • השפעה זרה
גרמניה	**מפלגות מקומיות קיצוניות משתמשות בכלים של רשתות השפעה זרה**[42] בגרמניה נמצא כי מפלגות אימצו דפוסי שימוש ב-AI אשר דומים למבצעי השפעה זרה, מה שהקשה להבחין בין תעמולה לגיטימית לבין השפעה זרה. נצפה שימוש בטכנולוגיות Deep Fake, סרטונים שנוצרו ב-AI ובהם התחזות לאמצעי תקשורת ולגורמים באקדמיה.	• פגיעה בשוויון ההזדמנויות בבחירות • פגיעה בהוגנות הבחירות

<u>הסיכונים להתערבות זרה באמצעות מבצעי השפעה במרחב הסייבר</u>

34. השימוש בדיפ-פייק הינו אמצעי מרכזי בכלי התקיפה של גורמים זרים המבצעים "התערבות זרה" באמצעות "מבצעי השפעה" במרחב הדיגיטלי.[43] מבצעי השפעה אלה עושים שימוש לרעה ביכולת להפיץ תכנים במרחב הדיגיטלי באמצעות ייצור תוכן כוזב, שימוש בזהויות שאינן אותנטיות (כולל כלים

[40] **Vision Times News, Vision Times, 02.11.2025.** *Beijing's Cyber Operations Allegedly Manipulate KMT Chair Election With Hundreds of Millions in Funding.* Vision Times. https://www.visiontimes.com/2025/11/02/beijings-cyber-operations-allegedly-manipulate-kmt-chair-election-with-hundreds-of-millions-in-funding.html (Vision Times)

[41] **Ichihara Maiko, Nippon.com, 27.10.2025.** *Japan's Upper House Election Reveals how Russian Influence Operations Infecting AI with Flood of Propaganda, Stoking Divisions.* Nippon.com. https://www.nippon.com/en/in-depth/d01170/ (Nippon.com)

[42] **ISD Germany, Institute for Strategic Dialogue, 10.07.2025.** *Country Report: Assessment of Foreign Information Manipulation and Interference (FIMI) in the 2025 German Federal Election.* ISD Global. https://www.isdglobal.org/publication/country-report-assessment-of-foreign-information-manipulation-and-interference-fimi-in-the-2025-german-federal-election/ (isdglobal.org)

[43] עופר פרידמן מציע להתבונן על "השפעה זרה" כחלק ממכלול האמצעים ששחקן בין־לאומי עושה כדי להשפיע על שחקן בין־לאומי אחר בזירה הבין־לאומית. על רקע זה הוא מגדיר **"השפעה זרה" כשימוש בכל המקורות הזמינים של עוצמה לאומית (דיפלומטיים, מידעיים, צבאיים, כלכלים, דתיים וכולי), על ידי שחקן בין־לאומי אחד כדי להשפיע על שחקן אחר, במטרה להשיג יעד פוליטי". בהמשך לכך מגדיר פרידמן "התערבות זרה" ("foreign interference") כסוג מסוים של השפעה זרה.** לפי גישה זו, "השפעה זרה" (foreign influence) תיחשב התערבות זרה (foreign interference) כאשר השימוש של שחקן אחד בעוצמה לאומית כדי להשפיע שחקן אחר נחשב על ידי השחקן האחר כמנוגד לערכים, לנורמות או לחוקים שלו. בהקשרי תכנים, ובפרט במרחב הסייבר, מאפיין מרכזי של השפעה זרה הוא באופן שהיא נעשית בו - שנועד להסוות את מקורה ולהיראות כמידע אותנטי שהוא חלק מהשיח הציבורי הלגיטימי. פעמים רבות השפעה זרה באה לידי ביטוי במרחב הסייבר על ידי "מידע כוזב". ר' עופר פרידמן, "השפעה והתערבות זרה – המשגה ומונחים", פירסום מיוחד, 2 בינואר 2024, המכון למחרי ביטחון לאומי והמכון לחקר המתודולוגיה של המודיעין. https://www.inss.org.il/he/publication/influence-and-interference

ממוכנים "בוטים") להפצת דיפ-פייק[44] או להדהד בצורה מלאכותית דיפ-פייק וכן גם תוכן אותנטי, כדי להטות את דעת הקהל או להגביר קיטוב. אתגרים אלה משלבים בין הסיכונים הקשורים בדיפ-פייק, לבין האתגרים הקשורים בהתמודדות עם הפצת מידע ברשתות החברתיות.[45]

35. ההתמודדות עם התערבות זרה במרחב הסייבר מחייבת מאמץ רב שכבתי הכולל כללי התנהגות אחראית עבור השחקנים הלגיטימיים, כללים ביחס לפלטפורמות הפצת המידע, אוריינות של הציבור, והיערכות של הגורמים הממשלתיים השונים, בהתאם לתפקידיהם, להתמודדות עם אירועים אלה.[46]

36. על פי מחקר של המכון למחקרי ביטחון לאומי INSS, ניתן להצביע על פעילות איראנית לא מבוטלת המכוונת לציבור בישראל תוך שימוש בבינה מלאכותית יוצרת. מטרת פעילות זו להתערב בענייני הפנים של ישראל, ללבות את המחלוקות, לזרוע חוסר אמון ודמורליזציה, ולהסית לאלימות נגד אזרחים ערבים.[47]

37. המחקר מתאר רשתות השפעה שונות שפעלו בין היתר בתקופת הבחירות לכנסת ה-25. המחקר מציין כי אחת הרשתות האיראניות פעלה במטרה להשפיע על קהל היעד של תומכי מחנה השמאל. היא קידמה מסרים נגד איתמר בן-גביר בניסיון להציגו כ-"מסוכן לישראל" ובד בבד קידמה קמפיין של "גניבת הבחירות" שלכאורה הובילו תומכי ימין ותומכי מפלגת הליכוד.

---

[44] בהקשר של מבצעי השפעה הכוללים הפצה של מידע, נציין את ההגדרות של ה- EU Disinformation Code of Conduct. הגדרות אלה נבחנו תקופה ארוכה בשיח בין גורמים שונים, והם תואמו גם עם פלטפורמות הפצת התוכן המהוות אחד השחקנים המשמעותיים בהפצה ומניעה של הפצת מידע כוזב. בהתאם לכך: "**misinformation**" הוא תוכן שקרי או מטעה המופץ ללא כוונה לגרום נזק, למרות שההשפעה יכולה היות מזיקה, למשל כאשר אנשים משתפים מידע כוזב עם חברים ומשפחה בתום לב. "**disinformation**" הינו מידע כוזב או מטעה המופץ בכוונה להטעות או לקדם יעד כלכלי או פוליטי שיכול ליצור נזק ציבורי. "**מבצע השפעה באמצעות מידע**" מאפיין מאמצים מתואמים על ידי שחקנים מקומיים או בינלאומיים להשפיע על קהל יעד באמצעות אמצעים מטעים שונים, כולל השתקת מקורות מידע עצמאיים בשילוב מידע כוזב. "**השפעה זרה במרחב הסייבר**", נעשית בדרך ככל כחלק ממבצע היברידי רחב יותר, מבטאת מאמצים לשבש את חופש הרצון והביטוי, באמצעות מדינה זרה או בידי סוכניה. European Commission, The 2022 code of practice on disinformation, https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation. האיחוד האירופי מקדם מסגרת קונספטואלית בנושא זה הידועה בשם FIMI, Foreign Information Manipulation and Interference.

[45] לתמונת המצב ביחס לרשתות חברתיות והאתגרים בעולם ובישראל בנושא זה ראו: אסף וינר, היערכות לאומית מול השפעה והתערבות זרה באמצעות רשתות חברתיות: ממצאים גלובליים ותמונת מצב ישראלית, פרסום מיוחד משותף של המכון למחקרי ביטחון לאומי והמכון לחקר המתודולוגיה של המודיעין במרכז למורשת המודיעין, 27 במארס 2024, https://www.inss.org.il/he/publication/national-preparation/. על מבצעי השפעה איראנים באמצעות הרשתות החברתיות ראו: ענבל אורפז, דוד סימן טוב, התערבות זרה והשפעה איראנית ברשתות החברתיות בישראל, פרסום מיוחד משותף של המכון למחקרי ביטחון לאומי והמכון לחקר המתודולוגיה של המודיעין במרכז למורשת המודיעין, 10 ביוני 2024, https://www.inss.org.il/he/publication/iranian-influence/

[46] להרחבה בנושא הצורך בתפיסה רב שכבתית כוללת להתמודדות עם התערבות זרה במרחב הסייבר, והתייחסות לתפקיד הייחודי של ועדת הבחירות המרכזית, ראו: עמית אשכנזי, "קווים מנחים להתמודדות ישראל עם השפעה זרה במרחב הסייבר והרשתות החברתיות" פרסום מיוחד משותף של המכון למחקרי ביטחון לאומי, משרד המודיעין והמכון לחקר המתודולוגיה של המודיעין במרכז למורשת המודיעין, 5 במארס 2024, https://www.inss.org.il/he/publication/israel-cyber/

[47] ניצן יסעור, דני סיטרינוביץ, "התערבות והשפעה זרה איראנית במהלך מלחמת חרבות ברזל", פרסום מיוחד משותף של המכון למחקרי ביטחון לאומי והמכון לחקר המתודולוגיה של המודיעין במרכז למורשת המודיעין, 11 באוגוסט 2024, https://www.inss.org.il/he/publication/iran-influence/. בנושא זה מציינים המחברים כי איראן עושה שימוש במימד הקיברנטי עקב מגבלות בהפעלת כוח בהקשרים האחרים. בהמשך לכך הם מציינים: "יתרה מזאת: ייתכן שהעובדה שאיראן מנהלת זה זמן רב מבצעי השפעה נרחבים אל מול ישראל מתחת לאותו סף הסלמה, רק הוסיפה לביטחון של איראן כי ביכולתה להמשיך ולהעמיק מבצעים אלו, בתוך שימוש באותו דפוס פעולה, באותה תשתית מאפשרת ובעיקר באותם כלים כדי להגשים את מטרותיה. נדמה שההצלחה האיראנית במבצעי ההשפעה הקודמים במהלך שלוש השנים האחרונות וביתר שאת בזמן ההפגנות נגד הרפורמה המשפטית, רק הוסיפה לביטחון של טהרן ביכולתה להמשיך ולהעמיק מבצעים אלו בתוך שימוש באותם כלים". ובהמשך לכך: "מאמצי ההשפעה האיראניים מזהים נקודות מחלוקת ושסעים אקטואליים שהם מנצלים ומעצימים. מכאן, אין זה מפתיע שהאיראנים לא יחמיצו את ההזדמנות לניסיונות התערבות והשפעה על רקע מלחמת "חרבות ברזל" תוך קידום מטרותיהם האסטרטגיות, ניצול השיח ושימוש במתחים הפנימיים בישראל כמו סוגיית החטופים, מספר הנרצחים ההרוגים והפצועים הרבים ,חוסר האמון הגובר בממשלה, במוסדות השלטון ובגורמי הביטחון והמשך הרחבת המתחים ותדלוק השסעים בין החלקים השונים בחברה הישראלית. **את מאמצי ההשפעה אפשר למצוא בכל סוגיה, בעיקר סביב נושאים ואירועים מעוררי מחלוקת, בעלי קשב ציבורי רב - מבצעי ההשפעה ורשתות הפרופילים עוסקים בשלל נושאים שונים, אינם בוחלים באמצעים ומשחקים על כל המגרש ולא בהכרח מקדמים מועמד או אג'נדה פוליטית ספציפית אלא את אלה שלהבנתם יקדמו את האינטרסים שלהם באופן המהיר והיעיל ביותר.** המבצעים שיתוארו במחקר זה פעלו באמצעות חשבונות ברשתות החברתיות שהתנהגותם אינה אותנטית ושאפשר לקבוע בסבירות גבוהה שהמפעילים שלהם הם גופים מדינתיים זרים, קרי איראניים. (ההדגשה אינה במקור). במסגרת זאת מתוארים מבצעים שמטרתם לא רק הפצת תעמולה ומידע כוזב אלא הנעה לפעולה כגון התחזות לגורמי ימין ועידוד התקהלויות אלימות והפגנות נגד אזרחים ערבים ישראלים, פעילויות שמתחזות להיות פעילות בנושא שחרור החטופים ברצועת עזה.

38. המחברים מציינים את הניסיון של איראן "להשפיע בו בזמן על כל הצדדים", ומדגימים זאת באמצעות מודעות בחירות שמטרתן השמצה של מועמדים שונים בקשת הפוליטית. המחברים מציינים כי: "כדי להעצים ולהרחיב שסעים חברתיים, פעלו הרשתות בקרב קהלי יעד מובחנים עוד לפני פרוץ המלחמה. רשתות אחדות קידמו קריאות ומסרים נגד הממשלה וקראו להדחתו של בנימין נתניהו; רשתות אחרות הפיצו מסרים מסיתים המאשימים את מתנגדי ההפיכה המשטרית ו"מפגיני קפלן" באירועי 7 באוקטובר. נוסף על המתחים הפנים יהודיים בחברה הישראלית פעלו הרשתות להתסיס ולהסית כנגד אזרחי ישראל הערבים ופעלו בניסיון לייצר אלימות של ממש בעולם האמיתי".

39. בהמשך לכך ולעניין שימוש בכלי AI מציינים המחברים כי: "השימוש הבולט ביותר בתוצרי AI בזמן מלחמת חרבות ברזל היה עיצוב כרזות והפצת סרטוני דיפ-פייק. תוצרי ה- AI עדיין אינם ברמת הפקה מושלמת אך איכותם גבוהה יחסית ויכולה להפיל בפח אזרחים, אנשי תקשורת ונבחרי ציבור בהיסח הדעת. איכות התוצרים, כמות התוצרים וקנה המידה של הפצתם, גדלים באופן אינטנסיבי והם אתגרים משמעותיים ביכולת הזיהוי, ניטור, בדיקה והפרכה של מידע. הרשת שהפיצה את הסרטונים פעלה תחת המותג "ישראל השנייה". המחברים מתארים סרטונים שהופקו, באחד מהם בנימין נתניהו (המזויף) פונה ליהודי העולם ומסביר מדוע אינו אחראי לאירועי 07.10, ובשני מופיעים לכאורה חמישה רבנים ישראלים (מזויפים) הטוענים כי נתניהו אשם בחיזוק החמאס.

40. בסיכום המחקר מציינים המחברים, בדומה למאפיינים שתוארו לעיל של שימושים בדיפ-פייק, לערעור האמון בזירת המידע, כי המטרה האסטרטגית של איראן לטווח הארוך היא "לייצר רתיעה וחוסר אמון מכל מידע שהוא, ובכך לפורר את היכולת הדמוקרטית לקבל החלטות מושכלות באופן חופשי על בסיס מידע ונתונים". המחברים מעריכים כי: "השימוש בתוצרים מזויפים או מניפולטיביים המיוצרים באמצעות טכנולוגיות AI צפוי להתרחב באופן שיגביר את האמינות של התוצרים, ישפר את רמת השפה, יגדיל את קנה המידה, יגביר את קצב ההפצה ויאפשר התאמה מדויקת יותר של מסרים על בסיס ניתוח של השיח המקומי." בהמשך לכך, בין המלצותיהם: "הטכנולוגיה של AI מגדילה את קנה המידה, את האיכות ואת המהירות של הפצת דיסאינפורמציה ומסרים של מבצעי השפעה זרה. לפיכך יש למצוא ולאמץ פתרונות טכנולוגיים אשר נותנים מענה לניטור ולניתוח כמויות מידע גדולות לזיהוי אנומליות התנהגויות ופיסות תוכן מפוברקות".

41. כפועל יוצא מכל אלה נדרש פתרון מערכתי הכולל מספר נדבכים. בהקשר שלפנינו חובת סימון התוכן הסינתטי, כולל באמצעים טכנולוגיים, יאפשר המשך סימונו וזיהויו בפלטפורמות ההפצה השונות, המשמשות להעברת מסרים אל הציבור, וכן יקל על גורמי הביטחון לזהות תכנים שאינם "תעמולת בחירות" לגיטימית.

<u>סיכום הסיכונים שעמם יש להתמודד</u>

42. לסיכום, שימוש בבינה מלאכותית יוצרת להפקת תכנים סינתטיים מערער את כללי המשחק הפוליטיים בבחירות ומשבש את השיח הציבורי, ומוביל לסיכונים הבאים:

א. **הטעיית בוחרים** - ניתן להעריך כי חלק משמעותי מנמעני המסר, הבוחרים, יניחו כי תוכן סינתטי מציאותי מסוים הינו אותנטי, ויתחשבו בו כחלק מגיבוש עמדתם. מחקרים עדכניים מלמדים על קושי של ממש של בני אדם לזהות תוכן סינתטי ככזה ולהבחין כי אינו אמיתי.

ב. **הכבדת הנטל על הבוחר** – הציבור המבקש לגבש את עמדותיו בשוק הרעיונות התוסס, נדרש עתה לעסוק לא רק במטלה (הנכבדה) של איסוף ועיבוד המידע הנדרש לגיבוש עמדתו, אלא להשקיע בבירור בסיסי האם המידע המוצג אותנטי או מזויף.

ג. **תחרות לא הוגנת** - שימוש באמצעים טכנולוגיים לשם הטעיה והפרעה בידי מפלגות יוצר יתרון בלתי הוגן ביחס למפלגות המקפידות על עקרונות ההוגנות ופוגע בשוויון בתחרות בין המפלגות השונות. זאת, גם כאשר מדובר בשימוש לצורך תעמולה עצמית, שכן כל תעמולה עצמית נועדה לקדם את מועמדותך, נגד מועמדותו של היריב.

ד. **חשש ל"מירוץ לתחתית" בין המפלגות בהפקת מידע כוזב -** מתן לגיטימציה להפקת תוכן סינתטי המתחזה לאותנטי בידי מפלגות העומדות לבחירת הציבור, כחלק ממהלכי השכנוע שלהן, עלול לחולל "מירוץ לתחתית" בין המפלגות לשימוש בטכנולוגיה ליצירת תכנים מניפולטיביים ככל הניתן, המערערים את היכולת להבחין בין מה שאותנטי למה שכוזב, ומהווים גניבת דעת.

ה. **ערעור האמון הציבורי במידע רלוונטי למערכת הבחירות וזיהום השיח הציבורי** – שכיחות גבוהה של תכנים סינתטיים המתחזים להיות אותנטיים תביא לפגיעה בנכונות של הציבור להאמין גם למידע אותנטי המופץ במסגרת תעמולה בידי המפלגות. זאת בפרט, ביחס למפלגות המקדמות דעות השונות מזה של הבוחר, ובכך לחתור עוד יותר תחת הרעיון של שוק רעיונות חופשי ותוסס.

ו. **פגיעה באמון הציבור באופן כללי במוסדות ובתהליכים הדמוקרטים** – מתן לגיטימציה להפצה והפקה של דיפ-פייק בידי השחקנים המרכזיים של הזירה הציבורית, המפלגות הפוליטיות ונבחרי הציבור, עלול לערער את האמון הציבורי במוסדות ובתהליכים הדמוקרטיים בכלל, לרבות הטלת ספק במהימנות של פרסומים מטעם גורמים רשמיים, כמו ועדת הבחירות המרכזית.

ז. **חשיפה להתערבות של גורמים זרים** - גורמים זרים עוינים שמטרתם המרכזית הגברת הקיטוב וחוסר האמון במוסדות (ללא קשר לתוכן בהכרח), מנצלים כלים טכנולוגיים אלו, כדי להציף את המרשתת בתכנים סינתטיים, כדי להתחזות לשחקנים פוליטיים ודוברים פוליטיים, מחמירים את ה- "ההפרעה הבלתי הוגנת" והכל באופן שמביא לפגיעה מערכתית נרחבת יותר ובכלל זה סיכון חמור לאיתנות המערכת הדמוקרטית ובפרט הליך הבחירות. בהקשר זה, היעדר חובת סימון של תכנים סינתטיים, מקשה על איתור מעורבות זרה.

<u>תחולתו של חוק דרכי תעמולה והשימוש בתכנים סינתטיים כ"הפרעה בלתי הוגנת"</u>

43. על רקע הסיכונים האמורים, נבקש להראות כי על מנת לעשות שימוש אחראי בבינה מלאכותית יוצרת באופן שמאפשר למפלגות למקסם את חופש הביטוי הפוליטי שלהם, על שימוש כאמור לכלול סימונים צורניים כמפורט להלן. סימונים אלה יאפשרו לצמצם את הסיכונים הנובעים מהשימוש בטכנולוגיה זו ולמקסם את התועלות ממנה עבור המפלגות.

44. לצורך כך נסקור בקצרה את חוק דרכי תעמולה ואת החלטות יושבי ראש הוועדה הנכבדה בעניין סעיף 13 לחוק, ולאחר כן מכן נבהיר מדוע שימוש בבינה מלאכותית יוצרת להפקת תכנים סינתטיים ללא סימון התכנים ככאלה מהווה פגם צורני המהווה הפרעה בלתי הוגנת.

45. נציג להלן מדוע עמדה זו נשענת על תורת המשפט של משפט וטכנולוגיה כפי שהוחלה על שימושים טכנולוגים לפי החוק ובאופן כללי; וכמו כן, היא מבוססת על עקרונות מקובלים לבינה מלאכותית מהימנה, אותם יש להחיל בהתאם להקשר ולסיכון, המהווים סטנדרט פרשני מקובל.

<u>חוק דרכי תעמולה ופרשנותו</u>

46. חוק דרכי תעמולה מסדיר את דרכי התעמולה לצורך הבטחת חופש הבחירה, טוהר הבחירות, והבטחת שוויון בין המפלגות בניסיונן להגיע אל הבוחר, שהיא ערובה למימוש הזכות החוקתית המעוגנת בחוק יסוד: הכנסת, לבחור ולהיבחר בבחירות שוות. החוק מבקש להשיג יעד זה בהתחשב בזכות החוקתית לחופש הביטוי הפוליטי.[48] תכליות אלה באות לידי ביטוי בין היתר בסעיף 2א1 לחוק שעניינו "שקיפות בתעמולת בחירות", ובסעיף 13 לחוק, שכותרתו "איסור הפרעה". החוק מבקש ליישב ולאזן בין חופש הביטוי הפוליטי לבין העקרונות החוקתיים של הבחירות.[49] ברור כי חופש הביטוי הפוליטי אינו מוחלט, והוא מוגבל בהוראות שונות בחוק דרכי תעמולה, ולעניינו בסעיף 13 לחוק כאשר יש בביטוי כדי ליצור "הפרעה בלתי הוגנת לבחירות".

47. במאמרם, "תעמולת בחירות – עבר הווה ועתיד", כותבים סלים ג'ובראן וגיא רווה כי סעיף 13 לחוק אוסר על תעמולת בחירות מטעה. במאמר, עמדו המלומדים על כך שההטעיה האסורה היא הטעיה צורנית בעיקרה (למשל שימוש מטעה באותו או בכינוי של רשימת מועמדים), ותוכנה של התעמולה עצמו אינו נבחן. יחד עם זאת, ישנן פעמים בהן ייבחן התוכן, כך שאם יש ודאות קרובה לפגיעה ממשית בדעתו החופשית של הבוחר, ייבחן התוכן התעמולתי אל מול ההטעיה שהוא גורם (ללא בחינה של אמת ושקר). למשל, אם מראים תמונה של מועמדים, ומשימים כאילו הם רצים יחדיו.[50]

48. בעניין תב"כ 5/25 קבע יו"ר ועדת הבחירות המרכזית, השופט עמית, כי כאשר מפלגה עושה שימוש במודעת בחירות זהה למודעת בחירות של מפלגה אחרת תוך שינויים של המודעה הכוללים תוכן אחר, יש בכך משום שינוי "צורני", כי מדובר בפרסום שנחזה להיות מטעם מפלגה אחת בעוד שהוא מטעם מפלגה אחרת. עקב כך פרסום כזה עלול להטעות את הבוחר הסביר, ולהפריע, באופן בלתי הוגן, לתעמולת הבחירות.[51]

49. בהקשר זה יודגש כי מובן כי לעניין דיפ-פייק, תוכן סינתטי הכולל תיאור ריאליסטי הנחזה להיות אותנטי, ובשים לב לנתונים המחקריים על האיכות הגבוהה של יכולות הבינה המלאכותית והקושי של הנמענים לזהות אותה, מתקיים מבחן זה.

50. אנו סבורים כי הכללים ביחס לשימוש בדיפ-פייק צריכים לחול גם אם מדובר בתעמולה עצמית. מועמד המפיץ מידע באמצעות תכנים סינתטיים מזויפים מפריע הפרעה אסורה לתעמולת הבחירות של יריביו הפוליטיים, שאינם מעוניינים להציף את הציבור בתכנים מזויפים יציר AI. שימוש כזה, כאמור, גם פוגע באופן רוחבי באמון של הציבור בתעמולת הבחירות וביכולת של הציבור לתת אמון בתכני תעמולה כלשהן, בשל הספקנות שמייצר השימוש בתכנים סינתטיים. לאור זאת, פרשנות של סעיף 13 לחוק התעמולה, הפוסעת בעקבות החלטות יושבי הראש הקודמים, ונותנת ביטוי לסיכונים כפי שהוצגו עד כה, מחייבת סימון כראוי של תכנים סינתטיים מכל סוג, גם כאלו שאינם עוסקים ביריבים הפוליטיים.

---

[48] תב"כ 8/21, בפסקה 68.

[49] שם

[50] סלים ג'ובראן וגיא רווה "דיני תעמולה – עבר, הווה, ועתיד" **ספר יורם דנציגר**, 531 (2019), בעמ' 554 (ל' זר-גוטמן וע' באום עורכים).

[51] תב"כ 5/25 **מפלגת חוסן ישראל נ' מפלגת הליכוד** (2.8.2022).

51. בעניין תב"כ 9/24 דן יו"ר ועדת הבחירות המרכזית, השופט פוגלמן, בסרטון שבו נראית דמות שנחזית להיות ח"כ יאיר לפיד כשהוא נושא נאום ואומר דברים שח"כ יאיר לפיד לא אמר. בעניין זה מציין השופט פוגלמן כי: "דיפ-פייק הוא כינוי לטכנולוגיה ליצירת תוכן קולי או חזותי או לשינוי תוכן קיים, כך שהצופה הסביר (ואף הצופה המתוחכם) יסבור כי פלוני ביצע פעולה או העביר מסר, אף התוכן אינו אמיתי. התוכן הוא באיכות גבוהה עד כדי כך שמשתמש מן היישוב יתקשה לרוב לגלות שמדובר בזיוף".[52]

52. מאחר שתוך כדי הליך בירור העתירה, נוסף לפרסום כיתוב ברור ביחס לכך שמדובר בתוכן שיוצר בטכנולוגיה של דיפ-פייק, קבע השופט פוגלמן כי אין הצדקה להתערבות נוספת שלו בתוכן הסרטון או בשימוש בו, וזאת לנוכח החשיבות של חופש הביטוי הפוליטי. הדברים יובאו במלואם מפאת חשיבותם:

> "המקרה שלפנייי איננו עומד במבחן הוודאות הקרובה. הסרטון העדכני הוא שנתון כעת לבחינה. **כאמור, בתחילתו מובהר שהוא נערך בטכנולוגיית "דיפ פייק". לכל אורך הצגת "דמותו" של ח"כ לפיד מצוין בבירור שהלה אינו הדובר. בתום הנאום מופיעה שקופית ובה כתוב "זה לא חייב להיות פייק". מייד אחריה מופיעה שוב הדמות, ופני הדובר הופכים לפנים של מי שגילם אותה.** יתר על כן, המסר לאחר מכן קורא לח"כ לפיד האמיתי לפעול כפי שהמשיבה 1 מעוניינת, כלומר ברור שבמציאות הוא אינו אומר (לפחות בשלב זה) ולא אמר לפני כן את מה שנאמר קודם לכן. כל בוחרת מן היישוב תדע לזהות מה אמת ומה לא. **גם אם היא תצפה רק בחלק מהסרטון שבו הנאום של "ח"כ לפיד", הרי שהכתובית שמציינת שאין מדובר בו, תסיר את החשש לזיהוי שגוי.** על כן אני סבור שאין ודאות, ודאי לא ודאות קרובה, להטעיית הציבור. משכך, בנסיבות אלו אין הפרה של הוראת סעיף 13 לחוק, אף בלי לבחון את התנאים האחרים המנויים בסעיף" (הדגשות לא במקור).

53. מכאן שהשופט פוגלמן, אימץ למעשה את העמדה כי לצורך מניעת הפרה של "איסור הפרעה" לפי סעיף 13 לחוק התעמולה, נדרש הפרסום להכיל חובת גילוי באמצעות סימון ברור שמדובר בפרסום שנעשה באופן טכנולוגי. יודגש כי השופט פוגלמן מצא לנכון לציין, כחלק מנימוקיו לדחיית העתירה, כי הסימון הופיע לכל אורך הסרטון.

54. על רקע דברים אלה, וכפי שננמק להלן, אנו סבורים כי **שימוש שיטתי בבינה מלאכותית יוצרת להפקת דיפ-פייק ללא סימון כראוי, צריך להיות אסור**. זאת משום שהוא מהווה הפרה של האיסור הקבוע בסעיף 13 לחוק ביחס ל"הפרעה בלתי הוגנת של תעמולת בחירות או רשימת מועמדים אחרת או למענה". זאת עקב הפגיעה שתוארה לעיל בהרחבה. פרשנות זו מתבקשת הן מקריאה תכליתית של הסעיף, הן מתורת המשפט של משפט וטכנולוגיה ויישומה כחלק מדוקטרינת ההגינות המגולמת בלשונו של סעיף 13 לחוק דרכי תעמולה, והן מהעקרונות המקובלים בעולם לבינה מלאכותית אחראית.

55. מובן שהפרשנות של הוראה זו כפופה, כמובן, לפרשנות הכללית של דיני התעמולה, כפי שפותחה בפסיקת יושבי ראש ועדות הבחירות המרכזית ובית המשפט העליון, לפיה ככל שמתקרבים לבחירות עצמן, המונח תעמולת בחירות יפורש באופן רחב, וככל שאנו מתרחקים מהבחירות – המונח יפורש בצמצום (מבחן התכלית הדומיננטית).

---

[52] תב"כ 9/24 **יש עתיד נ' עמותת "כן לשלום"** (18.1.2021)

<u>פרשנות סעיף 13 בכלים של משפט וטכנולוגיה</u>

56. נקודת מוצא ראשונה לבחינת אופן תחולת סעיף 13 על שימוש שיטתי בכלים טכנולוגיים להפקת דיפ-פייק הינו הנורמה הכללית החלה בישראל, הקבועה בסעיף 3 לחוק המחשבים, התשנ"ה-1995.[53] סעיף זה קובע עבירה על "**הפקת מידע כוזב באמצעות מחשב**". כפי שעולה מהנחיית פרקליט המדינה בעניין זה: "על פי לשונו המרחיבה של סעיף 3 לחוק, למעשה כל דבר שקר מכוון עשוי לבסס את יסודות העבירה".[54]

57. בהתאם לכך קובעת הנחיית הפרקליט שיקולים מנחים למקרים המתאימים להעמדה לדין. שיקולים אלה נגזרים מחומרת האיום על הערך המוגן, כוללים את הסיכון הכרוך בתוכן שהופק, מידת התכנון והתחכום שבמידע, היקף תפוצת המידע הכוזב, ומידת הנזק שנגרם או עלול היה להיגרם כתוצאה מהעברת המידע הכוזב או הפצתו. שיקולים אלה מצביעים על כך ששימוש בבינה מלאכותית להפקת תכנים סינתטיים המתחזים לאותנטיים נכללים במעשים שיש אינטרס ציבורי במניעתם. עוד יודגש כי הפקת תוכן כוזב והונאה באופן כללי עשויות להיות עבירות על דינים אחרים, אולם בחירתו של המחוקק לקבוע עבירה קונקרטית ביחס לשימוש במחשב להפקת תוכן כוזב שהוא "פלט", מלמדת על הצורך לקבוע איסור קונקרטי כנגד הסיכון הקונקרטי. כלומר קיומה של עבירה ייעודית בחוק המחשבים מלמדת על החשיבות שראה המחוקק לעסוק באופן ספציפי בשימוש לרעה במחשבים.[55] זאת משום שעולה החשש כי עקב הזמינות של מחשבים, האפשרויות השונות לשיבוש מידע בהם עלול להיות סיכון משמעותי יותר לשימוש לרעה בהם, העלול להביא לשיבוש מידע. להתממשותו של סיכון זה יש השפעה על היכולת להסתמך על מידע ממוחשב, ועל אמינות הפלטים הממוחשבים. לנוכח חשיבותו הייחודית של המחשב והמידע הממוחשב בחברה המודרנית, יש צורך לקבוע איסורים קונקרטיים שמטרתם הגנה על המחשב והמידע הממוחשב.[56] דומה כי מכלול היבטים אלה רלבנטיים לשימוש בבינה מלאכותית יוצרת.

58. עקב כך, אנו סבורים כי כאשר מדובר בשימוש בכלי של בינה מלאכותית יוצרת, באופן שנועד להקשות על הנמען לזהות את האותנטיות של התוכן, האמור מהווה הפרה לכאורה של הנורמה הספציפית בתחום חוק המחשבים. כאשר מוסיפים ביחס לסוג השימוש את המאפיינים של שדה התעמולה הפוליטית, עולה כי מדובר במעשים שיש בהם פוטנציאל סיכון גבוה יותר, המצדיק התערבות של הדין.

59. מובן שיושב ראש ועדת הבחירות המרכזית אינו מוסמך ליתן צו המונע ביצוע עבירה על חוק המחשבים (פרשת **טיבי**). יחד עם זאת, פשיטא שהפרת דין ובפרט בנסיבות הסיכונים שתוארו לעיל, הופכת את הפרסום, כפשוטו – להפרעה לא הוגנת, ולכן אסור גם לפי סעיף 13 לחוק דרכי תעמולה.

60. בהמשך לכך, הרובד של שימוש בכלי בינה מלאכותית, הינו ברובד הצורני של הביטוי, בכך שהוא עוסק בהיבטים טכניים. בהתאם לכך, מאחר שמדובר בשימוש באמצעי טכנולוגי שיש בו סיכון רב, דומה כי

---

[53] סעיף 3 לחוק המחשבים, שכותרתו "מידע כוזב או פלט כוזב" קובע כי: (א) העושה אחד מאלה, דינו – מאסר חמש שנים: (1) מעביר לאחר או מאחסן במחשב מידע כוזב או עושה פעולה לגבי מידע כדי שתוצאתה תהיה מידע כוזב או פלט כוזב; (2) כותב תוכנה, מעביר תוכנה לאחר או מאחסן תוכנה במחשב, כדי שתוצאת השימוש בה תהיה מידע כוזב או פלט כוזב, או מפעיל מחשב תוך כדי שימוש בתוכנה כאמור. (ב) בסעיף זה, "מידע כוזב" ו"פלט כוזב" – מידע או פלט שיש בהם כדי להטעות, בהתאם למטרות השימוש בהם".

[54] הנחיית פרקליט המדינה מספר 2.25 - מדיניות העמדה לדין על עבירה לפי סעיף 3 לחוק המחשבים - "מידע כוזב" או "פלט כוזב" (10.11.2016).

[55] יודגש כי ההתייחסות לעבירה לפי סעיף 3 לא נועדה להכריע האם שימוש בבינה מלאכותית יוצרת להפקת תכנים סינתטיים שקריים הינה בהכרח עבירה פלילית, אלא להצביע על כך שהדין הקיים מכיר בסיכון שנוצר עקב פוטנציאל השימוש במחשב להפקת מידע כוזב.

[56] ראו בעניין זה: מיגל דויטש, חוק המחשבים במבחן העיתים – זווית המבט של מציע החוק, **שערי משפט** ד(2), התשס"ו, עמ' 257, בעמוד 265-263.

גם התרופה מצויה ברובד הטכנולוגי הצורני, התרופה לסיכון זה היא קודם כל **סימון התוכן** יציר בינה מלאכותית באופן בולט. סימון התוכן מאפשר לאזן בין התועלות הרבות בשימוש בבינה מלאכותית יוצרת, לבין החשש הגבוה להטעיה, מרפא את הסיכון ברובד הצורני, ומאפשר חופש ביטוי פוליטי מלא (בכפוף למגבלות אחרות של הדין). למען שלמות התמונה יודגש כי כל האמור פה עוסק בהפקת תכנים יציר בינה מלאכותית, ואינו עוסק בשימוש בבינה מלאכותית יוצרת באמצעות בוטים אינטראקטיביים, שלהם רובד סיכונים אחר וכלי ניהול סיכונים שונים.

61. גישה זו המאתרת את אמצעי ניהול הסיכונים ברובד הצורני טכנולוגי, ומעצבת אותו באופן שמאפשר לאזן בין השימושים השונים בטכנולוגיה, עולה בקנה אחד עם תורת המשפט של משפט וטכנולוגיה, וכן עם עקרונות מקובלים, בעלי תוקף גם בישראל, ביחס לבינה מלאכותית מהימנה.

62. כך, כאשר עולה שאלה של החלת נורמה קיימת על שימושים חדשים בטכנולוגיה, תורת המשפט בתחום המשפט והטכנולוגיה מדריכה אותנו לאתר את התכלית של הנורמה, איתור הסיכונים הקונקרטיים הנובעים מהשיבוש הטכנולוגי לאותה תכלית, ועיצוב כללי התנהגות במסגרת הנורמה שמטרתם צמצום סיכונים אלה. גישה זו מבוססת על כך שהחקיקה הינה ניטרלית טכנולוגית, ונועדה לחול על מגוון שימושים ויישומים.

63. יפים לעניין זה דבריו של כבוד השופט סולברג בעניין **איגוד האינטרנט**, אשר גם צוטטו בידי יו"ר ועדת הבחירות בתב"כ 8/22:

> "בידוע, כי המשפט מדדה בעצלתיים אחר חידושי העולם, וכי החקיקה אינה מדביקה את קצב התקדמות המדע וישׂומיו. מפרי-החוק מסתגלים לקידמה מהר יותר מאוכפיו. זו אקסיומה. לראשונים אין עפּבות; לאחרונים יש. שנים רבות חלפו מעת המצאת המחשב ועד לחקיקת חוק המחשבים, התשנ"ה-1995. לא חלפו אז דור או שניים במונחי מחשב, והחוק כבר נמצא מיושן, כי המחוקק לא צפה – ולא יכל לצפות – את חידושי הטכנולוגיה. [...]
>
> גישה אקסטריטוריאלית שכזו אין לקבל. אמנם אין להרבות בחקיקה באופן שתפגע בתועלת הרבה הצפוּנה באינטרנט, וגם אין טעם בחקיקה שלא ניתן יהיה לאכפה לאור מאפייני הרשת. ברם, **למרחב הוירטואלי יש השפעה מוחשית בעולם הממשי – לטב ולמוטב – ועולמנו זה לא יסכון ולא יסבול את פריקת עול המשפט מן המרחב הוירטואלי**. [...]
>
> עם זאת, אין לכחד: הסדרה חוקית של הנעשה במרחב הוירטואלי היא מורכבת ומסובכת. קשה להגדיר את הדין הרצוי, וגם לא בנקל ניתן להחיל את הדין המצוי. לא בכדי יש שהגיעו למסקנה כי מדובר בתחום שראוי להסדרה חוקית; **יש הסבורים כי הפסיקה היא המסגרת הראויה לעשיית ההתאמות הנדרשות לעידן האינטרנט**; ואלה ואלה מתלבטים גם לגבי מידת השיתוף של קהילת משתמשי האינטרנט בעיצוב הכללים במרחב הוירטואלי וביישומם [...]
>
> דומה כי הסדרה כוללת של הנושא בחוק, באופן מדוייק ובהתאמה לעידן הוירטואלי היא עדיפה. השאלה היא האם בהעדר הסדר חקיקתי עדכני וממצה, יש בחוק, כפי נוסחו לעת הזאת, מענה מספּק על מנת להניח את הדעת באשר לסמכותה של המשטרה להוציא את הצווים הנדונים. בית המשפט לעניינים מינהליים פסק כי יש להמתין למוצא פיו של

המחוקק, וכי כל עוד לא יתוקן החוק אין ניתן להכיר בסמכותה הנ"ל של המשטרה. פסיקה זו צופנת בחובּה קשיים.

'תקופת המתנה' הריהי מגבילה, לעיתים מסכלת, מתן מענה הולם בתחום אכיפת החוק ועשיית המשפט. גישה שכזו, ביחד עם קצב התקדמות הטכנולוגיה כמתואר, צפויה להביא לכך שדברי-חקיקה רבים לא יהיו רלוונטיים ולא ישימים. גם לאחר שייצא דבר-חקיקה מתוקן מתחת ידו של המחוקק, אין זה מן הנמנע שעד מהרה לא יהיה אף הוא רלוונטי עוד, ולכן המתנה למחוקק לא בהכרח תביא מזור

[...]

הפשיעה באמצעות האינטרנט הולכת ומשתכללת, מחד גיסא; חוק העונשין מתקדם לאַטו, מאידך גיסא. יש צורך לגשר ביניהם. הגישור נעשה בכנסת באמצעות חקיקה, ובבית המשפט באמצעות פסיקה. מציאות החיים אינה מאפשרת להמתין לתיקון חוק העונשין ולקביעה סטטוטורית איזו עבירה ניתן לבצע באמצעות האינטרנט ואיזו לא. גם אין צורך משפטי להמתין עד שהמחוקק יסרוק את כל הוראות הדין הפלילי ויאמר לאיזו מהן יש תחולה גם באינטרנט. בית המשפט נדרש ליתן מענה לנושא קונקרטי שהובא לפניו, ולפסוק לכאן או לכאן. **לא ב'חקיקה שיפוטית' עסקינן, אלא ב'יצירה שיפוטית'. אותן עבירות פליליות שנחקקו לפני שנים רבות, כשביצוען נעשה ברחובה של עיר, מבוצעות עתה בתפוצה רחבה וביתר שאת באמצעות האינטרנט. לעיתים היסוד העובדתי זהה, היסוד הנפשי זהה, התכלית החקיקתית זהה; והנזק, לעיתים מזומנות, רב וחמור עוד יותר במימד הוירטואלי**.

למותר לציין, כי עודנו כבולים במגבלות הלשון, ולא נוכל לחרוג מגבולותיה על מנת לפרוש מצודתנו על כל מה שנתפשׂ כחטא, עוון, פשע או עוולה בעולם ה'ממשי', ונחזה להיות ככזה גם במרחב הוירטואלי. יחד עם זאת, תכלית החקיקה, המשותפת בדרך כלל לאותן עבירות, בין אם הן נעשות פה, בין אם שם, מחייבת לעשות מאמץ פרשני על מנת למנוע יצירת פרצות מלאכותיות באכיפת הדין הפלילי, שנְזקן רב" (ההדגשות במקור).[57]

64. החשיבות בפרשנות עדכנית של הדין הקיים, תוך הפנמה של איומים שמייצרות טכנולוגיות חדשות, שלא נצפו בעת חקיקתו, הוא חיוני כאשר על הכף מונחים אינטרסים ציבורים כבדי משקל כגון ביטחון לאומי, סדר ציבורי, טוהר הבחירות וזכויות יסוד של האזרחים, ובמקרה זה - הזכות לבחור ללא השפעה בלתי הוגנת וגניבת דעת. זאת בניגוד למצבים בהם השלטון הוא זה שמבקש להוסיף באמצעות פרשנות יצירתית סמכויות שלטוניות הפוגעות בזכויות אדם, המתאפשרות לו בשל חידושים טכנולוגיים, תוך עקיפת הצורך בהסמכה מפורשת בחוק (ר', למשל, בג"ץ 8298/22 **הסנגוריה הציבורית נ' היועצת המשפטית לממשלה** (נבו 31.8.2025), בפסקה 95 ואילך לפסק דינו של המשנה לנשיא, השופט סולברג).

65. דברים אלה מקבלים משנה תוקף כאשר בוחנים את העקרונות המתהווים בישום תפיסות אלה ביחס לבינה מלאכותית. עקרונות אלה כוללים פיתוח זהיר של כללי התנהגות אחראית ביחס לבינה מלאכותית, בהתבסס על העקרונות המקובלים במדינות המפותחות, ותוך התאמה של העקרונות

[57] עע"מ 3782/12 **מפקד מחוז תל אביב-יפו במשטרת ישראל נ' איגוד האינטרנט הישראלי** (24.3.2013), בפסקה 23.

להקשר השימוש והסיכון לזכויות יסוד ואינטרסים ציבורים.[58] גישה זו הינה בהלימה מלאה להתפתחות של תורת המשפט בתחום המשפט והטכנולוגיה.

66. כך, וכפי שניתן לראות מהפעילות הממשלתית בנושא זה, אופן יישום העקרונות נשען על תורת המשפט המתפתחת בתחום זה כפי שהיא באה לידי ביטוי בפסיקת בית המשפט העליון בעניין יישום הוראות הדין הקיים על תופעות טכנולוגיות חדשות, והנחיית היועץ המשפטי לממשלה 1.2500.[59] העקרונות המוצעים ליישום בתחום הבינה המלאכותית המהימנה שמים דגש על ניהול סיכונים ענפי בידי הגורם האחראי, באופן שכללי ההתנהגות לא יוחלו באופן גורף.[60] גישה זו תיושם להלן בכל הקשור להמלצות המוצעות בעניין סימון בינה מלאכותית.

67. ניתן לומר כי גישה זו מעודדת את הגישה הפרשנית התכליתית, **תוך שהיא מציידת אותה במידע ובכלים נדרשים על מנת להתאים את הוראות הדין לשימושים הטכנולוגים החדשים.** הדגמה לגישה זו ניתן לראות גם בדו"ח הצוות הבין-משרדי לבחינת שימושי בינה מלאכותית בסקטור הפיננסי.[61]

<u>עקרונות בינה מלאכותית מהימנה, אחריותיות, גילוי וחובת סימון</u>

68. כפי שמציין מסמך עקרונות המדיניות, מערכות "בינה מלאכותית יוצרת" מסוגלות ליצור תוכן מסוגים שונים, "המאתגרת את היכולת האנושית להבחין בין תוכן אמיתי לבין כזה אשר נוצר על ידי מכונה."[62]

69. בהתייחס לעקרונות בינה מלאכותית "מהימנה"[63], שמטרתם התמודדות עם הסיכונים מבינה מלאכותית באופן שמאפשר מיצוי התועלות ממנה, עיקרון רלבנטי ראשון הוא עיקרון ה-"אחריותיות". בהתאם לעיקרון זה "**מפתחי בינה מלאכותית, מפעיליה או המשתמשים בה יגלו אחריות לתפקודה התקין, ולקיום העקרונות האתיים האחרים בפעילותם, בין היתר בשים לב לתפיסות מקובלות של ניהול סיכונים ולאפשרויות הטכנולוגיות הזמינות.**" עיקרון האחריותיות נועד לאפשר חדשנות ושימוש בבינה מלאכותית, על אף הסיכונים, אולם לצד זאת מצפה כי השימוש ייעשה באופן אחראי ותוך ניהול הסיכונים שנוצרים כתוצאה מהשימוש בטכנולוגיה.[64]

70. עיקרון רלבנטי שני הינו "עיקרון השקיפות והסברתיות" הקובע כי "בפיתוח בינה מלאכותית ובשימוש בה, יש להביא בחשבון, בשים לב למגוון השיקולים הנוגעים לעניין, את הצורך ליידע מי שבא במגע עם

---

[58] ראו **במסמך העקרונות**, הערת שוליים 4 לעיל, לגבי קידום רגולציה ענפית ולא חקיקת מסגרת רוחבית (עמ' 92), התחשבות ברגולציה במדינות המפותחות ובארגונים הבינלאומיים (עמוד 94), אימוץ עקרונות אתיים בהתבסס על עקרונות ה- OECD (עמוד 102).

[59] שם, ביחס לגישה הפרשנית הכללית בעמ' 24-23, המצטטת אף היא את עמדתו של המשנה לנשיא סולברג בעניין איגוד האינטרנט.

[60] שם, בעמוד 94.

[61] **דוח הסקטור הפיננסי**, הערת שוליים 4 לעיל.

[62] מסמך העקרונות, הערת שוליים 4 לעיל, בעמ' 8.

[63] כמפורט שם, בעמ' 103, עקרונות אלה מבוססים על המלצות ה- OECD בנושא זה, שהינן הסטנדרט הראשון והמוביל במדינות המפותחות.

[64] שם, בעמ' 109-108: "עיקרון האחריותיות נועד לבסס את קיומו של "גורם אחראי" לבינה מלאכותית, שעשוי להיות למשל המפתח או המפעיל של מערכת מבוססת בינה מלאכותית... אחריות זו משמעה בין היתר הצורך לנקוט אמצעים לתפקודה התקין, לפעול לצמצם את החשש לפגיעה בזכויות יסוד ובאינטרסים ציבורים, ולשאוף לקיום העקרונות שתוארו לעיל. אופן ההתמודדות עם הסיכונים ויישום העקרונות מבוסס על אמצעים סבירים ויכולת צפיות סבירה, לפי מידת ההתפתחות הטכנולוגית. העיקרון משרת גם את היכולת לבצע בקרה בידי גורם בלתי תלוי, במקרים הרלוונטיים, בנוגע לאופן הפיתוח של בינה מלאכותית והשימוש בה. עיקרון זה מבטא אמון בגורמים האמורים, ומאפשר להם לפעול בתנאי שינהגו כמי שאחראיים לבינה המלאכותית. הוא מאפשר לגשר בין העקרונות המפורטים לעיל, לבין החדשנות המבוצעת בפועל בידי ארגונים המפתחים או משתמשים בבינה מלאכותית. עיקרון האחריותיות משקף את התפיסה הבסיסית שלפיה ראוי כי מי שיוצר את הסיכון יתמודד איתו ויהיה אחראי לו, בלי להחציו את הסיכון לצדדים אחרים".

בינה מלאכותית על כך, בהתחשב במידת השפעה, השלכותיה על מי שמושפע ממנה והאפשרויות הטכנולוגיות הזמינות."[65]

71. כדוגמא ליישום גישה זו, בדוח על בינה מלאכותית במגזר הפיננסי, שנערך בין שנת 2024 ל–2025, כבר מתייחסים הכותבים באופן מפורש לסיכונים הנובעים מבינה מלאכותית יוצרת ליצירת תכני deepfake. בהקשר זה בחרו עורכי הדוח לייחד עיסוק בנושא זה בשל החששות שתוכן סינתטי מוביל להשפעה על החלטות של גורמים שונים על סמך אותו מידע, ובכלל זה החלטות של משקיעים.

72. בפרט, עורכי הדוח מציינים כי אמנם "בעלי אינטרס פעלו להפצת דיסאינפורמציה באמצעות ידיעות שקריות, לעיתים תוך שימוש בזהויות בדויות, **אולם זמינותם ותפוצתם של כלי בינה מלאכותית מקלים על יצירתה, מגבירים את היקפה ואת התאמתה לנמענים, ולכן גם את האפקטיביות שלה בהשגת מטרות הגורם המשבש**".[66] (הדגשה אינה במקור).

73. בהמשך לסקירת הסיכונים מציינים עורכי הדוח את הצורך להיערך בשורה של כלים מערכתיים להתמודדות עם סיכון זה, ובכלל זה ניטור הפצה של דיסאינפורמציה, ובחינת כלי תגובה והיערכות של הרגולטורים הפיננסים עצמם.[67]

74. להשלמה של הסקירה נציין כי עקרונות בדבר אחריותיות ארגונים המשתמשים בבינה מלאכותית יוצרת, וסימון תכנים יצירי בינה מלאכותית, מופיעים במסמכים בינלאומיים נוספים שישראל צד להם, ויכולים לסייע בחיבור שבין איתור הסיכונים לבין כלל ההתנהגות הנדרש לצמצום הסיכון.

75. כך, אמנת מועצת אירופה בנושא בינה מלאכותית, זכויות אדם, דמוקרטיה ושלטון החוק כוללת בסעיף 8 לאמנה את עיקרון ה-"שקיפות והפיקוח" ("Transparency and Oversight").[68] בהתאם לעיקרון זה על צד לאמנה לנקוט צעדים, בהתאם להקשר ולסוג הסיכון, בקשר לשקיפות ביחס למערכות בינה מלאכותית, ובכלל זה "**זיהוי של תוכן שנוצר בידי מערכות בינה מלאכותית**".[69]

76. האמנה הינה ההסכם הבינלאומי המחייב הראשון לאסדרת השימוש בבינה מלאכותית.[70] ישראל הוזמנה לקחת חלק בגיבוש האמנה והייתה מעורבת בניסוחה. ישראל חתמה על האמנה, אולם טרם אישררה אותה.[71]

77. לבסוף נציין את תהליך עקרונות הבינה המלאכותית של הירושימה ביחס **לבינה מלאכותית יוצרת** (Generative AI).[72] בשונה מהמסמכים הבינלאומיים שנסקרו עד כה, מסמך זה ממוקד בסוג מסוים של טכנולוגיה, "מערכות בינה מלאכותית מתקדמות", כלומר בעיקר מערכות של מודלי שפה גדולים (LLM) ו-"בינה מלאכותית יוצרת", שהן רלבנטיות במיוחד להקשר הנוכחי. זו המסגרת הבינלאומית הראשונה הכוללת עקרונות מנחים וקוד אתי שמטרתם לקדם בטיחות, אבטחה ואמינות במערכות בינה

---

[65] שם, בעמ' 106.
[66] דוח הסקטור הפיננסי, הערת שוליים 4 לעיל, בעמ' 149.
[67] שם, בעמ' 154.
[68] ר': Council of Europe Framework Convention on Artificial Intelligence and Human rights, Democracy and the Rule of Law. https://rm.coe.int/1680afae3c (להלן - COE AI).
[69] COE AI - "Article 8 – Transparency and oversight: "Each Party shall adopt or maintain measures to ensure that adequate transparency and oversight requirements tailored to the specific contexts and risks are in place in respect of activities within the lifecycle of artificial intelligence systems, including with regard to the content generated by artificial intelligence systems."
[70] האמנה מחייבת את מדינות האמנה ליישם את הוראותיה ביחס לשימושים בבינה מלאכותית בידי גופים ציבורים, ואילו אופן היישום של הוראות האמנה על גופים פרטים נתון לשיקול דעת המדינות והן נדרשות לדווח על כך בלבד. ראו סעיף 3 לאמנה.
[71] https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence
[72] https://www.soumu.go.jp/hiroshimaaiprocess/en/index.html

מלאכותית יוצרת. המסגרת פותחה על ידי מדינות ה- G7, והצטרפו אליה 56 מדינות נוספות הכוללות גם את ישראל. כפי שנראה להלן, מסגרת זו כוללת במפורש התייחסות לסימון תכנים בידי בינה מלאכותית יוצרת, והובילה לכך שהיצרניות של המודלים הגדולים כוללים מרכיב זה ביכולות שהן מפתחות.

78. על מנת לקדם את ההשפעה של כללי התנהגות אלה בעולם טכנולוגי גלובלי, המדינות החברות נקטו צעדים מעשיים נוסף לקידום היישום של עקרונות בינה מלאכותית מהימנה. במסגרת תהליך זה פותח "קוד התנהגות" לארגונים המפתחים מערכות בינה מלאכותית.[73] "קוד התנהגות" זה כולל גם חובה לקדם **"פיתוח והטמעה של כלים לזיהוי מקור ומהימנות התוכן, כמו סימוני מים או טכניקות אחרות שיאפשרו למשתמשים לזהות שמדובר בתוכן שנוצר בידי בינה מלאכותית"**.[74] בהמשך לכך, מדינות ה-G7 השיקו מסגרת דיווח וולונטרית כדי לעודד שקיפות ואחריותיות בקרב ארגונים המפתחים מערכות בינה מלאכותית מתקדמות.[75]

79. בשנת 2025 כ-24 חברות דיווח על אודות אופן עמידתם בקוד ההתנהגות ובין היתר אמזון, גוגל, OpenAI, מיקרוסופט, אנטרופיק, ו-salesforce.[76] כפי שניתן לראות בטבלה שלהלן, חברות מרכזיות אלה כוללות יכולת **סימון של התוכן כיציר בינה מלאכותית כחלק מהאלגוריתם**, הידועה כ- "סימן מים" ("watermark").[77] כאמור סימן מים טכנולוגי הינו תכונה חשובה המאפשרת למחשבים לזהות את התוכן כדיפ-פייק, ובהתאמה לכך לגורמים נוספים בשרשרת הפצת המידע להציג מידע זה.

80. בנוסף, חלק מהחברות דיווחו כי על מנת למנוע שינוי של הסימנים הטכנולוגים, הן נוקטות בגישה רב שכבתית, הכוללת סימנים מסוגים שונים שימנעו הסרה של סימני המים. בהקשר זה יודגש כי קיומם של סימני מים טכנולוגים מאפשרים גם לפלטפורמות ההפצה, כגון פייסבוק ואינסטגרם להציג את המאפיין המלמד על כך שמדובר בתוכן יציר בינה מלאכותית.

---

[73] Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems) https://www.soumu.go.jp/hiroshimaaiprocess/en/documents.html

[74] Develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content.

[75] OECD, G7 reporting framework – Hiroshima AI Process (HAIP) international code of conduct for organizations developing advanced AI systems, https://transparency.oecd.ai/

[76] https://transparency.oecd.ai/reports

[77] על ההיבטים השונים של סימונים טכניים, ראו במסמך ההמלצות של NIST, הערת שוליים 22 לעיל.

81. בטבלה שלהלן נסכם את המצגים של חברות מרכזיות רלבנטיות בדיווחיהן.[78]

חברה	מנגנונים וטכנולוגיות מרכזיים	דוגמאות ליישום ומוצרים	תקנים ושיתופי פעולה
Amazon	**סימני מים בלתי נראים** לווידאו, תמונות ואודיו; **Content Credentials** (מטא-דאטה חתום קריפטוגרפית).	מיושם ב-**Amazon Nova** (דגמי Reel, Canvas, Sonic) ו-**Titan Image Generator**; שילוב ב-AWS Elemental MediaConvert עבור גופי שידור.	חברה בוועדה המנהלת של ה-**C2PA**; מעורבת ביוזמות של NIST, AISIC ו-FMF.
Google	**SynthID** (סימון מים לטקסט, תמונה, וידאו ואודיו); כלי **"About this image"** להקשר ויזואלי; **Content Credentials**.	אפליקציית **Gemini** (בווב ובמובייל); שילוב תקני C2PA בחיפוש (Search), בפרסומות (Ads) וביוטיוב.	חברה בוועדה המנהלת של ה-**C2PA**; שחררה את **SynthID-Text** כקוד פתוח לשימוש התעשייה.
OpenAI	**Content Credentials** (מטא-דאטה חתום); סימני מים עמידים לשינויים לאודיו; **סימני מים גלויים** לווידאו.	בשימוש ב-**DALL·E 3**, יצירת תמונות ב-**GPT-4o**, **Sora** (וידאו) ו-**Voice Engine** (אודיו).	חברה בוועדה המנהלת של ה-**C2PA**; חברה בכוח המשימה **NIST 2.1** בנושא תוכן סינתטי.
Microsoft	**Content Credentials** (מטא-דאטה); סימני מים בלתי נראים; **כלי שלמות תוכן** (Content Integrity Tools) לאימות מדיה מקורית.	**Bing Image Creator**, **LinkedIn** (תיוג אוטומטי), **Azure OpenAI** (DALL-E); מיושם ב-Copilot, Designer ו-Paint.	חברה מייסדת של ה-**C2PA**; שיתוף פעולה עם TruePic לפיתוח אפליקציות צילום מאובטחות.
Anthropic	שקיפות באמצעות **דיסקליימרים**, תזכורות בתוך השיחה ותיוג המבהיר כי מדובר בתוכן שנוצר על ידי AI.	ממשק ה-**Claude.ai**, חומרי שיווק ותיעוד למפתחים המדגישים את טבעו של קלאוד כעוזר AI.	משתתפת בקידום **תקנים טכניים גלובליים** למקור (Provenance) ומודעות ציבורית דרך ה-Transparency Hub.

[78] הטבלה יוצרה באמצעות notebook LM על בסיס דיווחי החברות המופיעות באתר הדיווח ונערכה לה בדיקה בידי אדם.

82. עד כאן נסקרו הוראות רלבנטיות ממסמכי המדיניות הישראלים ומסמכים בינלאומיים שישראל הצטרפה אליהם. המשפט ההשוואתי כולל הוראות נוספות ביחס לסימון תכני בינה מלאכותית יוצרת, ובפרט תכנים סינתטיים, הן באופן כללי, והן ביחס לשימוש בתוכן סינתטי בתעמולה פוליטית.

83. כך, מדינות דמוקרטיות רבות דורשות סימון מודעות בחירות כדי להעלות את מודעות האזרחים למידע סינתטי או מידע אחר שיש בו כדי להטעות. בהקשר זה יצוין כי הדין האירופי כולל חובה משפטית מפורשת לסמן תכנים שנוצרו בידי בינה מלאכותית בסעיף 50 לחוק ה-AI האירופי.[79] בכל הקשור לתכנים חזותיים גם פותח תקן של התעשייה המאפשר לסמן באופן אחיד תכנים כאלה,[80] ובהתאמה לפלטפורמות התוכן לאכוף באופן אוטומטי חובת גילוי זו. סעיף 50 מפרט חובות "שקיפות" (transparency) של ספקים (Providers) ומפעילים (Deployers) של מערכות בינה מלאכותית המצוינות בסעיף. למעשה הסעיף מטיל חובות ללא קשר לחובות אחרות המופיעות בחוק, ובנוסף לחובות "שקיפות" אחרות שעשויות לחול, כגון על מפתחים של מערכות בינה מלאכותית בסיכון גבוה.[81] לצורך יישום סעיף זה, ובדומה לעקרונות הירושימה שתוארו לעיל, האיחוד האירופי קיים הליך היוועצות ארוך לעניין אופן קיום החובה על יצרני מערכות בינה מלאכותית ועל משתמשים במערכות אלה, ופרסם בדצמבר 2025 טיוטת כללי פעולה בנושא זה.[82]

84. בהקשר של דיני הבחירות, הנושא מופיע הן בצו האירופי הקונקרטי הקשור לאחריות הפלטפורמות שצוין לעיל, וכן בחקיקה ביחס לחובות גילוי ושקיפות בפרסום פוליטי.[83]

85. הוראות דומות להוראות אלה נכללות גם בחוקי בחירות ב- 20 מדינות ארה"ב.[84] דרישות הגילוי הנפוצות הן סימון הקובע שהתוכן "עבר מניפולציה" או מכיל "תוכן שנוצר על ידי AI" או "תוכן סינתטי". הסימון צריך להיות ברור ובולט כך שצופה סביר יכול להבין אותו.[85]

86. להשלמה יצוין עוד כי לצד ההסדרים החקוקים, חלק מהפלטפורמות הטכנולוגיות המשמשות **להפצת תכנים**, מחייבות בהתאם לכללי השימוש שלהן סימון תכנים סינתטיים המופצים על ידן. כלומר הנורמה הגלובלית המתפתחת בתעשייה הינה לתמוך את יכולת הזיהוי של הציבור לגבי דיפ-פייק, כחלק מהתנהלות אחראית שמטרתה שימוש בטכנולוגיה באופן אחראי.

87. עם זאת, כללים אלה אינה אחידים ומפעילים לעיתים מבחנים מהותיים מסוגים שונים. למשמעות החלת תנאי השימוש של הפלטפורמות בהקשר של תעמולת בחירות ראו: תב"כ 27/21 **שמיר מערכות נ' ישראל ביתנו**, בפסקה 22 ולאסמכתאות שם.

---

[79] EU AI Act, Section 50, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689#cpt_IV

[80] Coalition for Content Provenance and Authenticity (C2PA), https://spec.c2pa.org/specifications/specifications/2.2/index.html

[81] *AI Act: A Practical Guide*, edited by Rolf Schwartmann, et al., C.F. Müller Verlag, 2025, p. 182.

[82] https://digital-strategy.ec.europa.eu/en/library/first-draft-code-practice-transparency-ai-generated-content

[83] ראו: Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising (Text with EEA relevance), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202400900#cpt_II

[84] See TechPolicy.Press, "Regulating Election Deepfakes: A Comparison of State Laws" (Jan. 8, 2025), https://www.techpolicy.press/regulating-election-deepfakes-a-comparison-of-state-laws/.

[85] See, e.g., Cal. Elec. Code (AB 2355) (requiring that political campaign committees use the disclaimer "This ad was generated or substantially altered using artificial intelligence"); Col. HB 24-1147 ("This (image/audio/video/multimedia) has been edited and depicts speech or conduct that falsely appears to be authentic or truthful").

88. להטלת חובת סימון גורפת על שימוש בטכנולוגיה יש תועלת גם במובן זה שהיא תאפשר פיקוח ואכיפה על כללים שווים גם בידי מפיצי התוכן השונים. בדומה לדברים שנקבעו בעניין תב"כ 8/21, קביעת כללים ברורים גם ביחס לפלטפורמות יש בו כדי להגביר את האפקטיביות של הסימון, צמצום ההטעיה, וסיוע משמעותי באיתור והתמודדות עם התערבות זרה.

<u>החשיבות בהטלת חובת סימון אחידה</u>

89. להטלת חובת סימון אחידה יש תועלת רבה ליישור מגרש המשחקים, ובהתאמה לכך לסייע גם לגורמי הביטחון לאתר מקרים של התערבות זרה. מנגד, תופעות הלוואי שלה, בהיותה אמצעי צורני בלבד, הינן מידתיות בשים לב לתועלות הרבות. אכן, יש להעדיף את הפשטות ואת האחידות, ולהטיל חובת סימון על כל תעמולה העושה שימוש בתוכן סינתטי, ולו כדי למנוע את הצורך לדון בכל מקרה ומקרה באופן פרטני האם יש בשימוש משום הפרעה אסורה לפי מבחן הבוחר הסביר.

90. כך, יצירת כללים מורכבים, שמותירים שיקול דעת רחב לגבי חובת סימון, משמעה שמסרים סינתטיים לא יסומנו, שכן מי שמפרסם יודע שגם אם תהיה תלונה על הפרסום, ועד שתתברר העתירה ליו"ר ועדת הבחירות, הפרסום כבר יעשה את שלו, ישפיע על דעת הבוחר, וכשהאיום על טוהר הבחירות יתממש, לא יהיה סעד אפקטיבי למנוע אותו. שנית – כי הרבה פעמים בלתי אפשרי לשים את האצבע על הנקודה שמפרידה בין תוכן המהווה הפרעה והונאה לבין תוכן שניתן להניח שהצופה בו יבין לבד כי מדובר בתוכן סינתטי. הניסיון מראה כי השימוש בתכנים סינתטיים מותירים לא פעם מרווח של תמיהה וחוסר ביטחון אצל הצופה האם מדובר בתכון מזויף או אותנטי. שלישית וזה עיקר - השימוש בדיפ-פייק עלול להיעשות כאשר אין זמן תגובה מתאים לצורך תיקון השפעתו הרבה.

<u>סיכום</u>

91. השימוש בטכנולוגיית בינה מלאכותית יוצרת להפקת דיפ-פייק מוביל לסיכונים משמעותיים לאינטרסים ציבורים ולזכויות יסוד, כגון ביטחון לאומי, טוהר הבחירות, אמון הציבור במידע בזירה הציבורית, הוגנות התחרות בין המפלגות, חופש הביטוי הפוליטי, וזכויות יסוד של הפרט הכוללות את יכולתו לממש באופן מלא את זכותו לבחור. מדובר בסיכונים אינהרנטיים לטכנולוגיה זו המביאים לחשש כי שימוש בה ללא נקיטת צעדים מתאימים יביא ל"הפרעה לא הוגנת לבחירות" מצד מדינות עויינות ומצד מפלגות. על רקע האמור, מוצע לפרש את החוק תוך הקפדה על חופש הביטוי הפוליטי של המפלגות והיכולת שלהן להשתמש בטכנולוגיה במסגרת הדין, תוך קביעת כללי התנהגות שמטרתם מניעת הפרת "אחריותיות" והחובה לסמן תוכן סינתטי יציר בינה מלאכותית.

92. ההגנה מפני הסיכונים של דיפ-פייק באופן כללי, והן מהשימוש בו לצורך התערבות זרה במרחב הסייבר, נדרשת להיות "רב שכבתית" כלומר לכלול אמצעים מסוגים שונים ופעילות של מוסדות שונים בהתאם לסמכותם. חובת הסימון מהווה נדבך מרכזי וראשון המאפשר לייעל ולמקד את מאמצי ההגנה השונים.

93. בהתאם לכך, מוצע כי שימוש של מפלגות או מי מטעמן בבינה מלאכותית יוצרת או אמצעים טכנולוגיים דומים להפקת תכנים סינתטיים יבוצע בהתאם לכללים הבאים:

א. תוכן סינתטי יסומן באופן בולט כיציר בינה מלאכותית/אמצעים טכנולוגיים, ויכלול גם סימן מים טכני, כלומר סימון טכנולוגי קריא מחשב, על גבי התוכן בדבר היותו יציר בינה מלאכותית.[86]

ב. במידה שהתוכן הסינתטי מתיימר להציג **תיאור עובדתי** של אירועים שהתרחשו טרם הפקת התוכן הסינתטי, ייכלל בתמונה או בסרטון בכל זמן הצגתם, כיתוב בולט כי: "התוכן אינו מהווה תיעוד של אירועים אמיתיים אלא נוצר באמצעים טכנולוגים לצרכי תעמולה".

ג. המפלגה שיצרה או מטעמה פורסם התוכן הסינתטי תכלול את שמה באופן בולט על גבי התוכן (כנדרש לפי סעיף 2א1 לחוק).

ד. חובת הסימון בתעמולת בחירות שכוללת תוכן סינתטי, תהיה אחידה ושוויונית כדי שהסימון יהיה משמעותי ובולט ויגשים את התכלית של מניעת הטעיה וגניבת דעת, ותכלול גם סימון מים טכני, בהתאם לכללים המקובלים. וזאת כדי שהסימון יהיה משמעותי ובולט ויגשים את התכלית של מניעת הטעיה וגניבת דעת ויקל על בעלי העניין השונים, גורמי הביטחון, המפלגות, כלי התקשורת, גורמי חברה אזרחית והבוחרים, לזהות את אותו תוכן.

94. בהתאם לכך מוצע לקבוע כי שימוש בבינה מלאכותית יוצרת ליצירת תוכן סינתטי, שלא בהתאם לתנאים המפורטים להלן, ייאסר, ובמיוחד בתקופה הסמוכה לבחירות ובכל מקרה, ייחשב "הפרעה לא הוגנת לבחירות" בניגוד לסעיף 13 .

95. יישום כללים אלה יאפשר לעשות שימוש אחראי בבינה מלאכותית תוך קידום חופש הביטוי הפוליטי וצמצום הסיכונים לפגיעה באינטרסים המוגנים על ידי החוק. לצד זאת, יישום הכללים יאפשר גם נקיטת גישה רב שכבתית להתמודדות עם הסיכונים השונים מתוכן סינתטי.

היום: כ"ה בשבט התשפ"ו, 12.02.2026

פרופ' ניבה אלקין – קורן
ראשת מרכז הנשיא מאיר שמגר למשפט דיגיטלי וחדשנות
הפקולטה למשפטים ע"ש בוכמן
אוניברסיטת תל אביב

עו"ד עמית אשכנזי
עמית מחקר בכיר, מרכז הנשיא מאיר שמגר למשפט דיגיטלי וחדשנות,
הפקולטה למשפטים ע"ש בוכמן
אוניברסיטת תל אביב

פרופ' יששכר רוזן – צבי
מנהל מרכז אדמונד ולילי ספרא לאתיקה
הפקולטה למשפטים ע"ש בוכמן
אוניברסיטת תל אביב

[86] סימן מים טכנולוגי מאפשר שמירה על היכולת לבדוק את האוטנטיות של הקובץ, וכן מאפשר לפלטפורמות הצגת התוכן להציג בעצמן כי התוכן הינו יציר בינה מלאכותית, ובכך לסייע בהבאת מידע זה לידיעת הציבור והגנה טובה יותר מפני שימוש לרעה של גורמים זדונים המבקשים להסיר את הסימן. ראו עוד להלן.