מבט על, גיליון 1205, 22 באוגוסט 2019

טכנולוגיות נגד פייק ניוז: כשל שוק או הזדמנות מפוספסת?

ענבל אורפז

מדינות, גופי תקשורת וארגונים מתמודדים בשנים האחרונות עם הקלות שבה ניתן להפיץ שקרים, עיוותים, ספינים, טעויות ותיאוריות קונספירציה (פייק ניוז). ועם זאת, מספר הסטארטאפים הפועלים לפיתוח פתרונות בתחום הוא מצומצם. קול קורא מטעם תוכנית ליפקין-שחק במכון למחקרי ביטחון לאומי מיפה כ-20 סטארטאפים ישראלים, הפועלים בתחום בשלוש קטגוריות שונות: זיהוי רשתות בוטים; זיהוי מידע שקרי ובדיקת עובדות; והקטגוריה שצוברת תאוצה לאחרונה – זיהוי ומניעת הפצתם של קטעי וידאו וקול מזויפים (Deepfake). נראה, שמיעוט החברות המסחריות הפועלות לפיתוח פתרונות בתחום נובע מכשל שוק המביא לכך שלמרות שקיים אינטרס ציבורי למנוע הפצת פייק ניוז, אין גופים שיש להם אינטרס חזק דיו לשלם עבור פתרונות טכנולוגיים. תחום שבו עשויה בכל זאת להיווצר הזדמנות עסקית, בין היתר להקמת סטארטאפים, הוא פתרונות שיאפשרו לתאגידים ומותגים למנוע הפצה של פייק ניוז, העלול לפגוע במוניטין שלהם ולהסב להם נזק פיננסי ותדמיתי.

מאמר זה מסכם ומנתח את ממצאי סקר שנערך לאחרונה במסגרת תכנית המחקר ע"ש ליפקין-שחק: ביטחון לאומי ודמוקרטיה בעידן של פוסט אמת ופייק ניוז. מטרת הסקר הייתה לברר מהם הפתרונות שכבר קיימים להתמודדות עם בעיית הפייק ניוז ולמפות סטארטאפים וגופים שפועלים בתחום זה בישראל. הסקר נערך, בין היתר, באמצעות קול קורא שהופץ ברשת ושיחות עם חברות ומשקיעים. הממצא העיקרי היה שלמרות שמדובר בתופעה שנמצאת בכותרות מזה שלוש שנים, מספר מצומצם יחסית של חברות פיתחו טכנולוגיות ופתרונות בתחום. המחקר שנערך בתוכנית התמקד בישראל, כמדינה שמרכזת פעילות יזמית טכנולוגית במגוון תחומים. אולם, זוהו בעולם מגמות דומות לאלו הנראות בישראל.

בסך הכל, זוהו כ-20 סטארטאפים ישראליים המפתחים טכנולוגיות ומוצרים רלבנטיים. ממצאי הבדיקה מעלים שישנם מספר אזורים בהם מתמחים הסטארטאפים הפועלים בתחום, והם כוללים: טכנולוגיות הקשורות לזיהוי הפצה של מידע שקרי כחלק מקמפיינים להפצת דיסאינפורנציה באמצעות זהויות מזויפות (ובכלל זה בוטים); טכנולוגיות ומיזמים המתמקדים בבדיקה האם מידע הוא אמיתי או שקרי, דוגמת מיזמי בדיקת עובדות (fact checking); וטכנולוגיות לזיהוי תוכן שקרי, בתחום הווידאו והתמונות (Deepfake) לצורך בניית חסינות נגד בינה מלאכותית התקפית (Defensive AI), המשמשת ליצירת שקרים וספינים.

לשלוש הקטגוריות הללו השלכות ברורות על היכולת לקיים שיח פוליטי נקי מהשפעות ומהתערבויות חיצוניות. בנוסף, הן גם משפיעות על היכולת של מקבלי החלטות (בדרג המדיני ובדרג הצבאי) לקבל החלטות הקשורות לביטחון הלאומי.

זיהוי רשתות בוטים

אחד התחומים הבולטים שהגיע לכותרות במערכות הבחירות האחרונות ברחבי העולם, הוא קמפיינים להפצת דיסאינפורמציה על אודות מועמדים במגוון אמצעים, ובכלל זה באמצעות הפעלת רשתות בוטים ופרופילים מזויפים ברשתות החברתיות. ואכן, חלק מהחברות הבולטות בתחום מפתחות פתרונות מבוססי טכנולוגיה לזיהוי רשתות בוטים. ביניהן Cyabra, חברה ששיתפה פעולה עם ועדת הבחירות המרכזית לבחירות לכנסת ישראל, שנערכו באפריל 2019. Cyabra מנסה לפתור את בעיית הפייק ניוז באמצעות זיהוי הזהויות הפיקטיביות שמפיצות אותם ברשתות החברתיות. המערכת של החברה מסתמכת על טכנולוגיות של למידת מכונה, בינה מלאכותית וניתוח שפה טבעית, כדי לנתח התנהגויות ותוכן ולהכריע האם משתמש הוא פיקטיבי או אמיתי. אם מזוהה משתמש פיקטיבי, החברה מתריעה בפני הלקוח - מותג או ארגון ביטחון לאומי - וכך למעשה משמשת ככלי מודיעין. המשך הטיפול במשתמשים הפיקטיביים תלוי בהחלטת הלקוח.

חברה נוספת שפיתחה אלגוריתמים לזיהוי רשתות בוטים היא Communit360 שעוסקת בניהול מוניטין ברשתות חברתיות עבור מותגים, ועבדה גם עם משרד החוץ הישראלי. בין היתר, החברה יצרה כלי לעיתונאים לדיווח על בוטים ומשתמשים פיקטיביים במהלך הבחירות לכנסת ישראל. זאת, כדי להתמודד עם ניסיונות הפצה של מידע כוזב באמצעות פניה לעיתונאים בטוויטר והזנתם במידע שקרי.

<u>זיהוי שקרים ושקרנים</u>

אשכול נוסף של חברות עוסק בבירור המציאות ובזיהוי שקרים ושקרנים. בקבוצה זו פועלים חברות ומיזמי תוכן, שאינם מבוססים על טכנולוגיה עמוקה ואשר מציעים מנגנוני דירוג לתוכן במטרה לאמת את רמת אמינותו. דוגמה לחברה כזו היא Rootclaim, שמאפשרת למשתמשים להזין עדויות בנוגע לסוגיה שנויה במחלוקת ולהכריע לגביה. אתר The Perspective שם לעצמו למטרה לאפשר לאנשים לצאת מתיבת התהודה בה הם נמצאים באמצעות הצגת שתי זוויות שונות לאותו סיפור ומתן אפשרות לגולשים לבחור בגישה שהם בוחרים לסיפור. עוד בקבוצה זו, ניתן למנות מיזמים בתחום בדיקת העובדות (facts checking), דוגמת "המשרוקית", ששולבה באתר **גלובס** ומספקת שירותים לפייסבוק בישראל, שבמסגרתם היא מנטרת ומתקנת הפצה של מידע כוזב מסוגים שונים בפלטפורמה בשפה העברית. הפתרון של משרוקית מבוסס על עבודת אנליסטים, המאמתים מידע שמופיע בכלי התקשורת.

סטארטאפים נוספים פיתחו פתרונות יותר מתוחכמים טכנולוגית לזיהוי שקרים ושקרנים. כך למשל, VineSight - סטארטאפ צעיר שמבצע בדיקת עובדות אוטומטית וכך מזהה פייק ניוז באמצעות מתן ציון אמינות (credibility) לכל פיסת תוכן. החברה מיישמת את הטכנולוגיה לבדיקת אמינות של ציוצים המפורסמים בטוויטר, והיא מתכוונת לבחון את הטכנולוגיה שלה לניטור מידע המתפרסם על מועמדים לקראת מערכת הבחירות לנשיאות ארצות הברית שייערכו ב-2020. זוהי למעשה גישה נוספת לזיהוי רשתות בוטים - באמצעות זיהוי מידע רשתי שמופץ על ידי משתמשים. הסטארטאפ AdVerif.ai ניגש לבעיה מגישה שונה ומציע פתרון למפרסמים ולרשתות פרסום באינטרנט, שאינם מעוניינים לבזבז את תקציבי הפרסום שלהם באתרים המיועדים להפצת פייק ניוז. החברה מבצעת זיהוי מבוסס בינה מלאכותית (AI), שבוחן את כתובת האתר והתוכן, מבצע ניתוח סמנטי של הסנטימנט וקובע האם מדובר באתר שמקדם פייק

– ואם כן מונע את הפרסום בו. כיום, היכולת הטכנולוגית לזהות ידיעות כוזבות מפגרת אחרי היכולת של המכונות שמייצרות ידיעות שקריות.

Deepfake

הקטגוריה השלישית והצומחת היא של סטארטאפים שמפתחים יכולות בתחום של Deepfake - הלחם של המונחים פייק ניוז ו-Deep Learning, טכנולוגיה עמוקה בתחום למידת המכונה. לקראת הבחירות לנשיאות ארצות הברית החל הקונגרס האמריקאי לחקור את התחום והשפעותיו על מערכת הבחירות. אחד האירועים שהעלה את הנושא לכותרות היה וידאו מזויף שהופץ, בו מופיעה ראש בית הנבחרים של ארצות הברית, ננסי פלוסי, כשהיא מדברת באיטיות, ספק שיכורה או דמנטית, ופייסבוק סירבה להסיר את הסרטון. וידאו זה לא עירב יכולות Deepfake, ונעשה בו שימוש פשוט בהאטה של סרטון, אך הוא המחיש את פוטנציאל הנזק של סרטונים מזויפים. לקראת השימוע בנושא, הכריז אדם שיף, יו״ר ועדת המודיעין בקונגרס שחוקרת את הנושא, ש-Deepfake יכול להיות ההסלמה החמורה ביותר של קמפיין ההתערבות הרוסי בבחירות באמצעות פייק ניוז. אחת הבעיות בתחום זה היא שיכולות הבינה המלאכותית ההתקפית ליצירת זיופים מתקדמות בקצב מהיר יותר מיצירת פתרונות ההגנה מפניהם.

בצד אחד של הקטגוריה, ישנן חברות שמפתחות יכולות ליצירת קטעי וידאו או קול מזויפים ברמה גבוהה, שנראים אמיתיים, ולמעשה מפתחות יכולות התקפיות בתחום. אחת הדוגמאות לכך היא Canny.AI, שבאמצעות הטכנולוגיה שלה כבר נוצרו קטעי וידאו מזויפים של מנהיגי העולם, כולל דונלד טראמפ, ולדימיר פוטין ובנימין נתניהו, וכן של מנכ״ל פייסבוק מארק צוקרברג. מנגד, בדומה לעולם אבטחת המידע, יש חברות שמפתחות יכולות הגנה נגדיות, המשמשות כעין ״אנטי וירוס״ – כלומר: כלים לזיהוי וידאו ותמונות מזויפים. חברת Serelay הישראלית, לדוגמה, מאפשרת למי שצילם תמונה או וידאו בטלפון להוכיח לנמען שהם אמיתיים, למשל גופי מדיה או חברות ביטוח - שני מקרים בהם יש ערך לאותנטיקציה.

מאין יגיע המימון למניעת פייק ניוז?

המחקר שנערך בתוכנית ליפקין שחק התמקד בסטארטאפים ישראלים. אולם, גם ברחבי העולם נצפות תופעות דומות – מיעוט יחסי של חברות בעולמות הפייק ניוז, שעוסקות בבירור אמת ושל השקעות שגייסו ממשקיעים. חלק מהחברות שפעלו בתחום נרכשו על ידי הפלטפורמות הטכנולוגיות, דוגמת פייסבוק וטוויטר. זאת, לצד התעוררות בפעילות ועניין גובר בתחום ה-Deepfake, שם מכוני מחקר וסטארטאפים מפתחים יכולות לזיהוי סרטונים והקלטות מזויפות, בעיקר לקראת הבחירות לנשיאות ארצות הברית.

נשאלת השאלה - מדוע אין אנו רואים יותר סטארטאפים פועלים בתחום? מערכות הבחירות בשנים האחרונות, ואירועים אחרים, הוכיחו את הצורך הבהול לייצר פתרון לבעיית הפייק ניוז, בפרט לקלות הפקתו והפצתו של מידע שקרי המשפיע על מדינות וארגונים. ברמה הציבורית - ברור שיש אינטרס להבטיח שלא יופץ מידע כוזב באופן מכוון על ידי גופי תקשורת ועל ידי גורמים פרטיים, שבימי הרשתות החברתיות יש להם שופר הפצה זמין לתוכן שיכול להגיע להמונים.

ואולם, מי הם הלקוחות שהיו יכולים לשלם עבור פתרונות כאלה? המועמדים הראשונים הם גופי התקשורת, שלהם צורך בכלים לאמת במהירות ובאמינות מידע שמגיע אליהם. אך גופים אלה

מחפשים בעצמם בשני העשורים האחרונים מודל עסקי חדש שייצר ערוצי הכנסה חדשים בעידן הדיגיטל. רוב גופי המדיה בעולם מתקשים לשרוד בעולם הדיגיטלי, ומבחינתם השקעה במערכות וטכנולוגיות למניעת פייק ניוז היא בגדר מותרות.

גם הפלטפורמות הדיגיטליות שמשמשות במידה רבה להפצת מידע שקרי, ובהן פייסבוק, טוויטר וגוגל, היו יכולות להתעניין בפיתוח פתרונות וטכנולוגיות בתחום. כפי שכבר הוכיחו מקרים בתקופה האחרונה, הפלטפורמות הטכנולוגיות יכולות לשמש רגולטור שמחליט שהוא אוסר ומונע הפצת תוכן שקרי ומזויף באמצעות מדיניות, ובלי להזדקק לטכנולוגיות מתוחכמות. אולם, עד כה נראה שסל הפתרונות וההשקעות שלהן בתחום הוא מצומצם, ושהן מתמקדות בסגירת פרופילים המזוהים כבוטים. בעניין זה מושמעת לעיתים הטענה הגורסת שלמעשה, הפלטפורמות לא מעוניינות במניעת הפצת פייק ניוז, כיוון שהן מרוויחות מתנועת גולשים אליהן.

אם לאף גורם אין אינטרס או יכולת לשלם עבור מניעת פייק ניוז, האם אנו נמצאים במצב של כשל שוק? כלומר, למרות הצורך בפתרונות לבעיית הפייק ניוז, אין גורם שמעוניין לממן אותם. האינטרס הציבורי הוא לכן לפתח כלים שיבטיחו שלא ייווצר מידע שקרי ויופץ להמונים, בוודאי בתחומים שמשפיעים על השיח הציבורי, בריאות הציבור, ביטחון לאומי ועוד. אך לגופים המסחריים ולשוק הפרטי אין כיום מספיק תמריצים להשקעה בתחום ולקדם את הנושא. תחומים אלה אינם מקבלים פתרון ומענה ראוי וגם לא נמצאים תחת פיקוח גופים ממשלתיים וציבוריים, שהיו יכולים להתערב במטרה לפתור את בעיית כשל השוק. יתכן שהתשובה תגיע מכיוון של רגולציה.

נראה שהתחום שבו בכל זאת ניתן לצפות להשקעות בסטארטאפים הוא בפיתוח פתרונות טכנולוגיים עבור ארגונים שמעוניינים להגן על המוניטין שלהם וירצו להגן על עצמם מפני הפצת מידע שקרי וקמפיינים שליליים נגדם. בעידן הדיגיטלי, מידע שקרי בנוגע לחברה יכול לגרום לה נזק פיננסי ניכר ולכן יש לה אינטרס להשקיע בתחום, בדומה להשקעות במערכי הגנת סייבר שנועדו להגן על הנכסים הדיגיטליים והמידע בארגונים. אולם, מסתבר שתחום זה עדיין בתחילת דרכו וכיום לא נמצא תחת סמכות של גורם ייעודי בארגונים ולא מוקצים לו משאבים.