בדיקת נאותות בנושא זכויות אדם בדבר השפעות Meta

בישראל ובפלסטין במאי 2021

1. מטרה

דו"ח זה מהווה בדיקת נאותות בנושא זכויות אדם, הבוחן את השפעת המדיניות והפעילויות של Meta[1] במהלך המשבר שאירע בחודש מאי 2021 בישראל ובפלסטין. המטרה העיקרית הינה לספק ל-Meta המלצות מתועדפות, אופרטיביות, ושימושיות בדבר מדיניות ופרקטיקות להגשמת התחייבויותיה של Meta על פי מדיניות החברה לזכויות אדם, ותחומי אחריותה על פי העקרונות המנחים של האומות המאוחדות לעסקים וזכויות אדם – UNGPs (United Nations Guiding Principles on Business and Human Rights), לרבות עקרונות 20-22.[2] לשם השגת יעד זה, BSR:

- מצביעה על ההשפעות של Meta על זכויות אדם,[3] לרבות תחומי אחריותה על פי ה-UNGPs.
- סוקרת את אחריותה של Meta לטפל בהשפעות אלה על זכויות אדם בהתאם ל-UNGPs.
- מספקת המלצות למדיניות ופרקטיקות בקשר עם תחומי אחריות אלה.

בדיקת נאותות בנושא זכויות אדם זו מסייעת למלא אחר המלצות מועצת הפיקוח (ה-Oversight Board) לפיהן על Meta להתקשר עם גוף עצמאי שאינו קשור לאף צד בסכסוך הישראלי-פלסטיני, על מנת לקבוע האם בקרת התוכן (content moderation) של Meta בערבית ובעברית במהלך התקופה האמורה בוצעה ללא הטיה.

BSR זיהתה פרקטיקה טובה, תחומים לשיפור ולקחים שהפיקה Meta במהלך חודש מאי 2021. לזכותה של Meta ייאמר כי היא קיבלה על עצמה את ביצוע הדו"ח הזה, כדי ללמוד מה פעל היטב ומה טעון שיפור והתייחסות. עוד יצוין, כי Meta כבר מראה התקדמות לגבי רבות מההמלצות של BSR.

---

[1] ביום 28.10.2021, Facebook, Inc. שינתה את שמה ל-Meta Platforms, Inc.. לשם עקביות, דו"ח זה משתמש בשם "Meta" כדי להתייחס לחברה הן ביחס לתקופה שקדמה ליום 28.10.2021 והן ביחס לתקופה המאוחרת לו. אזכורים של "Facebook" מתייחסים רק לפלטפורמת המדיה החברתית, ולא לחברה בכללותה. בנוסף, דו"ח זה מפנה לפעולות שננקטו על ידי Meta כחברה ביחס לישות ספציפית. אין באמירה זו כדי ללמד כי Meta נקטה בפעולה מסוימת ביחס לכל הישויות. לדוגמה, למעט אם מדיניות מוגדרת ככזאת שחלה על WhatsApp, היא אינה חלה על WhatsApp.

[2] לדוגמה, עיקרון 20 קובע כי חברות צריכות לעקוב אחר האפקטיביות של תגובתן להשפעות על זכויות אדם, באמצעות התקשרות עם בעלי עניין פנימיים וחיצוניים, ושילוב הממצאים בתהליכי דיווח פנימיים רלוונטיים וקידום שיפור מתמיד. עיקרון 21 קובע כי מקום בו התעוררו חששות מטעם בעלי עניין הנוגעים לזכויות אדם, על חברות להתקשר עם גורם חיצוני ולספק מידע מקיף דיו לשם הערכת נאותות התגובה של החברה. עיקרון 22 קובע כי על חברות לפעול לתיקון השפעות שליליות או לשתף פעולה לצורך זה, ובכלל זה באמצעות נקיטת פעולה כדי לוודא שפגיעות קודמות לא יישנו.

[3] השימוש במונח "השפעות על זכויות אדם" ("human rights impacts") נעשה בהתאם לאמור ב-UNGPs, המגדיר "השפעה שלילית על זכויות אדם" ככזאת המתרחשת כאשר פעולה מונעת או מפחיתה את יכולתו של אדם ליהנות מזכויות האדם המוקנות לו. המונח "השפעה שלילית על זכויות אדם" אינו שווה-ערך – ואין משתמעים ממנו – חובה חוקית, הפרתה או קשר סיבתי לפי דין. ראו https://www.ohchr.org/sites/default/files/Documents/Publications/HR.PUB.12.2_En.pdf.

BSR ו-Meta מבקשות להודות לכל בעלי העניין שהשקיעו מזמנם וממרצם בהשתתפות בבדיקות נאותות זו.

2. היקף ומגבלות

היקפה של בדיקת נאותות זו מוגבל למשבר של מאי 2021 והיא אינה בוחנת את תפקידה של Meta בישראל ובפלסטין באופן רחב יותר. כל קבוצת הפלטפורמות הרלוונטיות של Meta – Facebook, Instagram ו-WhatsApp – נכללות תחת הבדיקה. דו"ח זה מסכם את בדיקת הנאותות שביצעה BSR ביחס לזכויות אדם, שהתחילה בספטמבר 2021 והושלמה באפריל 2022.

3. מתודולוגיה

סקירה זו של בדיקת נאותות ביחס לזכויות אדם מבוססת על: מידע הזמין באופן פומבי; ראיונות עם נושאי משרה רלוונטיים ב-Meta ועיון בחומרים קשורים; נקודות מבט של מגוון רחב של בעלי זכויות ובעלי עניין מושפעים בישראל, בפלסטין ובעולם שמקורם בראיונות שניתנו ל-BSR ובתכתובות פומביות בין בעלי עניין מושפעים לבין Meta; נתונים רלוונטיים, לרבות ניתוחים של "jobs" (כלומר, מקרים) בערבית ובעברית שמקורם בישראל או בפלסטין; ניתוח פעולות אכיפה שבוצעו בבחינה אנושית ובחינה אוטומטית; וסקירת מקרים פרטניים של הסרת תכנים והגבלת חשבונות.

שיטת העבודה של BSR כללה בחינה של נתונים פנימיים של Meta. אולם, חשוב לציין כי לא קיימת "אמת יסודית" ("ground truth") מבוססת באשר למה צריכים להיות שיעורי אכיפת התוכן המוחלטים או היחסיים בישראל ובפלסטין, משום שלא קיימים נתונים לגבי שכיחותו של תוכן אלים ברמה המדינתית (בניגוד לרמת השוק, למשל שוק השפה הערבית). הנתונים שנבדקו על ידי BSR התקבלו מ-Meta במטרה לסייע ל-BSR להבין את התגלמות משבר מאי 2021 על גבי הפלטפורמה בישראל ובפלסטין, ולאמת מגמות ותובנות שאספנו באמצעות מחקר איכותני; אולם, אף שנדרשנו לשימוש בנתונים כבסיס לניתוח שלנו, אין הנתונים עומדים בדרישות האיכות, הוודאות והמהימנות לצורך פרסום חיצוני מפורט ורשמי.

4. ההקשר

- כללי המדיניות של Meta ויישומם התרחשו בהקשר של מערך מורכב של דינמיקות חברתיות והיסטוריות. דינמיקות אלו כוללת פוליטיקה בינלאומית ואזורית ואסימטריה של כוח מצד אחד, ואת מדיניות התוכן של Meta[4] ואכיפת מדיניות התוכן מצד שני.

- אין להתייחס לפעולותיה של Meta באירועי מאי 2021 באופן מבודד, אלא חייבים להבינן בהקשר של הסכסוך המתמשך בישראל ובפלסטין, המתאפיין בנרטיבים היסטוריים מתחרים של קורבנות ורדיפה; מיזוג של אוכלוסיות אזרחיות וקהילות גולות עם המנגנונים והפעולות של ממשלות וארגוני טרור; אסימטריה של הכוח המודרני, שבמסגרתה למדינה הישראלית יש עוצמה מנהלית, פיננסית וצבאית רבה יותר לעומת המוסדות

[4] BSR משתמשת במונח הגנרי "מדיניות תוכן" כדי לכלול את כללי הקהילה של Facebook ואת הנחיות הקהילה של Instagram.

הפוליטיים הפלסטיניים; והשלכות גלובליות הן של התפיסה והן של היחס לקהילות יהודיות ולפלסטינים גולים מחוץ לאזור.

- ההסלמה בסכסוך הישראלי-פלסטיני המתמשך התרחשה בחודש מאי 2021, אשר הוצתה על ידי מחאות במזרח ירושלים נגד פינוי משפחות פלסטיניות בשכונת שייח' ג'ראח. התפרצות זו אירעה בהקשר של הכיבוש הישראלי בגדה המערבית והמתחים הגוברים ביחס להרחבת ההתנחלויות הישראליות והפינוי של קהילות פלסטיניות.[5] האלימות כללה מחאות, מהומות ואכיפה משטרתית כלפי מהומות בתוך ישראל, מתקפות טילים חסרות הבחנה של חמאס על ישראל, ומתקפות אוויריות של ישראל על רצועת עזה. האלימות עוררה מחאות ברחבי העולם, והסתיימה ברובה ב-21 במאי לאחר הסכם הפסקת אש בין ישראל לחמאס שנחתם יום קודם לכן.[6]

- אנשים מכל צדי הסכסוך, הן ברמה האזורית והן ברמה הגלובלית, סבלו ורואים בעצם קורבנות של אירועים היסטוריים שונים. תפקידה של Meta הוא לא לשמש כבוררת בסכסוך זה, אלא לגבש ולאכוף מדיניות להפחתת הסיכון שהפלטפורמות שלה יחמירו את הסכסוך על ידי השתקת קולות, הגברת פערי כוחות, או מתן אפשרות להפצת תוכן שמסית לאלימות.

- ה-UNGPs מתווים את הציפייה מ-Meta להימנע מהפרת זכויות אדם של אחרים ולהידרש להשפעות שליליות על זכויות אדם (כהגדרתן בה"ש 3) שבהן היא מעורבת. בהקשר המושפע מסכסוך, כמו ישראל ופלסטין, הדבר כולל הבנה כיצד מצטלבת דרכו של הסכסוך המתמשך עם הפלטפורמות של Meta (לדוגמה, דרך נרטיבים מתחרים על ההיסטוריה ועל אירועים אקטואליים כאחד), כיצד פעולותיהם המקוונת של מגוון שחקנים, לרבות Meta, עשויות לעצב אירועים מחוץ לזירה המקוונת (לדוגמה, באמצעות דברי שטנה והסתה לאלימות שמקדמת פגיעה מחוץ לרשת), ואילו קבוצות פגיעות במיוחד לאור ההקשר הזה.

להלן יובאו אירועים ואבני דרך עיקריים במהלך תקופה זו:

ההקשר הרחב	Meta
אמצע אפריל: סכסוך על בתים בשייח' ג'ראח, פינויים, ומחאות	23 באפריל: נקבע כי ירושלים הוא "מיקום בסיכון גבוה באופן זמני".[7]
6 במאי: עליית המתיחות בגדה המערבית ובמזרח ירושלים עקב מותם של שני פלסטינים	5-6 במאי: תקלה טכנית גלובלית מונעת פרסום "סטוריי" ("story") ב-Instagram המכילים שיתופים חוזרים של פוסטים (לרבות, בין היתר, בישראל ובפלסטין) מדווחת ומטופלת
7 במאי: משטרת ישראל נכנסת למסגד אל-אקצא במהלך תפילות	

[5] מדיניותה של ישראל ליישוב אזרחים בשטחים הפלסטיניים הכבושים ועקירת קהילות פלסטיניות הוכרזה כהפרה של החוק הבינלאומי על ידי מספר גופים של האו"ם.

[6] ראו Israel and Hamas Begin Cease-Fire in Gaza Conflict, Israel-Gaza ceasefire holds despite Jerusalem clash, Timeline of the Israeli—Palestinian conflict in 2021, ותיאור של המשבר הישראלי-פלסטיני 2021.

[7] "מיקום בסיכון גבוה באופן זמני" ("temporary high-risk location") הוא ציון שבו משתמשת Meta ביישומה של כללי מדיניות תוכן מחמירים מסוימים על אזור גיאוגרפי מוגדר, ובפרט הגבלת קריאות להכנסת חימוש למקומות שיש בהם סימנים זמניים לסיכון מוגבר לאלימות או לפגיעה ממשית (מחוץ למרחב המקוון), כמו מחאה ידועה, מחאת-נגד, או אלימות שאירעה לאחרונה. בנוסף, המדיניות של Facebook ביחס לאיומים איומים מוסווים, המשתרעת על התבטאויות מוצפנות שבהן שיטת האלימות או הפגיעה אינה מבוטאת בצורה ברורה, ואשר דורשת הבנה משמעותית של ההקשר לשם ביצוע פעולות אכיפה, מיושמת באופן מחמיר יותר ב"מיקום בסיכון גבוה זמני".

ההקשר הרחב	Meta
8 במאי: התנגשויות בין מפגינים למשטרת ישראל 10 במאי ואילך: הסלמה נוספת, הכוללת הפגנות, אלימות, ומתקפות טילים 16 במאי: בניין אל ג'זירה בעזה מותקף במתקפה אווירית של ישראל 20 במאי: ישראל וחמאס מסכימות על הפסקת אש	11 במאי: מתקבל דיווח על חסימת התגית ("hash-tag") אל-אקצא, ומתחילה בחינה של הגורם לחסימה; Meta מפרסמת את הגורם לשגגה ב-12 במאי 12 במאי: מופעל צוות ייעודי המופקד על תגובה למשבר 13 במאי: הסטטוס של 'מיקום בסיכון גבוה זמני' מורחב לכל ישראל, הגדה המערבית ועזה 14-20 במאי: מתקבלים מכתבים מארגוני חברה אזרחית המביעים דאגה לגבי הגבלות על חופש הדיבור הפלסטיני ועלייה בתוכן אנטישמי, ותוכן התורם לאלימות בכל הצדדים 21 במאי: עיתונאים פלסטינים מדווחים על חסימת חשבונות ה-WhatsApp שלהם 27 במאי: הצוות הייעודי שהופקד על תגובה למשבר נסגר 24 ביוני: Meta מפרסמת מידע חדש על האופן שבו דברי שבח, תמיכה וייצוג מוגדרים במדיניות החברה לגבי אנשים וארגונים מסוכנים – DOI (Dangerous Individuals and Organizations), וכיצד יש לפרשם

5. ניתוח

<u>פעולות בהלימה עם ה-UNGP</u>

- Meta נקטה פעולות ראויות רבות במהלך המשבר של מאי 2021, כולל הקמת מרכז תפעול/צוות תגובה מיוחד, תיעדוף סיכונים לפגיעה ממשית מיידית בעולם הלא-מקוון, גיבוש גישה לנראות והסרת תוכן המבוססת על הגבלות הכרחיות ומידתיות שעולות בקנה אחד עם סעיף 19(3) לאמנה הבין-לאומית בדבר זכויות אזרחיות ומדיניות – ICCPR (International Covenant on Civil and Political Rights), ותיקון טעויות אכיפה בתגובה לערעורים של משתמשים. חלק מההמלצות של BSR ל-Meta מתבססות על יסודות חשובים לגישת ניהול תוכן (content governance) המכוונת זכויות אדם, דוגמת אלה שכבר בוססו על ידי Meta.

<u>אתגרים ותובנות</u>

- מהראיונות שנערכו ומהנתונים שנסקרו, עולה כי על גבי הפלטפורמות של Meta פורסמו דברי שטנה והסתה לאלימות נגד פלסטינים, ערבים-ישראלים, יהודים-ישראלים, וקהילות יהודיות מחוץ לאזור.

- על סמך המידע שנסקר, בחינה של מקרים פרטניים וחומרים קשורים, ויצירת קשר עם בעלי עניין חיצוניים, דומה כי הפעולות של Meta [8] במאי 2021 השפיעו באופן שלילי על זכויות האדם (כהגדרתן בה"ש 3) של משתמשים פלסטינים לחופש ביטוי, חופש הפגנה, השתתפות פוליטית ואי-אפליה, ומכאן על יכולתם של פלסטינים לשתף מידע ותובנות על חוויותיהם בזמן התרחשותן. דברים אלה באו לידי ביטוי בשיחות עם בעלי עניין שהושפעו, שרבים מהם שיתפו עם BSR את דעתם כי Meta נדמית כישות עוצמתית נוספת אשר מדכאת את קולם ואין ביכולתם לשנותה.

- הנתונים שנסקרו, בחינה של מקרים פרטניים וחומרים קשורים וראיונות עם בעלי עניין פנימיים וחיצוניים, הצביעו כולם הן על אכיפת-יתר (הסרת תוכן שגויה והטלה שגויה של סנקציות על חשבונות) והן על אכיפת-חסר (אי-הסרה של תוכן מפר ואי-הטלת סנקציות על חשבונות שביצעו עבירות) של מדיניות התוכן של Meta במהלך מאי 2021, ובפרט מדיניותה ביחס לאנשים וארגונים מסוכנים – DOI (Dangerous Individuals and Organizations) ומדיניותה בדבר אלימות והסתה – V&I (Violence and Incitement).[9]

- הנתונים שנבדקו העלו כי תוכן בשפה הערבית היה נתון לאכיפת-יתר (לדוגמה, הסרה מוטעית של תוכן פלסטיני) במידה רבה יותר פר-משתמש (כלומר, בהתאמה להבדל בגודל האוכלוסייה בין דוברי ערבית לדוברי עברית בישראל ובפלסטין). הנתונים שנסקרו על ידי BSR גם הראו כי שיעורי הניטור היזום של תוכן בערבית העלול להימצא מפר היו גבוהים במובהק משיעורי הניטור היזום של תוכן בעברית העלול להימצא מפר, דבר שניתן קרוב לוודאי לייחס בחלקו הגדול למדיניותה של Meta המשלבת חובות חוקיות מסוימות הקשורות בארגוני טרור זרים ייעודיים, והעובדה שהיה קיים מסווג (classifier) ביטוי עוין בערבית אך לא מסווג ביטוי עוין בעברית. על סמך אינדיקציות בנתונים ובחינת מקרים פרטניים, נמצאה גם אכיפת-חסר (לדוגמה, במקרים של הסתה לאלימות נגד ישראלים, דברי שבח לחמאס, כולל על ידי רשויות פוליטיות פלסטיניות), אם כי BSR מציין כי אכיפת-חסר הינה מאתגרת יותר למדידה מאשר אכיפת-יתר.

- אין ל-Meta מדדים ספציפיים לאכיפת-חסר מכיוון שמדידה חד-משמעית היא מאתגרת. עם זאת, החומרים שנבחנו וראיונות פנימיים הצביעו על כך שתוכן בעברית היה נתון לאכיפת-חסר רבה יותר, במידה רבה עקב היעדר מסווג בעברית ועזיבה של עובדים דוברי עברית המועסקים במשרה מלאה ובמיקור חוץ של בקרת תוכן בשבועות שקדמו למאי 2021. ואולם, בדיקה של מקרים פרטניים הראתה כי התרחשה גם אכיפת-יתר, והתקיימו הסרת תוכן שגויות מחשבונות ישראליים והגבלות הוטלו עליהם בשגגה.

---

[8] בהתאם לעיקרון 13 של ה-UNGP, BSR מפרש "פעולות" כ"פעולות ומחדלים".

[9] כאמור לעיל, שיטת העבודה של BSR לא ניסתה לזהות מה צריכים להיות השיעורים המוחלטים או היחסיים של אכיפת תוכן בישראל ובפלסטין. הביטויים אכיפת-יתר ואכיפת-חסר משמעם אפוא רק כי זוהו מקרים של הסרה שגויה של תוכן, הטלה שגויה של סנקציות על חשבונות, אי-הסרה של תוכן, או אי-הטלת של סנקציות. BSR לא בדק ולא הסיק מסקנות ביחס לשאלה האם האכיפה של Meta עמדה בחובות כלשהן לפי כל דין.

- על סמך החומרים שנסקרו, BSR זיהתה את סיבות השורש האפשריות הבאות לאכיפת-יתר, שעל Meta להמשיך ולחקור:
  - ייתכן כי מסווגים[10] בשפה העברית נוטים לשיעורי טעות גבוהים יותר ביחס לערבית-פלסטינית.
  - ייתכן כי תוכן מפר פוטנציאלית בשפה הערבית לא נותב לבודקי תוכן שמדברים או מבינים את הניב הספציפי של התוכן.

- עלייה מהותית בכמות המקרים במהלך מאי 2021 (עד פי 10 בימי השיא כפי שדווח על ידי צוותי השווקים ל-BSR והשתקף בנתונים שנבדקו על ידי BSR) עוררה אתגרים משמעותיים לאכיפה אפקטיבית של מדיניות התוכן במהלך המשבר. לדברי בעלי עניין פנימיים, Meta לא העסיקה מספיק בודקי תוכן דוברי ערבית ועברית כדי לטפל בעלייה החדה.

- על סמך בדיקת BSR של פניות (tickets) ומשוב מבעלי עניין פנימיים, מתברר כי התפתחה בעיה משמעותית של אכיפת-יתר במאי 2021, כאשר משתמשים צברו פסילות "כוזבות" שהשפיעו על הנראות והמעורבות (engagement), לאחר שפוסטים הוסרו בשגגה בגין הפרת מדיניות התוכן.[11] ההשפעות על זכויות אדם (כהגדרתן בה"ש 3) כתוצאה משגגות אלה היו חמורות יותר על רקע ההקשר, שבו נודעה חשיבות מוגברת לזכויות כמו חופש הביטוי, חופש ההפגנה וביטחון, במיוחד עבור אקטיביסטים ועיתונאים, ועל רקע הבולטות של הפרות חמורות יותר של מדיניות ה-DOI. בנוסף, פסילות אלה נותרו על כנן עבור אותם משתמשים שלא ערערו על הסרות התוכן השגויות.

- החומרים שנסקרו ומשוב מבעלי עניין פנימיים חשפו היעדר פיקוח ב-Meta, שאפשר טעויות באכיפת מדיניות התוכן שבצדן השלכות משמעותיות. אחת מהדוגמאות המרכזיות שנמסרה ל-BSR במהלך משבר זה הייתה שעובד בשירותי מיקור החוץ של Meta הוסיף את #AlAqsa לרשימת התגיות החסומות, לאחר שלוקט מתוך רשימה מעודכנת של ביטויים ממחלקת האוצר של ארה"ב שכללה את 'גדודי אל-אקצא', והתוצאה הייתה שהתגית #AlAqsa הוסתרה מתוצאות החיפוש. למעשה, בתגית #AlAqsa נעשה שימוש רחב בפוסטים שהתייחסו למסגד אל-אקצא, שהוא אחד מהאתרים הקדושים ביותר לאיסלאם.

- בעלי עניין רבים דיווחו ל-BSR כי משתמשים דיווחו שהם חווים צמצום של תפוצת תוכן במהלך תקופת המשבר. על אף שהערכות בדבר נראות תוכן הן מורכבות, BSR דן בשאלה זו עם בעלי עניין פנימיים וזיהה את הגורמים האפשריים הבאים שעשויים היו לתרום לכך: (1) עונשי אכיפה על הפרות לכאורה של מדיניות התוכן שהשפיעו על יכולת החיפוש או הנראות של התוכן; (2) Meta יישמה מספר אמצעי "שבירת זכוכית" ניטראליים-לתוכן עבור ישראל ופלסטין במהלך מאי 2021 שנועדו להפחית את הסיכון של הסטת פגיעה מהרשת לעבר המרחב הממשי

---

[10] מסווג הוא אלגוריתם שמזהה וממיין תוכן לסוגים של תוכן – לדוגמה, זיהוי אקטיבי של תוכן הכרוך שבסבירות גבוהה מפר את אחד מכללי המדיניות של Meta.

[11] בגישתה הכללית של Meta לאכיפת מדיניות תוכן, הפרות הנחשבות חמורות יותר (כמו הפרות של מדיניות DOI) גורמות להגבלות ממושכות יותר או להגבלות נוספות, כמו הגבלות על יצירת פרסומות, ואילו הפרות הנחשבות פחות חמורות (כמו הסתה לאלימות, דברי שטנה, או בריונות והטרדה) כרוכות בפחות הגבלות, כמו צמצום היכולת לחפש את החשבון (כלומר, המשתמשים נדרשים להקליד את השם המדויק של חשבון המשתמש כדי למצוא אותו, במקום חיפוש רגיל של מילות מפתח) או צמצום הנראות של התוכן (כלומר, מיקום התוכן במקום נמוך יותר בפיד). Meta מודיעה למשתמשים כאשר היא משפיעה על יכולת החיפוש אחריהם, אך לא כאשר היא מצמצמת את הנראות של התוכן. כאשר ערעור של משתמשים על הסרת תוכן מתקבל ו-Meta הופכת את החלטתה, הפסילות והעונשים הקשורים בכך מוסרים.

(online-to-offline), שהפחיתו במכוון את הנראות של כל התוכן שקיבל שיתופים חוזרים ונשנים; (3) משתמשים חוו שתי תקלות ב-Instagram שהשפיעו על הנראות הגלובלית של סטוריז (stories) ב-5-6 במאי.

- המשבר של מאי 2021 הביא אל קדמת הבמה שאלות סביב קווי המתאר ותוכנם של דברי שבח והאדרה של אלימות – במנותק ממדיניות Meta לעניין אלימות והסתה ומכללי מדיניות דומים, אשר אוסרים על "שפה המסיתה או מביאה לאלימות חמורה". בפרט, Meta צריכה לשקול האם המדיניות שלה נדרשת באופן מספק לדברי שבח והאדרה של אלימות חסרת הבחנה, דהיינו אלימות שאינה מכוונת לאדם או קבוצה מסוימים.

- בעלי עניין העלו חששות לגבי תוכן אנטישמי בפלטפורמות של Meta. תוכן אנטישמי הוא סוג של דבר שטנה, ועל כן הוא נופל על פי רוב בגדרי מדיניות דברי השטנה של Meta עבור Instagram ו-Facebook – אולם, מדיניות זו חלה על כלל הסוגים של דברי שטנה, ואינה מתווה בבירור את ההבחנה בין קטגוריות שונות של דברי שטנה, ואינה כוללת הגדרה מלאה שלהם. בעת הנוכחית, Meta אינה עוקבת (כלומר, מתייגת או סופרת) סוגים ויעדים ספציפיים של דברי שטנה, אלא רק את ה"דרגה" ("tier") שלהם, ועל כן אין לה מדדים המאפשרים להבין את שכיחותו של תוכן אנטישמי – לרבות האם חלה עלייה בהימצאותו במאי 2021, אם לאו. אף ש-BSR לא יכולה היתה לאמת באופן עצמאי את הנתונים, ארגונים שמקדישים את פעילותם לאנטישמיות מצאו תוכן אנטישמי שהפר את מדיניות Meta באופן מובהק, שלא אותר ולא הוסר במהלך תקופה זו. הניתוח של BSR מציע כי לא הייתה כשירות תרבותית מספקת מצדם של בקרי התוכן, וכי בקרב קובעי המדיניות הייתה יכולת לשונית בלתי מספקת בטווח השפות (כולל שפות אירופיות קטנות) שבהן הופיע תוכן אנטישמי.[12]

- בעלי עניין חיצוניים שרואיינו על ידי BSR דיווחו על מקרים שבהם נעשה שימוש ב-WhatsApp על ידי ישראלים מהאגף הימני בישראל כדי להסית לאלימות ולתאם מתקפות נגד ערבים ויהודים-ישראלים כאחד, כמו גם נגד עיתונאים ישראלים.

- היו עיתונאים ואקדמאים שחשבונות ה-WhatsApp שלהם הושבתו בשגגה כתוצאה מפעילות אכיפה נכונה נגד קבוצות של ארגוני טרור מוכרזים בגין הפרה של מדיניות WhatsApp. כאשר הובא לידיעתה של WhatsApp כי חשבונות אלה הושבתו בטעות, הם הושבו לשירות.

- בעלי עניין דיווחו ל-BSR כי לדעתם מדיניות ה-DOI ומדיניות ה-V&I של Meta אינן מובנות כראוי על ידי המשתמשים – לדוגמה, ייתכן שמשתמשים מתקשים להבין מה הם דברי שבח לארגוני טרור ומהי הסתה לאלימות.

- גורמים אלה מצביעים כולם על תמות רחבות שמוכרות בתחום של ניהול תוכן בפלטפורמות מדיה חברתית, כמו החשיבות של הבנת ההקשר הגיאוגרפי וניטור האופן בו הוא משתנה, הצורך בכמות מספקת של עובדים בעלי כישורי שפה רלוונטיים והבנה תרבותית, והמתח המתמשך בין הזכות לחופש הביטוי לבין המאמצים לשיפור הביטחון ולמניעת נזק פיזי בפלטפורמה ומחוצה לה. לשם התמודדות עם אתגרים אלה, תהיה חשיבות מרכזית לכך שיעמדו לרשות Meta אמצעים לבקרת תוכן בשפות הרלוונטיות ולווידוא שהן מדיניותה והן ההנחיות לאכיפתה מבוססות על מומחיות אזורית המשקפת פרספקטיבות מגוונות. BSR מדגישה, כי סביר שרבים

[12] ראו, לדוגמה, דיווחים של הקונגרס היהודי העולמי, Fighting Online Antisemitism, ו-CST.

מהגורמים שזיהתה שכיחים בתעשיית המדיה החברתית כולה ובאזורים אחרים מוכי סכסוך, וכי גישות כלל-ענפיות ודיאלוג בין ריבוי בעלי עניין יועילו לטיפול באתגרים אלה.

- על פי כל המקורות שנבחנו כחלק מסקירת בדיקת נאותות זו ביחס לזכויות אדם, ההשפעות השליליות העיקריות על זכויות אדם (כהגדרתן בה"ש 3) כללו את חופש הביטוי (לדוגמה, כתוצאה מאכיפת-היתר של מדיניות התוכן), חופש ההפגנה וההתאספות (לדוגמה, הפחתת היכולת להתארגן ולהתאגד באופן מקוון), חופש מהסתה (לדוגמה, אכיפת-חסר של תוכן המסית לאלימות), ביטחון הגוף (לדוגמה, אכיפת-חסר של תוכן שנועד לארגן אלימות), אי-אפליה (לדוגמה, השפעות שונות על דוברי ערבית), וגישה לסעדים (לדוגמה, איבוד גישה לתוכן ש-Meta אינה נתונה לכל חובה חוקית לשמרו, אך יכול לסייע לבעלי זכויות בתהליכים עתידיים).[13]

<u>הטיה</u>

- מועצת הפיקוח ביקשה לקבוע האם בקרת התוכן של Meta בשפה הערבית ובשפה העברית, ובכלל זאת השימוש שלה באוטומציה, יושמה ללא הטיה. מועצת הפיקוח לא הגדירה "הטיה" ("bias") בהמלצותיה, ולא סיפקה מסגרת כללים לכך.

- לצורך הערכה זו, BSR מתייחסת הן להטיה מכוונת (שבמסגרתה אנשים מסוימים מקבלים במכוון יחס שונה מאחרים) והן להטיה בלתי מכוונת (שבמסגרתה מדיניות ותהליכים עשויים להיות ניטראליים על פניהם או מוקמים מטעמים של ציות לחוק, אך הם משפיעים על אנשים מסוימים באופן שונה מאחרים). BSR גם שקלה את ההבחנה בין הטיה ברמת מדיניות התוכן, להטיה ברמת מערכת בקרת התוכן.

- מתוך הראיונות שנערכו על ידי BSR והנתונים שנבחנו, BSR לא זיהתה הטיה מכוונת אצל Meta ככלל או בקרב עובדים ספציפיים. BSR לא מצאה ראיות לקיומה של פעילות מכוונת על רקע גזע, אתניות, לאום או דת בצוותים האחראים ומציינת כי קיים ב-Meta גיוון של עובדים המייצגים מגוון של נקודות מבט, לאומים, גזעים, קבוצות אתניות ודתות הרלוונטיים לסכסוך זה. כמו כן, BSR לא זיהתה ראיות לכך שבפיתוח או ביישום מדיניותה, Meta ביקשה במכוון להיטיב עם או להזיק לקבוצה מסוימת כלשהי בשל הגזע, הדת, הלאום, המוצא האתני שלה, או כל מאפיין מוגן אחר.

- עם זאת, BSR כן זיהתה מקרים שונים של הטיה בלתי מכוונת שבהם המדיניות והפרקטיקות של Meta, בשילוב עם דינמיקה חיצונית רחבה יותר, כן הובילו להשפעות שונות על זכויות אדם של משתמשים פלסטינים ודוברי ערבית.

- כחברה אמריקאית, Meta מחויבת לציית לחוקי ארה"ב, ובכללם אלה הנוגעים למתן "תמיכה ממשית" ("material support") או משאבים לארגוני טרור זרים מוכרזים, כאשר "ארגון טרור זר" ("foreign terrorist organization") הוא ארגון זר שהוכרז על ידי מזכיר המדינה של ארה"ב בהתאם לסעיף 219 לחוק ההגירה והאזרחות (Immigration and Nationality Act) (18 U.S.C. §2339B). להכרזות משפטיות על ארגוני טרור ברחבי העולם קיים מיקוד בלתי מידתי ביחידים וארגונים שזוהו כמוסלמים, ולפיכך קיימת סבירות גבוהה יותר

[13] לדוגמה, ראו: UN Berkeley School of Law, Digital Lockers, June 2021.

שמדיניות ה-DOI של Meta והרשימה ישפיעו על משתמשים פלסטינים ודוברי ערבית, הן על סמך פרשנותה של Meta את החובות החוקיות החלות עליה, והן בשגגה.[14]

- הסבירות שפלסטינים יפרו את מדיניות ה-DOI של Meta גבוהה יותר בגלל נוכחות של החמאס כישות שלטונית בעזה ושל מועמדים פוליטיים הקשורים לארגונים מוכרזים. הפרות של ה-DOI מלוות גם בעונשים חמורים במיוחד, שמשמעם כי סביר יותר שפלסטינים יעמדו בפני השלכות חמורות יותר הן של אכיפה נכונה והן של אכיפה שגויה של מדיניות. בניגוד לישראלים ולאחרים, פלסטינים מנועים מלשתף סוגים מסוימים של תוכן פוליטי מכיוון שמדיניות ה-DOI של Meta אינה מחריגה דברי שבח לישויות מוכרזות בכובען השלטוני.

- כמו כן, ייתכן כי הסטטוס המנוגד של הערבית בהשוואה לעברית במערכת בקרת התוכן של Meta גרם להטיה בלתי מכוונת באמצעות אכיפת-יתר רבה יותר של תוכן בערבית בהשוואה לתוכן בעברית, אפילו לאחר התאמה לגודל האוכלוסייה. (יצוין כי זה לא מסביר הבדלים פוטנציאליים בשיעורי ההפרות של תוכן בעברית לעומת תוכן בערבית, שעבורם לא היו נתונים).

- BSR מציינת את הגורמים הבאים: (1) ניתוב בלתי מספק אפשרי של תוכן בערבית לפי ניב או מומחיות אזורית, שעשוי היה להבטיח שהבודקים החיצוניים מבינים בבירור את הניב של התוכן שהם בוחנים, וכן את ההקשר התרבותי שעשויה להיות לו חשיבות כבסיס להחלטות מושכלות; (2) השימוש במסווגים עבור תוכן בערבית, בשעה שלא היה מסווג מתפקד עבור תוכן בעברית; וכן (3) המסווגים לערבית כנראה פחות מדויקים עבור ערבית פלסטינית מאשר עבור ניבים אחרים, הן מפני שהניב פחות נפוץ, והן מכיוון שנתוני אימון התוכנה – המבוססים על הערכות של בודקים אנושיים – ככל הנראה משכפלים את הטעויות של הבודקים האנושיים בשל היעדר כשירות לשונית ותרבותית. להבדיל, עברית היא שפה סטנדרטית יותר והמדוברת ביותר בישראל, ולכן בודקי התוכן בעברית רהוטים בשפה וגם סביר יותר שהם מבינים את ההקשר. כתוצאה מגורמים אלה, ייתכן שמערכת בקרת התוכן אינה מדויקת עבור תוכן בערבית כפי שהיא עבור תוכן בעברית.

6. המלצות

BSR דנה בהבחנות ובנקודות לבדיקה עם Meta בהלימה עם האחריות המוטלת עליה לפי ה-UNGP לנקוט בפעולה הולמת לטיפול בהשפעות השליליות האמורות על זכויות אדם. המלצותינו כוללות:

1. בחינה האם על Meta לגבש אמצעי מדיניות לגבי תוכן שמשבח או מאדיר אלימות (כולל התקפות חסרות הבחנה, כמו אלימות שאינה מכוונת לאדם או קבוצה מסוימים).
2. בחינה האם על Meta להגביל את מדיניות ה-DOI ל"תמיכה" או "ייצוג" בלבד.
3. בחינת הפרקטיקה של סימון דמויות היסטוריות שנפטרו תחת מדיניות ה-DOI והערכת ההיתכנות של גישות מדיניות חלופיות לשיפור השקיפות וההוגנות.

---

[14] נציין כי ארגוני חברה אזרחית העלו בפני BSR מקרים שבהם קבוצות קיצון אלימות ישראליות לא נוספו לרשימת ה-DOI, כמו גם מספר אנשים שביצעו מעשי טרור בעבר, מה שמעלה שאלות באשר למתודולוגיה של הוספת קבוצות ויחידים לרשימה. אף שאין משתמע מתובנה זו שצריך להיות מספר שווה של קבוצות פלסטיניות וישראליות מוכרזות, העובדה שקבוצות ויחידים ישראלים ידועים אינם נכללים עלולה להוביל לאפליה.

4. ריבוד מערכת הסימון והפסילות עבור הפרות ה-DOI, כדי להביא בחשבון את זהותו של הארגון או האדם ואת מהות ההפרה (שבח, תמיכה, או ייצוג) כך שהפסילה תהיה מידתית להפרה.

5. מתן נימוק ספציפי ומפורט יותר למשתמשים בקשר עם הרציונאל שבבסיס המדיניות כאשר מוטלות פסילות. עליו לכלול לא רק את קטגוריית ההפרה, אלא באיזה אופן הפרסום מפר, על מנת שהמשתמשים יוכלו להבין טוב יותר את ההצדקה, להגיש ערעור מושכל, ולהקטין את הסיכוי שיפרסמו תוכן מפר בעתיד.

6. הגברת השקיפות לגבי פעולות האכיפה של Meta – כמו הגבלת מאפיינים והגבלת חיפוש – ותקשור פעולות האכיפה באופן ברור למשתמשים.

7. פרסום האלמנטים העיקריים של משאבי תפעול הקהילה הפנימיים של Meta המסייעים לאחראי בקרת התוכן לפרש וליישם את מדיניות התוכן של Meta, על מנת שמשתמשים יוכלו להבין טוב יותר את המדיניות ולציית לה, ולהשמיט תוכן במחלוקת.

8. קביעת הרכב השוק הנדרש (לדוגמה, מצבת כוח אדם, שפה, מיקום) עבור מצבי כוננות או יכולות תגובה מהירה עבור השווקים של עברית וערבית.

9. המשך ביסוס המנגנונים לניתוב משופר של תוכן בערבית העלול להימצא בהפרה לפי ניב/אזור.

10. הערכה האם ישים ורצוי לגבש מסווג ערבית ספציפי עבור ניב, תוך עבודה בשותפות עם בלשנים לערבית ומומחים במודלים של השפה.

11. המשך העבודה על הפעלת מסווגים מתפקדים בשפה העברית.

12. שינוי התהליך המאפשר לעובדי ספקים חיצוניים להוסיף מילות מפתח לרשימות החסימה, כדי לוודא שאלה יאושרו על ידי עובדים-במשרה-מלאה רלוונטיים של Facebook.

13. פיתוח תהליך בדיקה/פיקוח מסוג בקרת איכות על הוספת פריטים חדשים לרשימות החסימה של האשטאגים/מילות מפתח.

14. המשך התכנית לחשיפה של מספר הדיווחים הרשמיים המתקבלים מגופי ממשל (כולל פרקליטות המדינה בישראל) על אודות תוכן שאיננו בלתי חוקי, אך עלול להפר את מדיניות התוכן של Meta. הדבר צריך להתבצע בתדירות רבעונית (כחלק מדו״ח אכיפת כללי הקהילה) או אחת לשישה חודשים (כחלק מהדו״ח על הגבלות תוכן).

15. הערכת הדיוק באכיפת מדיניות ה-DOI בערבית בקרב צוותים פנימיים וחיצוניים כאחד, לרבות ביחס למנגנונים מבוססי מכונה ואנושיים גם יחד, וטיפול בממצאים (BSR מציינת כי זהו מאמץ שוטף).

16. פיתוח מנגנון למעקב אחר שכיחות של תוכן המכיל מתקפות על בסיס מאפיינים מוגנים ספציפיים (לדוגמה, תוכן אנטישמי, איסלאמופובי, אנטי-ערבי, הומופובי). הדבר יכול לכלול, למשל, הנחיות למשתמשים לסמן תוכן שטנה רלוונטי באמצעות תיוג.

17. קביעה של מבנה, פרוטוקול או צוות שתפקידו למדוד אכיפת-יתר ואכיפת-חסר של מדיניות התוכן באופן שיטתי במהלך משבר.

18. הרחבת הקיבולת של ערוצי ההסלמה המיוחדים של Meta באמצעות הגדלה של כוח אדם והקצאת משאבים שיאפשרו תגובה מהירה דיה לדיווחים דחופים שהובאו על ידי שותפים מהימנים, מגופי ממשל ומשחקנים אחרים, בזמנים רגילים ובעתות משבר כאחד.

19. קידום מעורבותם של בעלי עניין והכנת הצהרות שקיפות ציבוריות העוסקות בהבנה של Meta לגבי חובותיה בכל הנוגע לארגון טרור זר (FTO) ולטרוריסט גלובלי המוכרז באופן מיוחד (SDGT).

20. מימון מחקר ציבורי אודות היחס האופטימלי בין חובות המאבק בטרור הנדרשות מכוח חוק לבין המדיניות והפרקטיקות של פלטפורמות מדיה חברתית. הדבר יידרש לשאלות כגון כיצד יש לפרש את התפיסה של תמיכה ממשית בטרור בהקשר של המדיה החברתית, והאם ממשלות צריכות לייחד חקיקה או פרשנות עבור חברות מדיה חברתית.

21. בנפרד מהמידע הקיים ומדיניות אכיפת החוק, פיתוח שיטות חדשות או כללי מדיניות שיאפשרו ל-Meta לאחסן תוכן במקרים שבהם לא חלה על Meta כל חובה חוקית לשמרו, אך היכן שעשוי להיות לתוכן שימוש פוטנציאלי בידיו של גורם בעל זכויות בעתירות עתידיות למתן סעד.

הצהרת פטור

המסקנות המוצגות במסמך זה מייצגות את השיפוט המקצועי המיטבי של BSR, על סמך המידע שעמד לרשותו והתנאים שהתקיימו במועד הבחינה.

בביצוע משימתו, BSR הסתמך על מידע זמין לציבור, על מידע שסופק על ידי Meta, ועל מידע שסופק על ידי צדדים שלישיים. בהתאם לכך, המסקנות במסמך זה תקפות רק במידה שהמידע שנמסר או שהיה זמין ל-BSR היה מדויק ומלא, והחוזק או הדיוק של המסקנות עשויים להיות מושפעים מעובדות, מנתונים ומהקשרים שלא היו ידועים ל-BSR.

כך, אין להתייחס לעובדות או למסקנות האמורות במסמך זה כאל ביקורת (audit), אישור (certification), או כל סוג של הסמכה (qualification). מסמך זה אינו מהווה ואין להסתמך עליו כייעוץ משפטי מכל סוג, ואין לראות בו סקירה ממצה של הציות לחוק או לרגולציה.

BSR אינו נותן כל מצגים או התחייבויות, במפורש או מכללא, אודות העסק או פעולותיו. מדיניותו של BSR היא שלא לפעול כנציג של חבריו, והוא אינו מביע את תמיכתו במדיניות או סטנדרטים ספציפיים. הדעות המובעות במסמך זה אינן משקפות את אלה של העמיתים החברים ב-BSR.

אודות BSR

BSR הוא רשת עמיתים עסקית לייעוץ מוכוון קיימות לעסקים המתמקדת ביצירת עולם שבו כל בני האדם יכולים לשגשג בעולם בריא. באמצעות משרדיה באסיה, אירופה וצפון אמריקה, BSR מספקת ליותר מ 300-החברות העמיתות בה תובנות, עצות ויוזמות שיתופיות כדי לסייע להן לראות את העולם המשתנה ביתר בהירות, ליצור ערך ארוך-טווח, ולהרחיב השפעה.