

פתרונות ודרכי התמודדות בעולם

מדינות, גופי תקשורת, חברות ואנשים פרטיים מתמודדים בשנים האחרונות עם האיום שמציבות זמינות ונגישותן ההולכות וגוברות של יכולות דיגיטליות. בפרק זה סקירת פתרונות ודרכי התמודדות מן העולם, הנחלקים לשלוש קטגוריות: מאמצי רגולציה, חקיקה ואכיפה; פתרונות טכנולוגיים; מאמצים לחינוך הציבור.

רגולציה, חקיקה ואכיפה

ככל שהשימוש בדיגיטליזציה ובדיגיטליזציה מתעצם, כך גוברים המאמצים לפתח כלים ושיטות לסיכול האיום. סוכנויות מודיעין ורשויות ממשל במדינות דמוקרטיות בעולם משתתפות באופן פעיל במאמץ זה על ידי הקצאת תקציבים לפיתוח הטכנולוגיות לזיהוי תוכן מזוין ולהתמודדות עם התופעה (Parkin, 2019). לצד מתן תקציבים למציאת פתרונות טכנולוגיים, רשויות ממשל במדינות דמוקרטיות ברחבי העולם מגלות גם מעורבות משפטית בנושא. יצוין כי תהליכים קשורים מתרחשים גם במדינות לא-דמוקרטיות שבהן עשויה להיות השפעה לדיגיטליזציה, ביניהן סין, אולם המידע בנוגע לאלה גלוי פחות, והאתגרים שונים במדינות שבהן התקשורת מוגבלת.

האתגר המשפטי והיישומי של התמודדות עם סכנות הפייק ניוז והדיגיטליזציה בדמוקרטיות בפרט נובע מקצב ההתפשטות הגבוה, הנרחב והמגוון שלהן ומהיכולת המוגבלת לזהותן ולהגבילן באמצעים טכנולוגיים. לאחרונה ניתנה בישראל אף הגדרה משפטית לדיגיטליזציה, הממחישה את הצורך של המשפט להתייחס אליה ולסכנותיה בצורה ייחודית. היא ניתנה במסגרת פסיקה של שופט בית המשפט העליון עוזי פוגלמן בנוגע לעתירה שהוגשה לוועדת הבחירות המרכזית לכנסת ה-24 ומגדירה דיגיטליזציה כך:

טכנולוגיה ליצירת תוכן קולי או חזותי או לשינוי תוכן קיים, כך שהצופה הסביר (ואף הצופה המתוחכם) יסבור כי פלוני ביצע פעולה או העביר מסר, אך התוכן אינו אמיתי. התוכן הוא באיכות גבוהה עד כדי כך שמשמש מן היישוב יתקשה לרוב לגלות שמדובר בזיוף (תב"כ 24/9, 2021).

לצד מאמצי חקיקה ואכיפה אפשר לראות ברחבי העולם מאמצי רגולציה מצד ממשלות, הן על פיתוח מוצרי דיגיטליזציה והן על הפצתם. בסין כבר מחייבים ואוכפים סימון ברור של תוצרי דיגיטליזציה, ובארצות הברית הונחה של שולחן של הקונגרס ביוני 2019 הצעת חוק שעיקרה חובת סימון ושקיפות כשמדובר בדיגיטליזציה, אשר נותרה בעינה מאז (ברון ואלטשולר, 2019; Statt, 2019).

סוגיה מטרידה נוספת הנובעת מאיום הדיגיטליזציה נוגעת לעובדה שקיומה של הטכנולוגיה שוחק את האמון בראיות וידאו ועלול לערער את הערך הראייתי שלהן בבית המשפט. תמונות וסרטונים הם אופן שכנוע עוצמתי, שכן ייצוג חזותי נתפס מאז ומתמיד על ידי בני אנוש כבטוח ואמין. במערכת המשפט האמריקאית, למשל, תמונות וסרטונים דיגיטליים יכולים לשמש עדויות וראיות קבילות כל עוד ניתן לאמת את מהימנותם. המשמעות היא שמערכת המשפט צריכה למצוא את הכלים המתאימים ולפתח טכנולוגיות מקבילות כדי לזהות ולאתר דיגיטליזציה, בעיקר בעולם שבו הטכנולוגיה משתכללת בהתמדה, ועד מהרה הסרטונים המזויפים יהיו כה אמין עד שעין אנושית לא תוכל להבחין בזיוף (Maras & Alexandrou, 2018).

איום מזוין או אמיתי? דיגיטליזציה והאתגרים לביטחון הלאומי / לירן ענתבי בהשתתפות נועם רחמים

גם בישראל, ראיות כגון תמונות וסרטונים דיגיטליים עשויות להיות קבילות בתנאי שאושרו על ידי בית המשפט, הן מבחינת אמינות והן מבחינת קבילות האופן שבו הושגו. לכן מדובר באתגר בכל הנוגע להקלטות ולראיות מתחום הווידאו, שניכר כי הגורמים הרלוונטיים עדיין לא נתנו עליו את הדעת.

ההתמודדות בארצות הברית

עיקר הקושי להתמודד עם התוצר הטכנולוגי נובע ככל הנראה מכך שהגבלתו עשויה לפגוע בחופש הביטוי. ב־2019 כתבו סיתרון וצ'סני כי אין בעולם משטר או חוק פלילי האוסר על יצירה או הפצת תוכן הנוצר באמצעות דיפ־פייק, ודנו בבעיות העקרוניות הכרוכות באיסור כללי כזה. בראש ובראשונה הם ציינו כי איסור כללי אינו רצוי משום שעצם המניפולציה הדיגיטלית על תוכן היא לא הבעיה, שכן יש לה מגוון יישומים חיוביים. שנית, איסור כללי ימנע התפתחויות חשובות נוספות בתחומי החדשנות הדיגיטלית. לבסוף, איסור כללי על יצירת תוכן דיפ־פייק והפצתו מתנגש עם התיקון הראשון לחוקה, המגן על חופש הביטוי.

פסקי דין קודמים בארצות הברית קבעו כי גם במקרה שהמידע המופץ כוזב, עדיין חלה עליו ההגנה של חופש הביטוי. ב־1964 קבע בית המשפט העליון בארצות הברית במשפט ניו־יורק טיימס נגד סאליבן כי דברי כזב נהנים מהגנה חוקתית, מהסיבה שאיסורם היה מהווה פגיעה בחופש הביטוי. ב־2012 בית המשפט העליון בארצות הברית הוסיף וקבע כי יש להגן על שקרים, שכן תפקידם לעורר הפרכה ושיח מנומק שהוא חיוני (Chesney & Citron, 2019).

מאז פרסום מאמר זה גילו מדינות שונות בארצות הברית מעורבות משפטית חקיקתית בנושא. טקסס הייתה המדינה הראשונה בארצות הברית שאסרה תוכני דיפ־פייק כאשר חוקקה בספטמבר 2019 את החוק SB751, שלפיו יצירה או הפצה של תוכני דיפ־פייק במטרה לפגוע במועמד לבחירות או להשפיע על תוצאותיהן בשלושים הימים שטרם הבחירות מהווה עבירה פלילית (Salazar, 2019). באוקטובר 2019 חתם מושל קליפורניה גאווין ניוסום על חוק AB730, שלפיו אין זה חוקי להפיץ סרטוני וידאו שעברו מניפולציה במטרה לפגוע בדמותו של מועמד פוליטי ולהונות מצביעים, בטווח של 60 יום ממועד בחירות. מארק ברמן, חבר המועצה של קליפורניה שאישרה את החוקים, אמר כי "לבוחרים קיימת הזכות לדעת מתי קטעי וידאו, קול ותמונות שמראים להם, שמטרתם להשפיע על בחירות מתקרבות, עברו מניפולציה ואינם לקוחים מהמציאות".¹ ברמן הוסיף כי "בהקשר של בחירות, היכולת לייחס מלל או התנהגות שאינם אמיתיים למועמד כלשהו הופכת את הדיפ־פייק לכלי עוצמתי ומסוכן"² (Berman, 2019).

נוסף על כך, בדצמבר 2019 חתם נשיא ארצות הברית דאז דונלד טראמפ על החוק הפדרלי הראשון הקשור בדיפ־פייק. הצו עסק בהיערכות האמריקאית הנדרשת לאיום זה וכלל התייחסות להערכת היכולות הטכניות של ממשלות זרות בנוגע לזיופים עמוקים, ניתוח האיום וההשלכות ומתן תמריץ כספי לפיתוח כלים לאיתור ולזיהוי של דיפ־פייק (Ferraro et al 2019).

1 "Voters have a right to know when video, audio, and images that they are being shown, to try to influence their vote in an upcoming election, have been manipulated and do not represent reality."

2 "In the context of elections, the ability to attribute speech or conduct to a candidate that is false — that never happened — makes deepfake technology a powerful and dangerous new tool in the arsenal of those who want to wage misinformation campaigns to confuse voters."

אולם הצעדים הללו, מעודדים ככל שיהיו, אינם מייצגים את שאר העולם (Feeney, 2021). כך למשל במערכת הבחירות האחרונה בישראל בשנת 2021 התמודדה מערכת המשפט הישראלית עם פרשיית די־פייק, ובחירה לסרב להורות על הסרת הסרטון בשל החשיבות הרבה לשמירה על חופש הביטוי (תב"כ 24/9, 2021). נוסף על חוק AB730 ניוסום חתם גם על חוק AB602, אשר מעניק לתושבי קליפורניה את הזכות לתבוע אדם שיצר די־פייק שבו מוטמעת דמותם בחומרים פורנוגרפיים שנוצרו בלא הסכמתם (Sheller, 2019). זוהי אחת המדינות הראשונות שנותנת כלי משפטי בידי נפגעי די־פייק פורנוגרפי. במקומות רבים אחרים בעולם אין ודאות כי אדם שיתלונן יזכה לסעד מכוחות אכיפת החוק ומערכת המשפט, שכן לרוב אין להם מודעות כמו גם כלים מתאימים לכך, ובהם חקיקה רלוונטית.

גם במקרים שיהיה רצון לתת סעד לפונה בעניין של די־פייק, איתור האשם והעמדתו לדין נתקלים בשני מחסומים עיקריים. הראשון הוא אתגר טכנולוגי הנובע מאופן יצירת די־פייק, המאפשר ליוצרו להישאר אנונימיים. לפיכך גם אם ימצא חשוד ביצירת הסרטון, הוכחת האשמה תהא מאתגרת מאוד באמצעים הקיימים, ואולי אף בלתי אפשרית. המחסום השני מתייחס לכך שגם אם אותר הנאשם ביצירת התוכן המזויף, היכולת להעמידו לדין אינה פשוטה כלל. בארצות הברית, התקנות שעל בסיסן ניתן לתבוע את הנאשם והן בעלות הפוטנציאל הגבוה ביותר להצלחה הן תביעה בגין לשון הרע, הוצאת דיבה וגרימה מכוונת למצוקה רגשית. לתביעות על רקע הפרת זכויות יוצרים, זכויות פרסום ופרטיות יש סיכויים מוגבלים יותר להצלחה. לבסוף, במקרים מסוימים לא תתאפשר העמדה לדין של יוצר או מפיק התוכן המזויף, אם הוא שוהה מחוץ לגבולות השיפוט האמריקאיים (Chesney & Citron, 2019).

לאור האתגר הגדול הכרוך בהוכחת האחריות ובהטלת אשמה על יוצרי התכנים, נראה כי הדרך היחידה להרתעה ולתיקון עוולות היא ייחוס האחריות לפלטפורמות הפרסום שבהן הם מופצים. אולם, סוגיה זו מעלה אף היא קשיים עקב סעיף 230 לחוק הגינות התקשורת, המעגן בארצות הברית את חסינות הפלטפורמות מפני נשיאה באחריות לתכנים המתפרסמים בהן, וזאת במטרה להגן על חופש הביטוי ולשמור על ריבוי דעות ברשת (Chesney & Citron, 2019). כך השילוב של התיקון הראשון לחוקה וסעיף 230 בחוק הגינות התקשורת פועל כחסם עוצמתי בפני מאבקם החוקי של קורבנות סרטוני די־פייק פורנוגרפיים בארצות הברית, כאשר הם מגינים יחד הן על יוצרי התוכן והן על הפלטפורמות שבהן התכנים מופצים. עד כה, קיטוב חברתי, ביטויי שנאה ופייק ניוז לא גרמו למחוקקים לבטל את הפטור שממנו נהנות הפלטפורמות, אך ייתכן שדי־פייק יהיה קו פרשת המים להטלת אחריות כזאת (ברון ושוורץ אלטשולר, 2019).

מאמרה של אן פצ'ניק גיִזְקָה משנת 2020 מציע תיקון לחוק הגינות התקשורת בארצות הברית, כך שיאפשר לתבוע יצרנים ופלטפורמות המשתמשים לרעה בטכנולוגיית די־פייק, תוך ניסוח החוק בצורה המבדלת אותו מהגנות התיקון הראשון ומתייחסת נקודתית לסרטוני די־פייק פורנוגרפיים, בשילוב מתן חסינות לפלטפורמות הנאבקות להגן על הקורבנות מפני הטרדות ומשתפות פעולה בהסרת הסרטונים הפוגעניים. תיקון זה עשוי לספק שתי דרכים יעילות להילחם בתופעה: הראשונה היא אפשרות לתבוע את יוצרי הסרטונים; השנייה היא הסרה מהירה של הסרטונים הפוגעניים על ידי תמריץ חיובי לשיתוף פעולה מצד הפלטפורמות (Pechenik Gieseke, 2020).

ההתמודדות באיחוד האירופי

על הנעשה באירופה אל מול איום הדיפ־פייק אפשר ללמוד מדוח שפרסם הפרלמנט האירופי ביולי 2021. בהחלטת הפרלמנט ממאי 2017 הוא קרא לנציבות האירופית לנקוט אמצעי רגולציה קשיחים להתמודדות עם פייק ניוז. ההחלטה הפרלמנטרית ממאי 2018 בנושא פלורליזם תקשורתי קיבלה מספר המלצות שלא היו קשורות לדיפ־פייק אך הן רלוונטיות גם לגביו, ביניהן: שקיפות מלאה בשימוש באלגוריתמים, ביניה מלאכותית ובקבלת החלטות אוטומטיות; סינון והסרה של תוכן אינטרנט פוגעני; חשיבותם של ארגונים לבדיקת עובדות עצמאית וללא משוא פנים; חובת אימות המקור; מתן אפשרות למשתמשים לדווח ולסמן דיסאינפורמציה פוטנציאלית; הצגה ותיוג של דיסאינפורמציה המתגלה ככזו כדי לעורר דיון ציבורי ולמנוע עליית התוכן מחדש. אזכור ספציפי של דיפ־פייק אפשר למצוא בעמדות פרלמנטריות שונות, ביניהן החלטת הפרלמנט מפרברואר 2019 הקוראת לנציבות לדרוש הצגת תיוג ליוצרי דיפ־פייק (van Huijstee et al., 2021). המסמך המקיף והעדכני ביותר שיצא באירופה בנושא דיפ־פייק הוא החלטת הפרלמנט האירופי ממאי 2021 בנושא 'בינה מלאכותית בחינוך ובתרבות'. נוסף על ההצעות שהוזכרו לעיל, החלטה זו מכילה הצעות שונות כיצד להיערך להתמודדות עם דיפ־פייק כ"איום מידי על הדמוקרטיה". אלה כללו את חשיבות העלאת המודעות הציבורית לסיכונים של דיפ־פייק ושיפור האוריינות הדיגיטלית; טיפול בקושי הגובר לאתר ולתייג תוכן כוזב ומניפולטיבי באמצעים טכנולוגיים; קריאה לנציבות להציג מסגרות משפטיות מתאימות לשליטה ביצירה, ייצור או הפצה של דיפ־פייק למטרות זדוניות; קידום פיתוח יכולות נוספות לאיתור ולזיהוי של דיפ־פייק; שיפור השקיפות ביחס לתוכן המוצג למשתמשי הפלטפורמות, אשר נותן להם חופש רב יותר להחליט אם ואיזה מידע הם רוצים לקבל (European Parliament, 2021).

ההתמודדות בבריטניה

בבריטניה, כמו באיחוד האירופי, החוק לוקה בחסר באשר ליכולת להתמודד עם איום הדיפ־פייק, כאשר כיום אין באיחוד חוקים המיועדים במיוחד לדיפ־פייק, וכן אין "זכות קניין רוחני עמוק" או חוק המגן על תדמית או אישיות של אדם. יתרה מכך, במקרה של דיפ־פייק שקורבנותיו הם ידוענים גם זכויות יוצרים לא יועילו, שכן זכויות היוצרים נתונות לאולפני סרטים ולצלמים, לא לסובייקטים עצמם המופיעים בסרטון. המשמעות היא שסוגיית הדיפ־פייק בבריטניה פרוצה לחלוטין. מומחים קראו לממשלה הבריטית לנקוט צעדים מהירים להסדרת הנושא ולפעול לרגולציה שתשמור על יתרונות הטכנולוגיה אך תמגר השלכות שליליות כמו פורנוגרפיה, הונאה ופגיעה בדמוקרטיה.³ לעת עתה הטכנולוגיה בבריטניה מקדימה את החוק (The rise of the deepfake, 2021). כך גם במרבית מדינות העולם.

ההתמודדות בסין

סין הייתה המעצמה הראשונה שהפכה את השימוש בדיפ־פייק לפשע פלילי מבלי לסמנו ככזה. על פי תקנות משרד הסייברספייס של סין, אשר נכנסו לתוקף בינואר 2020, כל סרטון שנוצר בעזרת בינה מלאכותית צריך

3 "The time is now to introduce regulation in this area in order to prevent negative uses of the technology and create an environment where positive use cases emerge."

להיות מסומן ככזה בצורה ברורה. לדברי המשרד, טכנולוגיית דיפ־פייק יכולה "להוות סכנה לביטחון הלאומי, להפריע ליציבות החברה, להפריע לסדר הציבורי ולפגוע בזכויות וברצונות של אחרים". הפרת התנאי תחשוף את היוצר שלו ואת האתר המארח של הסרטון לדין פלילי. אף שהחוק הסיני אוסר על יצירה או הפצה של דיפ־פייק ללא סימון, הוא אינו ברור בעניין העונש למי שיפר אותו. בהודעה שפרסמה ממשלת סין לאתרים נאמר רק כי יוצרי הסרטון וכן מארחיו צפויים לעמוד לדין פלילי (Statt, 2019).

פתרונות טכנולוגיים

'זיהוי ידני' של דיפ־פייק מחייב תשומת לב רבה לשינויים כגון מציאת אי־התאמה בקול, בתמונה או בווידאו. זיהוי ידני מבוסס על יכולתה של העין האנושית למצוא חוסר עקביות באמצעות ניתוח דפוסי התנהגות בסרטונים לעומת אלו המוכרים מהמציאות, או לחלופין, מציאת עדויות לשינוי או להתערבות בקובץ הדיגיטלי, ניסיון לאתר את מקורו, את מועד יצירתו ועוד. החיסרון העיקרי של שיטות עבודה ידניות קשור בהיקף החומר שניתן להתמודד עימו. לעומת זאת, כלים אוטומטיים יכולים להתמודד עם כמות גדולה של קבצים ולאחר זיופים בקלות (van Huijstee et al., 2021). מכאן נובעת חשיבותן הרבה של שיטות הזיהוי האוטומטיות או האוטומטיות למחצה, העושות שימוש בכלי בינה מלאכותית, למשל: זיהוי קולי מבוסס בינה מלאכותית (דוגמת השוואה בין "חתימת הקול" של האדם שאותו מבקשים לחקות לבין הקובץ החשוד); ניתוח מבוסס בינה מלאכותית של תווי פנים; או איתור אזורים "מטושטשים" שלעיתים נוצרים כאשר נעשה שינוי בקובץ מדיה מקורי. לשיטות זיהוי אלה יש מגבלות, וייתכנו מקרי אי־זיהוי של קובצי דיפ־פייק גם באמצעים טכנולוגיים מתקדמים.

בין הכלים הטכנולוגיים הקיימים בשוק נמנה מאמת הווידאו של מיקרוסופט, שהושק ב־2020. התוכנה פועלת על בסיס שיטת ניקוד, הקובעת באחוזים את מידת הסיכוי של קובץ דיגיטלי להיחשב אותנטי או מזויף. גישה נוספת לזיהוי קבצים אותנטיים או מזויפים היא הטמעת "חותמת איכות" של יוצרי הקבצים. מיקרוסופט עושה שימוש במערך הענן שלה Azure בחותמת דיגיטלית המוצמדת לכל קובץ, כך שאפשר להבחין במהירות אם מדובר בקובץ מקורי על פי הפרמטרים של החברה.

חברת פייסבוק הודיעה אף היא ביוני 2021 כי פיתחה יישום לזיהוי דיפ־פייק בשיתוף עם אוניברסיטת מישיגן. הכלי עושה שימוש בזיהוי דפוסים של קבצים שנוצרו על ידי בינה מלאכותית. כך יכולה פייסבוק לאתר את מקור הקובץ, לחסום אותו ואף למנוע הפצה עתידית של קבצים נוספים. אולם איכות הזיהוי ככל הנראה עדיין אינה טובה דיה. בשנת 2020 ערכה פייסבוק תחרות שבמסגרתה הוגשו אלגוריתמים שונים לזיהוי דיפ־פייק, והכלי המוצלח ביותר השיג שיעור זיהוי של כ־65 אחוזים. מאז ינואר 2020 מיישמת פייסבוק את מדיניות האכיפה עבור קבצים המכונה manipulated media – קבצים שפייסבוק מזהה כי הם נוצרו על ידי בינה מלאכותית או למידת מכונה, אשר מטרתם להיראות אותנטיים ומטעים. מאז יישום המדיניות פייסבוק הסירה תכנים שאותם הגדירה ככאלה. פייסבוק גם יצרה שיתוף פעולה עם סוכנות הידיעות רויטרס לצורך קורס מקוון חינוכי לציבור, במטרה להכשירו לזיהוי קובצי דיפ־פייק.

חוקרי מעבדת המחקר של צבא ארצות הברית פיתחו עם אוניברסיטת דרום קליפורניה כלי חדשני לזיהוי תוצרי דיפ־פייק הנקרא DeFakeHop ומבוסס על טכניקות זיהוי פנים ביומטריות. החידוש הבולט בכלי זה

הוא תאוריה ומסגרת מתמטית חדשנית שפיתחו החוקרים, שקראו לה Successive Subspace Learning או SSL. בעזרת ה־SSL מחלצים באופן אוטומטי תכונות מחלקים שונים של תמונות פנים, מנתחים את הממצאים ומזהים אם הסרטון הוא זיוף. לפי פרופסור קאו מאוניברסיטת דרום קליפורניה, מדובר במסגרת מתמטית חדשה לחלוטין ושונה מהגישה המסורתית בתחום ארכיטקטורת הרשת העצבית. השיטה הוצגה לראשונה במאמר 2021 והיא מראה הצלחה של יותר מ־90 אחוזי דיוק בכל מערכי הנתונים, ובחלקם אף של 100 אחוזים (U.S. Army DEVCOM, 2021).

עד כה, רוב מאמצי המחקר הנוגעים לטכנולוגיות לזיהוי ואיתור דיפ־פייק מתמקדים בפתרונות איתור אוטומטי שייסיעו לזהות דיפ־פייק בשנים הקרובות, בהתאם להתפתחות הטכנולוגית הקיימת והצפויה. אך שיטות זיהוי אוטומטי עשויות להפוך בלתי יעילות כבר בעתיד הקרוב, שכן הטכנולוגיות ליצירת הזיוף משתפרות במידה ניכרת. כמו כן, מול ההתפתחות וההשקעה המסיבית של ענקיות האינטרנט בתחום נצפית מעורבות מדינית וחיקיקתית דלה בלבד, המתמצה בעיקרה בהגבלה ובפיקוח על פרסומים תעמולתיים בעת בחירות, ללא השקעה במחקר (Vizoso et al., 2021). לפיכך, נוסף על תמיכה בפיתוח פתרונות לטווח קצר, על גורמי ממשל להשקיע בגילוי ובפיתוח פתרונות לטווח הרחוק. אנג'לר (Engler, 2019) ממליץ לגורמי הממשל לתמוך ולממן תוכניות העוסקות במאמצי זיהוי מתמשכים ובהכשרת עיתונאים ובודקי עובדות שישתמשו בכלים אלו; לקיים תחרויות המעודדות חברות לפתח כלים חדשניים, למשל מערכות אימות מבוססות בלוקצ'יין⁴, שעשויים לפעול בצורה מהימנה יותר מול זיופים עמוקים; לעודד את פרסום מאגרי הנתונים הגדולים של המדיה החברתית לחוקרים לשם לימוד ומחקר אקדמי של פתרונות למניעת התפוצה של פייק ניוז, ובכלל זה דיפ־פייק.

חינוך הציבור

לצד מאמצי רגולציה, חקיקה ואכיפה ופתרונות טכנולוגיים, מדינות, ארגונים וחברות בעולם עוסקים בהסברה ובהעלאת המודעות בקרב הציבור לאתגר הדיפ־פייק. ההסברה לציבור מתמקדת בהבהרות כי יצירה או הפצה של תכנים באמצעות דיפ־פייק (ללא סימון) עלולה להוות עבירה פלילית; בקמפיינים להעלאת מודעות לקלות ולזמינות ייצורם של סרטוני דיפ־פייק אמניים למדי; וכן בחינוך הטלת ספק מתמדת ולצריכת מידע ממקורות מהימנים. אחת המדינות החלוצות והמובילות בתחום זה היא פינלנד, שבה קיימת תוכנית חינוכית רחבה לצריכה מושכלת של תקשורת וליטיטוט חכם ברשתות החברתיות (Barber, 2021). הפרלמנט האירופי המליץ שוב בשנים האחרונות לפעול לחינוך הציבור ולשיפור האוריינות הדיגיטלית (European Parliament, 2021), ומבריסל פרסמו לציבור האירופאי אסטרטגיה להתמודדות עם דיסאינפורמציה, הכוללת הנחיות ספציפיות בעניין דיפ־פייק.

בישראל, במסגרת המאמץ להעלאת מודעות הציבור לנושא, איגוד האינטרנט הישראלי פרסם לאחרונה רשימה של תוכנות ויישומים חינוכיים ונגישים לציבור לזיהוי דיפ־פייק. בין היתר הציג איגוד האינטרנט לציבור בספטמבר 2020 את מאמת הווידאו של חברת מיקרוסופט העולמית (Microsoft Video Authenticator),

4 טכנולוגיה המאפשרת פעילות עסקית מאובטחת ללא צורך בישות ניהול מרכזית.

המסוגל לנתח קובצי קול ווידאו על מנת להעריך באחוזים את ההסתברות שתכנים אלו עובדו או שונו בצורה מלאכותית (קאהאן, 2020).

כלי נוסף שעליו המליץ איגוד האינטרנט הישראלי שייך לחברת sensity.ai ההולנדית. הוא מתבסס על שיטת DCT (Discrete Cosine Transform) המאפשרת לגלות תוצרי דיפ־פייק שנוצרו על ידי שימוש ב־GANs. הכלי מתמקד בזיהוי פנים ששוננו, אך עדיין אינו מזהה חפצים שנערכו לתוך חומרי דיפ־פייק. חברת אבטחת המידע Zemana השיקה אף היא כלי טכנולוגיה לשימוש הציבור הרחב בשם Deepware, המאפשר למשתמשים להזין ישירות לאתר האונליין של החברה סרטוני וידאו המנותחים ונסרקים בו על מנת לזהות אם עברו עיבוד או מניפולציה חזותית, לרבות באמצעות דיפ־פייק (איגוד האינטרנט הישראלי, ל"ת). נוסף על חינוך האוכלוסייה הכללית, קיימת חשיבות גבוהה לחינוך בנוגע לדיפ־פייק לאוכלוסיות ייעודיות כגון אנשי מקצוע בתחום אכיפת החוק והמשפט, מקבלי החלטות ובעיקר עיתונאים, העשויים להוות גורם נוסף להפצה (שלא בידיעה) של ידיעות מזויפות. עוד לפני הופעתן של טכנולוגיות דיפ־פייק היה על עיתונאים להתמודד עם האתגר הכרוך בתארוך של מידע המגיע אליהם או אימות אמינותו. הדיפ־פייק הופך את עבודתם למסובכת יותר ולכן נודעת חשיבות להגברת המודעות בקרבם לסוגיה, וכן למתן כלים בידם לזיהוי ולאימות חומרים המגיעים אליהם.

במצב אידיאלי כלים לאימות של דיפ־פייק צריכים להיות זמינים לכל אדם, אולם משום שחלק מהטכנולוגיה הזו נמצא בתהליכי פיתוח מוקדמים, יש חוקרים הדואגים לסייע כבר כיום לעיתונאים מתוך תפיסה שהם "קו ההגנה הראשון מפני התפשטות מידע מוטעה" (Sohrawardi & Wright, 2020). זאת ועוד, מעת לעת מומחים מפיצים מדריכים המסייעים גם לאדם הסביר לזהות באופן לא־טכנולוגי תוצרי דיפ־פייק, כזה שפורסם על ידי מעבדת המדיה של MIT ובו צעדים מומלצים כגון: שימת לב למצח וללחיים של המופיעים בסרטון, האם הם חלקים מדי? האם העיניים קרובות מדי? האם יש בוהק מוזר במשקפיים? (Groh, n.d.). עם זאת, מוצלחות ככל שהיו, מדובר ביוזמות נקודתיות והשפעתן מוגבלת.