

19.12.2023

לכבוד
ועדת משנה בנושא בינה מלאכותית וטכנולוגיות מתקדמות
ועדת המדע והטכנולוגיה
כנסת ישראל

שלום רב,

הנדון: השימוש בבינה מלאכותית לצורך זיוף תמונות והפצת מידע כוזב במלחמת ישראל- חמאס בעזה (סקירה מקצועית לבקשת ועדת המשנה)

במענה לפניית ראש תחום מדע וטכנולוגיה במרכז המחקר והמידע של הכנסת, איגוד האינטרנט הישראלי מתכבד להגיש סקירה מקצועית וחומרי רקע. מחברים: עידן רינג (סמנכ"ל קהילה וחברה), ד"ר אסף וינר (סמנכ"ל מחקר ומדיניות), ניצן יסעור (חוקר רשת ודיסאינפורמציה).

1. פעולות האיגוד בתחום המוגנות ומניעת מידע כוזב או דיסאינפורמציה ברשת

איגוד האינטרנט הישראלי מפעיל מזה עשור את [קו הסיוע לאינטרנט בטוח](#) המספק לציבור הרחב סיוע ומענה אישי לטיפול בפגיעות, איומים ותוכן פוגעני ברשת בכלל וברשתות החברתיות בפרט. האיגוד נותן מענה אישי באמצעות הרשת ליותר מ-3000 פניות בשנה ומסייע למשתמשי רשת ישראל בדיווח והסרה על תכנים פוגעניים, מתקפות והטרדות רשת מסוגים שונים. במסגרת פעילות זו איגוד האינטרנט הישראלי מוכר על ידי מרבית פלטפורמות המדיה החברתית המובילות ושירותי רשת אחרים כגוף שהוא "מדווח רשמי" (Trusted Flagger) המוסמך במסגרת תכניות הבטיחות של הפלטפורמות להעביר דיווחים ובקשות הסרה על תוכן פוגעני ברשתות באופן רשמי וישיר באמצעות ערוצי דיווח מוסדרים ומוכרים למחלקות הבטיחות והמוגנות של הפלטפורמות והשירותים השונים.

לאחר הטבח בעוטף עזה ב-7/10 עם פתיחת הלחימה בעזה החל קו הסיוע לאינטרנט בטוח לפעול באופן מיידי כדי לתת מענה לפנייות מהציבור הרחב בנוגע לתוכן פוגעני ומידע כוזב המופץ ברשתות החברתיות על רקע המלחמה עם חמאס בעזה ופעל כדי לאסוף ולדווח לפלטפורמות השונות על תכנים כוזבים ומסיתים נגד ישראל ברשתות החברתיות ואפליקציות המסרים ופעל לפרסום סדרה של מדריכי חירום חיוניים לסייע במוגנות הציבור בפני חשיפה לתוכן פוגעני או תוכן גרפי קשה ומתקפות סייבר ברשתות על רקע המלחמה. לצד קו הסיוע לאינטרנט בטוח איגוד האינטרנט הישראלי פועל גם לקידום הגנת סייבר אזרחית של משתמשי הרשת בישראל, בין היתר באמצעות פורטל block.org.il ומדריכי מוגנות סייבר שונים המתפרסמים באתרי האיגוד וערוצי הסושיאל שלו.

2. היקף ואופי הפצת מידע כוזב ותוכן פוגעני מתחילת מלחמת חרבות ברזל

בחודש הראשון של המלחמה מספר הפניות לקו הסיוע לאינטרנט בטוח של איגוד האינטרנט הישראלי גדל פי 3 ביחס לתקופת השגרה טרום ימי הלחימה, והחלו לזרום לקו פניות רבות מהציבור על תכנים פוגעניים מסוגים שונים שהציפו את הרשתות על רקע המלחמה. ביניהם ניתן למנות מספר סוגים ונרטיבים עיקריים:

- תכנים גרפיים קשים מתוך מתקפת החמאס והטבח בעוטף עזה שצולמו על ידי המחבלים

- **תכנים תומכי טרור ותכנים של החמאס שהופצו על מנת לזרוע בהלה וחרדה בציבור**
- **ידיעות שקריות ומסיתות נגד קבוצות מיעוט או קבוצות שונות בתוך ישראל שכביכול תומכים או מברכים על הטבח**
- **תאוריות קונספירציה על "בוגדים מפנים" שטענו כי החמאס קיבל סיוע מגורמים בתוך ישראל**
- **דיווחים שקריים על פיגועי טרור או מתקפות עתידיות שנועדו להבהיל את הציבור**
- **תכנים כוזבים ומסיתים על פעילות צה"ל בעזה או על המצב בישראל**
- **מתקפות סייבר ונסיונות הונאה או פישונג שניסו לנצל את המשבר כדי לפגוע בציבור**

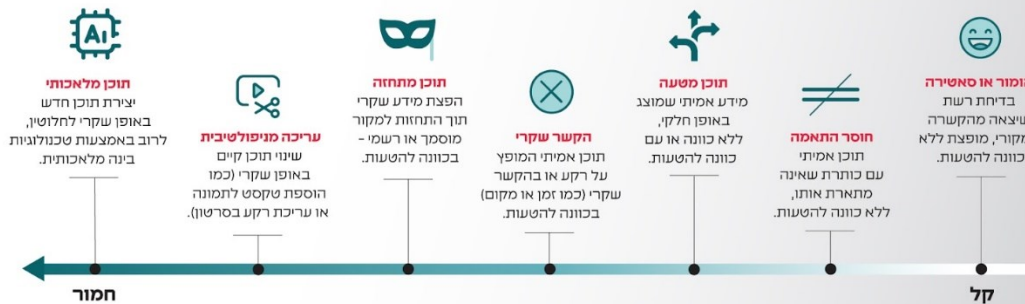
לצד קו הסיוע לאינטרנט בטוח של איגוד האינטרנט הישראלי פעלו מימיה הראשונים של המלחמה גם שורה של ארגונים אזרחיים קיימים כמו "פייק ריפורטר" וארגון "בודקים" או יוזמות אזרחית של אנשי הייטק ומתנדבים כמו "IronTruth" או "Fake-Off" שפעלו כדי לפתח מענים טכנולוגיים במטרה "לנקות את הרשת" מתכנים פוגעניים, שקריים ומסיתים. אזרחים רבים דיווחו לארגונים וליוזמות האזרחיות הרבות בשבועות הראשונים של המלחמה על עשרות אלפי פרסומים, פרופילים ותכנים שונים - חלקם שקריים, מטעים ו/או מסיתים וחלקם מקדמי שנאה נגד ישראל ומצדיקי טרור. לצד האינדיקציות הברורות של הדיווחים והפניות ליוזמות וגופי החברה האזרחית ניתן לראות את העלייה בהיקף התוכן השקרי והפוגעני שהופץ ברשתות בישראל, בתוכו גם תוכן לא אותנטי ומניפולטיבי, דרך הפרסומים הרשמיים של הפלטפורמות השונות על הסרות תוכן לא חוקי או פוגעני ודיווחים רשמיים של הפרקליטות, שפרסמה כי ביקשה להסיר היקף חסר תקדים של תכנים ופרסומים תומכי טרור ברשתות החברתיות, חלקם שקריים.

3. שימוש בכלי בינה מלאכותית ותוכן לא אותנטי במסגרת הפצת התוכן

במסגרת העלייה המשמעותית בהיקף וסוגי התוכן הכוזב שהופצו ברשתות החברתיות על רקע ואודות המלחמה והטבח בעוטף עזה ניתן לזהות בבירור גם מגמה בולטת של שימוש בטכנולוגיות בינה מלאכותית כדי ליצור פרופילים, פוסטים, תמונות ותכנים שונים במטרה לקדם ולהפיץ תכנים אלו, לרוב במטרה לפגוע בישראל ובאזרחים ישראלים. תוכן מסונתז הנוצר באמצעות בינה מלאכותית יכול להיות טקסט, תמונה, אודיו או וידאו - אך גם פרופילים המתחזים לגורמים אנושיים ברשתות החברתיות. יצירת טקסט מסונתז באמצעות בינה מלאכותית מסייעת בשיפור איכות וגיוון התוכן המתחזה והשקרי - בעיקר כשהיוזמים מפיצי התוכן הם גורמים זרים לא דוברי השפה המעוניינים ליצור תוכן מקומי מהימן הקשה לזיהוי.

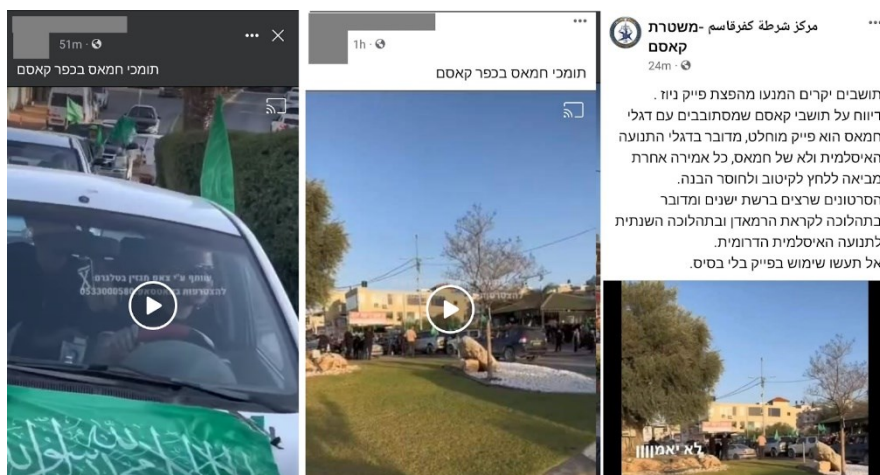
חוקרי איגוד האינטרנט הישראלי ביצעו מיפוי של סוגי התוכן והפרסומים הכוזבים בחודשיים הראשונים למלחמה וקטלגו אותם על פי פרמטרים מוכרים בספרות המקצועית של שבעה סוגי דיס-ומיסאינפורמציה הנפוצים במצבי חירום גם בכלי התקשורת. אחד מתוך שבעת הסוגים הללו הוא שימוש בתוכנות עריכה ותוכנות בינה מלאכותית כדי ליצור תוכן כוזב מלאכותי שלא מתבסס על תוכן מקורי הניתן לאיתור ברשת:

7 סוגים של מידע כוזב



www.fakenews.org.il | CC by 4.0 רישיון 477, FIRST DRAFT ב Claire Wardle (2020)

על אף החשיבות של זיהוי והבנת האיום של יצירת תוכן כוזב ושקרי באמצעות בינה מלאכותית, חשוב לציין כי על פי המידע שנאסף באיגוד האינטרנט הישראלי בחודשיים הראשונים למלחמה, מרבית התוכן והמידע הכוזב שהופץ בחודשיים הללו לא היה סינתטי, גנרטיבי או מעובד באמצעות בינה מלאכותית. מרבית התוכן הכוזב שאותר כלל תכנים, פרסומים, תמונות או סרטונים ממועד מוקדם יותר או מקומות אחרים שלא קשורים במלחמה בעזה, שמופצים בהקשר או בתיאור כוזב על מנת לקדם נרטיבים שקריים הנוגעים למלחמה. למשל – תמונות של הרס והרוגים במלחמת האזרחים בסוריה המוצגים כאילו הם בעזה, סרטונים מהפגנות בחברה הערבית בשנים קודמות המוצגות כאילו הן מתרחשות כעת בתגובה לתקיפות חמאס או תמונות של פועלים ונותני שירות ערבים המצלמים את עבודתם במרחב הציבורי המוצגים כאילו הם נועדו לשרת מטרות טרור או חתרנות נגד ישראל. מרבית הפרסומים הללו לא מתבססים על טכנולוגיה מתקדמת אלא דווקא על חוסר הבנה ועירנות של צרכני מדיה בתקשורת. העלייה הגדולה בהיקף המידע והתוכן הכוזב כללה ברובה בעיקר תכנים כאלה, אך בתוכם היה ניתן להבחין גם בתכנים שנוצרו באמצעות תוכנות בינה מלאכותית – בעיקר תוכנות גנרטיביות ליצירת תמונות שנראות מציאותיות.



עיקר התמונות הגנרטיביות שזוהו ונאספו על ידי חוקרי איגוד האינטרנט הישראלי נוצרו והופצו ברשתות חברתיות בעולם על רקע המלחמה במטרה להציג את ישראל כפושעת מלחמה הפוגעת באופן עיוור ואכזרי באזרחים פלסטינים ובעיקר בילדים ובזקנים, כדי להקצין את הרושם של היקף ואופי הנזק והפגיעה באזרחים בעזה מתוך רצון לזעזע ולגרום לדעת הקהל העולמית לתמוך בחמאס ולבקר את ישראל. תמונות גנרטיביות רבות שנוצרו באמצעות AI הציגו ילדים פלסטינים מתים ומוכתמים בדם בהריסות בעזה או תמונות דרמטיות קשות של הרס ופגיעה באזרחים. תמונות אחרות נועדו להציג את כיבוש ויישוב רצועת עזה בידי ישראל.



זיהוי תכנים מג'נרטיבים (מלשון generated) על ידי בינה מלאכותית הוא עדיין אתגר טכנולוגי שאין לו פתרון ברור וחד משמעי, בין השאר בגלל שטכנולוגיות היצירה מתפתחות כל הזמן. בשוק קיימים כיום מספר כלים שכביכול מזהים תמונות מג'נרטיביות, אך הם עשויים לטעות ולא תמיד נותנים חיזוי מדויק. בתמונות או וידאו אפשר לחפש ולמצוא תקלות (המכונות גליציים) או חוסר הגיון בדמויות (למשל מספר לא נכון של אצבעות ביד או שיבוש פרצופים) - אבל זיהוי טעויות באודיו או בטקסטים הוא קשה יותר לפענוח. הרבה כלי בינה מלאכותית מוסיפים סימני מים - ועדיין יהיו כלים שיאפשרו לייצר בלי או להסיר את סימן המים מתוצרי כלים קיימים. מצד שני, במהלך המלחמה הנוכחית כבר היו מספר פעמים בהם כלי זיהוי כאלה תייגו תמונות אמיתיות כמסוננות (מה שמכונה False Positive) - ותרמו בכך דווקא להפצה של דיסאינפורמציה ומידע שגוי במקום להאבק בו.

שימוש נוסף שזוהה בכלי בינה מלאכותית במסגרת המלחמה הוא במסגרת יצירת פרופילים, עמודים, שחקנים ובוטים סינתטיים המתחזים לגורמים אנושיים, שנועדו להפיץ תכנים שקריים שונים וקשים יותר לזיהוי. במסגרת הנסיון לחדור לשיח הציבורי הישראלי ולזרוע בתוכו הסתה, פילוג ובהלה נעשה שימוש ברשתות זרות של פעילות לא אותנטית מתואמת (CIB), שנוצרה בחלקה באמצעות כלי בינה מלאכותית. רשתות אלו, לעתים ביוזמת מדינות זרות או ארגוני טרור, נועדו לייצר גלים של הפצה ויראלית של תכנים המיועדים לפגוע בחוסן החברתי בישראל בתקופת הלחימה. במובן זה, השימוש בכלי בינה מלאכותית לא נועד לשפר את איכות התוכן הכוזב והשקרי, אלא ליעל ולהאיץ את היקף וקצב ההפצה של התכנים הללו, ועל ידי כך להציף את הרשתות בתכנים רבים ככל האפשר ולהקשות על זיהוי שלהם עקב העומס על המערכות. עד כה

לא זיהינו במסגרת המלחמה הנוכחית רמה לא מוכרת של תוכן הקשה לזיהוי או הפרכה מבחינה איכותנית (על אף שבהחלט ייתכן שהוא קיים והופץ ברשתות אך עדיין לא זוהה) אלא יכולות מוגברות ומסוכנות שמאפשרות יצירה אוטומטית של היקפים וסוגים רבים יותר של תוכן שקרי באמצעות כמויות גדולות יותר של גורמים ושחקנים שגם הם קשים יותר לזיהוי והפצה בהיקף ומהירות גדולה יותר.

כלומר גם אם סוג התוכן השקרי לא מתוחכם או מורכב יותר, ההיקף והמהירות הגבוהים של ההפצה מהווים איום משמעותי ומקשים על ההתמודדות עם התופעה ככלל. נכון לכתובת שורות אלה ניתן לומר כי מידע כוזב שמיוצר על ידי בינה מלאכותית מחריף בעיות מוכרות אך לאו דווקא מייצר בעיה חדשה או איום בסדר גודל שלא הכרנו קודם. אולם העלות הנמוכה, ההיקף ומהירות הייצור וההפצה של התכנים הכוזבים מציבים כעת אתגרים חדשים מבחינת כמות, איכות, התאמה אישית של תוכן ומסרים והפצת מידע מסוננת שגוי או מטעה ללא כוונה והיכולת של רשויות ומערכות מדינתיות לזהות ולהתמודד עם היקפים גבוהים ולא מוכרים של תכנים כאלה. אחת הדרכים להתמודד עם התופעה יכולה להיות בחינה מרובת פלטפורמות של מנגנוני ורשתות ההפצה, בנוסף לבדיקת אמיתות התוכן הבודד עצמו. הסתכלות זו עשויה לאפשר זיהוי מקור ההפצה והתמודדות עם מנגנונים לא אותנטיים או זדוניים.

4. אתגרים בזיהוי, התמודדות והסרה של תוכן כוזב ברשתות

כל הפלטפורמות והרשתות החברתיות המרכזיות משתמשות כיום בשילוב של טכנולוגיה וכוח אדם אנושי כדי לאכוף באופן פעיל את הנחיות הקהל שלה ולהסיר תכנים או חשבונות שמפירים אותם; ומעודדות את המשתמשים שלהן את חברי הקהל שלה להשתמש בכלים שהיא מספקת כדי לדווח על תוכן וחשבונות שמפירים את כללי הקהילה, ובפרט התנהגות כמו שימוש במספר חשבונות כדי להשפיע ולהטות את דעת הקהל, תוך הטעיית אנשים, הקהל, או מערכות הפלטפורמה לגבי זהות, מיקום, קשרים, פופולריות, או מטרה של החשבון.

מאז 2019, ראינו עלייה מהירה במספר הרשתות שהשתמשו בתמונות פרופיל שנוצרו באמצעות טכניקות של בינה מלאכותית כמו רשתות יריבות יצירתיות (GAN). טכנולוגיה זו זמינה בקלות באינטרנט, ומאפשרת לכל אחד - כולל גורמי איומים - ליצור תמונה ייחודית. כיום, הפיתוח המהיר של בינה מלאכותית כמו ChatGPT מעורר דאגה לגבי הפצת מידע כוזב המאיימת על הדמוקרטיה ברחבי העולם, כהגברת האיום המוכר של הפצת תעמולה דרך אלגוריתמים פגיעים ומיקרו-טרגטינג בשנים האחרונות. כפי שמסבירים החוקרים כהנא ושוורץ-אלטשולר, טכנולוגיות תקשורת ופלטפורמות כמו TikTok, פייסבוק וטוויטר בנו מערכות אינטיליגנציה מלאכותית רגישות באופן מרשים והשאירו אותן ללא הגנה.¹ אין צורך להסתמך באופן בלעדי על תוכן שנוצר על ידי אינטיליגנציה מלאכותית כדי לבצע מסעות תעמולה יעילים. הנקודה החשובה ביותר אינה מסתרת בתוכן שנוצר על ידי כלים של אינטיליגנציה מלאכותית כמו ChatGPT אלא באופן שבו אנשים מקבלים, מעבדים ומבינים את המידע שמאופשר על ידי מערכות האינטיליגנציה המלאכותית של פלטפורמות הטכנולוגיה.

בניגוד לעבר, במקביל לשימוש המדינתי-בטחוני בכלי בינה מלאכותית, כעת גם גורמים זדוניים – פרטיים או מדינתיים - יכולים גם להשתמש בהם למטרות שכנוע ומניפולציה. זאת, מכיוון שכלים כגון ChatGPT של

¹ אמיר כהנא ותהילה שוורץ-אלטשולר אדם, מכונה ומדינה (המכון הישראלי לדמוקרטיה, 2023)

חברת OpenAI או LLaMa2 של חברת Meta נמצאו יעילים בקידום נרטיבים בצורה משכנעת ובהשפעה על אמונות של בני אדם ועל עיצובן באמצעות דיאלוג איתם או טכניקות אחרות, כמו השפעה על אלגוריתם של הפצת מידע ברשתות חברתיות. למעשה, גם חברות הטכנולוגיה שמפתחות כלי Generative AI לשימוש הציבור מגדירות אותן בעצמן כאיום פוטנציאלי מדינות דמוקרטיות, כתוצאה מיכולתן להפיק תעמולה שמוחקת את האבחנה בין עובדה לדמיון, וקוראות לקדם רגולציה שתחסן ותסדיר אותן.

מול העלייה בהיקף, כמות וסוגי התוכן הכוזב והמסית המופץ ברשתות, קיימת בישראל היערכות חסר והיעדר סמכויות מדינתיות לזיהוי וטיפול בתופעה, גם ברמה הציבורית וגם ברמת הפלטפורמות. על אף שרשויות וגופי מדינה ובטחון שונים עוסקים באיומים במרחב המקוון אין הגדרה והערכות ברורה של גוף שאמור לעסוק בהגנה על הציבור בפני איומי דיסאינפורמציה והפצת תוכן כוזב או מסית שנועד לפגוע בחברה הישראלית באמצעות הפצה ברשתות החברתיות בעברית, גם כזה הנוצר באמצעות כלי בינה מלאכותית. משימה זו דורשת הכרות מעמיקה עם סוגי האיום אך גם עם הפלטפורמות ועם המדיניות שלהן, שלעתים רבות לא מאפשרות הסרה מהירה של תכנים שלא מוגדרים כתוכן מסוכן או פוגעני אלא נתפסים כתוכן שנוי במחלוקת שיש עליו לעתים הגנה של חופש הביטוי, אלא אם הוכח בצורה ברורה כי הוא מופץ על ידי גורמים עוינים (למשל טרור או מדינות אויב) המשתמשים בו כדי לייצר נזקים ולפגוע באזרחים.

5. האם כלי הרגולציה הקיימים בישראל מספיקים לצורך המאבק בהפצת פייק ניוז?

אל מול החדירה העצומה של הפלטפורמות הגלובליות בישראל, תמונת המצב של הפיקוח והרגולציה על הסיכונים החברתיים של פלטפורמות תוכן ורשתות חברתיות גלובליות אינה חיובית. כיום, במדינת ישראל קיימת שורה ארוכה של גורמים העוסקים באיתור ומניעה של הסיכונים השונים מהרשתות החברתיות עבור הפרט והחברה: שירות הבטחון הכללי, מערך הסייבר הלאומי, משטרת ישראל, מחלקת הסייבר בפרקליטות המדינה ואף רגולטורים ורשויות אכיפה מינהליות כמו הרשות להגנת הפרטיות והרשות להגנת הצרכן. דוגמה עדכנית לבעייתיות העולה מריבוי הגופים המדינתיים המופקדים על איומי המרחב המקוון, היא שאלת גבולות הגזרה בין מערך הסייבר לשירות הבטחון הכללי בכל הנוגע להתמודדות עם התערבות של גורמים זרים בבחירות לכנסת. בעוד ששני הגורמים מכירים באיום היחוס של השפעה זרה על הבחירות באמצעות הרשתות החברתיות כממשי וקל יחסית למימוש, מערך הסייבר הלאומי רואה בשב"כ כאחראי הבלעדי והטבעי לסיכולו. כפי שהראו הימין ואח', הטלת האחריות על השב"כ כאחראי לסיכול חתרנות נראית טבעית, אך אינה מספקת מענה בטחוני מלא מכיוון שהשב"כ מתמחה בסיכול פעילות זדונית באזורי הפעולה החשאיים וכמעט ואינו פועל באזורים הגלויים.

היעדר הסדרה חקיקתית ברמה המדינתית היא בעייתית במיוחד בזירה הישראלית, מכיוון שמערכות הניטור והאכיפה שמפעילות הפלטפורמות נגד תוכן ופעילות אסורים מופנים בעיקרם לתוכן בשפה האנגלית ולזירה האמריקנית. התמקדות זו בשפה האנגלית מובילים להזנחה של האכיפה בכל הנוגע לתוכן בשפות אחרות, ובפרט בעברית. לראיה, בבדיקה שנערכה, אשר ממציאה הועברו ל-Meta במכתב מטעם איגוד האינטרנט הישראלי, נמצאו כשלי אכיפה רבים ומובהקים לגבי תוכן בשפה העברית, שכללו התעלמות למעשה מהפרות

מובהקות, חוזרות ונשנות של כללי הקהילה.² לדוגמה, תועדו מאות חשבונות בשפה העברית המפירים באופן מובהק את הכללים בנושא אמינות ואותנטיות, אשר לא מוסרים על ידי הרשת החברתית גם לאחר דיווח אקטיבי של משתמשים. פייסבוק אף אינה מפרסמת נתונים לגבי היקף האכיפה שביצעה ביחס למשתמשי הפלטפורמה שמקורם בישראל או בשפה העברית, למרות שהיא מפרסמת נתונים אלו לגבי מדינות כמו הודו וגרמניה. מכאן ניכר כי האכיפה של Meta אינה זהה ושוויונית בין משתמשיה, ובהקשר הישראלי אף לוקה בחסר באופן משמעותי.³

נוכח מרכזיותן של הרשתות החברתיות בחיי החברה והכלכלה של ישראל והתבססותן כערך מרכזי של שיח ציבורי ומידע מקור מידע חליפי עבור האזרחים, ישראל חייבת להחיל אסטרטגיה לאומית מקיפה כדי למנוע החדרה נוספת שמכוונת לשבש את התקשורת החברתית והיציבות הפוליטית. בעוד שישראל מתמודדת היטב עם זירת האינטרנט בהגנה מפני מתקפות סייבר, טרם גובש מענה תכנוני או מבצעי אל מול המקרים של פעולות השפעה זדוניות מצד גורמים מדינתיים שמכוונות לאזרחי ישראל במדיה החברתית. בזירת החקיקה והרגולציה, ישראל נדרשת ויכולה להסדיר ולחדד את מערכת היחסים בין הפלטפורמות הדיגיטליות לבין ציבור המשתמשים ורשויות הבטחון והאכיפה המקומיות – תוך תאימות לכללים ומסגרות האכיפה לאומיות שנקבעו בשנה האחרונה בזירה הגלובלית.

האתגרים משמעותיים, אך הם ניתנים להתגברות באמצעות כלים טכנולוגיים ומשטר אחריות ורגולציה על פלטפורמות הדיגיטליות בישראל. ראוי לעגן בישראל הסדר תואם להסדרת ממשקי הפלטפורמות עם רשויות האכיפה והרגולציה המדינתיות, לרבות כפיפות הפלטפורמות לדין המקומי ויצירת מנגנון להחלת צווים שיפוטיים על מפעילות הפלטפורמות; את האבחנה הרגולטורית בין הנורמות ומשטרי האסדרה המתאימים לסוגים שונים של מתווכים דיגיטליים. זאת, על בסיס ההבנה שכיום אלגוריתמים הם שמתווכים בין המציאות ובין המשתמשים ברשתות החברתיות – וכיוון שהם אלו אשר שולטים ב"פיד" הידיעות של המשתמשים, הם מחזיקים בכוח משמעותי לעיצוב עמדות ורחשי לב, ולעיתים עשויות אלו להיות עמדות קיצוניות שמובילות להסתה, אלימות וקיטוב.

את הכללים שיוחלו בהסדר מסוג זה יש להטיל מתוך עמידה בחמישה עקרונות לעיצוב מדיניות ציבורית ישראלית כלפי תוכן מזיק ברשתות חברתיות: (1) חיוב הפלטפורמות באכיפה אפקטיבית ושוויונית של כללי התוכן והשימוש שהן מפרסמות; (2) תיקון פערי מידע ומיקוח בין הפלטפורמות למשתמשים וחיוב הפלטפורמות הגלובליות לפרסם דוחות שנתיים הנוגעים לאפקטיביות פעולות ניהול התוכן שלהם בישראל, ולשקף למשתמשים מידע על אודות פרסומות או תוכן מקודם; (3) זכויות פרוצדורליות והליך הוגן לחייב את הפלטפורמות לעמוד בפרקטיקות של הליך הוגן ושקיפות כאשר הן אוכפות את כלליהן נגד משתמשים; (4) יש לייצר ערוצי דיווח ומסירת מידע לפלטפורמות מצד רשויות ונפגעים; (5) יש להביא לכך שפלטפורמות התוכן

² פניית איגוד האינטרנט לפייסבוק ישראל ו-Meta: כשלי אכיפה של כללי הקהילה ביחס לתוכן ומשתמשים בעברית (13.9.2022). [\(קישור\)](#).

³ לדיון נרחב בכללי הקהילה של הפלטפורמות ומנגנוני האכיפה שלהם נגד תוכן בלתי-חוקי ונגד תוכן חוקי-אך-מזיק, ראו: אסף וינר, תהילה שורץ-אלטשולר ואייל זילברמן מתווה לאסדרת רשתות חברתיות בישראל (2023) [\(קישור\)](#).

יפעלו בשקיפות בכל הנוגע להתמודדות עם הפצה של תוכן מזיק ברמה החברתית-לאומית (כגון דיסאינפורמציה או הסתה לטרור).⁴

נשמח לעמוד לרשותכם בכל הבהרה נוספת שתידרש. מידע נוסף לעדכונים על התפתחות יוזמות חקיקה ורגולציה להתמודדות עם דיסאינפורמציה והיבטים נוספים של רגולציה על פלטפורמות ושירותים מקוונים, קיים בעמוד ייעודי באתר איגוד האינטרנט: isoc.org.il/regulating-digital-services

⁴ להרחבה על חשיבותם של כל אחד מעקרונות אלו וכיצד ניתן ליישם אותם בדין הישראלי, ראו: אסף וינר, תהילה שוורץ-אלטשולר ואייל זילברמן מתווה לאסדרת רשתות חברתיות בישראל (2023) (קישור).