



"אין יותר איפה להתחבא. אנחנו עירומים"

איך מצותתים לאלפי גיגה בשנייה? כמה זמן ייקח עד שכל שיחות הטלפון בעולם ינותחו למטרות מודיעיניות? ואיך מלמדים מחשב להבין אורדו? ראש המוסד לשעבר **שבתי שביט**, ופרופ' עמי מויאל שמפתח אלגוריתמים לזיהוי קולי, מספקים הצצה אל תעשיית המודיעין הממוחשבת. G בעקבות חשיפתה של מערכת הציתות הגלובלית ← **שחר סמוחה**

פרופ' עמי מויאל. השתתף בתחרות של משרד ההגנה האמריקאי לפיתוח טכנולוגיה

ביום ראשון בבוקר, יומיים לאחר שממשלת ארצות הברית האשימה אותו בריגול המור ודרשה מהונג קונג להסגירו, עלה האיש המבוקש בעולם על טיסה מספר 213 של אירופלוט הרוסית. רק לאחר כמה שעות, כשהמטוס יצא משמי סין, שחררה ממשלת הונג קונג הודעה לעיתונות, שלפיה ארוארד סנאורן עזב את שטחה. עד שהריפולומטים הסוכנים האמריקאים המומים הבינו לאן מועדות פניו, עברו עוד כמה שעות. כשנחת האייירבוס 330 בשרה התעופה שרמטייבו שבמוסקבה, המתינו על המסלול חמש מכוניות גדולות ולבנות ששלח נשיא רוסיה, ולדימיר פוטין. ביטחוננו וכבודו של מי שמצליח להביך את המערב מובטחים ברוסיה של פוטין, וסנאורן, שלפני שבועיים חשף לעולם את קיומה של מערכת הציתות החשאית הגדולה ביותר בהיסטוריה, בהחלט נמצא ראוי לחסות נשיאותית. לפני שתתברר כנכונה, נראית האינפורמציה ששחרר דובר וויקיליקס באשר לצעדיו הבאים של סנאורן כמו תרגיל בדיסאינפורמציה. לדבריו, איש המחשבים המבריק שעבד בחברת ייעוץ שסיפקה שירותים ל-NSA - הסוכנות שאמונה על פעולות איסוף המודיעין האלקטרוני של זרועות המודיעין האמריקאיות - ימריא בשבוע הבא לקובה. משם הוא ימשיך לוונצואלה או לאקוודור. והיה גם דיבור על איסלנד.

כך או כך, סנאורן מודע למגבלות המקלט העתידי שלו. בראיון הראשון שנתן הוא הצהיר שהוא מודע לכך שהוא עשוי לשלם בחייו על מעשיו. ברדלי מאנינג, החפ"ש שהעביר לוויקיליקס מאות אלפי מסמכים שהביכו עד אימה את ארצות הברית, נעצר במאי 2010, החזק מאז בכידוד מוחלט וככל הנראה ירצה עשרות שנות מאסר. המידע שחשף סנאורן נפיץ פי כמה. הוא לא רק מביך את ארצות הברית ובעלות בריתה לקואליציית ההאונות הגלובליות (בריטניה, קנדה, אוסטרליה וניו זילנד), אלא גם נתפס כפגיעה חמורה בביטחונה הלאומי של המעצמה החזקה בעולם וביכולתה להילחם בטרור הבינלאומי.

סיפור חשיפתה של מערכת הציתות הגלובלית - PRISM שמה - בנוי מכמה וכמה סיפורים. הוא מפנה ורקור אל שאלת מחויבותו של הפרט כלפי עצמו וכלפי האדיאולוגיה הפרטית שלו מול נאמנותו לקולקטיב שבקרב הוא חי; הוא מעלה שאלות כבדות משקל בנוגע לאיזון העדין במשטרים דמוקרטיים בין הרצון לשמור על חיי אזרחיהם לבין המוכנות לפגוע לשם כך בזכויותיהם הבסיסיות ביותר; הוא עוסק ביחסי הגומלין הנסתרים בין ממשלות לחברות מסחריות שהולקות אינטרסים משותפים, שעיקרם הרצון לדעת עליכם הכול; הוא משרטט את קווי המתאר של עתיד קרוב, שבו שאריות פרטיותו של האדם הפשוט יוקרבו למען מדיניות כוזב או אחרת, והוא מספר את סיפור ההתפתחות המדעית-טכנולוגית המרהימה שעברה בעשור האחרון קהילת המודיעין הבינלאומית.

כמה קל להאזין?

טכנולוגית, שני כוחות מנוגדים השפיעו בשנים האחרונות על המאמצים המודיעיניים של מדינות המערב. מצד אחד, עליית האינטרנט יצרה תעבורת תקשורת אדירה, שטורם נראתה כמוה בהיסטוריה של המין האנושי. בעולם האינטרנט, התקשורת לא רק הפכה לזולה להפליא אלא גם לאנונימית. טכנולוגיות הסתרה כאלו ואחרות גרמו לבלבול מפחיד: אימייל שלכאורה נשלח ממינסוטה, עשוי היה באמת להישלח מפקיסטן ושיחת סקיפ שלכאורה מתנהלת בין שני משתמשים ממוגרישו שבסומליה, עשויה במציאות להתנהל בין שכנים בפרבר של לונדון. בעולם מרכזיות הטלפונים של פעם, כל זה לא היה אפשרי, וארגוני הביון עמדו בפני בעיה שצריך היה לפתור.

מהצד השני, ההזולה הניכרת בעלות האחסון, אפשרה לארגוני הביון לשמור אצלם כמויות לא נתפסות של



למה נגד טרור ולא גם נגד פשיעה?

← מיד לאחר חשיפתה של מערכת PRISM על-ידי אדוארד סנאודן יצאו זרועות הממשל השונות למתקפת יחסי ציבור. העובדה שזרועות המודיעין האמריקאיות צותתו לאזרחים אמריקאים ולא רק לתושבי הכפר הגלובלי הכעיסה מיליוני אמריקאים ואחד האנשים שנקראו לדגל היה כמובן ראש ה-NSA, הגנרל קית' אלכסנדר. בעדותו בפני הקונגרס אמר אלכסנדר שבזכות מערכת ההאזנות נמנעו עשרות פיגועים. באופן פרטני יותר, הוא הזכיר פיגוע שנמנע בוול סטריט. השתכללות הטכנולוגיה מעלה שאלות מוסריות לא מעטות. מצד אחד ניתן לטעון שאם הממשל סבור שמערכות מודיעיניות דוגמת PRISM יעילות במניעת פיגועים, אזי מדוע לא לאפשר להן לפעול גם בתחומים אחרים כמו מלחמה בפדופילים, בארגוני פשע מאורגן, בשודדי בנקים וכו'.

מצד שני, חוששים ארגוני זכויות אזרח שהיכולות הטכנולוגיות המוכחות של מערכות כמו PRISM יאפשרו בעתיד הקרוב למדינות לבצע מעקבים יעילים אחר אזרחים שפשוט אינם מרוצים מהתנהלותן של ממשלותיהם. "אני לא רוצה להשוות את הטכנולוגיה הזו לכלי נשק כי לכל דבר יש שימושים טובים ורעים", אומר פרופ' מויאל כשהוא נשאל אם הוא חש דילמה מוסרית כשהוא מפתח את הטכנולוגיה שכה מעניינת אותו. "כבן אדם וכאזרח אני חושב שבשימוש נאות אין נזק, אבל כמובן שיש משטרים טוטליטריים בעולם שמשתמשים בכלים כאלה כדי לעקוב אחר כל אחד מהתושבים. במדינות המערביות הנאורות אני רוצה להאמין שעושים בזה שימוש רק כדי להגן על התושבים. התחושה הנוחה שלי יש עם זה היא שגם אם מקליטים את כל השיחות, הרוב המכריע שלהן נותר במחשב ואין שמיעה אנושית. אני רוצה להאמין שרוב השיחות שאותרו על-ידי מנגנוני חיפוש מילות המפתח ונשמעו על-ידי בן אדם היו ראויים להישמע והן באמת מסביב לאותו קהל יעד שצריך להאזין לו, אבל ברור שיש התראות שווא ושבתהליך הזה יישמעו שיחות שלא היו צריכות להישמע".

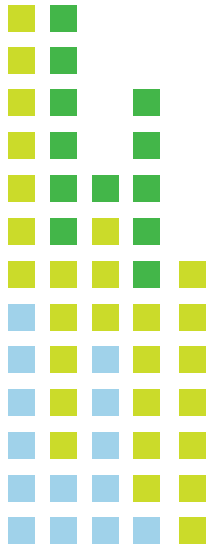
מויאל מסכים שהפיתוי לשנות את הקריטריונים שעל פיהם מאוינים לאנשים בהחלט קיים ואומר שארגוני אכיפה כמו משטרות, מאוד מתעניינים בטכנולוגיה. "ברמה האישית מניעה אותי סקרנות מדעית ואהבה של התחום. התחושה החזקה שלי היא שהייתי רוצה לעזור לארגונים הביטחוניים שיש לנו פה. אני יודע שאחוש סיפוק רב מאוד אם אצליח לעזור לכוחות המודיעין שלנו ואדע שמשוהו שפיתחתי הציל חי אדם. זה יחפה על כל יתר הדברים ואני מקווה מאוד שלא יעשו שימוש לא נאות ביכולות הללו, אבל זה בהחלט יכול לזלוג. אין ספק".

אבל לא בשלה - חלו בה שינויים מרחיקי לכת. "התחום של זיהוי דיבור מאוד מורכב ובעבר היו בו הרבה חטאים של שימוש", הוא אומר. "היו הרבה שירותים שהיו להם למשל ביצועים של 70% שנכנסו לשוק. בעבר ניסו לקחת את זה למערכות כמו How may I help you? בטלפון וכולנו מכירים את זה מסיינפלד", הוא צוחק.

כל זה השתנה בשנים האחרונות כדי לסייע לפתור בעיות שאיתן מתמודדות כמה וכמה תעשיות. "כולנו יודעים, למשל, שכיום אנחנו לא יכולים לחפש בווידיאו. תחשוב שיש לך, למשל, גאוס של אובמה שנמשך שלוש שעות ואותך מעניינת רק הרקה וחצי שבה הוא דיבר על הסכסוך הישראלי-פלסטיני. הגישה המקובלת היא לתמלל את הפסקול ואז אתה יכול לעשות חיפוש בטקסט שיביא אותך למקטע הקטן של התוכן שרצית".

מויאל מסביר שכדי שמחשב יוכל לזהות מילים מסוימות בשיחה, משתמשים בשיטה מסודרת

שביט. אינו מוכן לפרט אילו יכולות בתחום זיהוי הדיבור היו בידי המוסד בתקופתו



צילום: חטי מילוד - עיתון הארץ

שביט: "אפילו הארגונים הגדולים ביותר לא מסוגלים להעמיד את כוח האדם שיוכל לכסות את כל התעבורה התקשורתית של התפקיד של הטכנולוגיה הוא לפתח תוכנות שיעשו את הסינון - להוציא את המוץ מהתבן"

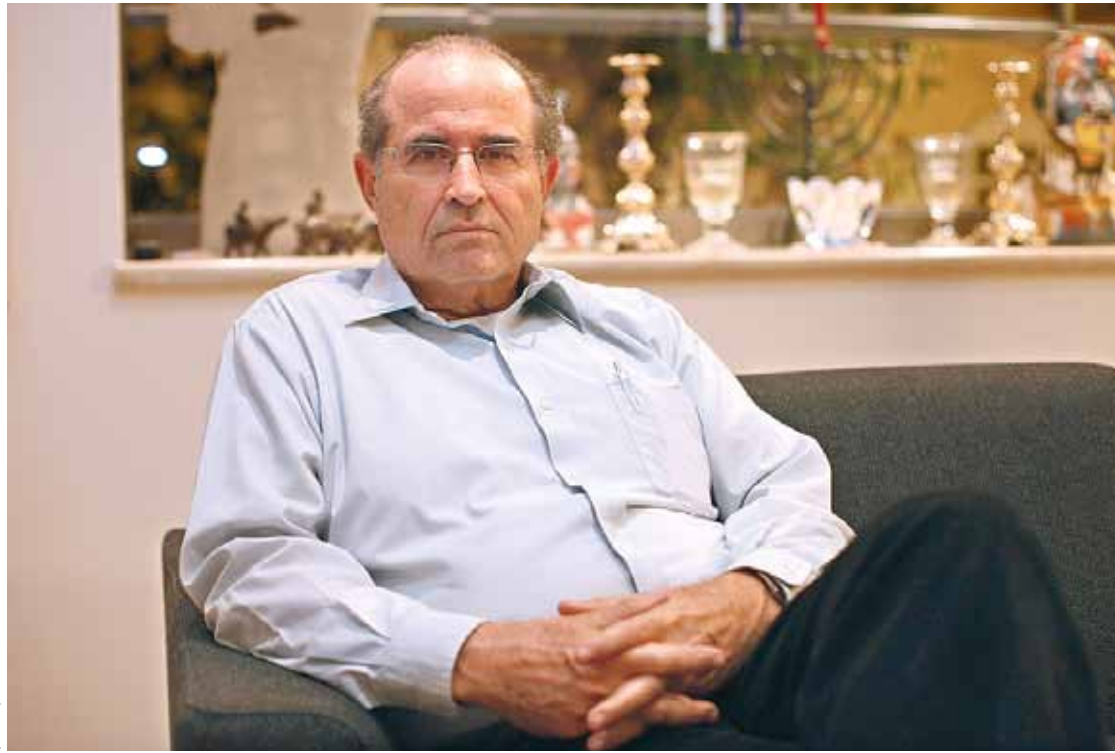
משפטים למחשב, ואז לסרוק את האינפורמציה. אם המילים שחפשת מופיעות בה, המערכת מוציאה את המידע לטיפול מיידי של אנליסטים אנושיים", מסביר שביט. טוב ויפה, אבל איך עושים את זה?

פרופסור עמי מויאל ממכללת אפקה להנדסה, הקדיש את כל חייו המקצועיים לתחום זיהוי הדיבור. הוא למד וחקר את התחום במסגרת שלושת התארים שעשה באוניברסיטת בן גוריון, עבר לתעשיית ההייטק המקומית והתקדם בה מתפקיד כותב אלגוריתמים ועד למנכ"ל חברה. בשלוש השנים האחרונות הוא עומד בראש מרכז המחקר והפיתוח לזיהוי דיבור שהקים במכללה ונמצא בקשר הדוק עם חברות מקומיות ועם ארגוני המודיעין הישראליים: אמ"ן, השב"כ והמוסד.

"בעד לי בעצמות להקים מרכז מו"פ בתחום כיוון שהתחום הלך ורעך באקדמיה בארץ", מסביר מויאל את המעבר מהשוק הפרטי לאקדמיה. "כדי לבצע מחקר משמעותי בתחום הזה אתה צריך להקים קבוצה גדולה כי זה תחום אינטרדיסציפלינרי - בעולם עובדים עם צוותים של עשרות אנשים. כשהייתי בתפקידי מחקר ופיתוח בשוק הפרטי, חיפשתי שותפים באקדמיה כדי לדאוג לפיתוחים לטווח הבינוני והארוך ולא מצאתי. כשכבר מצאתי שותפים, המנגנונים של החתימה על זכויות קניין רוחני היו כאלה מורכבים שאף פעם לא הצלחתי ליצור שיתוף פעולה אמיתי".

את המשיכה לתחום מסביר מויאל בכך ש"יש בו שילוב בין תוכנה וביצוע סימולציות לבין שימוש באלגוריתמים מאוד-מאוד חכמים, שנושקים לתחומים של אינטליגנציה מלאכותית וחיפוש - תהליכים של מחשב שלומד לבר. זה קוסם לי כי אני נמשך מאוד לתחום של תקשורת אדם-מכונה. אני מאוד מאמין בהתפתחות של סוג התקשורת הזה. בחזון שלי אני אתקשר עם רובוטים כפי שאני מתקשר עם אנשים. הרובוט יזהה את הדיבור שלי, הוא יזהה את האינטונציה שלי, הוא יבין אם אני כועס או מתרגש. תהיה לו גם יכולת לדבר איתי. זה חזון שיתממש בצורה מדורגת. בעוד כמה שנים יהיה רובוט בכל בית עם פונקציונליות בסיסית. הבעיה תהיה תמחור ולא טכנולוגיה. בתכנון יפה, מערכת זיהוי קולי שתוכל לתת מענה מושלם למשל במוקד עירוני יכולה לעבוד כבר היום. אם תנתת את מוקד 106 תראה שלא צריך שם יותר מכמה אלפי מילים. במידה של חוסר זיהוי - של נניח 5% - השיחה תועבר למוקד אנושי. זו לא בעיה".

מויאל מסביר שמאז הימים שבהם טכנולוגיית זיהוי הדיבור הופיעה לראשונה בחיינו - מבטיחה



מידע. המידע הושג באמצעות שיתוף פעולה עם ענקיות הטכנולוגיה האמריקאיות, אבל מרגע שהגיע לידי ארגוני הביון נדרש היה לסנן ממנו את החומר בעל הערך המודיעיני.

רק כדי לסבר את האוזן, הנה כמה מספרים שהופיעו במסמכי NSA שחשף סנאודן: ב-2011 עמדה תעבורת האינטרנט בין צפון אמריקה לאירופה על 4,972 גיגה-בייט לשנייה. בין צפון אמריקה לאסיה הוחלפו מדי שנייה 2,721 גיגה-בייט, ובין צפון אמריקה לדרומה טסו בכל שנייה נתונה כ-2,946 גיגה-בייט. תעבורת המידע בין אירופה לאפריקה, למשל, נמוכה בהרבה ועמדה באותה שנה על כ-343 גיגה-בייט לשנייה. הנתונים הללו כוללים את כל סוגי התקשורת המוכרים לנו: שיחות טלפון על גבי תשתית האינטרנט, אימיילים, צ'טים, שיחות קוליות בתוכנות כמו סקיפ, שיחות וידיאו בכל המסגרות למיניהם והעברות קבצים. כך או כך, מדובר בהיקף לא נתפס של מידע, שמאז 2011 רק הלך וגדל.

"בסביבה התקשורתית שאנו חיים בה יש התפוצצות של אינפורמציה", אומר ראש המוסד לשעבר שביט, שעומד כיום בראש חברת אתנה שמספקת שירותים ביטחוניים ומודיעיניים לגופים פרטיים ולממשלות. "אפילו הארגונים הגדולים ביותר והעשירים ביותר לא מסוגלים להעמיד את כוח האדם שיוכל לכסות את כל התעבורה התקשורתית מצד אחד, ומצד שני ברור שגם אי אפשר לחסום את התקשורת. לא טכנולוגית ולא מבחינת הנורמות הציבוריות. אז איך מתגברים על הכמות הזאת של האינפורמציה? התפקיד הראשון של הטכנולוגיה הוא לפתח תוכנות שיעשו את הסינון - להוציא את המוץ מהתבן".

התוכנות שאליהן מתכוון שביט, שעמד בראש ארגון הביון הישראלי בין 1989 ל-1996, עוסקות בפענוח דיבור אנושי באמצעות זיהוי של מילות מפתח מסוימות. זה נשמע מסובך, אבל בפועל זה הרבה יותר מסובך. כל מי שנתקל במערכות אוטומטיות לזיהוי דיבור שנפוצות כיום במגוון רחב של אפליקציות - מסירי של אפל ועד מוקדי הזמנות כרטיסים לקולנוע - יודע שהן רחוקות מלהיות מושלמות, ולא סתם: לאנשים שונים יש מבטאים שונים, אינטונציות שונות וכמובן שהם גם דוברים שפות שונות. השירותים האחרים שאנו חשופים אליהם מכירים מספר מוגבל של מילים שרלבנטיות לתחום העניין שלהן. ארגוני מודיעין, לעומת זאת, רוצים לנתח דיבור ספונטני.

"הרעיון של טכנולוגיית איתור מילות המפתח הוא להזין מילה אחת או מספר מילים או מספר

ראש ה-NSA, הגנרל קית' אלכסנדר. מנעו פיגוע בוול סטריט



צילום: רויטרס



הכנס במכללת אפקה

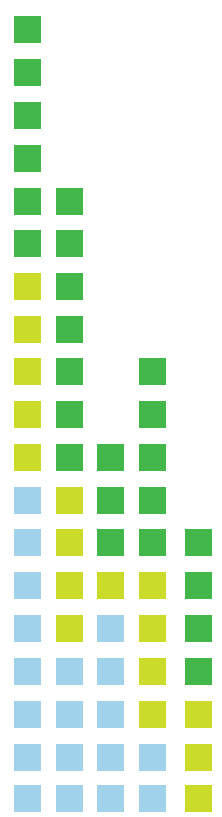
← ביום שני הקרוב יתקיים במכללת אפקה כנס שיפגיש את הכוחות החדשניים ביותר בתחום זיהוי הדיבור בישראל עם כמה מומחים בינלאומיים שיגיעו לארץ. בכנס יוצג לראשונה המחקר של מויאל עם ניסוי על זיהוי שפות אקזוטיות וכן מחקרים אחרים בנושאים כמו זיהוי דיבור מעוות ולא ברור, זיהוי רגשות, אתגרים של זיהוי דיבור בעברית, פיתוחים של זיהוי דיבור בתחום הבריאות, שירות לקוחות ורכב ועוד. בין באי הכנס יהיו שניים מהחוקרים החשובים בעולם בתחום - פרופ' לורנס רבינר ופרופ' צ'ין הו לי. מויאל: "ישנן בישראל חמשי"ש חברות שעוסקות בפיתוח טכנולוגי נטו ועוד עשרות חברות שמשמשות בזיהוי קולי, במיוחד בעידן הטלפונים החכמים. משום שהקמנו פה מרכז ידע שעוסק בתחומים כמו ניתוח דיבור, זיהוי מין הדובר ועיבוד שפה, כל הסטרטגיה האפיק הקטנים באים הנה כדי להיות איתנו בקשר. אנחנו עוזרים לתעשייה לבחור מנוע, לעשות פיילוט וגם לפתח את הטכנולוגיה למוצר. אני בונה להם גרסה אלגוריתמית. האפליקציות לא מעניינות אותי."

נרות תמיכה בסנאודן בהונג קונג. ידע שעלול לשלם בחייו

להסתמכות חלקית על חברות אורחיות", הוא מסביר. "מה שמאיץ את התהליך הוא שבחברות אורחיות שעוסקות בטכנולוגיות רלבנטיות לנושאים מודיעיניים יש הרבה מאוד אנשים שהם יוצאי המערכות הבטחוניות. בישראל העניין הזה בולט במיוחד. קשה היום למצוא חברת הייטק אחת שהיא רצינית שאין בה יוצאי מערכת והעובדה הזאת מקלה על מקבלי החלטות לסמוך על החברות הללו."

שביט אינו מוכן לפרט אילו יכולות בתחום זיהוי הדיבור היו בידי המוסד בתקופתו, אולם אומר שהוא מכיר את הנושא "למעלה מעשר שנים". לדבריו, תהליך הפיתוח של הטכנולוגיה היה ארוך והדרגתי. "ארגונים כמו זה ששירתי בו מעוררים אנשים לחשוב מחוץ לקופסה - לבוא עם הרעיונות הכי פרוצים. לאחר מכן עושים מיון וסינון ומבחינים בין מה שנשמע אמיתי לבין מה שהווי לחלוטין. כשבא אליי לראשונה אדם שמציע הצעה שעל פניה עשויה להישמע כהווייה, אני לא נופל מהכיסא. כשהאינטרנט בא לעולם והפך את התקשורת לזמינה וזולה מאוד בהשוואה לאמצעים אחרים, היה ברור לי שהמדינה הזו תלך ותשתלט על העולם והיה לי גם ברור שנמצא פתרון להתמודד איתה. מה שיפה בטכנולוגיה הזו הוא שהמשתמש הסופי קובע את הסרגל של מילות המפתח ויכול להכתיב לתוכנה באיזה ימים או שעות או נסיבות המחשב יחפש תעבורה שמילות המפתח יופיעו בה. אפשר לגרוע מילים, להוסיף מילים. זה נהדר וזה משנה מאוד את עבודת המודיעין שהייתה בעבר."

שביט מספר שבנוסף לתחומים הקלאסיים שבהם עוסקת חברת אתנה שהקים, היא כוללת גם בית תוכנה עצמאי שפיתח מערכות העוסקות בניהול מידע. "בהתחלה הטכנולוגיה הייתה לביטחון ולמודיעין והיום היא משרתת גם בנקים וגופי אישור. כמו כן, אנחנו משקיעים בפיתוח טכנולוגיה לזיהוי קול וזיהוי דובר, שזו למעשה הרחבה של טכנולוגיית זיהוי המילים. המטרה היא שנרדע להוציא מאוקיינוס האינפורמציה את המידע הרלבנטי לצרכים שלנו ושנרדע לזהות קול ודובר. אני מאמין שזה יהיה אפשרי טכנולוגית תוך שנה עד שנה וחצי והמשמעות היא קפיצה משמעותית מאוד באיכות המודיעין. בעולם הזה אין יותר איפה להתחבא. אנחנו עירומים."



מויאל: "המטרה של התחרות הייתה למצוא פתרון לזיהוי דיבור ב-300 שפות שונות שמעניינות את זרועות המודיעין של ארצות הברית"

הביטחוניות של ניס, שמקבלת כסף מהמדינה הראשי. "אם רוצים לעשות זיהוי דיבור בשפה אקזוטית מבלי ליצור בסיס נתונים קלאסי, האלגוריתם נדרש להשתמש בשפות אחרות הקיימות ולפתח טרנספורמציה שמבוססת על הידע הקיים. תוך שבוע אתה כבר מקבל דמו - אמנם לא באותה איכות של השפות הקיימות - אבל מספיק כדי לעבוד איתו ולשכלל אותו."

אבל גם בין השפות ה"אקזוטיות" ישנם הברלים. "בתחרות אתה לא יודע אפילו איזו שפה תקבל והתפללתי שלא נקבל שפה טונאלית שבה לאותה מילה יש משמעויות שונות לפי השימוש בטונים שונים. מה שכמובן קרה זה שקיבלנו וייטנאמית - שהיא שפה טונאלית. ארבעה שבועות לא ישנו ואימנו את האלגוריתמים שלנו ולאחר מכן קיבלנו מהאמריקאים חומר לזיהוי ושבוע לעבוד עליו. אתה מקבל קובץ עם 75 שעות מוקלטות, שממנו הם רצו שאזוהו 4,000 מילים. מה שאתה צריך לתת להם זה מסמך שבו כתוב איפה מצאת את המילים על גבי הטיימליין של הקבצים ועד כמה אתה בטוח בכונות שלהם."

הצוות של מויאל אמנם לא ניצח בתחרות, אולם הוא משוכנע שהמחקר שלו מתקדם בכיוון הנכון. "החלום שלי הוא לקחת קטע דיבור טבעי, ולעשות עליו Diarization, מלשון יומן - מי דיבר בכל אחד מן המקטעים. זה יאפשר שמפתחות החיפוש העתידיים יהיו משהו כמו 'תן לי נאום של ברק אובמה שבו הוא דיבר על מכללת אפקה והייתה אינטונציה נורא שמחה בדיבור שלו'. כמה אנחנו רחוקים משם? מבחינת זיהוי רגשות בדיבור ברמת החיובי או השלילי, אנחנו יכולים לתת מענה של 80%-90%; מבחינת תמלול של תוכן אפשר להגיע לביצועים של 85%. בסך הכול אפשר לקבל תוצאות זיהוי טובות. באמצעות משוב מהשוק ומהלקוח אפשר להתקדם במהירות."

כשמויאל מדבר על משוב מהשוק, הוא מתכוון לא רק לשוק הפרטי, אלא גם לשוק הביטחוני. העובדה שהן ארגוני ביון והן חברות ענק כמו גוגל או פייסבוק מתעסקות בבסיסי נתונים אדירים ושעיקר עיסוקן הוא מציאת דרכים להציף מידע בעל משמעות, הפכה אותם לשותפים טבעיים. שביט, ראש המוסד לשעבר, שמכיר את שני העולמות מקרוב, אומר כי בכל העולם המערבי מתקיים בשנים האחרונות תהליך הדרגתי של הסתמכות ארגוני ביון על טכנולוגיות שמפתחות בגופים אורחיים. "בעבר ארגוני מודיעין היו מפתחים את כל צורכי הטכנולוגיה שלהם בבית, אך בשנים האחרונות אנחנו רואים מעבר הדרגתי

למדי. "יש מכניזם אימון, כמו שלילד יש כשהוא לומד שפה. ההגדרה הייתה ליצור בסיס נתונים של לפחות 200 שעות מוקלטות של הרבה מאוד דוברים בתנאים שונים; משהו שמע את כל ה-200 שעות ותמלל אותן ואז כל המידע הזה הוזן לאלגוריתם שניתח אותו". עם זאת, אומר מויאל, הקמת בסיס נתונים נפרד וייחודי לכל שפה הוא עסק יקר ומורכב - לפחות 100 אלף דולרים והרבה שעות אדם. וגם אז מדובר רק בשלבים הראשונים של לימוד המחשב.

ומה עם פרסית?

באופן אירוני, בעוד שקיימים פתרונות טובים למדי לשפות שבהן יש פוטנציאל עסקי - שפות מערביות שאותן דוברות אוכלוסיות בעלות כושר קנייה - קיימת בעיה קשה ביכולת לנתח שפות שמעניינות באמת את ארגוני המודיעין בעולם. דארי, פשוט, פרסית, אורדו ואחרות הן רק כמה מהן. תחרות מקצועית שערכה לאחרונה הזרוע המודיעינית של דארפה - סוכנות המחקר והפיתוח של הפנטגון - משרטטת את הכיוון שאליו התחום ככל הנראה יצעד בשנים הקרובות. מויאל: "בפברואר האחרון היה אצלי אורח מארצות הברית שהציע לי להשתתף בתחרות. המטרה שלה הייתה למצוא פתרון לזיהוי דיבור ב-300 שפות שונות שמעניינות את זרועות המודיעין של ארצות הברית מבלי להקים מאגרי מידע נפרדים לכל אחת מהן. הרעיון הוא להקטין תוך חמש שנים את פרק הזמן שנדרש כדי ללמוד שפה חדשה ב-80%. אם השנה עברנו ארבעה שבועות כדי לאמן את המחשבים, המטרה היא שבועות חמש שנים ניתן יהיה להיכנס לשפה חדשה תוך שבוע עם בסיס נתונים מינימלי". מויאל כבר עבד על פרויקט דומה יחד עם אנשי החטיבה הביטחוניית של חברת ניס, אולם אומר שכשראה אילו גופים משתתפים בתחרות שארגנה דארפה הוא התלבט מאוד אם להשתתף בה. "אנחנו ארגון קטן מאוד וראיתי שיהיו שם גופים כמו קיימברידג', ימ ואוניברסיטאות עשירות מאוד מסין ומסינגפור. צריך להבין שצוות של גוף כזה מונה בדרך כלל בסביבות 15 איש בעוד שאני יכול לגייס לעניין הזה שניים וחצי או שלושה אנשים ולהעביר אותם מסביב לשעון. בעוד אני מתלבט התקשרו אליי ממפא"ת (מינהל פיתוח אמצעי לחימה ותשתית טכנולוגית במשרד הביטחון) ושאלו אותי אם אני מתכוון להשתתף בתחרות. אמרתי להם שעם מימון אני מוכן ללכת על זה וכך יצאנו לדרך."

השיטה שבה פעל מויאל התבססה על אלגוריתם ייחודי שהעבודה עליו מומנה על-ידי החטיבה