

שימושים בדיפ־פייק ומקרי בוחן מן העולם

במסגרת 'אתוס הקוד הפתוח' של חלק ניכר מהקהילה של למידת מכונה, כל התקדמות מחקרית בתחום הזיוף העמוק מתפרסמת מייד בפורומים ייעודיים והופכת פומבית וזמינה (Toews, 2020). כמו ביישומי בינה מלאכותית רבים אחרים, עד לפני מספר שנים היה קשה מאוד להשתמש ביכולות דיפ־פייק, משום שהדרישות לכמות מידע גדולה וכן לכוח מחשוב רב ויקר יצרו מכשול. על כן עיקר השימושים בוצע בגופים גדולים – חברות ומדינות (רשות החדשנות, 2019). אולם ההתקדמות הטכנולוגית בשנים האחרונות הפכה את הטכנולוגיה זמינה לקבוצה רחבה של יוצרי תוכן, חובבים ואנשי מקצוע כאחד (Biggs & Moran, 2021). כיום גם משתמשים ללא כל רקע בתחום המחשוב יכולים ליצור סרטון מזויף קצרצר תוך שניות ספורות, וסרטונים ארוכים יותר תוך מספר דקות (Panyatham, 2020). לפיכך שחקנים שונים יכולים לעשות שימוש בטכנולוגיה זו – אלה הפועלים בחסות מדינה ולצידם קבוצות פוליטיות ואנשים פרטיים (Toews, 2020). קיימות שלוש דרכים עיקריות ליצירה של זיופים עמוקים בתחום הווידאו והקול. הראשונה היא ייצור עצמי על בסיס אפליקציות ותוכנות מובנות ייעודיות למחשב, שמרביתן מספקות יכולות החלפת פנים בסרטונים קיימים, ואחת מהן אף מאפשרת שיבוט קול סינתטי. שימוש יעיל בתוכנות אלו מצריך לרוב ידע בתכנות ומעבד גרפי רב־עוצמה, מה שהופך אותן נגישות פחות לחובבים. למרות זאת הופצו מספר מדריכים מפורטים לשימוש ביישומים הפופולריים הללו, ועם השנים בוצעו בחלק מן התוכנות הנפוצות עדכונים ששיפרו את הגישה למשתמשים רבים. עם זאת, עדיין נדרשים ידע ויכולת תפעול של יישומים מתקדמים יחסית. הדרך השנייה היא רכישה של סרטונים דרך פורטלים מקוונים המספקים את השירות. הדרך השלישית היא הזמנה של שירותים כאלה מיוצרים פרטיים המפרסמים את שירותיהם בפורומים שונים וב'אתרי שוק' ברשת. שלוש הדרכים המפורטות מתמחרות בטווח שבין שלושה לכ־30 דולר בממוצע עבור סרטון, אך יכולות להיות גם עלויות גבוהות יותר. כמו כן, המזמין נדרש לספק כמות גדולה של תמונות וסרטונים של האדם שהוא מבקש שיופיע בסרטון (Ajder et al., 2019). למרות זאת ניכר כי בשונה מן העבר גברה הזמינות של סרטונים כאלה, תוך שהמחיר ירד ופשטות היצירה או ההזמנה של השירות השתפרו מאוד.

הטכנולוגיה מאפשרת מגוון רחב של שימושים בסרטוני דיפ־פייק, חלקם נפוצים כבר היום. לצד שימושים שליליים בטכנולוגיית הזיוף העמוק כמו פורנוגרפיה, סחיטה והונאה, נעשה גם שימוש לצרכים חיוביים מגוונים, החל בקולנוע וטלוויזיה דרך נגישות, רפואה, פסיכולוגיה ועד אומנות, תרבות ושימור היסטורי. להלן תיאור של השימושים השונים וכן מספר מקרי שימוש שלילי שהתרחשו בעולם, והם בעלי פוטנציאל השפעה על הביטחון הלאומי.

קולנוע וטלוויזיה, אומנות ותרבות

דיפ־פייק יכול להיות שימושי לשם הפיכת סרטים בשחור לבן לסרטים בצבע, שיפור איכות הווידאו והפיכת סרטוני חובבים למקצועיים. נוסף על אלה, השימוש הנפוץ ביותר בדיפ־פייק הוא יצירת סרטונים המבוססים על "השתלת" פנים של אדם אחד בסרטונים של אדם אחר. כך לדוגמה, יוצרי סדרת הטלוויזיה South Park

יצרו ב־2020 את התוכנית Sassy Justice – תוכנית קומית ששודרה ביוטיוב העשויה כל כולה דיפ־פייק סאטירי, בכיכובו (לכאורה) של נשיא ארצות הברית לשעבר דונלד טראמפ ודמויות נוספות. לפרק הראשון נרשמו מעל מיליון צפיות, והתוכנית הפכה במהרה לפופולרית במיוחד בארצות הברית (Itzkoff, 2020).



Sassy Justice, 2020

מקור: Press/YouTube

יתרה מכך, בזכות יכולות דיפ־פייק אפשר להשתמש בשחקנים בסרטים ובפרסומות גם אם אינם יכולים להגיע פיזית לאתר הצילומים, כאשר כל שנדרש הוא צילומי בסיס וכפיל גוף. כך לדוגמה, אתר הסטרימינג HULU הפיק בזמן משבר הקורונה פרסומת עתירת כוכבי ספורט ידועים מכל העולם, ללא צורך בתיאום בין לוחות הזמנים שלהם והגעתם לסט (Roettgers, 2020). עם זאת יש לזכור כי שימוש זה אינו שונה במידה משמעותית מצילום האנשים באתרים שונים וחיבור קטעי הווידאו בצורה אמינה תוך שימוש בכלי עריכה טכנולוגיים מתקדמים.

עוד ב־2020 יצא לאקרנים הסרט התיעודי Welcome to Chechnya, המתמקד ברדיפת הלהט"בים בידי הרשויות בצ'צ'ניה בעשור השני של המאה ה־21, ובו רואיינו פליטים להט"בים שברחו מצ'צ'ניה על חוויותיהם הקשות. הפקת הסרט הייתה מחויבת לביטחונם האישי של המרואיינים ולכן נמנעה מחשיפת זהותם, אך בה בעת רצתה להצמיד פנים אנושיות לסיפור. בעזרת יכולות דיפ־פייק מתקדמות הציגה הפקת הסרט מרואיינים "מזויפים" בעלי פנים אנושיות והבעות שנראו אותנטיים לחלוטין, ואפשרו לצופים להזדהות עם כאבם של המציגים (Thomson, 2020).

אפשר להשתמש בטכנולוגיית דיפ־פייק גם בקטעי קול: המכונה לומדת את קולו של השחקן ויכולה לדמות אותו באופן מלאכותי, למשל במקרה ששחקן מאבד את יכולתו הקולית בשל מחלה. הטכנולוגיה המתקדמת אף יכולה ליצור הלימה אמינה בין תנועת השפתיים והבעות הפנים של השחקן למוצא פיו. מדובר ביישום שימושי ביותר בהתמודדות עם בעיות נפוצות כמו שינוי מאוחר בטקסט או בתסריט או צורך בצנזורה (Hao & Heaven, 2020).



Welcome to Chechnya — a film by David France (2020)

השימוש ביצירת קולו של אדם באמצעות דיפ־פייק עורר השנה סערה בתחום הקולנוע התיעודי, עת נעשה שימוש בטכנולוגיה במסגרת הסרט 'שף: סרט על אנתוני בורדיין'. בשלוש סצנות מהסרט על אודות השף שהתאבד נעשה שימוש בדיפ־פייק לביצוע מניפולציה בקולו של השף המנוח, על מנת להקליט משפטים שהשף כתב אך לא אמר ולא הוקלטו בקולו. במאי הסרט בחר ליצור את קולו של השף באמצעות טכנולוגיית בינה מלאכותית במקום להיעזר בקריין, ובחירה זו עוררה ויכוח גדול לגבי עתיד תעשיית הסרטים התיעודיים (רוה, 2021).

שימוש בעל פוטנציאל רחב ביותר הוא ליצירת מצג של אמינות במקרים של דיבוב תוצר וידאו לשפות מרובות. בתעשיית הקולנוע והטלוויזיה הטכנולוגיה מפחיתה את עלות העסקת שחקנים מדוברים ויכולה לסייע ביצירת חווית צפייה איכותית יותר עבור קהלים ברחבי העולם, שיוכלו ליהנות מהם בשפה המקומית (Panyatham, 2020). השימוש ביכולת זו נעשה גם לשם הנגשת מידע חשוב לציבור הרחב בכל העולם, דוגמת הודעתו של דיוויד בקהאם על סוגיית המלריה בתשע שפות שונות, אשר הופצה בקולו המקורי ובתנועות שפתיים התואמות את המילים הנאמרות (ThinkAutomation, n.d.).

כך גם הפוליטיקאי ההודי מאנוג' טיווארי (Manoj Tiwari) נעזר ביכולות דיפ־פייק על מנת "לדבר" בשפתם של כל תומכיו הפוטנציאליים. השפות הרשמיות בהודו הן הינדי ואנגלית, אך קיימים גם אלפי ניבים במאות השפות המדוברות בקרב אזרחים בחלקים גיאוגרפיים שונים של הפדרציה. כדי לא להחמיץ אף מצביע פוטנציאלי, טיווארי הפך את אחד מסרטוני התעמולה שלו לרב־לשוני באמצעות יכולות דיפ־פייק. התוצאה הייתה אמינה במיוחד והיה קשה עד בלתי אפשרי להבחין שמדובר בסרטון מזויף, שכן תנועות השפתיים תאמו במדויק את הדיבור, והסרטון גרף מעל 15 מיליון צפיות (חיימוביץ', 2020).

ב־2019 הציגה מיקרוסופט בכנס Microsoft Inspire בלאס וגאס כלי שמייצר הולוגרמה ההופכת אדם לדובר של שפה אחרת, שהוא כלל לא מכיר. בכנס נשאו דברים באנגלית מנכ"לית Azure ג'וליה וייט שהרכיבה את המשקפיים המהפכניות של מיקרוסופט, וכך ההולוגרמה אמרה את דבריה של וייט ביפנית שוטפת באופן אמין במיוחד (Focaloid, 2019). סביר להניח כי בעתיד אף יוכלו להקליט ולהפיץ או למכור

הרצאות שלמות של מומחים בשפה שאינם דוברים כלל, וכך להנגיש תכנים הן למרצים והן למאזינים. מדובר בפוטנציאל לקפיצת מדרגה ניכרת בתחום בשילוב עם שירותי תרגום מבוססי בינה מלאכותית, שהולכים ומשתפרים (Lu, 2020).



הדגמת ההולוגרמה של וייט הדוברת יפנית, מקור: Microsoft/YouTube

באופן דומה, אתרי אומנות ותרבות בחרו להשתמש ביכולות דיפ־פייק למטרותיהם. לדוגמה, מוזיאון דאלי בסנט פטרסבורג פלורידה יצר תערוכה בשם 'דאלי חי', ובאמצעות יכולות דיפ־פייק "השיבו לחיים" את האמן הסוריאליסטי סלבדור דאלי ואפשרו למבקרים במוזיאון לקיים עימו אינטראקציה באופן שנראה אותנטי למדי. כך מעבדת הבינה המלאכותית של סמסונג במוסקבה השיבה לחיים את המונה ליזה (Jaiman, 2020).

נגישות, רפואה ופסיכולוגיה

שימושים חיוניים נוספים ביכולות דיפ־פייק החלו להתפתח גם בתחום הנגישות. חולים במחלות כמו ניוון שרירים, אשר מאבדים את יכולת הדיבור, יוכלו בזכות יכולות דיפ־פייק להשתמש במכשיר שיקריא בקולם האישי את הטקסט שיזינו לו, אם המחלה תאפשר להם לבצע זאת. כך אפשר לתרום רבות לעצמאותם של החולים ואף לשפר את מצבם הנפשי, בשל היכולת לתקשר עם קרוביהם וסביבתם גם לאחר שאיבדו את יכולת הדיבור (Jaiman, 2020).

גם בתחום הטיפול הפסיכולוגי נעשה שימוש חיוני ביכולות דיפ־פייק. כך למשל הטכנולוגיה מאפשרת למי שאיבדו את יקיריהם לנהל שיחת וידאו מלאכותית עם אהובם המנוח, בהנחיית מטפלים מוסמכים באבל ושכול. ניסוי שנעשה בהולנד ותועד בסרט בבימויה של רושן נג'ל הוכיח כי אף על פי שבני המשפחה השכולים ידעו שמדובר באשליה, הם דיווחו על שיפור משמעותי במצבם (Nejal, 2020).

נוסף על אלה נעשה שימוש ביכולות דיפ־פייק בהכשרת רופאים ובאימון אלגוריתמים של בינה מלאכותית. יכולות דיפ־פייק מאפשרות לשמור על פרטיותם של המטופלים כאשר נעשה שימוש בחולים מזויפים ובסריקות

MRI סינתטיות לחלוטין, המאפשרות לשפר את חדות האבחנה של הרופאים וכן לאמן אלגוריתמים לזיהוי גידולים ללא צורך במידע פרטי של חולים אמיתיים (ThinkAutomation, n.d.).

חינוך ושימור היסטורי

יכולות דיפ־פייק מאפשרות לגורמי הוראה, חינוך ותרבות להשיב לחיים דמויות היסטוריות באופן מסקרן ואינטראקטיבי, כך שהתלמידים יוכלו ללמוד בצורה ייחודית והידע יישאר חקוק בזיכרונם זמן רב. כך למשל מוזיאון השואה ומרכז החינוך באילינוי יצר ב־2018 סדרת ראיונות הולגרמטיים עם שורדי שואה. המבקרים יכלו לשוחח עם שורדי השואה, לשאול שאלות ולשמוע את סיפוריהם. עם התקדמות טכנולוגיית הזיוף, היסטוריה וירטואלית מסוג זה עשויה להיות בת־השגה בקנה מידה רחב בהרבה (ThinkAutomation, n.d.).



הדמיית עדות, מקור: USC Shoah Foundation/YouTube

עוד בהקשר זה, חברת MyHeritage מציעה כלי חינוכי מקוון ונגיש בשם Deep Nostalgia, שאליו ניתן להעלות תמונה משפחתית עתיקה ולקבל סרטון מונפש קצרצר של המופיעים בה (MyHeritage, n.d.).

משחקי מחשב

גם בתעשיית המשחקי המחשב נעשה שימוש בטכנולוגיית הדיפ־פייק. ניתן ליצור דמויות מציאותיות במשחקי וידאו ללא צורך בהשקעה בצילומן במשך שעות וימים, אלא להשתמש בחומר גלם מצולם של השחקן וליצור על בסיסו מצבי משחק רבים ומגוונים (De Agostini, 2020). נוסף על כך השחקנים יכולים להשתמש בקול מזויף התואם לדמותם הווירטואלית, בלי לאבד את האותנטיות והאנושיות של הקול. כך יכולים השחקנים לשמור על זהותם וגם לשפר את חוויית המשחק ואת אמינותו (Somers, 2020). אפליקציות חברתיות פופולריות גם מאפשרות למשתמשים לתפעל הבעות פנים ולבצע החלפת פנים זה עם זה (Panyatham, 2020).

לצד השימושים החיוביים קיימים מספר שימושים שליליים, שניכר כי הם בעלי פוטנציאל לפגוע בצורה ישירה או עקיפה בביטחון הלאומי. להלן שלושה סוגים של שימושים כאלה ושל מקרים בולטים, כפי שפורסמו בתקשורת, שבהם נעשה שימוש זדוני בדיפייק.

אלימות, סחיטה והונאה

פעילים נגד אלימות במשפחה סימנו את טכנולוגיית הזיוף העמוק כאיום ותיארו אותה כ"כלי מושלם למישהו המבקש להפעיל כוח ושליטה בקורבן" (Hao, 2021). ואכן נעשה בה שימוש ליצירת תוכן מזויף ואמין למטרות זדוניות כגון סחיטה, הונאה, השפלה והטרדה. הטכנולוגיה אף מאפשרת סחיטה של אנשים, עסקים, ארגונים או ממשלות על ידי איום בהפצת תוכן מזויף המכיל מידע כוזב כמו סרטון פורנוגרפי, הקלטות קוליות וסרטונים שבהם מוצג אדם האומר או עושה דברים שלא אמר או לא עשה, ועוד (Citron & Chesney, 2019). דוגמה להונאה מסוג זה היא הודעה קולית מזויפת שקיבל אחד מעובדי חברת הייטק קטנה בארצות הברית, שבה ביקש לכאורה מנכ"ל החברה מהעובד לבצע עסקה דחופה ולהעביר תשלום לחשבון בנק. במקרה זה העובד זיהה כי ההודעה חשודה והעביר אותה למחלקה המשפטית, ובכך מנע את הצלחת ההונאה (Libby, 2020). למרות שלא מדובר בניסיון שונה במהותו מכזה שאפשר היה לבצע באמצעות דואר אלקטרוני למשל, השימוש בהקלטה או בסרטון המזויפים באיכות גבוהה עשוי להוביל להצלחה רבה יותר בתחום ההונאה, הן בשל מודעות נמוכה יותר מזו הקיימת היום להונאות באמצעות דואר אלקטרוני והן משום שבמצבים מסוימים של קושי לאמת או לשלול את אמיתות ההקלטה, תהיה נטייה לבצע את המתבקש בה.

מקרים מן העולם של מרמה באמצעות דיפייק

המקרה הראשון המתועד של הצלחה בהונאה בעזרת טכנולוגיית דיפייק קולי התרחש במארס 2019. מנכ"ל של חברת אנרגיה באנגליה קיבל שיחת טלפון שבה נשמע לכאורה מנהלו, מחברת האם הגרמנית, מבקש ממנו לבצע העברה דחופה לספק הונגרי על סך 243 אלף דולר. המנכ"ל שרומה תיאר כי הקול בשיחה נשמע זהה לחלוטין לזה של מנהלו, הוא זיהה את המבטא הגרמני ואת האינטונציה בדיבור של מנהלו ולכן פעל בהתאם להוראותיו. לאחר ההעברה הראשונה התקבלו עוד שתי שיחות טלפון, ובשלישית, לפני העברת כספים נוספת, מנכ"ל החברה שם לב כי המספר שממנו השיחות מתקבלות חשוד. הוא סיים את השיחה והתקשר למנהלו, שטען כי לא שוחח עימו טלפונית בסיטואציות שהוזכרו וכי מעולם לא ביקש ממנו לעשות עסקאות אלו. המקרה הועבר למחלקה המשפטית בחברה, אך עד שהבינו כי מדובר בהונאה מבוססת דיפייק של קול היה מאוחר מדי. הכסף הועבר ועד היום לא אותרו הפושעים ולא הכסף (Stupp, 2019). מקרה אחר אירע בינואר 2020, אז הונאה שהתבססה על טכנולוגיית דיפייק הצליחה להוביל לגניבה של 35 מיליון דולר מבנק באיחוד האמירויות הערביות. החוקרים בדובאי טענו כי מנהל בנק שקיבל שיחת טלפון ממי שזיהה לכאורה את קולו בתור דירקטור של חברה שעמו שוחח בעבר, התבקש להיערך לביצוע העברות כספים בסכום של מיליוני דולרים בעקבות רכישה שהחברה מבצעת. חלק ממה שהסיר חשד ממנהל הבנק הוא העובדה שהוא קיבל מיילים שהתחזו גם הם לאותו מנהל ולעורך דין שהיה אמון על העסקה. בעקבות

השיחה והמיילים הכספים הועברו לכמה חשבונות בנק במדינות שונות. מאז נחשפה זהותם של חלק מן הנאשמים בהונאה, אולם התיק לא נסגר (הלפרן, 2021).

הטכנולוגיות הנדרשות ליצירת הונאות מסוג זה זמינות ונגישות למגוון רחב של משתמשים ואף אינן מצריכות הקלטות קוליות ממושכות של המקור כדי להתבסס עליהן. ההתפתחויות האחרונות בטכנולוגיות רשתות נוירונים עמוקות, ובייחוד בטכנולוגיית GAN, סללו את הדרך ליצירת קולות סוריאליסטיים באיכות גבוהה ביותר, שאינם מובחנים באוזני אדם (David, 2021). חוקרים טוענים כי שילוב של הטכנולוגיות והיכרות עם מבנה החברה יחד עם שימוש במניפולציה רגשית כמו הפעלת לחץ זמן יכולים לגרום לכל אדם כמעט בלוע את הפיתיון (Noone, 2021).

מקרים אלו צריכים לשמש נורות אזהרה לעסקים וזאת בעקבות השימושים הרבים שיכולים להיות לטכניקה זו של הונאה, מעבר ליכולת הפגיעה הכלכלית שכבר הודגמה. בדרך שבוצעה הונאה זו אפשר היה להוציא במרמה ממנכ"ל החברה נתונים רגישים על עסקאות עתידיות, שמות לקוחות ועוד נושאים שעלולים להוביל לפגיעה כלכלית. בעולם של היום מאגרי נתונים יכולים להיות בעלי שווי רב; חברות כמו חברות ביטוח מחזיקות במסדי נתונים גדולים ורגישים שפרסומם יכול לסכן את לקוחותיהן. כן ניתן להשתמש במידע גם לשם סיכול פעילותה של החברה או לפגיעה בשמה הטוב (Chesney & Citron, 2019).

התבססות של גופים שונים כמו בנקים או שירותים ממשלתיים על טכנולוגיות זיהוי קולי למטרות אימות זהות, מסירת מידע ומתן שירותים עשויה להיות נקודת כשל רצינית, אם לא יבוצעו התאמות לזיהוי דיפ־פייק ולמניעת מרמה. זאת ועוד, עקב העבודה בחברות גלובליות ובעסקים חובקי עולם ולצד מגפת הקורונה, עסקים נעשו תלויים יותר ויותר בתקשורת ממרחק – טלפונית ומקוונת. שינויים אלו הופכים חברות וארגונים לפגיעים להונאות ומעלים את סיכויי ההצלחה של הפושעים (Graham, 2021). ניכר כי החקיקה הקיימת תוכל לסייע אם הזיוף שימש למרמה או לגניבה, אולם יש צורך בכלים המתאימים למניעה של תפוצת סרטוני דיפ־פייק או ליכולות של החוק להורות על בירור מקרים כאלה. כמו כן, לא ברור אם יתאפשר טיפול בתלונות על זיוף מסוג זה במקרה שלא ייגרם נזק, אלא רק תיגרם לאדם השפלה או שאמונו במציאות יתערער.

במקרה נוסף של דיפ־פייק למטרות מרמה שאינה כלכלית בוצע שימוש בסרטונים על ידי רפאלה ספון מארצות הברית. ספון השתמשה בטכנולוגיית דיפ־פייק ליצירת סרטונים של חברות קבוצת המעודדות של בתה. בסרטונים נראו הבנות מעשנות, שותות אלכוהול ואף בעירות – כל זאת במטרה להביא להדחתן כדי לשפר את מעמדה של בתה של ספון בקבוצה. את הסרטונים שלחה האם ממספרים חסויים למאמני הקבוצה, אך לאחר שהבנות נחשפו לסרטונים והכחישו כל קשר אליהם נפתחה חקירה, שבסופה נתפסה ספון והתגלה כי סרטונים אלו מזויפים (Ynet, 2021). גם מקרה זה מעלה את השאלה אילו כלים קיימים בידי הרשויות השונות, אם יידרשו לכך, להתמודדות עם מקרים המפריעים לסדר הטוב. ניכר כי שימוש בסרטונים מזויפים עלול להופיע כשאדם מתמודד על מקום עבודה או על משרה ציבורית, או בכל סיטואציה אחרת שעשויה להוביל לרווח של אחד על חשבון השפלתו או פגיעה בשמו הטוב של האחר. גם עצם האיום להפיץ סרטון עשוי לגרום לאדם שאינו מאמין ביכולתה של מערכת החוק והמשפט להגן עליו או לנקות את שמו, בפרט בסיטואציות שבהן יש רגישות לעניין הזמן, להיענות לסוחר ולו כדי למנוע את הפצת הסרטון, למרות היותו מזויף.

פורנוגרפיה

כיום השימוש הנפוץ ביותר בדיפ־פייק הוא בתעשיית הפורנו. משנת 2018 ועד דצמבר 2020, בין 90 אחוזים ל-95 אחוזים מהסרטונים שהופקו באמצעות יכולות דיפ־פייק והופצו ברשת הם בעלי אופי פורנוגרפי, ו-90 אחוזים מתוכם לא בהסכמת המופיעים בהם. ארבעת אתרי האינטרנט הפופולריים ביותר מבין אלו המפיצים פורנוגרפיה מזויפת זכו ליותר מ-134 מיליון צפיות (Patrini, 2019). למעשה אפילו המונח דיפ־פייק נטבע לראשונה ב-2017 על ידי מתכנת שהשתמש בטכנולוגיות בתחום הלמידה העמוקה (deep learning) כדי להטמיע פנים של ידועניות על גבי סרטונים פורנוגרפיים (אורפז, 2020). ללא התרעה מוקדמת, ידועניות רבות מצאו עצמן "משתתפות" בסרטים פורנוגרפיים ללא ידיעתן. אמה ווטסון הייתה אחת מהקורבנות הראשונים של תופעה זו, כאשר התפרסם סרטון פורנוגרפי עם פניה ללא ידיעתה (Stroud & Hayden, 2018). טכנולוגיה זו הפכה את היכולת ליצור תכנים פורנוגרפיים מזויפים לקלה ונגישה. אפליקציות למטרות אלו הופיעו שוב ושוב אף שנאסרו על ידי החוק במהירות בחלק מהמדינות, למשל בווייטנאם ובניו יורק. אפליקציית DeepNude שעלתה בשנת 2019 ובוט מסוג דומה שהופץ בטלגרם בשנת 2020 הם רק שתי דוגמאות ידועות לאפליקציות כאלה. הקוד המאפשר את יצירת האפליקציות "המפשיטות" נשים מבגדיהן עדיין מופץ במאגרי מידע שונים (Hao, 2021). אף על פי שיצירת סרטוני דיפ־פייק התמקדה בעיקר בידוענים, אנשים פרטיים הופכים גם הם למטרה בעקבות ריבוי התמונות המתפרסמות ברשתות החברתיות ומאפשרות יצירת סרטונים אלו (Delfino, 2019). בשל כך היקף סרטוני הפורנו המיוצרים באמצעות דיפ־פייק הלך וגדל, וכעת היעדים הם לא רק ידוענים אלא גם אנשים אנונימיים לחלוטין העומדים בפני סחיטה. לדוגמה, הבוט שהופץ באפליקציית טלגרם אחראי לבדו לכ-100 אלף קורבנות, בהם גם קטינות (Hao, 2021).

מקרים מן העולם של דיפ־פייק פורנוגרפי לשימוש פוגעני

מקרה ידוע של זיוף פורנוגרפי בוצע נגד עיתונאית ידועה בהודו בשם רנה איוב, המוכרת בזכות פרסומיה על אודות מקרים רגישים הנוגעים להפרות זכויות אדם ופשעים שבוצעו על ידי פקידי ציבור וממשל. ייחודיותה של איוב נובעת מהעובדה כי מלבד היותה עיתונאית חוקרת במדינה שבה הפער המגדרי עדיין גדול מאוד, היא גם מוסלמית ואנטי־ממסדית. באפריל 2018, לאחר שראיון עימה בנושא התעללות בילדים עורר זעם רב בקרב תומכי המפלגה הלאומית ההינדית השלטת (BJP), פורסם סרטון שנערך בעזרת טכנולוגיית דיפ־פייק ובו מופיעה איוב לכאורה כשהיא מככבת בסרטון פורנוגרפי. בסרטון שהועלה לרשת האינטרנט והופץ באמצעות רשתות חברתיות נראים פניה של איוב, כאשר בעצם מככבת בסרט שחקנית פורנו צעירה. תומכי ה-BJP שזעמו על איוב שיתפו את הסרטון המזויף באתר הרשמי שלהם, והדבר הוביל לכך שהוא נצפה והופץ על ידי עשרות אלפי אנשים. דווח כי תוך 48 שעות הופיע הסרטון ביותר ממחצית מהטלפונים הסלולריים בהודו (Citron, 2020).

למרות שיש מי שיטענו כי אין בדבר נזק אמיתי, בוודאי אם הוכח שמדובר בזיוף, רוברט צ'סני ודניאל סיטרון (Chesney & Citron, 2019) טוענים כי הלחץ הפסיכולוגי שמופעל על קורבנות של זיוף פורנוגרפי יכול להיות הרסני למדי. הקורבנות יכולים להרגיש מגוון של רגשות, החל מפחד והשפלה ועד התעללות פסיכולוגית. הדבר ניכר גם במקרה של איוב. מעבר לדיווח על הידרדרות בריאותה הפיזית בימים הראשונים

לפרשה, פרסום הסרטון גרם לה תחושת השפלה כה קשה שמנעה ממנה ככל הנראה את היכולת לתפקד. היא נמנעה מיציאה מביתה ואף הפסיקה את פעילותה כעיתונאית ואת הפעילות ברשתות החברתיות למשך חודשים ארוכים. לאחר ההתאוששות מהמקרה חזרה בהדרגה לעבודתה, אך בניגוד לעבר היא החלה להיזהר בכל פרסום ולצנזר את עמדותיה מתוך חשש מתמיד שהסרטון יופץ בשנית (Ayyub, 2018).

בראיון נוסף עימה היא סיפרה כי לאורך כל הקריירה העיתונאית שלה היא עומדת מול ניסיונות השתקה ופגיעה במוניטין שלה בדרכים שונות, אך פרסום הסרטון היה הפעם הראשונה שהצליחו לשבור אותה; הוא הצליח להחזיר בה פחד שלא הכירה לפני כן (The World, 2019). הדבר נובע מתפיסת האיום המוחשי שנוצרת אצל קורבנות דיפ־פייק פורנוגרפי – הסרטון ממחיש לקורבן באופן ויזואלי את התממשות האיום ואת הפגיעה בו (Chesney & Citron, 2019).

איוב סיפרה כי מהרגע שראתה את הסרטון, היא ידעה שבמדינה כמו הודו תהיה לו השפעה הרסנית על מעמדה. ואכן פרופיל הפייסבוק וחשבון הטוויטר של איוב הוצפו מייד באיומים קשים הקוראים לפגוע בה, לאנוס ואף לרצוח אותה (Citron, 2020). ההשפעה החזקה של הסרטון נבעה בין היתר מתהליכים בהודו, המתאפיינים בעליית הלאומנות והעוינות כלפי מיעוטים בכלל ומוסלמים בפרט. אקלים פוליטי זה, המנציח נרטיבים העולים בקנה אחד עם האידיאולוגיה הפוליטית השלטת, אפשר לסרטון הפורנו המזויף של איוב להפוך לפופולרי ואמין בקרב הציבור, כך טוענת איוב עצמה (Ayyub, 2018). נוסף על כך, הודו אינה המדינה הליברלית ביותר בכל הנוגע לזכויות נשים ולחירויות מיניות. לכן, כאשר איוב ניסתה להגיש תלונה במשטרה ולקבל הגנה מרשויות המדינה, התעלמו מבקשתה במשך חודשים ארוכים עד שהוגש לממשלת הודו דוח מיוחד מהא"ם, הדורש מהמדינה לספק לאיוב את ההגנה המגיעה לה (Citron, 2020).

סרטוני דיפ־פייק פורנוגרפיים שנועדו להשתיק, להפחיד או לנצל נשים הפכו לתופעה מקוונת חדשה בכל רחבי העולם ולא רק בהודו. גם בחברות דמוקרטיות שבהן יש שוויון מגדרי ברמה גבוהה יחסית, נשים מתקשות מאוד להילחם בתופעה (Farago, 2019). ה"ל מורט", כתבת ושרדנית בריטית נפגעה גם היא מתופעה זו. בסוף שנת 2020 גילתה מורט כי סרטוני פורנו וסרטוני אלימות מינית עם שמה ופניה מופצים באינטרנט, ואף מופץ מאגר תמונות שלה בקריאה לעוד יוצרי סרטוני דיפ־פייק ליצור סרטונים מסוג זה.

היא מתארת כי ההשפעה של מסע ההשפלה שאותם אנשים העבירו אותה הייתה עמוקה מאוד. היא שקלה למחוק לגמרי את כל חשבונותיה ברשתות החברתיות ובאינטרנט, והיא גם סובלת מפגיעה פסיכולוגית ומסיוטי לילה בעקבות חשיפתה לסרטונים שבהם נראה כי היא עוברת התעללות קשה. גם היא נתקלה במבוי סתום כאשר ניסתה לקבל את תמיכת הרשויות. בדומה לאיוב ניגשה מורט למשטרה, אך נאמר לה כי אין ביכולתם לעשות דבר לסייע לה. זאת מכיוון שהפצת סרטוני פורנו מזויפים אינה נחשבת כיום עבירה על החוק באנגליה (Hao, 2021; Jacson, 2021).

בשני מקרי הבוחן לא נמצא האדם העומד מאחורי יצירת הסרטון, ולכן גם אי אפשר למצוא פרטים על הטכנולוגיות שבהן השתמשו לשם יצירת הסרטונים. עיתונאי ניסה לבחון מה נדרש לשם יצירת פורנו דיפ־פייק. הוא מצא כי פורומים של יוצרים בעזרת טכנולוגיה זו נמצאים כבר בעמוד הראשון בחיפוש בגוגל, ובעזרת סרטון קצר של הקורבן והסכום הזעום של 30 דולר, ניתן להזמין סרטון פורנו מזויף של כל אדם (Jacoby, 2019). כמו כן, רוב סרטוני הזיוף שנערכו על ידי יוצרים עצמאיים מתבססים על אבולוציה של קוד המקור

ליצירת סרטוני זיוף פורנוגרפיים. קוד זה נוצר ופורסם על ידי משתמש בפורום Reddit בשנת 2017, שעל שמו אף נקראה הטכנולוגיה. משתמש זה התבסס על הספרייה הפתוחה של גוגל בנושא למידת מכונה ופיתח קוד ליצירת סרטוני פורנו מזויפים. לפיכך אפשר להעריך כי על בסיס קוד זה נוצר גם הסרטון במקרה של איוב ובמקרים דומים נוספים (Adee 2020).

שני המקרים שהוצגו כאן מעידים על חוסר היערכות של גורמי חוק במדינות שונות לטיפול בסוגיות כאלה. ניכר כי גם החקיקה בנושא חסרה במדינות רבות וכן חסרים כלי בדיקה ואכיפה, חלקם טכנולוגיים שאינם נמצאים בידי האמונים על שמירת החוק, למרות שניכר כי חלקם זמינים ממש כמו אלו המשמשים ליצירת הזיוף עצמו.

השפעה על פוליטיקה וביטחון לאומי

טכנולוגיית הזיוף העמוק מאפשרת שימושים שלחלקם השלכות בתחום הביטחון הלאומי והפוליטיקה. מדובר בסכנה מטרידה לפגיעה בחברה האזרחית או במדינות, על ידי יצירת מצגים שקריים למטרות עיצוב תודעה ודעת קהל כדי להשפיע על המערכת הפוליטית, על תהליכי בחירות ואולי אף על יציבותן של מדינות (Saylor & Harris, 2021). הזיוף העמוק הוא למעשה יישום טכנולוגי נוסף המעצים את אתגרי התקופה הקיימים בתחום הביטחון הלאומי, כאשר סרטוני הדיפֿייק יכולים לייצר מצג כוזב שעלול להטעות את דרג מקבלי ההחלטות ולהשפיע על דעת הקהל והציבור בהתאם לאינטרסים של יוצריהם (Schiff, 2019). לאור כל אלה, השימוש ביכולות דיפֿייק סומן על ידי ארגוני המודיעין האמריקאיים בשנת 2019 כאיום האסטרטגי הגדול ביותר על הביטחון הלאומי (Konkel, 2019; Ng, 2019).

השפעת סרטונים אלה על דעת הקהל והפוליטיקה העולמית עשויה להיות רבה, משום שלמרות ההתפתחויות הטכנולוגיות המתקדמות בהפקת סרטונים או קטעי שמע סינתטיים בשנים האחרונות, אנשים עדיין נותנים אמון רב בדברים מוחשיים שראו או שמעו, בייחוד כאשר מדובר בפנים ובקולות של אנשים מוכרים להם (Keitzmann et al., 2020). עם זאת, רק חלק קטן מתוך סרטוני הדיפֿייק נועד לתמך מצב פוליטי; רוב הסרטונים של דמויות פוליטיות כיום משמשים למטרות פרודיה או לחינוך הציבור לזהות סרטונים סינתטיים הנוצרים בטכנולוגיה זו (Dunn, 2021).

לפי דוח חברת Deeptrace שפורסם באוקטובר 2019, מבין סרטוני הזיוף העמוק שאינם פורנוגרפיים רק 12 אחוזים מהדמויות שהופיעו בהם היו של פוליטיקאים (Patrini, 2019). מכיוון שנדרשת כמות גדולה של נתונים לשם יצירת תוכן שלא יזוהה כזיוף, כיום עיקר השימוש בדיפֿייק מתמקד בדמויות מוכרות, אך עם התפתחות הטכנולוגיה הצפי הוא שהטכנולוגיה תהיה נגישה יותר וגם אנשים אנונימיים יהפכו לקורבנותיה.

מקרים מן העולם של ניסיון להשפעה פוליטית באמצעות דיפֿייק

בסתיו 2018 דווח כי נשיא מדינת גבון עלי בונגו חלה במהלך ביקור בערב הסעודית ונעלם מהחיים הציבוריים, בזמן שהמצב הפוליטי והכלכלי במדינה היה שברירי. חוסר במידע ודיווחים לא עקביים על מצבו יצרו חרושת שמועות על מצבו הבריאותי ואף על מותו של הנשיא. בערב ראש השנה 2019 פרסם הממשל בגבון סרטון שבו הנשיא נראה לכאורה נואם לאומה, לראשונה מזה שלושה חודשים (Brady, 2020). בסרטון נראה כי

תנועותיו של הנשיא מעט מכניות וקפואות ותנועות עיניו מוגבלות. מייד לאחר מכן נמסר בחדשות כי הסרטון זויף בעזרת טכנולוגיית דיפ־פייק, וכי השמועות בדבר מותו ככל הנראה נכונות. כשבוע לאחר מכן מתנגדיו של בונגו ניצלו את המצב וביצעו הפיכה צבאית שכשלה. בבדיקות שנעשו לאחר הפרשה לא נמצאה הוכחה שהסרטון היה מזויף, ועלו חשדות כי הופעתו השונה של הנשיא נבעה אולי מאירוע מוחי שעבר בתקופה זו (אורפז, 2020).

מקרה זה מדגיש את הפוטנציאל של השלכות טכנולוגיית דיפ־פייק על יציבותם של משטרים בעזרת ערעור תפיסת האמת. המקרה הוכיח כי ניתן ליצור השפעה לא רק באמצעות שימוש בטכנולוגיה, אלא עצם הידיעה כי הטכנולוגיה קיימת ואפשרית עשויה להוביל לערעור האמון גם במיצגים אמיתיים. במקרה של גבון ניכר כי תפיסה מעוותת זו תרמה למשבר פוליטי ולניסיון הפיכה צבאית (Delcker, 2019). עם זאת יש לשים לב לכך שנדרשות נסיבות ייחודיות על מנת שלסרטון מזויף תהיה השפעה פוליטית. במקרה של גבון הייתה השפעה משמעותית להיעדרות הממושכת של מנהיג מן הזירה הציבורית. נוסף על כך בונגו נחשד בעבר בעבירות הונאה ושחיתות – מה שהשפיע על אמון הציבור בפרסומיו (Brady, 2020). בד בבד עם המצב המדיני הרעוע במדינה מלכתחילה, ניכר כי במקרה מבחן זה היה מצע מצוין להשפעה שלילית של סרטון מזויף או לחשדות בסרטון מהימן.

מקרה אחר התרחש בישראל במסגרת הבחירות לכנסת ה־24, כשנעשה שימוש בקטע מזויף שבו נראה יאיר לפיד, מנהיג מפלגת יש עתיד, אומר דברים שלא אמר במציאות ואשר עלולים היו לפגוע בבחירה בו, משום שציירו אותו כמי שישתף פעולה עם פלגים מסוימים בחברה הישראלית. בסרטון הופיע סימון קטן שהעיד על היות הקטע המדובר מזויף, אולם לא היה קל להבחין בו. מפלגת יש עתיד עתרה לוועדת הבחירות המרכזית נגד הסרטון אולם לאחר העתירה, עוד בטרם פרסום הפסיקה בעניין, בוצע שינוי נוסף בסרטון על ידי יוצריו ונוספה בו שקופית גדולה המבהירה כי מדובר בזיוף לצורך המחשת רעיון. לאור זאת פסקה ועדת הבחירות כי אין בסרטון משום הפרה של הוראת סעיף 13 לחוק, המדברת על הטעיה מכוונת של הציבור, ולכן לא נדרשה הסרתו מרשת האינטרנט (תב"כ 24/9, 2021).

מקרה של זיוף נוסף לצורך השפעה פוליטית עירב שימוש בפורנוגרפיה. המפרסמים ניסו להשפיע על הבחירות לתפקיד מושל סאו פאולו, ברזיל. בסרטון שהופץ נראה המועמד המוביל באותה עת, ג'או דוריה, מקיים לכאורה יחסי מין עם חמש נשים צעירות. הסרטון הופץ ברשתות החברתיות ימים ספורים לפני מועד הבחירות. דוריה הכחיש כל קשר להאשמות וטען כי זהו סרטון מזויף שמטרתו פגיעה פוליטית. מדובר היה בפוטנציאל לפגיעה קשה במועמד, בעיקר בגלל הדעות השמרניות וחשיבות ערכי המשפחה שהבליט לאורך מסע הבחירות. דוריה נבחר כשם שניבאו הסקרים לפני הופעת הסרטון (Birchall, 2018). מקרים כאלה יכולים גם ללמד כי הטענה ל"זיוף", כאשר לציבור קשה להבחין או לבדוק בעצמו אם מדובר בזיוף או באמת, עשויה להוות מפלט קל גם למי שסרטון או הקלטה אמיתיים שלו יפורסמו ויביכו אותו. לשם כך נדרשים כלים לאימות ולמעורבות מצד העיתונות החוקרת ומוסדות המדינה הרלוונטיים, שיוכלו לאמת את המידע.

זאת ועוד, ראוי להתייחס גם להשפעה פוליטית בינלאומית פוטנציאלית. אף שכיום אין תיעוד גלוי של שימוש בדיפ־פייק שהוביל למשל להשפעה צבאית על הזירה הבינלאומית, על פי טוד מאירס, מוביל אוטומציה של מנהלת הטכנולוגיה במודיעין הגיאומרחבי הלאומי של ארצות הברית, סין היא המובילה העולמית ביצירת

סרטוני דיפ־פייק באמצעות שימוש בטכניקת GANs. ב־2017 השתמשו חוקרים סינים ב־GANs כדי לזהות כבישים, גשרים ועצמים אחרים בתמונות לוויין והנדסו את תצלומי הלוויין כך שישתרו אינטרס – כמו יצירת גשרים מזויפים לחלוטין כחלק מתמונת לוויין שנראית אותנטית לחלוטין. מאירס טען שהצבא והמודיעין האמריקאי יכולים להביס את טכניקת ה־GANs, אך לשם כך דרושים מאגרים כפולים ומשוכפלים של תמונות לוויין וראיות מחזקות נוספות (Tucker, 2019).

נכון לשנת 2021, עדיין יש צורך במחשוב יקר וביצרנים מנוסים על מנת לייצר סרטון אמין או זיוף של צילומי לוויין בצורה אמينة (Toews, 2020). אך כשמדובר במקרים של פוטנציאל להשפעה פוליטית, עשויים להיות בעלי אינטרסים שברשותם האמצעים הנדרשים ליצירת סרטונים כאלה. השאלה הנותרת היא מה תהיה רמת ההשפעה של סרטון כזה. התשובה היא ככל הנראה שגם אם סרטון יהיה אמין דיו, יש צורך בהפצתו בסיטואציה מתאימה כמו למשל חוסר יכולת של הגורם המופיע בסרטון להכחיש, או בסיטואציה שבה יש קריאה לפעולה שתוביל לפעולה מיידית, או במקרים שבהם יש חשיבות לזמן (למשל זמן קצר לפני ההצבעה במערכת בחירות, בצורה שאינה מאפשרת בירור). בשלל סיטואציות כאלה עשויים המפיצים להשיג את מטרותיהם, יהיו אשר יהיו, במחירים נמוכים יחסית. עם זאת, עדיין לא ידוע על מקרה מוצלח שבו ניסתה מדינה אחת להשפיע על אחרת בעזרת סרטון מזויף שנוצר באמצעות למידה עמוקה.

לבסוף, האתגר שמציב דיפ־פייק לביטחון הלאומי נוגע בעיקר לסכנה הפוליטית הנשקפת. לאור זמינות הטכנולוגיה ופשטות היצירה של תוצרי דיפ־פייק שונים, קיימים שלל תרחישים שעלולים להוביל לערעור הסדר הציבורי, לאלומות או אפילו לאובדן אמון הציבור גם במסרים אמיתיים שדרושים לניהולה התקין של חברה בכלל, ושל חברה דמוקרטית בפרט.