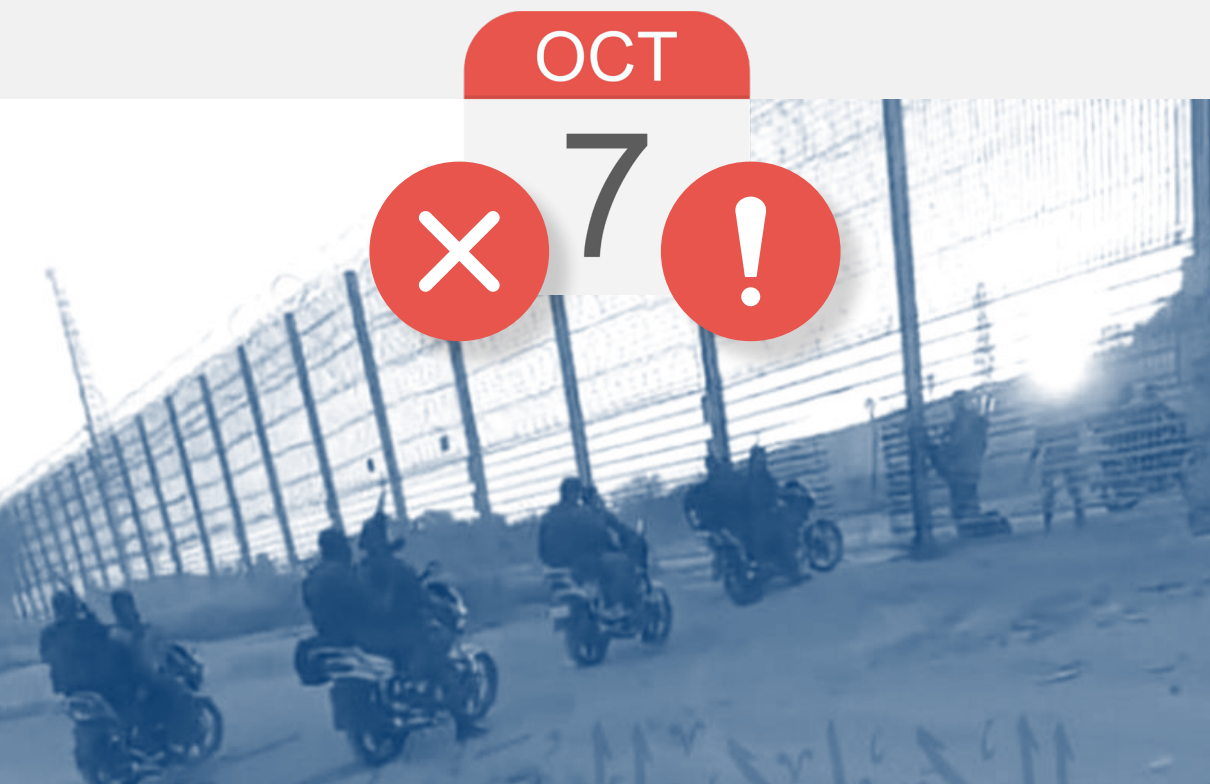


# לא עומדות בפרץ

התמודדות הרשתות החברתיות עם תוכן פוגעני ותומך טרור  
במתקפת ה-7 באוקטובר ומלחמת חרבות ברזל



ניצן יסעור ועידן רינג

– מאי 2024 –

איגוד  
האינטרנט  
הישראלי  
ISOC-IL



**איגוד האינטרנט הישראלי** (ע"ר) הוא ארגון ללא כוונת רווח אשר פועל מאז תחילת שנות התשעים להטמעת השימוש באינטרנט לטובת הציבור בישראל ומפעיל תשתיות אינטרנט חיוניות בישראל: מרשמי שמות המתחם (Domain Names) המדינתיים ".il" ו-"ישראל" ומחלף האינטרנט הפנים מדינתי (IIX). הידע והניסיון המקצועי של איגוד האינטרנט הישראלי במחקר, בפיתוח ובהפעלה של טכנולוגיות אינטרנט עבור הציבור הישראלי משמש בסיס לפעילות מחקר, מדיניות והעצמת הקהילה שמבצע האיגוד. בכלל זה, איגוד האינטרנט הישראלי מפעיל את קו הסיוע לאינטרנט בטוח ואת מיזם Block.org.il, המספק לציבור הרחב סיוע, ידע וכלים לשיפור המוגנות והבטיחות ברשת; מיזמים ופעילויות שונות לצמצום פערם דיגיטליים וקידום מיומנות ואוריינות רשת בחברה הישראלית; את מיזם data.isoc.org.il האוסף ומנגיש נתונים כמותיים-סטטיסטיים על האינטרנט הישראלי ומשתמשו; ומפרסם מחקרי מדיניות מקצועיים המיועדים לידע ולטייב את זירת האינטרנט הישראלית והגלובלית בצמתים מגוונים של משפט וטכנולוגיה

## לא עומדות בכרף

התמודדות הרשתות החברתיות עם תוכן פוגעני ותומך טרור במתקפת ה-7 באוקטובר ומלחמת חרבות ברזל

כתיבה וניתוח נתונים: ניצן יסעור – חוקר רשתות חברתיות ודיסאינפורמציה  
עריכה וניהול מקצועי: עידן רינג – סמנכ"ל חברה וקהילה, איגוד האינטרנט הישראלי  
איסוף נתונים: יונתן בן חורין – מנהל קו הסיוע לאינטרנט בטוח של איגוד האינטרנט הישראלי  
ניתוח נתונים: רתם חשדי  
ייעוץ ועריכה: ירדן אמיר  
ייעוץ ועריכה באנגלית: נינה מישרקי  
ייעוץ אקדמי: ד"ר אסף וינר  
עריכה לשונית: חבצלת שפירא  
עיצוב: סטודיו לי ותמר

מחקר זה זמין גם בשפה האנגלית באתר איגוד האינטרנט הישראלי

איגוד האינטרנט הישראלי – [www.isoc.org.il](http://www.isoc.org.il)

קו הסיוע לאינטרנט בטוח – [www.safe.org.il](http://www.safe.org.il)

לפרטים נוספים ויצירת קשר – [disinfo@isoc.org.il](mailto:disinfo@isoc.org.il)

צילום כריכה: שימוש על פי סעיף 27 לחוק זכויות יוצרים.  
כל הזכויות שמורות לאיגוד האינטרנט הישראלי (ע"ר), 2024.

# תוכן עניינים

5	תקציר המחקר וההמלצות
7	א. רקע ומטרות המחקר
16	ב. מתודולוגיה
19	ג. ממצאים
19	היקף וסוג הדיווחים שבהם טיפל קו הסיוע לאינטרנט בטוח
21	בחינת זמן התגובה לדיווחי קו הסיוע
24	אפקטיביות המענה של הפלטפורמות לדיווחים
28	מאפיינים וכשלים ייחודיים בטיפול הפלטפורמות במידע כוזב ודיסאינפורמציה
29	איכות המענה והמידע שהפלטפורמות מספקות לדיווחי Trusted Flaggers
31	סימון וחיווי למשתמשים על פעולות אכיפת התוכן שמבצעות הפלטפורמות
35	ד. ניתוח המשמעויות עבור הזירה הישראלית
41	ה. סיכום והמלצות לתיקון המצב
46	ביבליוגרפיה

---

## תקציר המחקר וההמלצות

המתקפה הקטלנית של חמאס על יישובים ובסיסים בעוטף עזה ב-7 באוקטובר 2023, והמלחמה שפרצה בעקבותיה ברצועת עזה, חשפו בצורה חסרת תקדים את השימוש ההרסני שעושים ארגוני טרור וגורמים עוינים בפלטפורמות מדיה חברתית וטכנולוגיות תקשורת מסחריות ברשת, כאשר הם הופכים אותן לנשק המיועד למניפולציה המונית, זריעת פחד, הפצת טרור והרחבת מעגלי הפגיעה בעולם האמיתי.

פלטפורמות המדיה החברתית, שמאפשרות ומעצבות את השיח וסביבת המידע הציבוריים של ימינו, הוצפו במסגרת המלחמה בשטף של מידע קשה, רעיל ופוגעני, והתקשו להתמודד עם ההיקף והאופי של התוכן הפוגעני שאליו נחשפו המשתמשים. בתוך רגעים מתחילת מתקפת החמאס הפכו הפלטפורמות לסביבה לא בטוחה עבור משתמשיהן בשל פעילות פוגענית מצד גורמים זדוניים, אשר ניסו להשפיע על תפיסת המציאות ועל המידע שבידי המשתמשים. אף שפלטפורמות המדיה החברתית טענו כי הן עושות מאמצים לנטר את הרשתות ולנקות אותן מהתכנים הללו, הצפת התוכן הפוגעני נמשכה עוד שבועות וחודשים רבים, והעמיקה את החרדה והטראומה המשנית שהתעוררו אצל אזרחים רבים בישראל ובעולם על רקע המלחמה.

בהיעדר רגולציה ישראלית או שקיפות מצד הפלטפורמות לגבי אמצעי הבטיחות שהן מפעילות בישראל, אין למוסדות המדינה ולמשתמשים תמונה עובדתית בנוגע להיקף התוכן הפוגעני, ולא ידוע כיצד הוא מטופל ומה היקף הטיפול בו מצד הפלטפורמות. על רקע זה יזם איגוד האינטרנט הישראלי מחקר שמטרתו לבחון באופן ביקורתי, אמפירי והשוואתי את מאפייני הטיפול של הפלטפורמות השונות בתוכן הפוגעני ובמידע הכוזב שהופץ באמצעותן. המחקר בדק את ההליך והאיכות של הטיפול והתגובה של פלטפורמות המדיה החברתית המובילות ובודק כיצד הן הגיבו לפניית ודיווחים אשר נגעו לפרסום תוכן המנוגד לקווי המדיניות שלהן שהועברו אליהן במהלך החודשיים הראשונים למלחמת חרבות ברזל בידי קו הסיוע לאינטרנט בטוח של איגוד האינטרנט הישראלי, המוכר כגורם מדווח מהימן (Trusted Flagger).

ממצאי המחקר מעלים כי התפקוד והמענה הכללי של הפלטפורמות לאורך תקופה זו לקו בחסר: הן לא נקטו פעולות אכיפה מספקות עבור רבים מהפניות והדיווחים, זמני התגובה שלהן למרבית הדיווחים היו ארוכים ובלתי סבירים ביחס למצב החירום וחומרת התוכן, וכמעט ולא סיפקו מענה הולם במהלך סופי השבוע. עוד עולה מהממצאים כי ישנם הבדלים ופערים ניכרים באופי הטיפול וזמני התגובה של הפלטפורמות השונות עבור תכנים מסוגים שונים: הטיפול בתכנים הדורשים מעורבות ובחינה אנושית היה לקוי במובהק, כאשר רמת הטיפול בדיסאינפורמציה ומידע כוזב הייתה נמוכה במיוחד.

ניתוח הממצאים מצביע על כך שהפלטפורמות ככלל לא עשו מספיק כדי לתת מענה שקוף, מפורט, הולם ומקיף לפניות ודיווחים בנוגע לתוכן פוגעני ולפעילות המפרה את המדיניות וכללי הקהילה שלהן במהלך המלחמה. רובן סיפקו מענה חלקי, איטי ולא אחיד גם לדיווחים דחופים ומאומתים שהעביר אליהן קו הסיוע לאינטרנט בטוח, אף על פי שהן מכירות בו רשמית כגורם מדווח מהימן, מנוסה ומועדף. אלה נתונים מדאיגים במיוחד בבואנו לשער כיצד טופלו דיווחים שהועברו בידי הציבור הרחב דרך מנגנוני הדיווח המובנים של הפלטפורמות. על אף כל אלה, ניכר כי בשבוע הראשון למלחמה הפלטפורמות פעלו במהירות וטיפלו כראוי בתכנים פוגעניים, מה שמראה כי כאשר הן משקיעות משאבים, כוח אדם ומאמצים נאותים בהסרת תכנים ויצירת סביבה בטוחה יותר – הן מסוגלות לעמוד במשימה.

על בסיס מסקנות המחקר יש לבצע מספר צעדים נרחבים כדי לשמור על מוגנות המשתמשים בשגרה ובחירום בעתיד. האתגרים שהתגלו במחקר זה דורשים פעולות מצד כלל השחקנים בשדה – פלטפורמות המדיה החברתית, המדינה והמחוקקים, וכן גופי מחקר, תקשורת וחברה אזרחית. ביצוע השינויים המתבקשים במשאבים המושקעים ובמידת השקיפות בנוגע אליהם מצד הפלטפורמות, כמו גם בפיקוח ובבקרה של המדינה על פעילות הפלטפורמות, יוכל להוביל לשיפור מהותי בביטחונות של משתמשי הפלטפורמות בישראל. המלצות אלה נכונות לימי שגרה, ועל אחת כמה וכמה לתקופות חירום, כמו מלחמות וקונפליקטים צבאיים וביטחוניים, המאפיינים לצערנו את החיים בישראל ובאזור, וגם במקומות רבים נוספים ברחבי העולם.

---

# א. רקע ומטרות המחקר

המתקפות הקטלניות של חמאס על יישובים, בסיסים ואזרחים בישראל בגבול רצועת עזה ב-7 באוקטובר, והמלחמה שפרצה בעזה בעקבותיהן, חשפו בצורה חסרת תקדים את השימוש ההרסני שעושים ארגוני טרור ומדינות עוינות בפלטפורמות מדיה חברתית וטכנולוגיות תקשורת מסחריות, כאשר הם הופכים אותן לנשק המיועד למניפולציה המונית, זריעת פחד, הפצת טרור והרחבת מעגלי הפגיעה של פעולות טרור בעולם האמיתי.<sup>1</sup> חמאס שידר ברשתות החברתיות בזמן אמת את מעשיו, הפיץ תוכן גרפי קשה ופרסם פרטים מניפולטיביים על גורל החטופים. הוא ניצל באופן מודע ומכוון את הפלטפורמות כדי לחשוף את הקהל הישראלי והבינלאומי למעשי הרצח והחטיפה, וגם לשם הכחשת הזוועות שביצע והפצת מידע שקרי על פעולות ישראל בעזה.

הזוועות שביצע חמאס ב-7 באוקטובר ובימים שאחריו כללו לא רק טרור ולוחמה פיזית בשטח, אלא גם מתקפה דיגיטלית חסרת תקדים של תוכן גרפי קשה (לרבות תיעוד מעשי רצח), דיסאינפורמציה, מניפולציות ותוכן זדוני שהופץ על גבי ובאמצעות פלטפורמות מדיה חברתית, כולל פייסבוק, אינסטגרם, X (לשעבר טוויטר), טיקטוק, יוטיוב, טלגרם, וואטסאפ ועוד. **אל מול האירועים הקשים שהתחוללו בשטח וברשתות במהלך 7 באוקטובר ובשבועות שאחריו, התברר כי פלטפורמות המדיה החברתית מתקשות להתמודד עם ההיקף והאופי של התוכן הפוגעני ותומך הטרור שהופץ באמצעותן.** פלטפורמות התוכן הדיגיטליות וטכנולוגיות התקשורת, ובראשן הרשתות החברתיות המסחריות, שהפכו בשנים האחרונות לכלי מרכזי שמעצב את המרקם החברתי והכלכלי בעולם כולו, ובמיוחד בישראל,<sup>2</sup> התגלו כקרקע פורייה במיוחד גם לפגיעה באזרחים תמימים.<sup>3</sup> בעקבות מתקפת

1. Ring, E (January 22, 2024). Op-Ed for Haaretz | How Telegram, Twitter and TikTok have become lethal tools of Hamas psychological warfare - Israel Internet Association - (ISOC-IL). [קישור](#)

2. שימוש ברשתות חברתיות ושירותים מקוונים בישראל: נתוני 2024 עם פילוחים דמוגרפיים – איגוד האינטרנט הישראלי. [איגוד האינטרנט הישראלי. קישור](#)

3. Greenwood, S. (2023, June 1). Internet, smartphone and social media use around the world | Pew Research Center. Pew Research Center's Global Attitudes Project. [קישור](#)

הטרור של חמאס על ישראל ופריצת מלחמת חרבות ברזל, הרשתות החברתיות נאלצו להתמודד עם כמות ומגוון עצומים וחסרי תקדים של תוכן פוגעני שמנוגד לכללי הקהילה שלהן, פוגע בביטחון ובחוויה של משתמשיהן ולעיתים גם עלול לסכן חיי אדם. עקב כך הן נאלצו לשנות ולהגמיש את כללי המדיניות הנוגעת להסרה וסימון של תוכן פוגעני, וגם את האמצעים והמשאבים המוקדשים לאכיפה, לזיהוי תוכן מסוג זה ולטיפול בו, במיוחד בחודשים הראשונים ללחימה.

אף שפלטפורמות המדיה החברתית טוענות בשנים האחרונות כי הן משתדלות לנטר ולנקות את הרשתות מתכנים פוגעניים ולא חוקיים, ומצהירות על כללי מדיניות ואמצעים לניטור וטיפול בתוכן אסור לפרסום, הן עדיין מוצפות בתכנים כאלה, בעיקר במרחבים ובמדינות עם שפות בעלות מספר משתמשים מצומצם יחסית או בעלות כתיב ייחודי, כמו בישראל. הרשתות אינן משקיעות מספיק משאבים במערכות סינון וניטור תוכן אוטומטיות בשפות אלו, ולא מצליחות לטפל בתכנים בהיקף ובקצב מספיקים.<sup>4</sup> **כך, בחודשי המלחמה הראשונים התברר כי הרשתות החברתיות נעשו לא בטוחות מספיק עבור משתמשים בישראל**, בשל פעילות פוגענית מצד גורמים זדוניים, ובשל ניצול המבנה האלגוריתמי שלהן על ידי גופים פוליטיים, מסחריים או עוינים, המנסים להשפיע על תמונת המידע שבידי משתמשים. הפגיעה בביטחון המשתמשים, הנוכחת גם בזמני שגרה, מורגשת ביתר שאת בעיתות חירום. למעשה, עוד לפני אירועי אוקטובר והמלחמה הנוכחית, היקף השימוש של הציבור הישראלי במנגנוני הדיווח של הרשתות החברתיות כדי לדווח על חשבוניות ותכנים פוגעניים היה מצומצם ביותר. לפי נתוני איגוד האינטרנט הישראלי מינואר 2022, רובו המוחלט של הציבור בישראל משתמש במנגנוני הדיווח לעיתים רחוקות או כלל לא, כאשר הסיבות העיקריות לכך הן שאינו מאמין כי הרשתות מתייחסות לדיווחים ברצינות, וכן חוסר ידע לגבי הפעלתם.<sup>5</sup>

בניגוד לחקיקה החדשה שנכנסה לתוקף במהלך 2023 במדינות האיחוד האירופי (DSA – Digital Services Act), וחקיקה דומה לה באוסטרליה ובאנגליה, בישראל אין רגולציה שמחייבת את הפלטפורמות הדיגיטליות והרשתות החברתיות לתת מענה לציבור ולרשויות

4. אסף וינר, תהילה שוורץ-אלטשולר ואייל זילברמן – מתווה לאסדרת רשתות חברתיות בישראל (המכון הישראלי לדמוקרטיה ואיגוד האינטרנט הישראלי, 2023) **קישור**

5. מאפייני השיח האלים והתמודדות מולו במרחב האינטרנט והרשתות החברתיות בישראל (דצמבר 2022) – איגוד האינטרנט הישראלי. איגוד האינטרנט הישראלי. **קישור**

או לחשוף את האמצעים שהן מפעילות כדי להגן על המשתמשים. בפועל, הפלטפורמות השונות נהנות מחסינות ואינן נדרשות לקבל אחריות לנזקים או לפגיעות שנגרמו למשתמשים עקב תוכן שהועלה בהן והופץ באמצעותן בידי משתמשים עצמאיים או גורמים בעלי אינטרס. עקב חסינות זו אין לפלטפורמות תמריץ לפעול ליצירת מרחב רשת נקי ובטוח למשתמשים בישראל.<sup>6</sup>

**בהיעדר כל רגולציה או חובת שקיפות, אין למדינה ולמשתמשים תמונה אמפירית ברורה לגבי היקף התוכן הפוגעני ותומך הטרור שמופץ בפלטפורמות השונות בישראל ונחשף לעיני המשתמשים, ולא ידוע כיצד הוא מטופל ומה היקף הטיפול בו מצד הפלטפורמות.** במציאות זו, מאז שנת 2013 פועל קו הסיוע של איגוד האינטרנט הישראלי, המספק סיוע, הכוונה, מידע וכלים למשתמשי האינטרנט והרשתות החברתיות בישראל, במטרה לתמוך בהתמודדות עם מגוון רחב של איומים מקוונים ופגיעות ברשת. קו הסיוע מטפל במאות פניות חודשיות של משתמשים לעזרה ותמיכה, ומוכר כגורם מדווח רשמי (Trusted Flagger) בידי פלטפורמות ושירותי רשת גלובליים מובילים כגון פייסבוק, אינסטגרם, טיקטוק, יוטיוב, סנאפצ'ט, X (טוויטר) ואחרים. מכיוון שקו הסיוע לאינטרנט בטוח מוגדר ומוכר כגוף דיווח רשמי ותיק, הנחת העבודה היא שהמענה שהוא מקבל לדיווחים פרטניים דרך מנגנוני הדיווח בפלטפורמות השונות אמור להיות טוב ויעיל יותר מהמענה שמקבלים ציבור המשתמשים. בזכות המיומנות הגבוהה, הניסיון הרב וההיכרות המעמיקה של צוות קו הסיוע עם סעיפי המדיניות וסוגי התכנים שבהם ניתן לטפל ברשתות החברתיות השונות, קו הסיוע מצליח להעביר לפלטפורמות מדי שנה מאות דיווחים מבוססי פניות מהציבור על אירועי פגיעה שונים, שנבדקו וסוננו בידי הצוות. המומחיות והניסיון המצטבר מאפשרים לקו הסיוע לדווח בהתאם לקריטריונים הברורים של מדיניות הפלטפורמות, כדי לצמצם למינימום דיווחים על תוכן שלא עומד בסף ההסרה או הסימון. לכן הדיווחים שלו לרוב זוכים לאחוזי היענות והסרה גבוהים מאוד. צוות קו הסיוע מקפיד לדווח לפלטפורמות במקרים דחופים על הפרות ברורות של סעיפי המדיניות הרלוונטיים בכל אחת מהן, ובעיקר כאלו שכוללות היבטים של נזק ממשי או מיידני ואיום על חיי אדם.

על רקע זה, המחקר הנוכחי מציע בחינה אמפירית וביקורתית של האופן בו פלטפורמות המדיה החברתית השונות הגיבו לפניות ודיווחים בנוגע לפרסום תוכן פוגעני ואסור המנוגד למדיניות ולכללי הקהילה שלהן במהלך החודשיים הראשונים למלחמת חרבות ברזל (7

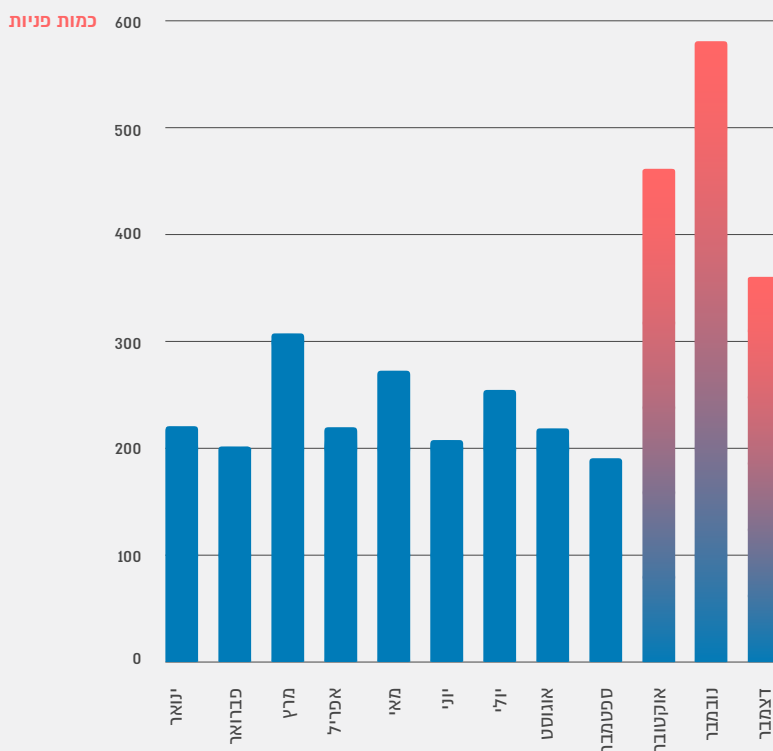
6. מתווה לאסדרת רשתות חברתיות בישראל (ר' הערה 4 לעיל).



באוקטובר-7 בדצמבר 2023), ובודק האם הן הסירו אותו או טיפלו בו באופן אחר. בדיקה זו תתבסס על הניסיון שנצבר בקו הסיוע לאינטרנט בטוח בדיווח על תוכן מסוג זה בחודשיים הראשונים של המלחמה ועל תוכן הדיווחים שנשלחו לפלטפורמות השונות. כאמור, לפני העברת הדיווחים לפלטפורמות, כל הפניות עברו בדיקה יסודית בידי הצוות המקצועי והמיומן של קו הסיוע, על בסיס היכרות מעמיקה עם מדיניות הסרת התוכן של הפלטפורמות השונות ועם מנגנוני הדיווח שלהן.

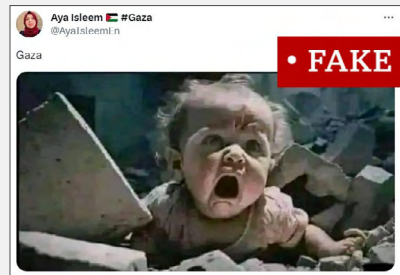
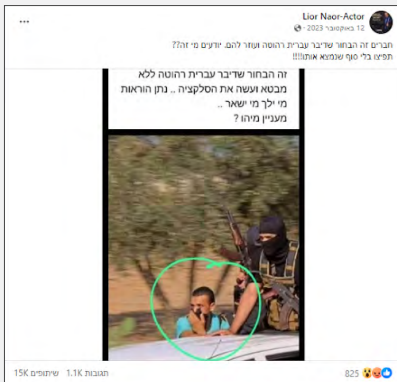
כמפורט בדו"ח הסיכום של שנת 2023 של קו הסיוע של איגוד האינטרנט הישראלי, החל מפרוץ המלחמה כמות הפניות החודשיות שהגיעו לקו הסיוע הייתה כפולה מהכמות הרגילה.<sup>7</sup>

### התפלגות הפניות לקו הסיוע לפי חודשים בשנת 2023



7. פעילות קו הסיוע לאינטרנט בטוח ומאפייני הפגיעות המקוונות במרחב המקוון הישראלי בשנת 2023 – איגוד האינטרנט הישראלי. [קישור](#)

פניות אלה כללו דיווחים על תכנים גרפיים ותומכי טרור, תוכן אלים, הונאות, שקרים וקונספירציות, וביניהם: הכחשת אירועי הטבח ופרטיהם, האשמת ישראל בקצירת איברי פלסטינים, האשמת צה"ל ברצח ישראלים בפסטיבל נובה וביישובי העוטף, דיווחים שקריים על החזקת ילדים ישראלים חטופים בכלובים או התעללות בהם, הפצת מידע כוזב על זיהוי שגוי של חטוף כקצין בכיר בצבא, עלילות וטענות שווא על "בוגדים מבפנים" שסייעו לחמאס בתכנון וביצוע הטבח, שמועות על אזרחים ערבים ישראלים המצלמים בנייני מגורים כדי לתכנן מתקפה נוספת, והודעות תפוצה שקריות על מתקפות סייבר והיערכות לפיגועי טרור נרחבים. כל אלה עוררו בהלה, מתח וחרדה בקרב הציבור.



דוגמאות לקונספירציות ותוכן כוזב ומזוייף ברשתות חברתיות (מימין - X, שמאל למעלה - פייסבוק, שמאל למטה - טיקטוק)

מאות הפניות שהתקבלו מהציבור בערוצים הרשמיים של קו הסיוע בחודשיים הללו ודווחו לפלטפורמות יוצרות מדגם של התכנים הפוגעניים שפורסמו בפלטפורמות במהלך המלחמה, ושל המענה שהן סיפקו לציבור. דו"ח זה מספק הצצה אמפירית, מבוססת ומעשית לאופן ההתנהלות של הפלטפורמות ולמידע שהן נמנעות מלספק לציבור הישראלי לגבי תפקודן במהלך מצבי משבר וחירום בכלל, ומלחמת חרבות ברזל בפרט.

הצפת התוכן הפוגעני והמסוכן ברשתות השונות החל מבוקר הטבח של 7 באוקטובר ולאורך חודשי המלחמה לא נעלמה מעיניהן של הפלטפורמות. מאז אירועי 7 באוקטובר, הפלטפורמות מיהרו לפרסם עדכונים על הפעולות שנקטו כדי לשפר את המוגנות והבטיחות של משתמשים בישראל ובאזור מפני תוכן מזיק, מטריד או לא רצוי, נוסף על המאמצים לאכיפת כללי הקהילה הקיימים שלהן נגד מידע מזויף או מטעה.<sup>8</sup> חברת מטא (Meta), המחזיקה בפייסבוק, אינסטגרם, וואטסאפ ועוד) הודיעה ב-13 באוקטובר כי הקימה חדר מצב מיוחד (Special Operations Center) עם מומחים דוברי עברית וערבית, כדי להגביר את קצב הסרת התכנים הפוגעניים ולהגן על משתמשים מקומיים מפני מידע מטעה אשר מפר את כללי הקהילה. החברה הוסיפה כי בשלושת הימים שלאחר 7 באוקטובר הוסרו 795 אלף פריטי תוכן פוגעניים, כולל הסרה של פי 7 יותר תכנים פוגעניים בהשוואה לחודשיים הקודמים בשל הפרת המדיניות לגבי "ארגונים ואנשים מסוכנים".<sup>9</sup> טיקטוק הכריזה שהקימה מרכז פיקוד (Command Center) מיוחד עם מומחים מקומיים כדי לפעול בזמן אמת נגד תוכן אלים ושנוי במחלוקת ולאכופ את מדיניותה נגד אלימות, שנאה ודיסאינפורמציה. ב-25 באוקטובר החברה מסרה כי במסגרת פעולות האכיפה המיוחדות הוסרו יותר מ-775 אלף סרטונים והופסקו יותר מ-14 אלף שידורים חיים שקידמו אלימות, טרור, שנאה ומידע כוזב.<sup>10</sup> בדומה לכך, מחלקת הסייבר בפרקליטות המדינה – שמייצגת הלכה למעשה את מדינת ישראל מול הרשתות החברתיות בהיבט של הגשת דרישות וולונטריות להסרת תוכן לא חוקי או תומך טרור – הודיעה ב-12 בדצמבר כי בעקבות מתקפת 7 באוקטובר תוגבר צוות הסרת תכנים במחלקה, והוא עובד בשיתוף פעולה עם גורמי ביטחון לשם איתור תכנים

8. אסדרת תוכן ומדיניות פלטפורמות עבור הזירה הישראלית במלחמת "חרבות ברזל" (סקירה מתעדכנת) – איגוד האינטרנט הישראלי. (1 בפברואר, 2024). [קישור](#)

9. [קישור](#). Meta's ongoing efforts regarding the Israel-Hamas war. (December 7, 2023).

10. Our continued actions to protect the TikTok community during the Israel-Hamas war. Newsroom. TikTok (April 7, 2024) [קישור](#)

פוגעניים ופועל להסרת תוכן תומך טרור הנוגע למלחמת ישראל-חמאס.<sup>11</sup> בהודעה נמסר גם כי הצוות בחן כ-39 אלף פריטי תוכן שפורסמו ברשתות החברתיות ובאתרי אינטרנט, והגיש לפלטפורמות השונות יותר מ-26 אלף בקשות הסרה. בתגובה לבקשות אלה הוסרו מעל 2% מהתכנים הבעייתיים שהוגשו לפלטפורמות של מטא, טיקטוק ויוטיוב.<sup>12</sup>

על אף כל אלו, מצבה של ישראל מבחינת אסדרה (רגולציה) מחייבת של אחריות הפלטפורמות לפעול במהירות וביעילות נגד תוכן בלתי חוקי או מסוכן אינו טוב בהשוואה לסטנדרטים המקובלים בעולם. מדינות וגופים בינלאומיים רבים מגדירים מחדש את היקף החסינות והאחריות של הפלטפורמות, ומתנים את החסינות בעמידה בתנאי סף מסוימים. בחזית מגמה זו עומד האיחוד האירופי, שחוקק בשנת 2022 את חוק השירותים הדיגיטליים (DSA). החוק מגדיר מחדש את מערכת היחסים בין הפלטפורמות לציבור ולממשל באמצעות קידום מדיניות הכוללת מספר מרכיבים, כאשר אחד הבולטים שבהם הוא חובת ההגיינות של הפלטפורמות כלפי משתמשים – הנהגת שקיפות מרבית בכל הנוגע לכללי התוכן שלהן ולפרקטיקות ניטור והסרת תוכן שהן מפעילות.

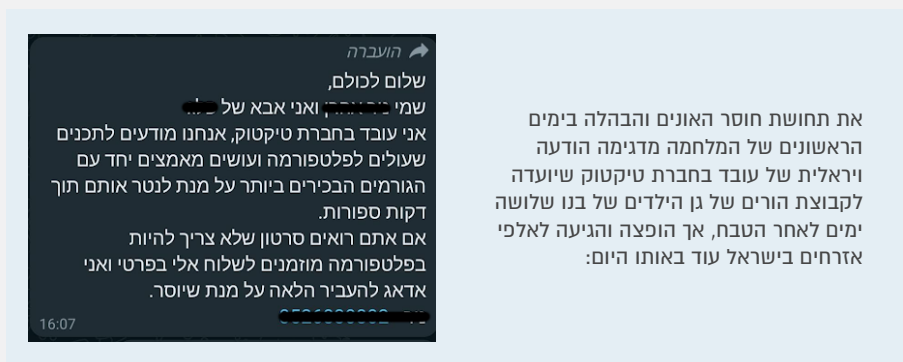
לעומת זאת, בישראל לא חלה על הפלטפורמות חובת שקיפות בנוגע לפרקטיקות ניטור והסרת תוכן. פרסום המידע והנתונים הוא וולונטרי, ולפיכך אין למדינה ולציבור נתונים אמינים על דיווחי המשתמשים, אופי התוכן או אופן הטיפול בו – דבר המהווה חסם לביקורת על התנהלות הפלטפורמות. בהתאם לכך, בשבועות שלאחר אירועי 7 באוקטובר הציבור והרשויות לא קיבלו דיווח לגבי הטיפול של הפלטפורמות בהצפת התוכן הפוגעני ברשתות.<sup>13</sup>

11. פעילות מחלקת הסייבר בפרקליטות להסרת תכנים מרשתות חברתיות: מגמות ונתוני עומק (אוקטובר, 2023) [קישור](#)

12. האם הפרקליטות מצליחה להביא להסרה של תכני טרור מהרשת? - אנשים ומחשבים - פורטל חדשות היי-טק, מיחשוב, טלקום, טכנולוגיות (12 בדצמבר, 2023). [קישור](#)

13. Komaitis, K., Malaret, J., & Jackson, R. (December 6, 2023). Learning more about platforms from the first Digital Services Act transparency disclosures. DFRLab. [קישור](#)

התגובות והיזמות הרבות של אזרחים ישראלים וגורמים שונים בחברה האזרחית ובחברות הטכנולוגיה בימים שלאחר מתקפת 7 באוקטובר, שיקפו את הבהלה והמצוקה של הציבור אל מול ההצפה חסרת התקדים של התוכן הפוגעני והטרור התודעתי ברשתות החברתיות. הן הבליטו את החוסר במענה מספק מצד הפלטפורמות, את החלל הריק שנוצר ואת חוסר האונים שחשו ישראלים רבים. יוזמות אזרחיות כמו [fake-off](#),<sup>14</sup> [IronTruth](#),<sup>15</sup> [digitaldome](#)<sup>16</sup> ואחרות – אשר קמו מיד לאחר 7 באוקטובר בשיתוף פעולה בין יזמי הייטק, טכנולוגים, חוקרים ופעילים חברתיים – שאפו לפתח ממשקי ניטור, זיהוי ודיווח ייעודיים לתוכן פוגעני ועוין ברשתות, במטרה "לנקות" אותן מתכנים פוגעניים ואנטי-ישראליים. רובן כללו מנגנונים חיצוניים ומקבילים למנגנוני הדיווח הפנימיים של הפלטפורמות ופעלו למען העברת דיווחים על תוכן פוגעני בדחיפה לפלטפורמות במטרה להסירו. יוזמות אלו אספו תכנים וחשבונות פוגעניים ופעלו במספר דרכים להסרת התוכן הפוגעני מהרשתות החברתיות, כמו הפעלת פעילים ומתנדבים לדיווח מתואם והמוני על תוכן, העברת התוכן לצוותי הבטיחות והמוגנות באמצעות גורמים רשמיים, או יצירת קשר עם עובדים ישראלים בפלטפורמות השונות, אשר דיווחו על התוכן הפוגעני בערוצים פנימיים הנגישים לעובדים בלבד.



את תחושת חוסר האונים והבהלה בימים הראשונים של המלחמה מדגימה הודעה ויראלית של עובד בחברת טיקטוק שיועדה לקבוצת הורים של גן הילדים של בנו שלושה ימים לאחר הטבח, אך הופצה להגיע לאלפי אזרחים בישראל עוד באותו היום:

14. FakeOff | Cleaning up the social networks. FakeOff. [קישור](#)

15. IronTruth. Telegram. [קישור](#)

16. DigitalDome - פלטפורמה טכנולוגית שפותחה לדיווח על תכנים זדוניים מכל הרשתות, עם חמל שעובד סביב השעון לאמת אותם ולדווח לרשתות למען הסרתם. Digital Dome [קישור](#)

למשל, גורמי דיווח רשמיים באוקראינה קיבלו תשומת לב חריגה ומוגברת מנציגי מטא לאחר הפלישה הרוסית בשנת 2022, וברוב המקרים זכו

לתגובה לדיווחים שהוגשו דרך ערוץ הדיווח הרשמי בתוך 24 עד 72 שעות. זמני התגובה המהירים לדיווחים על תוכן פוגעני באוקראינה בזמן המלחמה עומדים בניגוד מוחלט להתנהלות ולמדיניות של מטא בחלקים אחרים של העולם, כולל אזורי עימות רבים וקשים, בעיקר בדרום הגלובלי.

לדוגמה, במלחמת תיגראי באתיופיה, נהרגו כ-600 אלף אזרחים בשנים 2021-2022, אך לדברי גורמי הדיווח הרשמיים, הפלטפורמות של מטא הגיבו רק לאחר שבועות, ואפילו חודשים, לדיווחים שהוגשו. זאת למרות שדיווחים אלו כללו, כמו באוקראינה, תכנים של איומים, דיסאינפורמציה, הסתה ותמיכה באלימות.

החוויה החיובית של גורמי הדיווח המהימנים המוכרים על ידי מטא באוקראינה מראה כי יש לפלטפורמה יכולת להגיב במהירות כאשר מוקצים לכך די משאבים – אז הן מספקות תגובה עקבית ויעילה בתוך שלוש יממות לכל היותר. הן עמדו ביעדים האלה באוקראינה והן יכולות לעמוד בהם בכל מקום, אם חברת מטא תבחר להשקיע בכך את המשאבים הנדרשים.

## מחקר ארגון INTERNEWS: פערים במענה של META למדווחים רשמיים ברחבי העולם<sup>17</sup>

ערוצי הדיווח של קו הסיוע, כמו של גורמי דיווח מוכרים אחרים בעולם, ייעודיים וייחודיים לכל פלטפורמה או שירות רשת, ואין אחידות או זהות באופן הפנייה וסגנונה. לדוגמה, תוכנית Trusted Partners של חברת מטא (המפעילה את פייסבוק ואינסטגרם) כוללת למעלה מ-400 ארגונים מ-113 מדינות. הארגונים המוכרים כגורמי דיווח מהימנים (Trusted Flaggers) מתוארים באתר החברה כשותפים חיוניים התורמים להבנה של ההקשרים המקומיים בקהילות ברחבי העולם על התוכן הפוגעני בפלטפורמה.<sup>18</sup>

במחקר שערך הארגון Internews נבדקה היענות של מטא לדיווחים של גורמי דיווח מהימנים (Trusted Partners) במדינות שונות בעולם. המחקר מצא כי זמני התגובה של מטא לא היו אחידים עבור מדינות שונות או בהקשרים שונים, ולעיתים לא עמדו בסטנדרטים סבירים לטיפול באירועי בטיחות, משברים ותוכן פוגעני. גורמי דיווח רשמיים במדינות שונות סיפרו כי אף שחלק מהמקרים טופלו באותו היום, דיווחים אחרים לא קיבלו תגובה במשך שבועות ואפילו חודשים.

17. Internews. (August 2, 2023). Safety at stake: How to save Meta's trusted partner program - Information Saves Lives. [קישור](#)

18. Bringing local context to our global standards. Meta [קישור](#)

## ב. מתודולוגיה

מחקר זה מתבסס על נתונים שנאספו בקו הסיוע לאינטרנט בטוח של איגוד האינטרנט הישראלי, וכוללים את כל הדיווחים שהעביר לפלטפורמות צוות קו הסיוע בין 7 באוקטובר ל-7 בדצמבר 2023, המענה שהתקבל מהפלטפורמות לדיווחי קו הסיוע בערוצים הרשמיים, והפעולות שהפלטפורמות ביצעו במסגרת הטיפול בתוכן המדווח. המחקר אינו מתייחס לכלל הדיווחים שהתקבלו מהציבור בקו הסיוע, אלא רק לדיווחים על תוכן שהבעייתיות שלו נבדקה ואומתה והועברו מקו הסיוע לפלטפורמות השונות,<sup>19</sup> ובכלל זה דיווחים על פגיעות מקוונות "שגרתיות" בנושאים מגוונים, כגון הונאות. הדו"ח כולל את הדיווחים השגרתיים בנושאים הללו בימי המלחמה כדי לבדוק כיצד מטפלות הפלטפורמות גם בדיווחים על נושאים שלא בהכרח קשורים ישירות למלחמה, אך התקבלו במהלכה וייתכן שהטיפול בהם מושפע ממנה.

קו הסיוע לאינטרנט בטוח **קיבל** במהלך החודשיים הראשונים למלחמה (7.01.2023-7.12.2023) 1,004 פניות ודיווחים מגורמים שונים – הציבור הכללי ומשתמשי הרשתות, יוזמות שונות, ארגוני חברה אזרחית ורשויות המדינה. מבין אלו, 581 פניות נגעו לפגיעות, תכנים או חשבונות בחמש הפלטפורמות שנבחנו במחקר זה – פייסבוק, אינסטגרם,<sup>20</sup> יוטיוב<sup>21</sup>, טיקטוק ו-X. חשוב להבהיר שמספר הדיווחים שהעביר הקו לפלטפורמות הוא לא נגזרת ישירה של הפניות שהתקבלו מהציבור. הפניות לקו התקבלו בדרכים שונות, וחלק מהדיווחים כללו מספר גדול של פריטי תוכן או חשבונות. דיווחים אלה עברו הליכי בקרה, בדיקה וסינון, והפרטים הרלוונטיים הועברו הלאה כדיווח (אסקלציה) לצוותי הבטיחות

19. במהלך חודשי המלחמה הראשונים קיבל צוות קו הסיוע אלפי פניות מאזרחים, ארגונים ויוזמות שונות, בהן כל דיווח בדקדקנות אל מול סעיפי המדיניות הרלוונטיים בכל פלטפורמה, והעביר לפלטפורמות רק דיווחים על תכנים או חשבונות שבבדיקה היסודית נמצא כי הם אכן מנוגדים לסעיפי מדיניות, או עלולים להוות פגיעה של ממש, ועל כן ישנה ציפייה שיטופלו.

20. פייסבוק ואינסטגרם שייכות לחברת Meta

21. יוטיוב שייכת לחברת Google

של פלטפורמות. חלק מהדיווחים הובילו לתחקירי המשך עצמאיים של צוות קו הסיוע בהם נחשפו פריטי תוכן או חשבונות נוספים הדורשים דיווח לפלטפורמות.

מחקר זה יתאר וינתח את אופן ואיכות המענה והטיפול שסיפקו הפלטפורמות השונות ל-447 הדיווחים שקו הסיוע של איגוד האינטרנט הישראלי **העביר לפלטפורמות** בחודשיים הראשונים למלחמה: פייסבוק (103 דיווחים), אינסטגרם (253 דיווחים), יוטיוב (8 דיווחים)<sup>22</sup>, טיקטוק (60 דיווחים) ו-X (23 דיווחים).

נושאי הדיווחים סווגו בידי צוות קו הסיוע לפי קטגוריות נושאיות של סוג התוכן. כל דיווח שויך לקטגוריה אחת לפחות, וייתכנו דיווחים הנכללים בשתי קטגוריות או יותר:

FALSE INFO	מידע מטעה, קונספירציות ושקרים בוטים
GRAPHIC CONTENT	תוכן גרפי הכולל תמונות וסרטונים של פציעות קשות, גופות, הפעלת אלימות או חטיפות
HATE SPEECH	שיח שטנה הכולל גזענות ואנטישמיות בוטה
TERROR	תוכן וערוצים של ארגוני טרור, תמיכה בפעולות של ארגוני טרור ועידודן, ותוכן הקשור לפעולות הטרור עצמן
SCAM / FRAUD	שימוש ברשתות לשם הונאות כלכליות ופישניג
INCITEMENT AND CALL TO VIOLENCE	הסתה וקריאה לאלימות
PRIVACY OR IMPERSONATION	גניבת תמונות, התחזות והפצת מידע אישי
SEXUAL HARASSMENT	פגיעות על רקע מיני - סחיטה מינית (סקסטורשן) ושימוש בתמונות אישיות ואינטימיות של נשים וגברים לשם הונאות והתחזות על רקע מיני

22. הדיווחים שהועברו ליוטיוב הם מעטים מכדי להפיק בעזרתם תמונת מצב ברורה.



לנוכח השינויים שחלו ב-X, כמו צמצום צוותי בטיחות ומוגנות וצוותי מודרציה ומדיניות<sup>23</sup>, הפלטפורמה נעשתה מתירנית יותר ומתערבת פחות בסימון והסרת תוכן, וממשק הדיווח הייעודי לגורמי דיווח מהימנים מוכרים (Trusted Partners) נסגר. כל אלו עוררו קושי בדיווח על תוכן בעייתי ופוגעני שהופץ ברשת זו. צוות הקו העביר מספר דיווחים באמצעות מנגנוני הדיווח למשתמשים רגילים, אך אלו אינם מאפשרים ניתוח של הפעילות ואופי התגובה, ועל כן אין אפשרות לבחינה השוואתית של ערוץ דיווח רשמי של X.

בנוסף לכך, לטובת השלמת התמונה והרחבת הרקע, נערכו במסגרת המחקר שיחות וראיונות עם נציגי יוזמות אזרחיות וטכנולוגיות לניטור הרשת, נציגי הפלטפורמות וצוותי בטיחות שלהן בישראל ובעולם, בודקי עובדות העובדים עם הפלטפורמות וצוות קו הסיוע של איגוד האינטרנט הישראלי.

במהלך החודשיים שנבדקו, קו הסיוע טיפל גם בתופעה בולטת של חשבונות אינסטגרם מזויפים, אשר השתמשו בתמונות שנגנבו מנשים ישראליות לפרסום אתרים עם תוכן מיני בתשלום. תופעה זו אינה קשורה ישירות למלחמה ולמצב החירום, אך קו הסיוע דיווח עליה והפלטפורמות הרלוונטיות טיפלו בה גם בתקופה זו. גם בימי החירום, קו הסיוע המשיך לעסוק בדיווחים על מקרים של פגיעה בשלום הציבור אשר עלולים להעמיד את הקורבנות בסכנת חיים ממשית או לגרום להם פגיעה מתמשכת על רקע מיני. לכן במחקר הנוכחי הוחלט שלא להשמיט דיווחים מסוג זה בנתונים הנבדקים, מתוך רצון להבין ביצד הפלטפורמות התמודדו גם עם דיווחי פגיעה שגרתיים בזמן המלחמה.

23. Report reveals the extent of deep cuts to safety staff and gaps in Twitter/X's measures to tackle online hate | eSafety Commissioner, Australia. [קישור](#)

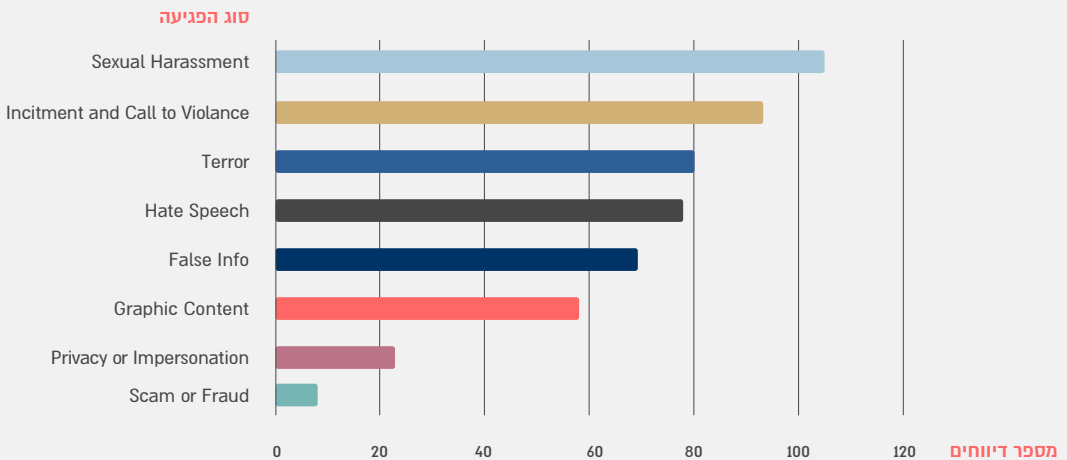
# ג. ממצאים

פרק זה יציג את הממצאים העולים מניתוח תגובות הפלטפורמות לדיוחי קו הסיוע לאינטרנט בטוח של איגוד האינטרנט הישראלי שהועברו בחודשיים הראשונים למלחמת חרבות ברזל (07.10.23–07.12.23) ויתאר זאת לפי סוגי הפגיעות בפלטפורמות השונות, משך זמן התגובה וימי המענה ואופי והיקף הטיפול בדיווחים.

## 1.1 היקף וסוג הדיווחים שבהם טיפל קו הסיוע לאינטרנט בטוח

ראשית, הדיווחים שהעביר קו הסיוע לפלטפורמות מוינו לפי סוגי הפגיעה: פגיעות על רקע מיני (105), הסתה וקריאה לאלימות (93), תוכני טרור (80), שיח שטנה (78), מידע שגוי (69), תוכן גרפי (58), פגיעות בפרטיות והתחזות (23), הונאות ועוקצים (8).

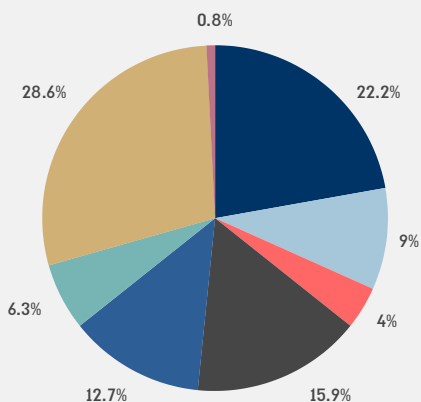
### דיווחים לפלטפורמות בחלוקה לסוג הפגיעה



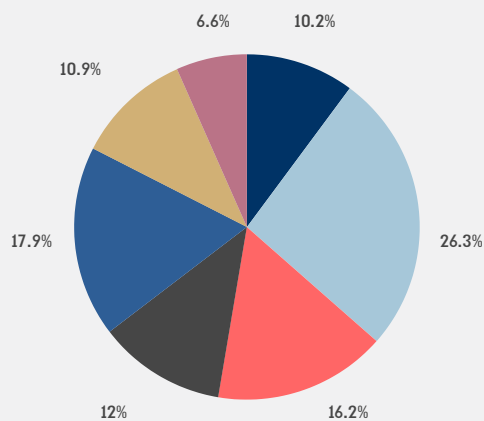
התרשים בעמוד הבא מצביע על שונות בין הנושאים המדווחים ביותר בפלטפורמות השונות, כאשר בפייסבוק נושאים אלה הם מידע שקרי והסתה וקריאה לאלימות; בטיקטוק – שיח שטנה והסתה וקריאה לאלימות; ובאינסטגרם הנושאים הבולטים הם פגיעות בעלות אופי מיני, תכנים גרפיים וטרור.

הנושאים המדווחים בכל פלטפורמה (אחוז סוגי הפגיעה)

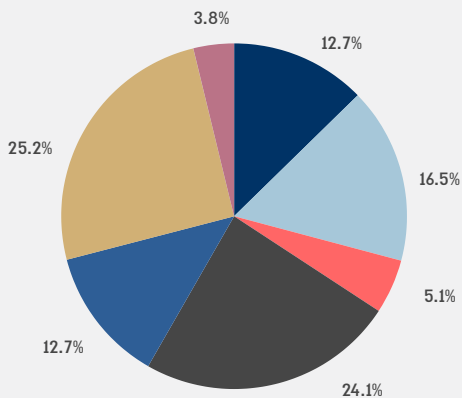
- False info
- Graphic content
- Terror
- Incitement and call to violence
- Sexual assault
- Hate speech
- Scam / Fraud
- Privacy or Impersonation



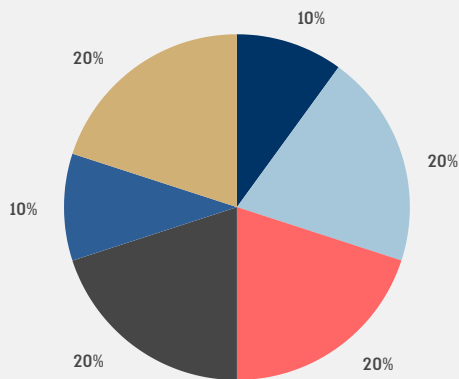
Facebook



Instagram



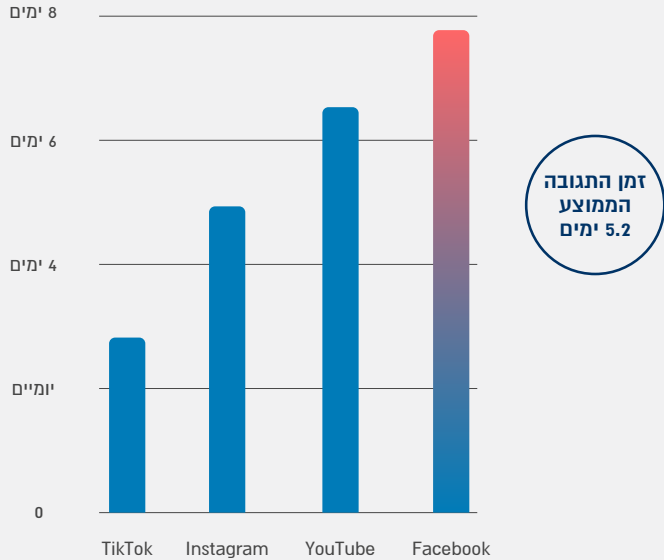
TikTok



YouTube

## 2.ג בחינת זמן התגובה לדיווחי קו הסיוע

זמן תגובה ממוצע לדיווחי קו הסיוע (בימים)



הזמן הממוצע לקבלת מענה מכלל הפלטפורמות לדיווחים ששלח קו הסיוע בחודשיים הראשונים למלחמה עמד על כ-5 ימים. חישוב הזמן הממוצע התבסס רק על דיווחים שזכו לתגובה עד לעת סגירת הנתונים לביצוע המחקר ב- 11.01.24, ואינו כולל דיווחים שלא נענו כלל. לפיכך, נוסף על זמן המענה הממוצע יש להביא בחשבון את אחוז הדיווחים שלא קיבלו מענה בפלטפורמות השונות.

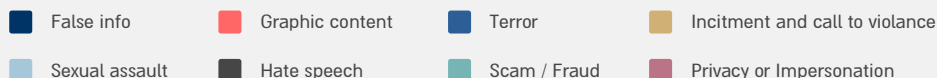
להלן אחוז הפניות של קו הסיוע שלא קיבלו מענה בפלטפורמות השונות עד למועד איסוף הנתונים למחקר זה – אמצע ינואר 2024 (למעלה מ-50 יום אחרי סוף התקופה הנבחנת): פייסבוק – 1 מתוך 103 (1%), אינסטגרם – 9 מתוך 253 (3.6%), טיקטוק – 6 מתוך 60 (10%), יוטיוב – התקבל מענה לכל הפניות, X – מענה אוטומטי. אין משמעות למדידת הזמן ולכן פלטפורמה זו הושמטה מהסעיף הנוכחי.

ניתן לראות כי ישנם הבדלים ניכרים בין הפלטפורמות השונות מבחינת זמני התגובה. בעוד טיקטוק הגיבה לדיווחי קו הסיוע בתוך שלושה ימים בממוצע, זמן התגובה הממוצע של פייסבוק היה למעלה משבוע. חשוב לציין כי נמצא גם הבדל בין זמני התגובה של פייסבוק ואינסטגרם. אף ששתי הפלטפורמות הן בבעלות חברת מטא, ולכאורה כפופות למדיניות תוכן זהה, זמן התגובה של פייסבוק לדיווחי קו הסיוע היה ארוך בממוצע בשלושה ימים כמעט – פער בולט עבור פלטפורמות שונות באותה בעלות.

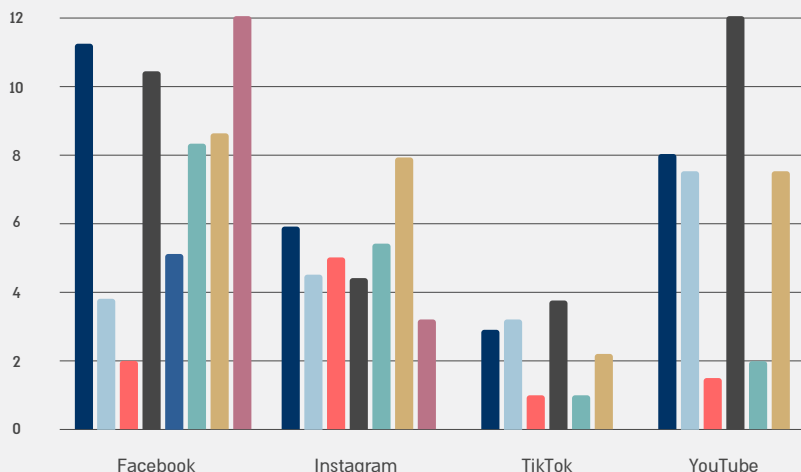
### זמני התגובה לדיווחי קו הסיוע בשבועות הראשונים למלחמה

בשבוע הראשון למלחמה, פייסבוק, אינסטגרם וטיקטוק נענו לדיווחי קו הסיוע תוך כ-24-72 שעות. בעוד שטיקטוק המשיכה לתת מענה לדיווחי קו הסיוע בתוך 24-48 שעות בשלושת השבועות הראשונים של המלחמה, זמני התגובה הממוצעים של אינסטגרם ופייסבוק התארכו בשבוע השני של המלחמה לכ 7- ימים. בתרשים הבא מוצג זמן התגובה של כל פלטפורמה בחלוקה לסוגי תוכן:

### זמן תגובה (בימים) לדיווחי קו הסיוע בחלוקה לפלטפורמות וסוגי הפגיעה



ימים



פלטפורמה

**בפייסבוק** דיווחים על מידע מטעה, שיח שטנה וסוגיות של התחזות ופרטיות קיבלו תגובה לאחר זמן ארוך במיוחד – למעלה מ-10 ימים בממוצע. **באינסטגרם** פניות בנושאי הסתה וקריאה לאלמות מטופלות לאט יחסית לשאר הנושאים, וזמן התגובה עומד על למעלה משבוע. ביוטיוב דיווחים בנושאי תוכן גרפי וטרור מקבלים מענה מהיר יחסית, תוך 48 שעות, אולם שאר הדיווחים מקבלים מענה לאחר למעלה משבוע. ואילו ב**טיקטוק** זמני התגובה הממוצעים קצרים, ומתקרבים ברובם למסגרת של 72 שעות.

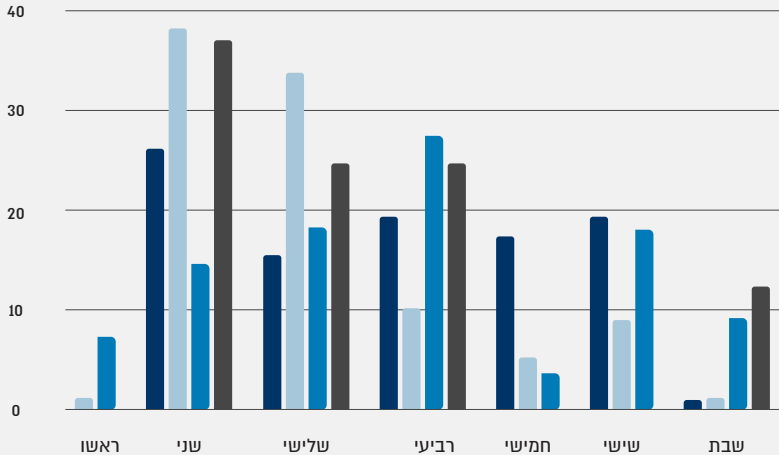
### תגובת הפלטפורמות לפי ימי השבוע בשבועות הראשונים למלחמה

בניגוד להצהרות על היערכות החירום של הפלטפורמות בשל המלחמה והעליה בתוכן הדורש טיפול, היקף המענה מצד הפלטפורמות בסופי שבוע (שבת-ראשון) הוא כ-10% מההיקף הממוצע באמצע השבוע. היקף הדיווחים לפלטפורמות שהגיש קו הסיוע לאינטרנט בטוח אמנם ירד במהלך סופי השבוע בישראל (שישי-שבת), אך רק ב-60%, מה שלא מסביר את הירידה הדרמטית של 90% שהציגו הפלטפורמות.

### התפלגות תגובות הפלטפורמות בחלוקה לימי השבוע

Facebook Instagram TikTok YouTube

אחוז מסה"כ  
התגובות  
שהתקבלו

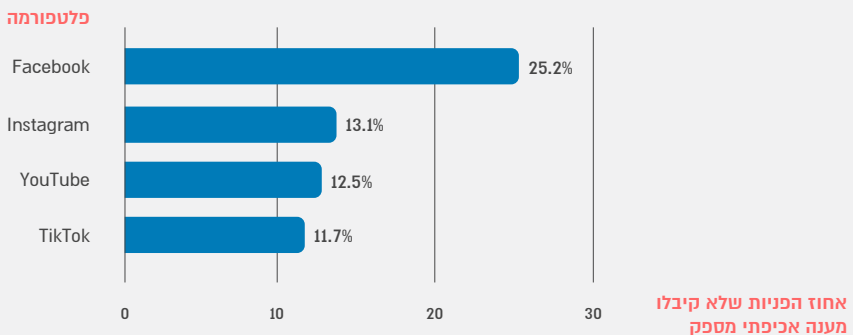


ניתן לראות כי בניגוד לפלטפורמות האחרות, טיקטוק ויוטיוב מספקות מענה חלקי גם בסופי שבוע. לעומת זאת, בפייסבוק ואינסטגרם המענה בסופי שבוע הוא אפסי, וניכר כי צוותי המודרציה והבטיחות זמינים רק בימות העבודה המקובלים בצפון אמריקה ואירופה.

### 3.ג אפקטיביות המענה של הפלטפורמות לדיווחים

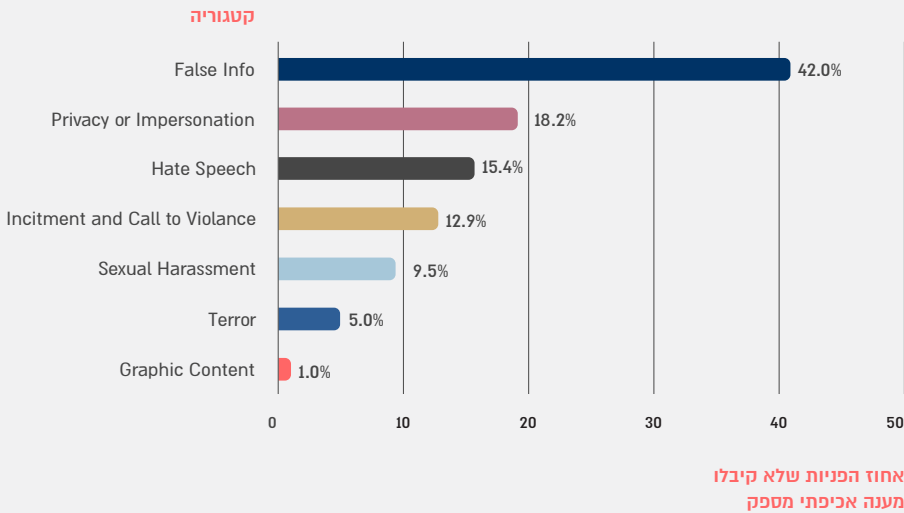
מאחר שכל דיווח נבחן לפני העברתו לפלטפורמות אל מול כללי המדיניות המוצהרים שלהן בידי הנציגים המיומנים של קו הסיוע, מקרים שבהם הפלטפורמה לא הסירה או סימנה את התוכן המדווח ללא הסבר או הצדקה מספקת מוגדרים כהיעדר פעולות אכיפה מספקות בטיפול בדיווח על תוכן פוגעני. במסגרת המחקר נערכה בחינה השוואתית של אחוז המקרים שבהם לא בוצעו פעולות אכיפה מספקות (מתוך כלל הדיווחים שהועברו לפלטפורמות לאורך החודשיים הראשונים של המלחמה) אל מול אחוז המקרים שזכו לטיפול ראוי. "טיפול ראוי" מוגדר כהסרה או סימון של התוכן המדווח, או לחלופין כהסבר משכנע לכך שהתוכן המדווח לא נחשב הפרת מדיניות. מקרים שבהם הועבר דיווח לפלטפורמה והיא לא טיפלה בו ללא מענה או הסבר מספק נחשבו לאירועים שבהם הפלטפורמה לא ביצעה פעולת אכיפה ראויה. בתרשים הבא מוצג היקף הדיווחים שלא קיבלו מענה אכיפתי מספק, בחלוקה לפלטפורמות:

אחוז הפניות שלא טופלו כראוי, בחלוקה לפלטפורמות



במבט השוואתי על אחוזי פעולות האכיפה הלא מספקות בטיפול בדווחים, פייסבוק בולטת לרעה עם 25% חוסר טיפול מספק (כלומר, 75% טיפול ראוי בלבד). לעומתה, הפלטפורמות האחרות לא נקטו פעולות אכיפה מספקות עבור 13-11 אחוז מהדיווחים (כלומר, 87-89 אחוז טיפול ראוי).

### אחוז הדיווחים שלא קיבלו מענה אכיפתי מספק, בחלוקה לנושא הדיווח



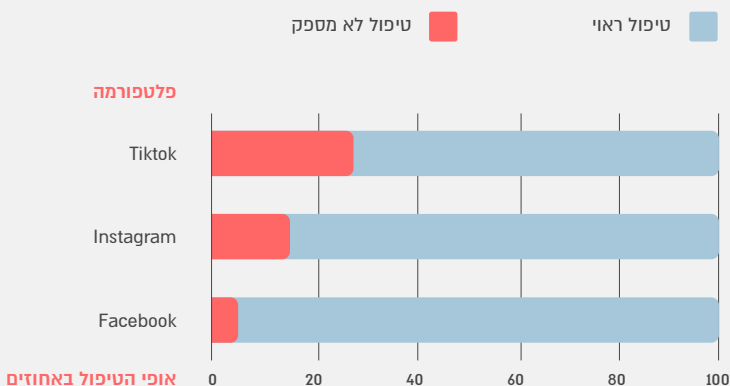
בהשוואה לפי סוגי תכנים פוגעניים, מתברר כי מידע כוזב (False Info) הוא סוג התוכן שבו אחוז הטיפול הלא מספק הוא הגבוה ביותר בממוצע, כאשר הפלטפורמות השונות לא נקטו פעולות אכיפה מספקות ב-42% מהדיווחים (ראו הרחבה בתת-הפרק הבא). אחוזי הטיפול הלא מספק גבוהים גם בדיווחים על פגיעה בפרטיות והתחזות, שיח שנאה והסתה וקריאה לאלימות (19-12 אחוז מהדיווחים). לעומת זאת, רוב רובם של הדיווחים על תוכן גרפי ותוכן תומך טרור זכו לטיפול ראוי (95-99 אחוז).



### אפקטיביות וכשלים מצד הפלטפורמות: תכנים של שיח שטנה (HATE SPEECH)

שיח שטנה כולל הפצת פרסומים או התקפות ישירות נגד אנשים על בסיס מאפיינים מוגנים כגון גזע, דת, מין, העדפה מינית וכו', באמצעות שפה תוקפנית, השמצות, שימוש בסטריאוטיפים פוגעניים, הצהרות על נחיתות מולדת וכיוצא בזה. שיח שטנה יוצר סביבת שיח של הפחדה וקיטוב ולעיתים אף מעודד אלימות פיזית ישירה. מרבית הרשתות החברתיות לא מאפשרות פרסום תכנים המוגדרים כשיח שטנה נגד קבוצות מיעוט בפלטפורמות שלהן.<sup>24, 25</sup>

### היקף הדיווחים על תכני שיח שטנה שלא טופלו כראוי, בחלוקה לפלטפורמות



טיקטוק היא הפלטפורמה בעלת השיעור הנמוך ביותר של טיפול ראוי בדיווחים על שיח שטנה, כאשר עבור 26% מהדיווחים בנושא לא ננקטו פעולות אכיפה מספקות. אינסטגרם לא נקטה פעולות אכיפה מספקות בכ-15% מהדיווחים, ופייסבוק – בכ-5% בלבד.

24. Countering hate on TikTok | TikTok. [קישור](#)

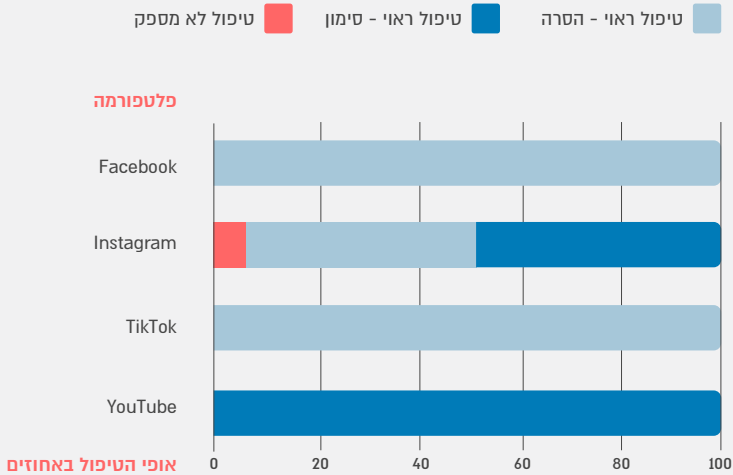
25. Hate speech | Meta Transparency Center. [קישור](#)

### אפקטיביות וכשלים מצד הפלטפורמות: תוכן גרפי (GRAPHIC CONTENT)

תוכן גרפי הוא תמונות או קטעי וידאו מזעזעים, המכילים תוכן אלים או בוטה במיוחד, כגון פציעות חמורות, גופות, הפעלת אלימות וגרימת סבל חמור, או לחלופין תכנים מיניים בוטים או קשים שאינם מתאימים לצפייה של הקהל הרחב. מאז פרוץ המלחמה הוצפו פלטפורמות המדיה החברתית בתכני זוועה מסוג זה, שתיעדו את אירועי הטבח והחטיפות ב-7 באוקטובר ביישובים ישראליים ואת הפצצות צה"ל ברצועת עזה.

בחינה של אופן הטיפול בתוכן גרפי מראה כי אינסטגרם נוקטת במדיניות משתנה לגבי פרסומים מסוג זה, בשונה מפלטפורמות אחרות. לעיתים היא מסירה את התוכן, ובמקרים שהתוכן אינו מוסר, לרוב מתווספים אליו סימון ותווית אזהרת צפייה, אשר מתריעה מפני התוכן הגרפי המוצג בפוסט. 94% מהדיווחים שהעביר קו הסיוע לאינסטגרם בנוגע לתוכן גרפי טופלו כראוי. מבין אלה, נמצא כי מחצית הוסרו מהפלטפורמה, ולמחצית השנייה נוספו סימון ותווית אזהרת צפייה. יש לציין כי בשאר הפלטפורמות מדגם הדיווחים על פרסומים המכילים תוכן גרפי היה קטן.

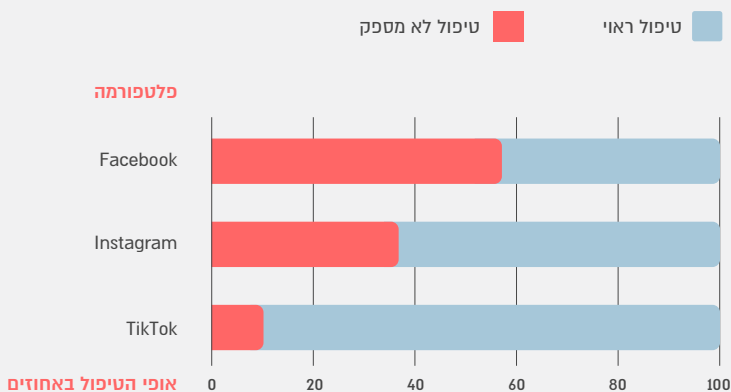
### היקף הדיווחים על תוכן גרפי קשה שלא טופלו כראוי, בחלוקה לפלטפורמות



## 4.ג מאפיינים וכשלים ייחודים בטיפול הפלטפורמות במידע כוזב ודיסאינפורמציה

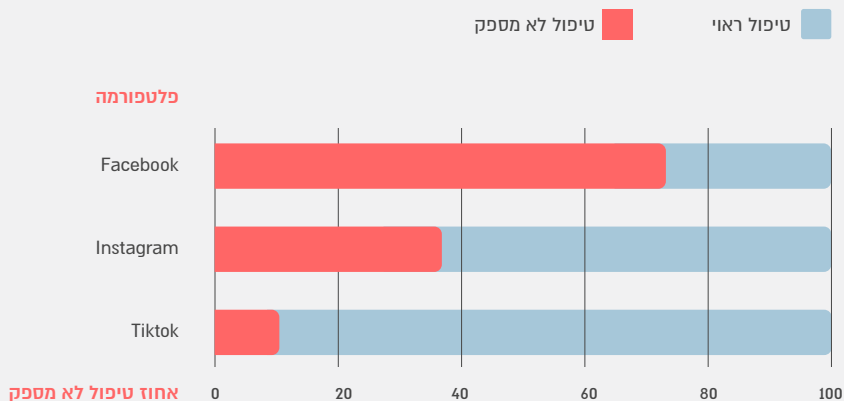
בטיפול בדיווחים על תוכן כוזב, פייסבוק בולטת לרעה – הפלטפורמה לא נקטה פעולות אכיפה מספקות ב-57% מהדיווחים שקיבלה מקו הסיוע לגבי תוכן שקרי ומידע מטעה.

היקף הדיווחים על תוכן כוזב (FALSE INFO) שלא טופלו כראוי, בחלוקה לפלטפורמות



בחינה מעמיקה של תגובת הפלטפורמות לדיווחים על תוכן שסומן כשקרי וכוזב (False Info) בלבד – בניכוי דיווחים על כמה נושאים במקביל, למשל מידע מטעה שהוא גם הסתה לאלימות – הראתה כי פייסבוק לא נקטה פעולות אכיפה מספקות ב-72% מהמקרים. לעומת זאת, עבור אינסטגרם וטיקטוק לא נמצא הבדל מהותי. נתון זה מחזק את הרושם כי ביחס לפלטפורמות האחרות, פייסבוק מתקשה לטפל או נמנעת מטיפול שיטתי בדיווחים על תכנים המוגדרים דיסאינפורמציה או תוכן כוזב.

## אי נקימת פעולות אכיפה מספקות לגבי מידע כוזב בלבד (ללא סעיפי דיווח נוספים)



### 5.1 איכות המענה והמידע שהפלטפורמות מספקות לדיווחי TRUSTED FLAGGERS

הבדיקות ההשוואתיות התייחסו גם לאופי ולתוכן של התגובה המילולית מצד הפלטפורמות לדיווחי קו הסיוע של איגוד האינטרנט הישראלי, שבו הן מכירות כמדווח אמין או שותף דיווח ייחודי בשנים האחרונות. תגובת הפלטפורמות לדיווח היא הדרך היחידה של קו הסיוע להתעדכן בנוגע לגורל הדיווחים שהעביר, וכך לנהל את הפניות המתקבלות, לסגור פניות שטופלו, וגם ללמוד ולהשתפר בסינון הפניות מהציבור והעברת דיווחים מדויקים לפי סעיפי המדיניות הרלוונטיים. ככלל, התגובות מגיעות לכתובת המייל הייעודית של קו הסיוע, שדרכה מתבצעים הדיווחים לפלטפורמות. מלבד מהות התגובה (התקבל/נדחה) והפעולה שננקטה (הוסר/סומן וכו'), נבדקו גם התוכן והנוסח של תגובות הפלטפורמות, על מנת לבחון את יחסן לדיווחים שהן מקבלות מגופי דיווח רשמיים המוגדרים Trusted Flaggers.

בהכללה, בבחינה ההשוואתית נמצא כי מטא (בעלת פייסבוק ואינסטגרם) מספקת תשובות קבועות, שבלוניות ושטחיות בתגובה לדיווחים, שאינן כוללות מידע או מתייחסות לתוכן הדיווח הספציפי.

## מענה חברת מטא לדיווחי קו הסיוע על תוכן פוגעני

נוסח תגובת מטא לדיווחי קו הסיוע במקרים שבהם התוכן המדווח הוסר או סומן:

**Thank you for your report. Please note that we have now reviewed the related content and taken appropriate action. Do not hesitate to let us know if you require further assistance.**

נוסח תגובת מטא אשר מצביע לרוב על החלטה שלא לנקוט צעדים בנוגע לתוכן המדווח:

**Our team has done an in depth investigation of the content in question, but has found that it does not violate our Community Standards.**

נוסח תגובת מטא אשר מצביע לרוב על כך שהתוכן הוסר לפני שהצוותים הספיקו לתת מענה לדיווח:

**It looks like the content you reported is no longer accessible on our site. If you're still seeing this content on the site, please reply to this email with a web address (URL) that links directly to the potentially violating material, in order for our teams to review it.**

נוסח תגובת מטא לדיווחי קו הסיוע במקרים שבהם התוכן המדווח הועבר לבודקי עובדות עצמאיים:

**We have reviewed your report. It unfortunately isn't clear to us that any of the reported content violates Facebook's Community Standards or Instagram's Community Guidelines. We have passed the content to third party fact checkers (3PFC) for review for potential misinformation.**

בהקשר של דיווחים על מידע שגוי או הטעיה, לעיתים מתקבלת תגובה כללית על העברת המקרה לבודקי עובדות עצמאיים. בדומה לכך, גם טיקטוק מתאפיינת בתגובות שבלוניות ולקוניות לדיווחים של קו הסיוע המוכרים על ידה כמהימנים.

לעומתן, יוטיוב מספקת תשובה מפורטת וייעודית לכל דיווח ומדגישה את סעיף המדיניות שהופר ועל בסיסו ננקטה הפעולה. התגובות של צוותי הבטיחות של יוטיוב בולטות בהיותן מפורטות ואישיות, ומתייחסות לתוכן המדווח ולסעיפי המדיניות הרלוונטיים. הן כוללות פנייה אישית למדווח (נציג קו הסיוע) בשמו, ועל המייל חתום בשמו אחד מעובדי הפלטפורמה. תגובות כאלה עוזרות לצוות קו הסיוע להשתפר, לאפיין, להבין ולדייק דיווחים עתידיים.

Hi [REDACTED]

Thank you for reaching out. Upon review, we have removed the video [REDACTED] in line with our [Violent extremism or criminal organizations policy](#).

Content intended to praise, promote, or aid violent criminal organizations is not allowed on YouTube. These organizations are not allowed to use YouTube for any purpose, including recruitment. Upon review, we have removed the video per our Violent criminal organizations policy as it contains content produced by violent criminal or terrorist organizations.

For further information, please visit our [Help Center](#).

Sincerely,

[REDACTED] from the YouTube Team

Hello [REDACTED]

Thank you for your patience as we conducted our review.

Upon review, we have age-restricted the video [REDACTED] under our [Violent or graphic content policies](#). The video will only be viewable for users who are logged in and over the age of 18.

Thank you and please feel free to reply to this email in case of any questions.

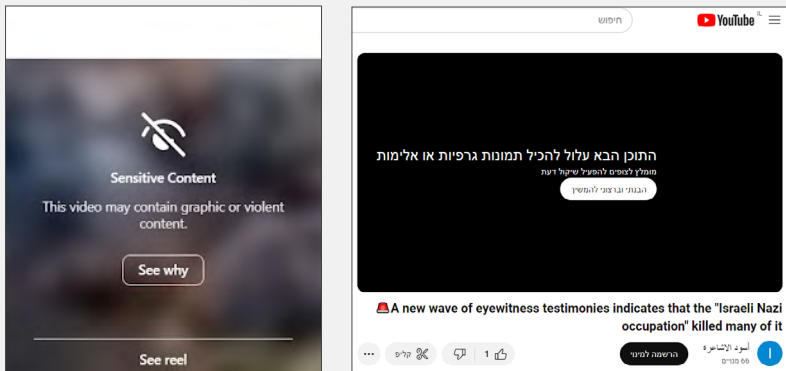
Regards,

[REDACTED] on behalf of The YouTube Team

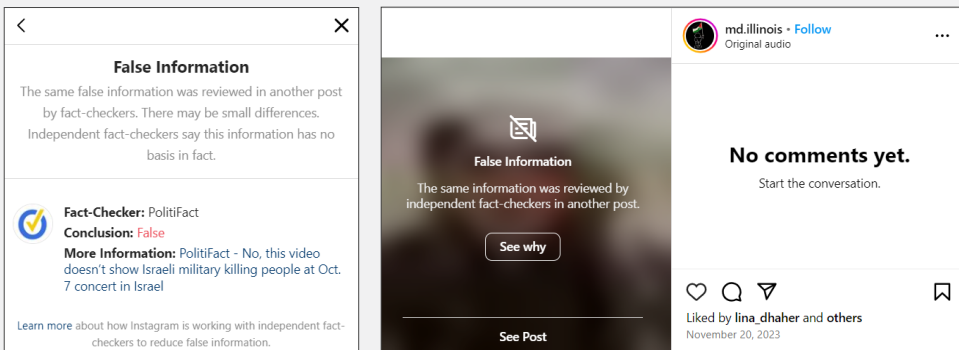
דוגמאות למענה שהתקבל מיזמיוב בנוגע לדיווחים של קו הסיוע

## 6.2 סימון וחייווי למשתמשים על פעולות אכיפת התוכן שמבצעות הפלטפורמות

הפלטפורמות נוקטות צעדים מגוונים כלפי תוכן גרפי אשר מפר את מדיניותן ועלול לסכן משתמשים או לפגוע בהם, אך הוחלט שלא להסירו. אחד הצעדים הללו הוא סימון התוכן באמצעות טשטושו והוספת הערת אזהרה או הערת הקשר לפני הצפייה, שמהווה תחליף להסרה מוחלטת ומאפשרת למשתמשים לקבל החלטה מושכלת לגבי צריכתו.



דוגמאות לסימון תוכן גרפי שלא הוסר (בתגובה ל-22 דיווחים באינסטגרם ו-2 ביוטיוב)



סימון של מידע כוזב בידי בודקי עובדות חיצוניים (בתגובה לדיווח אחד בלבד באינסטגרם)

בשונה מההערות והתוויות המתוספות לתוכן מסומן, כאשר התוכן מוסר מהפלטפורמה מוצגת הודעה למשתמש כי הוא הגיע לתוכן שאינו זמין. אולם, רוב הפלטפורמות אינן מסבירות או מציגות מה היה התוכן שהוסר, מדוע הוסר, על סמך איזה סעיף מדיניות, והאם הסרתו הייתה פעולת אכיפה או פעולה שביצע המשתמש שהעלה אותו, כמו הגבלת פרטיות והרשאות הצפייה או מחיקת הפוסט.

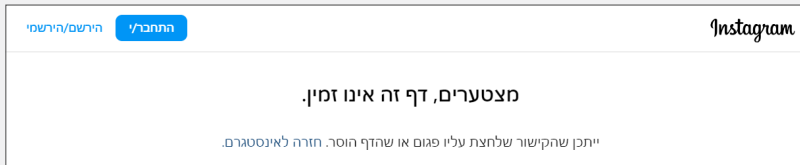
חוסר השקיפות והפירוט הדל אינם מאפשרים למשתמשים לדעת מה עלה בגורל התוכן ומדוע, ומונעים גביית מחיר ציבורית מחשבונות שננקטו כלפיהם פעולות אכיפה, או סימון משתמשים שהתחרטו על הפרסום והסירו אותו בעצמם, והעלאת מודעות אצל עוקבים ומשתמשים אחרים בנוגע לפרסומים פסולים שנמחקו.

חוסר הפירוט והשקיפות גם מונע מגופי הדיווח המוכרים (Trusted Flaggers) לבחון את פעולות האכיפה שביצעו הפלטפורמות ואת הטיפול בדיווחים שהעבירו להן. הדבר פוגע בתהליכי העבודה של גופים אלה, בעיקר לנוכח הדלות, השבלוניות והגנריות של התגובה לדיווחים במייל.

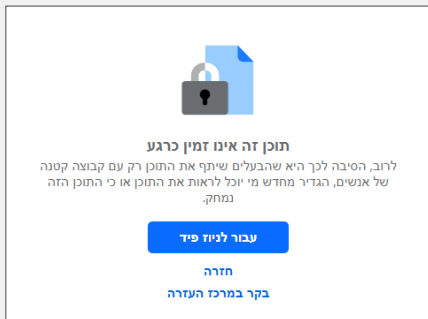
**בהכללה, הבחינה ההשוואתית מצאה כי טיקטוק והפלטפורמות של מטא מציגות מידע מעורפל על גורלו של התוכן.** לא ניתן לדעת אם התוכן הוסר בשל פעולת אכיפה, מאיזה סוג ועל סמך איזה סעיף, או אם בכלל מדובר בפעולה של המשתמש עצמו. טיקטוק אף מגדילה לעשות ומעודדת את המשתמשים להמשיך ולצרוך תוכן נוסף בפלטפורמה. לעומתן, יוטיוב מפרטת מעט יותר ומציינת כי ננקטה פעולת אכיפה כנגד התוכן או החשבון, אך אינה מפרטת מה הן העבירות ועל סמך אילו סעיפי מדיניות נקבע שנעשתה עבירה.



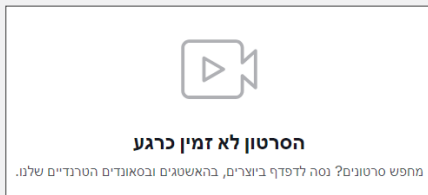
## דוגמאות לאופן הצגת תוכן שהוסר מהפלטפורמה



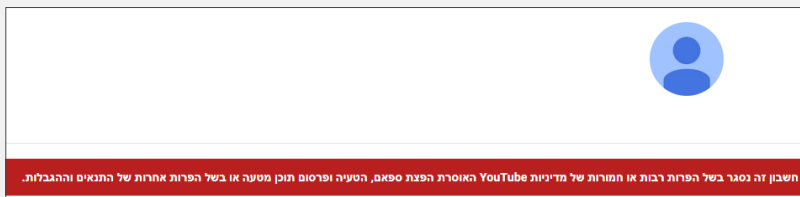
**אינסטגרם** – מצטערים דף זה אינו זמין (ייתכן שהקישור שלחצת עליו פגום או שהדף הוסר)



**פייסבוק** – תוכן זה אינו זמין כרגע (לרוב, הסיבה לכך היא שהבעלים שיתף את התוכן עם קבוצה קטנה של אנשים, הגדיר מחדש מי יכול לראות את התוכן או כי התוכן הזה נמחק)



**טיקטוק** – הסרטון לא זמין כרגע מחפש סרטונים? נסה לדפדף ביוצרים, בהאשטגים ובסאונדים המרנדיים שלנו)



**יוטיוב** – הבהרה שהחשבון נסגר ופירוט הסיבות האפשריות לכך

## ד.

# ניתוח המשמעויות עבור הזירה הישראלית

מחקר זה בחן את התנהלותן של פלטפורמות המדיה החברתית במהלך החודשיים הראשונים למלחמת חרבות ברזל (7 באוקטובר-7 בדצמבר 2023), באמצעות בדיקת ההליך והאיכות של הטיפול והתגובה לדיווחים של קו הסיוע לאינטרנט בטוח של איגוד האינטרנט הישראלי על תוכן זדוני, פוגעני או שקרי. הבדיקה נעשתה באמצעות ניתוח השוואתי של זמני ואופי התגובה ושל הטיפול בפועל בתוכן שדווח בידי קו הסיוע, המוגדר גורם מדווח מהימן (Flagger Trusted) במרבית הפלטפורמות המרכזיות ובשירותי רשת רבים.

## טיפול לא מספק בתוכן פוגעני במלחמה

המסקנה המרכזית העולה מהממצאים היא **שהפלטפורמות לא הצליחו להתמודד באופן מספק עם שטף התוכן הפוגעני בחודשיים הראשונים למלחמה**. במקרים רבים הפלטפורמות לא נקטו פעולות אכיפה מספקות, זמני התגובה שלהן למרבית הדיווחים היו ארוכים ובלתי סבירים ביחס למצב החירום וחומרת התוכן, והן לא סיפקו מענה הולם בסופי שבוע, אפילו בתחילת המלחמה. ממצאי המחקר מראים כי התפקוד והמענה הכללי של הפלטפורמות בתקופה הנבדקת לקו בחסר. נמצא כי:

הפלטפורמות לא נקטו פעולות אכיפה מספקות (הסרה, סימון או הסבר מספק להימנעות מפעולת אכיפה) בתגובה לדיווחי קו הסיוע, ולמעשה קיימים פערים מטרידים בין מדיניות התוכן המוצהרת ובין ההתנהלות בפועל. הממצאים מצביעים על שוני בשיעורי ההיענות בין הפלטפורמות השונות ועבור סוגים שונים של תוכן פוגעני. **פייסבוק בולטת לרעה במיוחד בכך שלפי נתוני המחקר היא לא טיפלה באופן מספק ב-25% מהדיווחים של קו הסיוע** (כלומר, רק 75% מהדיווחים קיבלו טיפול ראוי) – הרבה יותר משאר הפלטפורמות, שבהן שיעור הדיווחים שלא קיבלו טיפול מספק נע סביב 13-11 אחוז (כלומר, 89-87 אחוז קיבלו טיפול ראוי).

זמן התגובה של הפלטפורמות המרכזיות לדיווחים של קו הסיוע לאינטרנט בטוח בחודשיים הראשונים למלחמה היה כ-5 ימים בממוצע. מדובר בזמן ארוך ובלתי מתקבל על הדעת גם בשגרה, ובוודאי בעת מצב חירום רגיש ונפיץ במיוחד, עבור דיווחים הנשלחים על רקע

הלחימה, שחלקם קשורים לתמיכה בטרור, הסתה לאלימות או מידע שקרי. הפלטפורמה שתגובתה הייתה האיטית ביותר בחודשיים הראשונים ללחימה הייתה פייסבוק, שזמן התגובה הממוצע שלה לכלל הדיווחים עמד על כ-7 ימים – זמן ארוך ובלתי סביר עבור משתמשים החשופים לתוכן פוגעני, מבהיל ומעורר חרדה בתקופת לחימה וחירום, שחלקם אף נמצאים בסכנת חיים.

ככלל, כמעט ולא התקבל מענה מהפלטפורמות במהלך סופי שבוע (שבת-ראשון). החוסר בנקיטת פעולות אכיפה מספקות בולט במיוחד על רקע שטף התכנים הפוגעניים בפלטפורמות, שבעה ימים בשבוע, בשגרה ובוודאי בזמן מלחמה. נראה כי המשאבים המוגבלים שהשקיעו הפלטפורמות במהלך סופי השבוע אינם עולים בקנה אחד עם הצורך שהתעורר ועם הפרסומים וההצהרות על היערכויות מיוחדות לטיפול בתוכן בעייתי.

טיקטוק, הגיבה לרוב הדיווחים במהירות, ובלטה בהיותה הפלטפורמה היחידה כמעט שהגיבה לפניות ודיווחים בסופי שבוע (שבת-ראשון). עם זאת, יש לציין כי רבים מהדיווחים לטיקטוק לא קיבלו תגובה עד לעת סגירת הנתונים, מה שפוגם במוצע זמני התגובה וההתייחסות הכוללת מצד פלטפורמה זו לכלל הפניות.

פלטפורמת X (טוויטר) היא נוכחת נפקדת בנתוני המחקר – במהלך החודשים שנבדקו הועברו אליה דיווחים רבים מקו הסיוע, אך מנגנון הדיווח עבור גורמי דיווח מוכרים אינו פעיל, והתגובות האוטומטיות והדלות של הפלטפורמה לא מאפשרות לעקוב אחר זמני התגובה ואופי הטיפול בדיווחים.

## נשלים במיפול במידע כוזב ובתכנים הדורשים מעורבות אנושית

בכל הנוגע להערכה של היענות הפלטפורמות והטיפול בדיווחים, ניכרים הבדלים עקביים בטיפול בין סוגי תוכן שונים. דיווחים על תוכן שניתן לטפל בו בפשטות באמצעות סריקה אוטומטית וממוחשבת – כמו תוכן גרפי וטרור – זכו לטיפול מהיר ויעיל יחסית. לעומת זאת, תכנים הדורשים בחינה באמצעות כוח אדם מיומן וקבלת החלטות אנושית, המבוססת על הבנה והקשר מקומי – למשל מידע כוזב או מטעה ושיח שטנה והסתה – סובלים מפעולות אכיפה לא מספקות ואיטיות.

בחינה השוואתית של המענה שניתן לדיווחים על סוגים שונים של תוכן פוגעני מראה כי התוכן שקיבל את רמת המענה הנמוכה ביותר היה מידע כוזב או מטעה. על פי סיכום הנתונים, כלל הפלטפורמות לא נקטו פעולות אכיפה מספקות עבור 42% מהדיווחים על מידע כוזב. אחוזי הטיפול הלא מספק היו גבוהים יחסית (19-12 אחוז) גם בדיווחים על התחזות ופגיעה בפרטיות, שיח שטנה והסתה וקריאה לאלימות. לעומת זאת, כמעט כל הדיווחים על תוכן גרפי וטרור טופלו כראוי (95-99 אחוז).

בחינה מעמיקה של התוכן שעליו דיווח קו הסיוע כמידע כוזב מראה כי בפועל הפלטפורמות של חברת מטא לא טיפלו כראוי ב-57% מהדיווחים עבור פייסבוק וב-36% מהדיווחים עבור אינסטגרם. בניגוד לכך, טיקטוק לא טיפלה כראוי רק ב-10% מהדיווחים על תוכן שקרי וכוזב.

דיווח על תוכן פוגעני יכול לכלול מספר סעיפים. כאשר נבדקו דיווחים על מידע כוזב בלבד, ללא סעיפים נוספים (כמו שיח שטנה או הסתה), לא נצפה שינוי מהותי עבור אינסטגרם וטיקטוק, אך בפייסבוק אחוז הדיווחים שבהם לא ננקטו פעולות אכיפה מספקות זינק ל-72%.

הממצאים מצביעים על כך שמטא לא נוקטת פעולות אכיפה מספקות בנוגע למידע שגוי או שקרי המתפרסם בפלטפורמות שלה, הן מבחינת אופי הטיפול והן מבחינת זמן המענה הנדרש. מטא הצהירה כי היא מסתמכת על בודקי עובדות עצמאיים, שבודקים ומדרגים את מידת הדיוק של תכנים,<sup>26</sup> ומעבירה אליהם בין היתר פניות ודיווחים מגורמי דיווח רשמיים. בודקי העובדות הם עצמאיים ואין למטא יכולת לכפות עליהם בדיקת תוכן מסוים. עם זאת,

26. Meta's Third-Party Fact-Checking Program. [קישור](#)

ממצאי מחקר זה מעוררים דאגה לגבי המנגנון של מטא להעברת הטיפול לבודקי עובדות חיצוניים בזירה הישראלית, כאשר מבין כלל הדיווחים של קו הסיוע שהועברו למטא בנוגע למידע כוזב, רק לפוסט אחד (מתוך 57 שדווחו) נוספה הערה של בודקי עובדות אלה. אין באן כוונה למתוח ביקורת על בודקי העובדות החיצוניים, שנדרשים לעמוד באמות מידה עיתונאיות מחמירות ואין להם יכולת להתמודד עם קונספירציות או מידע שקרי המוצג כדעה או כשאלה, אך אופי הטיפול של מטא במידע כוזב מצטרף למסקנות הנוגעות לזמן התגובה הארוך במיוחד של פייסבוק לדיווחים על תוכן כזה. כאמור, בפייסבוק זמן התגובה הממוצע לדיווחים על סוג התוכן הזה היה למעלה מ-11 ימים, לעומת כ-3 ימים בטיקטוק וכ-6 ימים באינסטגרם. הדבר עשוי ליצור דינמיקה של "טלפון שבור", שבה גורם דיווח רשמי מתריע על מידע שקרי, שגוי או מטעה, ומדווח עליו בשל חומרתו, ומטא אינה מטפלת בדיווח בעצמה אלא מעבירה אותו (ואת האחריות עליו) לבודקי עובדות המשמשים כצד שלישי עצמאי.

עובדה זו מדגישה את הצורך בהשקעת משאבים והעסקת בוחני תוכן (מודרטורים) מותאמי אזור ושפה, וכן את הצורך בהגברת השקיפות בנוגע למשאבים שמפנות הפלטפורמות לצוותי הבטיחות המקומיים ומידת הזמינות שלהם.

## חוסר בשקיפות בנוגע לפעילות הפלטפורמות

לצד כשלי האכיפה שנמצאו בכל הפלטפורמות שנסקרו, בולטים גם יידוע ושקיפות בלתי מספקים בנוגע לצעדים שננקטו כלפי התוכן המדווח אל מול גורמי הדיווח הרשמיים והמוכרים. התגובות לגורמי הדיווח הרשמיים הן גנריות, שבלוניות וקבועות, אינן כוללות התייחסות לתוכן או נושא הדיווח, ובמקרים של הסרה או סימון אינן מספקות מידע מהותי, אלא מסתפקות לרוב בנוסח הלקוני "Taken appropriate action". לעיתים הפלטפורמות מצהירות על הסרה או טוענות שלא הסירו כיוון שהתוכן או החשבון המדווח לא נוגד אף סעיף בכללי הקהילה.

השוואה של זמני התגובה בין הפלטפורמות העלתה כי ביוטיוב זמני התגובה היו ארוכים יחסית, אך יש לציין לטובה את פלטפורמה זו, העושה מאמץ ניכר ליצור שקיפות לגבי הצעדים שנקטה ומציינת במקום התוכן המוסר מה עלה בגורלו ומדוע. יוטיוב גם מספקת מענה המתייחס לתוכן הדיווח ומסביר את הסעיף שעל בסיסו התוכן הוסר, תוך פנייה אישית לגורם המדווח, כשלו רוב ההודעה גם כוללת חתימה בשם של אחד מעובדי הפלטפורמה. מענה מפורט כזה תורם רבות ליכולת של גורמי דיווח רשמיים לעקוב אחר הדיווחים, לוודא

שהם מטופלים כראוי וליידע את ציבור המשתמשים. כמו כן, הסבר מפורט אשר מבחיר את ההחלטה ואת הצעדים שננקטו יכול לעורר שיח ולשפר את הפירוט של ההקשר המקומי, ככל שהדבר נדרש, ואת האיכות והדיוק של הדיווחים העתידיים מטעם אותם גורמים.

בחינת הקישורים שהועברו בדיווח חושפת כיצד נראה תוכן או חשבון שהוסר לאחר ביצוע פעולת ההסרה. גם במקרה זה, הפלטפורמות אינן נוהגות בשקיפות – רובן לא מציינות שננקטה פעולת אכיפה כלפי התוכן, וודאי שלא את סעיף המדיניות הרלוונטי. חוסר השקיפות מקשה על גורמי דיווח רשמיים לעקוב אחרי הטיפול בפניותיהם, הן כדי לדעת מה עלה בגורל הדיווחים והן כדי לשפר את דיווחיהם העתידיים. בדומה לכך, גם משתמשים שנפגעו לא מקבלים מידע מספק מהפלטפורמות על אופן הטיפול בפניותם או גורלו של התוכן המדווח.

## מענה הולם ומהיר מתאפשר בזכות השקעת משאבים ואכיפת מדיניות אחידה

עם פרוץ המלחמה הפלטפורמות מיהרו לפרסם את פעילותן הענפה למען התמודדות עם תוכן פוגעני וטענו כי פעלו להסרה מוגברת של תכנים בעייתיים. **ממצאי המחקר מלמדים כי יש לפלטפורמות יכולת לספק מענה מהיר, בתוך 72-24 שעות, במצב חירום כגון מלחמה או קונפליקט אזורי, ככל שיש להן תמריץ כלכלי-תדמיתי לעשות זאת.** ואכן, ניתוח הממצאים הנוגעים לזמני התגובה של הפלטפורמות בתקופה שנבדקה מראה כי בשבוע הראשון ללחימה הן הצליחו לתת מענה כזה והגיבו לדיווחי קו הסיוע בתוך 72-24 שעות, בדומה לזמן התגובה המתועד של מטא לדיווחים שהעבירו גורמי דיווח מוכרים במהלך מלחמת רוסיה-אוקראינה ב-2022. הנתונים מראים כי זמני תגובה מהירים ונאותים לגורמי דיווח רשמיים ומוכרים בזמן משבר הם יעד בר השגה כאשר הפלטפורמות מעוניינות בכך ומפנות למטרה זו משאבים וכוח אדם. כך עולה מניסיון העבר במקומות אחרים בעולם, וכן מימי המלחמה הראשונים, כאשר הפלטפורמות עצמן הודיעו כי פתחו חדרי מצב וחדרי פיקוד והסירו מאות אלפי פריטי תוכן. עם זאת, במרבית השבועות מאז פרוץ המלחמה ועד 7 בדצמבר 2023, הטיפול בדיווחים לא היה מספק.

ככלל, **ממצאי המחקר עולה כי למרות האתגרים והאיזמים שניצבו בפני הפלטפורמות בעקבות הטבח של ה-7 באוקטובר והמלחמה שפרצה מיד אחריו, הן לא השקיעו מספיק משאבים כדי לתת מענה הולם ומקיף לפניות ודיווחים בנוגע לתכנים ולפעילות המנוגדים למדיניות ולכללי הקהילה שלהן,** וכדי לשמור על בטיחות המשתמשים בשעת חירום. בפועל,

הפלטפורמות השונות סיפקו מענה חלקי, איטי ולא אחיד גם לדיווחים דחופים ומאומתים שהעביר אליהן קו הסיוע לאינטרנט בטוח, אף שהן מכירות בו כגורם מדווח מהימן, מנוסה ומועדף. אלה נתונים מדאיגים במיוחד בבואנו לשער כיצד טופלו דיווחים שהועברו בידי הציבור הרחב.

המסקנה הכללית המתבקשת היא שאופן התנהלות זה הוא בחירה מודעת ותוצר של מדיניות, ולא רק כורח בלתי נשלט התלוי בנסיבות. בהיעדר מדיניות ומשאבים הולמים, זמן התגובה, מועדי התגובה ואיכות הטיפול אינם מספקים.

---

# ה.

## סיכום והמלצות לתיקון המצב

### פלטפורמות ורשתות חברתיות

הניתוח של תגובת הפלטפורמות לדיווחים על תוכן פוגעני בזמן המלחמה, והמסקנות המרכזיות העולות ממנו, מצביעים על צורך בביצוע מספר צעדים ושינויים חשובים כדי לשמור על המוגנות והבטיחות של משתמשי הפלטפורמות אל מול האיומים והאתגרים הקיימים. ההמלצות נכונות לשגרה, ועל אחת כמה וכמה לתקופות חירום (סכסוכים אלימים, מלחמות ואסונות). הן נכתבו על סמך ניסיון שהצטבר בחודשים ובשנים האחרונים בישראל ובאזור, אך סביר להניח שהן מתאימות גם לאזורים אחרים בעולם ובהקשרים אחרים, וראוי לשקול אותם בכובד ראש וליישם אותן בהקדם.

### 1. היערכות מיוחדת לקונפליקט ומלחמה

מצבי חירום, מלחמות ועימותים אלימים מחייבים את פלטפורמות המדיה החברתית להיערך לפעולה ולתגובה באופן מהיר, החלטי ומקיף יותר מאשר בימי שגרה. ההיערכות צריכה לכלול:

**הרחבת שעות הפעילות וימי המענה** – יש להקצות משאבים להרחבת שעות הפעילות וימי המענה, כדי לספק מענה רציף לעלייה בכמות ובמגוון התכנים בשעת חירום באמצעות הוספת כוח אדם מקומי ומיומן, אשר מבין את השפה, ההקשרים והניואנסים של המצב בזמן אמת.

**התאמת המנגנונים להתמודדות עם שטף מתמשך של תוכן** – יש להתאים את מנגנוני האכיפה והטיפול בתכנים מזיקים להתמודדות עם עלייה חדה ומתמשכת בהיקפים ובחומרה של תכנים אלו בעת משבר. כפי שעולה מהממצאים, הפלטפורמות חזרו במהירות לשגרה ולא הקצו די משאבים לערוץ הדיווח של קו הסיוע לאינטרנט בטוח וגורמים אחרים.



**קיצור זמני התגובה** – יש לקצר את זמני התגובה ולספק מענה קבוע, אחיד ומהיר, גם בסופי שבוע, על מנת לטפל בבעיות ומשברים בזמן אמת. הטיפול צריך לכלול הן ניטור אקטיבי והסרה של תכנים מזיקים ביוזמת הפלטפורמות, והן תגובה מהירה לדיווחים מהציבור ומגופי דיווח מהימנים.

**מתן עדיפות לגורמי דיווח מהימנים** – יש לתת עדיפות גבוהה לערוצי הדיווח של גורמים מוכרים ומהימנים, הן במהירות הטיפול והן במידת הרצינות של היחס לדיווח וברמת השקיפות של המענה.

**שיפור מנגנוני הדיווח של המשתמשים** – יש לשפר ולהתאים את המנגנונים לדיווח על אירועים אקטואליים ומצבי חירום, בדגש על תכנים מסוג מידע כוזב, הסתה ותמיכה בטרור. כמו כן, יש לספק למשתמשים הסבר ברור על הסיבות לסימון או הסרת התוכן, כדי לאפשר בקרה יעילה יותר על תהליכי האכיפה.

## 2. התמודדות עם דיסאינפורמציה ובדיקת עובדות

המחקר מצא קושי ניכר ורוחבי אצל הפלטפורמות השונות לטפל ביעילות במידע כוזב או מטעה (דיסאינפורמציה), אשר בעת מלחמה או עימות אלים עלול לסכן חיי אדם. שיפור ההתמודדות עם בעיה זו צריך לכלול:

**הרחבת שיתוף פעולה עם גורמים מקומיים** – יש להרחיב ולייצר שיתופי פעולה עם מגוון גופי בדיקת עובדות וגופים עצמאיים מקומיים מומחים אשר מבינים את השפה וההקשר המקומי.

**שיפור מנגנוני הדיווח והטיפול בתוכן כוזב הקשורים לאירועי משבר וחירום בזמן אמת** מנגנוני הדיווח הנוכחיים אינם מאפשרים דיווח מדויק ויעודי על תכנים כוזבים הקשורים לאירועים אקטואליים, ויש להתאימם לדיווח על דיסאינפורמציה ספציפית בהקשר של מלחמות, אסונות ומשברים.

**הנגשת מידע וכלים לגופי מחקר ואקדמיה** – יש לתמוך ולספק מידע וכלים לגופים העוסקים בזהויה והבנה של דפוסי תוכן ופעילות מתואמת ולא אותנטית של הפצת תוכן כוזב ומטעה. על הפלטפורמות לחשוף לקהילת המחקר ולציבור את המחקרים הפנימיים שהן

עורכות, לטובת למידה הדדית ושיפור ההבנה בנוגע לטיפול הנדרש בתופעה, ולאפשר להם גישה למידע ונתונים שיאפשרו בחינה חיצונית ועצמאית של המצב הקיים.

### 3. חיזוק הקשר עם גורמי דיווח מוכרים ומקומיים (TRUSTED FLAGGERS)

נוכח החשיבות של גורמי דיווח רשמיים מקומיים, כגון קו הסיוע לאינטרנט בטוח, על הפלטפורמות להשקיע משאבים בחיזוק הקשר עימם – העברת הכשרות, מתן כלים, עדכונים בשעת חירום, מתן עדיפות לטיפול בדיווחיהם, הגברת שקיפות התגובות, והקפדה על הליך הוגן ופרוצדורה תקינה ומהירה.

**הכשרות ומשוב לגורמי דיווח מקומיים** – יש להרחיב את מעגל גורמי הדיווח באמצעות יצירת קשר עם ארגוני חברה אזרחית נוספים בתחום ולערוך הכשרות תקופתיות לגופי הדיווח גם בימי שגרה, תוך הסברת מדיניות התוכן וקווי האכיפה השונים. כמו כן, יש לספק משוב שוטף על דיווחים שנדחו והסיבות לדחייה, ועדכונים רצופים על שינויי מדיניות ועל אכיפה בשעת חירום.

**פיתוח, שיפור ואספקת כלים טכנולוגיים** – יש לספק לגורמי הדיווח המוכרים כלי מחקר, ניתוח ודיווח מתקדמים, שיאפשרו להם לייעל את עבודתם. כמו כן, יש לשפר את ממשקי הדיווח, ולאפשר שימוש בטפסים מקוונים נגישים ופשוטים למילוי (במקום דיווח במייל), לצד אפשרות למעקב אחר תהליך הטיפול ויכולת ערעור.

**תשומת לב מוגברת לדיווחים** – יש להגדיר יעדי זמן מחייבים, שיפורסמו לציבור, למענה על דיווחים, ולמנות נציגים אישיים שישמשו כתובת לגורמי הדיווח במקרי חירום. כמו כן, יש לפרסם הנחיות ברורות בנוגע לזמני התגובה הצפויים ולאפיקי הפעולה הפתוחים למשתמשים במקרה של עיכובים או ערעור על מענה לא מספק.

**הגברת שקיפות ציבורית** – על הפלטפורמות לפרסם מדיניות ברורה בנוגע לתהליכי הטיפול בדיווחים, לרבות זמני תגובה ויעדי טיפול מוגדרים. כמו כן, יש לספק לגופי הדיווח מידע שוטף המפרט את הפעולות שננקטו בעקבות כל דיווח ואת סעיף המדיניות או קו האכיפה המנחים.

## המחוקק והממשלה בישראל

כדי שהמדינה תוכל לוודא שהציבור אינו נפגע מהתוכן בפלטפורמות, עליה לבדוק שהן פועלות כראוי ולחייבן לעמוד בסטנדרטים ובדרישות באמצעות חקיקה מקומית.

**דרישה לרמות גבוהות של שירות וזמינות מהפלטפורמות** – המחקר מראה כי רמת התגובתיות והזמינות של הפלטפורמות הטכנולוגיות בחירום אינה מספקת, ופוגעת בביטחון ובאינטרס הציבורי. על המחוקק לחייב את הפלטפורמות לספק רמות שירות וזמינות גבוהות, הן בימי חול והן בסופי שבוע, וביתר שאת בתקופות חירום.

**דרישה להבנת השפה וההקשר המקומי** – יש לחייב את הפלטפורמות להעסיק כוח אדם מקומי איכותי שמבין היטב את השפה וההקשר בישראל, כתנאי לפעילות במדינה.

**לשקיפות** – על המחוקק לחייב את הפלטפורמות לפעול בשקיפות בנוגע להיקפם ואיכותם של המשאבים המוקצים לטיפול בבטיחות ומוגנות הגולשים בישראל ובנוגע לנקיטת פעולות האכיפה וההסרה, הן בשגרה והן בחירום.

**השקעה בחינוך והסברה לאוריינות דיגיטלית** – מאחר שקיים קושי לסמוך על הפלטפורמות בלבד, המדינה צריכה להפנות משאבים באופן שיטתי לחינוך והסברה על שימוש נכון ובטוח ברשת – הן לציבור הרחב והן לגופי ממשל, אקדמיה ותקשורת. הדבר יתרום לחיזוק החוסן בשגרה ובחירום.

## גופי מחקר, תקשורת וחברה אזרחית

לא רק המדינה והפלטפורמות עצמן יכולות לפעול לחיזוק המוגנות של המשתמשים ברשת, אלא גם חוקרי וחוקרות תקשורת ואקדמיה, מכוני מחקר, עמותות, קרנות, אקטיביסטים, כלי תקשורת וגופי בדיקת עובדות.

**חיזוק היכולות בתחומי הבדיקה והניטור** – חשוב לפתח מיומנויות וכלים מתקדמים לבדיקת עובדות, חקר דיגיטלי וניטור תוכן ודיסאינפורמציה בקרב גופי אקדמיה, כלי תקשורת והחברה האזרחית – יכולות שכיום יש בהן מחסור בישראל. יש לשפר את ממשקי העבודה והחיבורים בין חברות טכנולוגיה העוסקות בפיתוח יכולות של ניטור ואיסוף מידע ברשתות, לבין גופי מחקר אקדמיים וארגוני חברה אזרחית.

**איסוף נתונים ומעקב אחר ביצועי הפלטפורמות** – יש לעודד את הרחבת הפעילות העצמאית של האקדמיה, אנשי מחקר וארגוני חברה אזרחית הכוללת מעקב ואיסוף נתונים על התנהלותן של הפלטפורמות השונות בהקשר הישראלי, לפתח שיטות מחקר מתאימות וקואליציות שיפעלו למען מטרה זו, ולהפיץ את המידע בציבור ובקרב מקבלי ההחלטות.

**הגברת הלחץ על הפלטפורמות והדרישות הציבוריות מהן** – על החברה האזרחית וגופי המחקר לנקוט עמדה אקטיבית וביקורתית מול הפלטפורמות הגלובליות הפועלות בישראל, כמו גם מול המחוקקים ומקבלי ההחלטות. עליהם להדגיש את הדרישה הציבורית לשיפור המענים והתאמתם למצב בישראל, הגברת השקיפות מול המשתמשים והרשויות, חיזוק המוגנות ושיפור רמת הטיפול והתגובתיות של הפלטפורמות בכל הנוגע להגנה על משתמשיהן בישראל.

**העצמת הפעולה האזרחית והתקשורתית נגד תוכן כוזב ברשתות** – יש להרחיב ולפתח את השדה האזרחי והתקשורתי הנותן מענה לפרסום והפצה של תוכן כוזב ומטעה ברשתות באמצעות תמיכה בארגוני מדיה עצמאיים, ארגונים וגופים אזרחיים של מחקרי רשת ובדיקת עובדות. חיזוק השחקנים הפועלים כיום, לצד הקמת גופים עצמאיים נוספים ויחידות לבדיקת עובדות ומחקרי רשת בכלי התקשורת, יובילו למתן מענים נרחבים ומספקים יותר.

## ביבליוגרפיה

Ring, E (January 22, 2024). Op-Ed for Haaretz | How Telegram, Twitter and TikTok have become lethal tools of Hamas psychological warfare - Israel Internet Association - (ISOC-IL).

<https://en.isoc.org.il/about/news-room/op-ed-for-haaretz-how-telegram-twitter-and-tiktok-have-become-lethal-tools-of-hamas-psychological-warfare>

שימוש ברשתות חברתיות ושירותים מקוונים בישראל: נתוני 2024 עם פילוחים דמוגרפיים – איגוד האינטרנט הישראלי. איגוד האינטרנט הישראלי.

[https://www.isoc.org.il/sts-data/online\\_services\\_index](https://www.isoc.org.il/sts-data/online_services_index)

Greenwood, S. (June 1, 2023). Internet, smartphone and social media use around the world | Pew Research Center. Pew Research Center's Global Attitudes Project.

<https://www.pewresearch.org/global/2022/12/06/internet-smartphone-and-social-media-use-in-advanced-economies-2022/>

מתווה לאסדרת רשתות חברתיות בישראל. (2023). המכון הישראלי לדמוקרטיה.

<https://www.idi.org.il/books/49130>

מאפייני השיח האלים והתמודדות מולו במרחב האינטרנט והרשתות החברתיות בישראל (דצמבר 2022) – איגוד האינטרנט הישראלי. איגוד האינטרנט הישראלי.

[https://www.isoc.org.il/sts-data/violent\\_discourse\\_survey\\_2022](https://www.isoc.org.il/sts-data/violent_discourse_survey_2022)

פעילות קו הסייע לאינטרנט בטוח ומאפייני הפגיעות המקוונות במרחב המקוון הישראלי בשנת 2023 – איגוד האינטרנט הישראלי. איגוד האינטרנט הישראלי.

<https://www.isoc.org.il/sts-data/helpline-isoc-2023>

אסדרת תוכן ומדיניות פלטפורמות עבור הזירה הישראלית במלחמת חרבות ברזל – איגוד האינטרנט הישראלי. (10 בכברואר, 2024). איגוד האינטרנט הישראלי.

<https://www.isoc.org.il/regulating-digital-services/israel/platform-policies-iron-swords>

Meta's ongoing efforts regarding the Israel-Hamas war. (December 7, 2023).

<https://about.fb.com/news/2023/10/metas-efforts-regarding-israel-hamas-war/#hebrew-translation>

Our continued actions to protect the TikTok community during the Israel-Hamas war. Newsroom. TikTok (April 7, 2024)

<https://newsroom.tiktok.com/en-us/our-continued-actions-to-protect-the-tiktok-community-during-the-israelhamas-war>

Komaitis, K., Malaret, J., & Jackson, R. (December 6, 2023). Learning more about platforms from the first Digital Services Act transparency disclosures. DFRLab.

<https://dfrlab.org/2023/12/06/learning-more-about-platforms-from-the-first-digital-services-act-transparency-disclosures/>

---

Bringing local context to our global standards. Meta

<https://transparency.fb.com/policies/improving/bringing-local-context/>

"Internews". Safety At Stake.

[https://internews.org/wp-content/uploads/2023/07/SafetyAtStake\\_Internews.pdf](https://internews.org/wp-content/uploads/2023/07/SafetyAtStake_Internews.pdf)

Report reveals the extent of deep cuts to safety staff and gaps in Twitter/X's measures to tackle online hate | eSafety Commissioner, Australia.

<https://www.esafety.gov.au/newsroom/media-releases/report-reveals-the-extent-of-deep-cuts-to-safety-staff-and-gaps-in-twitter/xs-measures-to-tackle-online-hate>

Meta's Third-Party Fact-Checking Program

<https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking>