



## ביג דאטה בעבודת המודיעין - תפקידים, יישומים ומגבלות

שירה פתאל - חוקרת במכון לחקר המתודולוגיה של המודיעין

השימוש הראשון במונח ביג דאטה נעשה ככל הנראה בשנת 1997 במהלך כנס של ה־IEEE, אגודה בין־לאומית של העוסקים בהנדסת חשמל ובאלקטרוניקה.<sup>1</sup> כאשר משתמשים במונח ביג דאטה מתכוונים לרוב למידע בהיקפים גדולים. עם זאת, נפח המידע הוא עניין יחסי התלוי בתקופת הזמן (סביר כי מה שנתפס בעבר כמידע רב לא ייתפס כך היום).

אף שאין הגדרה אחת מוסכמת לביג דאטה, יצוינו שתי הגדרות שהופיעו במחקרים שונים:

- **McAfee ו־Brynjolfsson** מגדירים ביג דאטה דרך שלושה מרכיבים: **היקף (Volume)**, **מגוון (Variety)** ו**מהירות (Velocity)**.<sup>2</sup>
- **Schonberger-Mayer ו־Cukier** מציעים שלושה עקרונות משלהם להגדרת ביג דאטה: **מעבר מאיסוף חלק מהמידע לאיסוף כל המידע, אי־סדר (messiness) ומתאמים (קורלציות)**. מטבע הדברים, יותר מידע כולל גם יותר טעויות, אך לשיטתם היעדר סדר אינו מהווה בעיה "we no longer need to worry about individual data points biasing overall the analysis".<sup>3</sup> הדיגיטציה של המידע היא תנאי הכרחי, החיוני להפיכת המידע ליחידות מדידות, המאפשרות את ניצולו בדרכים חדשות (Datafication).<sup>3</sup>

**אם כן, ביג דאטה הוא מידע בהיקפים גדולים מאוד ממקורות מידע שונים ומגוונים המאוחסן לאחר שעבר תהליך דיגיטציה המאפשר אחזורו במהירות, באופן שמאפשר ליצור ערך משמעותי ולהצביע על מתאמים בין מרכיבי המידע.**

מאמר זה מתמקד בבחינת השימושים של ביג דאטה בעבודת המודיעין על בסיס המאמר Beyond the buzzword: Big data and national security decision-making ב־Chatham House ונכתב על ידי Damien Van Puyvelde, Stephen Coulthart, M. Shahriar Hossain.<sup>4</sup> המאמר מגדיר מהו ביג דאטה ומהם תפקידיו בסוגיות הנוגעות לביטחון

1 Gil press, "Very Short History of Big Data", Forbes, 21.12.2013 <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#3dbce0fe65a1>

2 John McAfee – תובניתן אמריקאי-בריטי, מייסד חברת אבטחת המידע McAfee. Erik Brynjolfsson – פרופסור אמריקאי בבית הספר לניהול ב־MIT, ראש היוזמה ל־Digital Economy ב־MIT ומנהל במרכז ל־Digital Business, גם הוא ב־MIT.

3 Viktor Mayer-Schönberger – פרופסור במכון האינטרנט של אוניברסיטת אוקספורד; Kenneth Neil Cukier – עיתונאי אמריקאי. יחד חיברו את הספר *Big data: A revolution that will transform how we live*.

4 Damien Van Puyvelde Dr – מרצה באוניברסיטת Glasgow שבסקוטלנד על מודיעין וביטחון לאומי, ותחומי המחקר שבהם הוא עוסק כוללים: לימודי מודיעין וביטחון, ביטחון לאומי בארצות הברית, אחריות בדמוקרטיה (Democratic Accountability) ואבטחת מידע. בין תפקידיו משמש גם במרצה אורח בבית הספר של הצי האמריקאי ללימודים מתקדמים.

לאומי. נעסוק גם בשאלה האם כלים אוטומטיים יכולים להחליף את הידע האנושי. זאת, זאת, על בסיס המאמר *Fusing algorithms and analysis: Open-source intelligence in the age of 'big data'* שפורסם בכתב העת *Intelligence and National Security* ונכתב על ידי Christopher Eldridge, Christopher Hobbs, Matthew Moran.<sup>5</sup>

## ביג דאטה בעבודת המודיעין

מאגרי מידע גדולים מאפשרים לזהות מגמות, תבניות וחריונות מהשגרה. הטענה היא שניתוח ביג דאטה עשוי לסייע לעבודת המודיעין, אך גם להציב אתגרים משמעותיים, ובכל מקרה לא יוכל להחליף את המומחיות והידע האנושיים. כדי לבחון את תרומתם של מאגרי ביג דאטה, המאמר ב-Chatham House מתייחס לכמה יכולות הניצבות בבסיסו של התהליך המודיעיני. בראש ובראשונה, ההתפתחויות הטכנולוגיות מאפשרות לאסוף מידע, גלוי או מסווג, בהיקפים גדולים בהרבה מאשר בעבר. לכאורה מדובר ביתרון משמעותי, אך קהילות מודיעין ברחבי העולם מצביעות על כך שנוצר "היצף מידע", העלול לגרום לפספוס מרכיבי מידע חשובים.

**עיבוד המידע ומחקר** – כותבי המאמר מתייחסים לעיבוד ומחקר כאל שתי פעולות נפרדות. עיבוד מידע, לשיטתם, היא פעולה טקטית יותר והמחקר הוא התהליך החשיבתי אשר מייצר תובנות. (בראיית, נכון להתייחס לשתי פעולות אלה ביחד, מכיוון שתופעת הביג דאטה מצמצמת את ההבדלים בין עיבוד המידע, שבדרך כלל נמצא באחריותם של גופי האיסוף, לבין המחקר, שעליו אמונים גופי המחקר וההערכה). לאור המגוון הרחב של מאגרי המידע, נדרשים כלי עיבוד שונים, למשל: תמלול שמע, זיהוי גורמים חשודים ומתן התרעה לגביהם (על ידי למידת מכונה, למשל), זיהוי מילות מפתח בטקסט, וכן הפיכה של מידע לא מובנה למובנה.

ניתוח מאגרי ביג דאטה מסייע במציאת מתאמים ומאפשר לקשור בין אירועים או פעולות מבודדים לכאורה, ובכך להגביר את המודעות להקשרים רחבים יותר. ניתוח מאגרי ביג דאטה מסייע מאוד במענה על שאלות מסוג: מי? מה? מתי? ואיפה? אך קשה לגזור ממנו תשובה לשאלות המכוונות למציאת סיבה (למה?).<sup>6</sup> טענתם של כותבי המאמר מתבססת על מחקר של Kevjn Lim<sup>7</sup> לפיו ניתוח מאגרי ביג דאטה מסייע בגיבוש הערכות לטווח הארוך, על ידי זיהוי טרנדים ומגמות, וכן בבניית תזות ובהפרכתן.

5 Dr. Christopher Eldridge - עמית מחקר במרכז ללימודי מדע וביטחון ב-King's College שלבולדון. חוקר את האינטראקציה שבין תרבות, מדע וטכנולוגיה.

Dr. Christopher Hobbs - מנהל במרכז ללימודי מדע וביטחון במחלקה ל-War Studies שב-King's College בולדון. תחומי המחקר שבהם הוא עוסק כוללים בקרת נשק, טרור בלתי קונבנציונלי, נשק להשמדה המונית, אוסינט, אבטחת מידע.

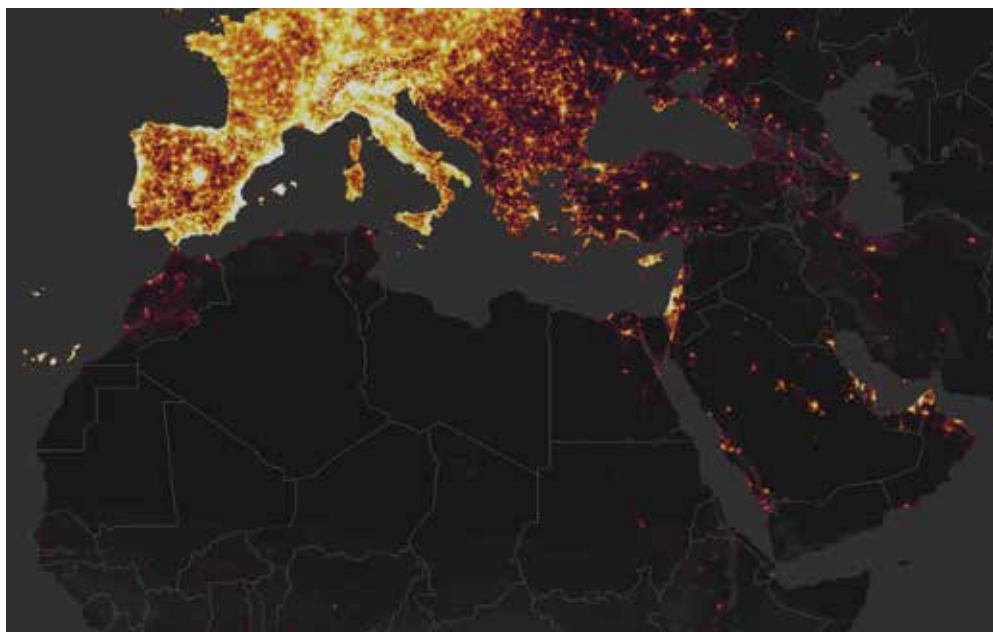
Dr. Mathew Moran - מנהל במרכז ללימודי מדע וביטחון במחלקה ל-War Studies שב-King's College בולדון. תחומי המחקר שבהם הוא עוסק כוללים מחקר של התקוממויות (collective disorder and rioting) ונשק להשמדה המונית, בדגש על גרעין, לצד אוסינט.

6 Mayer-Schonberger ו-Cukier שחוזר קודם לכן, טוענים בספרם כי החיפוש אחר קורלציות זול ומהיר יותר בהשוואה לחיפוש אחר הסיבה.

7 חוקר בתחומי מדיניות חוץ ומדיניות ביטחונית במזרח התיכון.

ניצול מאגרי ביג דאטה מחייב את אנשי המודיעין ליצירתיות, להיכרות עם מאגרי מידע שונים וליכולת לשאול את השאלות הנכונות. לאחרונה, למשל, נחשף כי אפליקציית הכושר הגופני Strava המאפשרת למשתמשים לעקוב אחר פעילותם הגופנית ולשתף אותה עם אחרים, פרסמה בנובמבר 2017 מפה המציגה את פעילות משתמשיה בכל רחבי העולם, בהתבסס על שלושה טריליון מיקומי GPS. בעקבות פרסום המפה זוהו בסיסים אמריקאיים סודיים ברחבי העולם כשבמדינות, כמו אפגניסטן, שבהן יש מעט מאוד משתמשים, זוהו במפות החום נקודות מוארות העשויות להעיד על כוחות זרים.<sup>8</sup>

### איור 36: מפת החום באזור המזרח התיכון המקור: מפת החום של Strava



דוגמה אחרת, ניתן למצוא בדבריו של ראש IARPA,<sup>9</sup> Jason Matheny, שהסביר כי אחת האינדיקציות לזיהוי התפרצות של מחלות, אתגר אשר קהילת המודיעין האמריקאית נדרשת אליו, היא ביטול הזמנת מקומות במסעדות. לכן, IARPA מנטרת אפליקציות זמינות, דוגמת OpenTable (חברה המספקת שירותי הזמנת מקומות במסעדות ברחבי העולם); כמו גם חיפוש תסמינים במנועי חיפוש, תיאור של תסמינים ברשתות חברתיות, או פרסומים של משתמשים ברשתות חברתיות בדבר אי-הגעתם לעבודה.

8 "Have you signed up for a tracking app by mistake?", The Guardian, "29.1.2018

<https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases>

9 Intelligence Advanced Research Projects Activity – IARPA – משקיע בתוכניות מחקר חדשניות כדי להתמודד עם אתגרים בין-סוכנותיים שונים. תחתיו פועלים ארבעה מרכזים: איסוף, מחקר, חיזוי (anticipatory intelligence) המפתח טכנולוגיות המסייעות בתחום זה ומרכז מחשוב המתמודד עם איומי סייבר.

## קביעת התעדוף המודיעיני

היכולת לזהות ממאגרי ביג דאטה מגמות וחריגות משגרה עשויה לאפשר זיהוי סוגיות מודיעיניות מתהוות או דעיכה של סוגיות אחרות, ובכך להשפיע על התעדוף המודיעיני ללא הכוונה חיצונית.

## הפצת התוצרים

הכותבים טוענים כי שיטות וכלים לניצול מאגרי מידע גלויים מאפשרים להנגיש את התוצרים. הם מציעים, למשל, הסבה של אפליקציות המציגות לגולשים באינטרנט תכנים שעשויים לעניין אותם על בסיס חיפושי עבר, אך שישמשו צרכני מודיעין ודרגי קבלת החלטות, על ידי התאמת התוצרים המוצגים להם על בסיס זיהוי הנושאים שמעסיקים אותם. לצד זאת, ישנם כלים המאפשרים הצגה חזותית חכמה של מידע, באופן המסייע לפשט סוגיות מורכבות ומסייע בקליטתן.

## האדם והמכונה

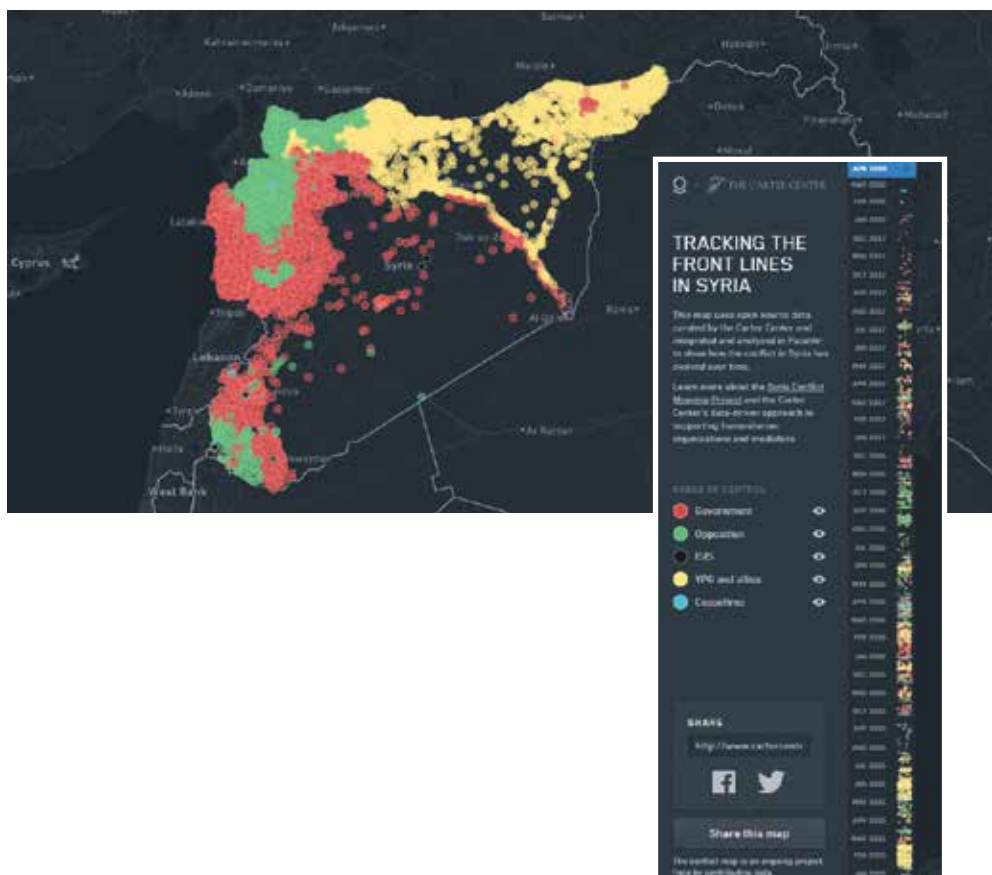
רבים דנים בשאלה האם תופעת הביג דאטה מחייבת להגדיר באופן שונה את החלוקה הבסיסית של המפעל המודיעיני, למשל בין גורמי איסוף, עיבוד ומחקר. מרבית העוסקים בסוגיה זו מסכימים כי התופעה מאפשרת להפוך פעולות רבות, שבעבר בוצעו ידנית, לאוטומטיות, ובכך להטיל על המחשב למלא פונקציות מסוימות אותן מילאו עד לאחרונה בני אדם. בה בעת, יש הסבורים שאף שהביג דאטה מהווה מכפיל כוח משמעותי, הוא צריך לשמש כלי משלים למומחיות האנושית ולא להחליפה. יש אף הטוענים כי לא רק שהידע האנושי נדרש, אלא גם שתופעת הביג דאטה מגבירה את הצורך במומחיות האנושית ולא מצמצמת את הצורך בה. כך, השיפוט של מומחי תוכן נדרש כדי להעריך האם מתאמים שזוהו על ידי כלי ביג דאטה, רלוונטיים או נכונים (ולא נובעים מהטיות). כותבי המאמר ב-Chatham House טוענים כי לחוקרים תפקיד מרכזי בהכנסת התובנות העולות מניתוח מאגרי ביג דאטה להקשר הנכון, ולשילובו עם "data small". במאמר שפורסם ב-Intelligence and National Security מוצגת תפיסה הגורסת כי עד שלא יתפתח תחום "הבינה המלאכותית הכללית" (Artificial General Intelligence), תידרש מומחיות אנושית בהגדרת תזות ובהצעת תשובות אפשריות במענה לשאלות מורכבות. תחום "הבינה המלאכותית הכללית" שנמצא בשלבי פיתוח ורחוק מלהגיע לרמת בשלות, אמור לאפשר למכונות לבצע כמה פעולות קוגניטיביות במקביל, בדומה לבני אדם. מערכות הבינה המלאכותית הקיימות כיום מסוגלות לבצע רק פעולות נקודתיות מסוימות.

בו בזמן מתפתחות גישות ותפיסות העוסקות בשילוב בין המכונה לבין האדם. אחת מהן התפתחה תחת התחום של הנדסה קוגניטיבית המכונה "Joint Cognitive System"<sup>10</sup>. תפיסה זו כוללת תחת הגדרת המערכת את בני האדם ואת המחשבים גם יחד – המערכת בכללותה נדרשת למלא פונקציה מסוימת, וכדי לממש אותה יש לבחון מהי הדרך הטובה ביותר לנצל את כלל מרכיביה. הדגש הוא על הפונקציה שממלאת המערכת ולא על המבנה והמרכיבים שלה. דוגמה לשילוב הגורם האנושי ניתן למצוא בתוכנה שפותחה על ידי מומחה למדעי ההתנהגות

10 הנדסה קוגניטיבית היא תחום בהנדסת אנוש העוסק בהתאמת עיצובם של כלים לאופי וליכולות הקוגניטיבית של המשתמש בהם.

ומומחה למדעי המחשב במכון Rand לצורך ניתוח טקסטואלי של מאגרי מידע גדולים. זאת מתוך התפיסה כי המומחה למדעי ההתנהגות יכול להשתמש בהבנתו את השפה כדי להצביע על תופעות ייחודיות, ואילו המומחה למדעי המחשב מבצע תהליך הפוך המתחיל בזיהוי המאפיינים הייחודיים של טקסט ספציפי, ולאחר מכן מגבש תובנות רחבות יותר. באופן זה הצליחו לאפיין את השפה המשמשת את תומכי המדינה האסלאמית ואת מתנגדיה, הצליחו לזהות אינדיקציות טקסטואליות להתחזקות המדינה האסלאמית במצרים על בסיס שימוש בטרמינולוגיה דומה לזו המשמשת את הארגון בכלל, וכן להבחין בין השפה המשמשת את המדינה האיסלאמית לבין השפה המשמשת את אל־קאעידה.<sup>11</sup>

**אזור 38: מפת שליטה בסוריה של "The Carter Center" הכוללת ניתוח של מודיעין גלוי.** המפה מתבססת על למעלה מ־100,000 אירועים, מתעדכנת באופן אוטומטי וקרוב לזמן אמת, תוך שימוש בתוכנה Palantir. המפה מאפשרת להשוות גם בין המצב בשטח כפי שהופיע במפות שליטה קודמות ברזולוציה גבוהה.



11. Doug Irving. "Big Data Big question". The RAND Blog. 16.10.2017 <https://www.rand.org/blog/rand-review/2017/10/big-data-big-questions.html>