

אחזור מידע מתקדם בעידן הביג דאטה

**ד"ר חיים אסא - ארכיטקט מערכות מבוססות
אינטליגנציה מלאכותית**

מקורות המידע

המודיעין ניזון ממקורות איסוף מסוגים שונים. מקורות אלו מובחנים זה מזה על פי קטגוריות שונות של מאפייני הנגישות למידע - חזותי, סיגינטי, אנושי, גלוי וכדומה - אך בכל קטגוריה גם נכלל מגוון רב של תת-מקורות הנבדלים על פי אופן הרכשת המידע. המודיעין החזותי, למשל, יכול להיות מושג באמצעות מצלמות המפיקות תמונות וסרטי וידיאו מפלטפורמות מוטסות כמו לוויינים או מטוסים עם טייס וללא טייס, באמצעות מצלמות המוצבות על כלי רכב, על בלונים או בעמדות תצפית קרקעיות, באמצעות תצפיות מאוישות בהן התצפיתנים כותבים בטקסט את מה שהם רואים, ועוד ועוד.

בדרך כלל האיסוף מתבצע על "ישויות" מוגדרות מראש - אנשים או מערכות מוגדרות, כמו סוללות טילים מסוגים שונים או מערכות שליטה ובקרה למיניהן. האיסוף יכול להתקיים בכמה "אורכי גל". כלומר, על ידי כמה גורמי איסוף בעלי אורך גל שונה כמו "טקסטואלי וסיגינטי" לסוגיהם.

לעיתים קרובות מתווספת למידע הפרשנות של הגורם האחראי להשגתו ולעיבודו הראשוני, אפילו דרך התרגום או סיכום המידע הגולמי. גם "הערכות מודיעין" שמבצעים חוקרים/קציני מודיעין על סמך המידע המוזרם על ידי אותם גורמי איסוף הן סוג של ישות מודיעינית שיש לה מעמד ייחודי, והיא מופיעה בדרכים שונות (טקסט חופשי, טקסט בפורמט קבוע, מצגות, סרטונים, מפות וכדומה).

אחת ההתפתחויות החשובות בהקשר של עיבוד המידע המודיעיני בשנים האחרונות ברחבי העולם טמונה בהבנה שנדרש שילוב של תוצרי גורמי האיסוף, משום שמקור סיגינטי מעניק מידע מסוג שונה ממידע אופטי, וכך גם לגבי מידע שנכתב בטקסט על ידי קצין מודיעין ומופץ ברשתות מסוגים שונים. כל גורם מביא את המידע מנקודת מבטו, שבדרך כלל שונה מזו של חברו. הסיגינט, למשל, אינו מזהה את הפלטפורמה שממנה משודרים הסיגנלים, שיכולה להיות מטוס או מכונת או בניין, ואילו גורם האיסוף האופטי מסוגל לעשות זאת, אבל אין לו את המידע הסיגינטי שמקורו באותה ישות שהוא מצלם. מרכיב משותף אפשרי ושכיח לכולם הוא הזמן, ולכן, הדיוק בהגדרת נקודת הזמן שבה התקבלו פריטי המידע השונים הוא קריטי. וכך, השילוב או ההיתוך של מרכיבי המידע הללו הוא קריטי ובעל ערך עצום להבנת התמונה המודיעינית הכוללת.

כבר לפני עשרים שנה פותחה שפת סימון שמגדירה אוסף של חוקים שיוצרים מעין "סימון" של מילים/מסמכים המאפשרים למכונה (מחשב) ואדם לקרוא אותם ולהבין אותם, וכן לצורך העברת מידע בין מכונה מסוג אחד למכונה מסוג אחר. למעשה, זהו תקן שמתאפיין בתחום שבו

מבקשים להתמקד ולהעביר מידע בין מכונות שונות כמו תחום הספורט, או תחום כלכלי/עסקי וכיוצא באלה.

כדוגמה נתבונן ביצירת שילוב בין מידע טקסטואלי ומידע אלקטרוני. מידע טקסטואלי נחשב כמידע לא מובנה. כלומר, טקסט חופשי שנכתב על ידי אנליסט או תצפיתן או קמ"ן. מידע דיגיטלי הוא מובנה. כלומר, כל חלקיק מידע דיגיטלי ניתן לאחסון באופן סדור וקבוע - זמן, תדר, מיקום המקור וכיוצא באלה. טקסט, לעומת זאת, אינו סדור, וכל אדם כותב כאוות נפשו. כדי ליצור שילוב בין שני סוגי המידע נדרש תהליך "הבניה". כלומר,

**מידע דיגיטלי -
בניגוד למידע
טקסטואלי - הוא
מידע מובנה.
כלומר, כל חלקיק
שלו ניתן לאחסון
באופן קבוע.**

מיצוי כל משפט לליבת המשפט ושליפת הליבה הזאת באופן סדור ובכך מתבצעת ההבניה שיכולה להשתלב עם המידע הדיגיטלי בתוספת שני נתונים: זמן ומקום.

כל טקסט בנוי ממשפטים. בכל משפט קיימת שלישיית מרכיבים שהיא ליבת כל שפה אנושית: סובייקט - פועל - אובייקט. בשפה העברית: נושא - נושא - מושא. שליפת מרכיבים אלו מתאפשרת כיום על ידי טכנולוגיות מתקדמות מבוססות יכולת ניתוח בסיסי (Analyzer)

וניתוח של חלקי דיבר (Parser). שילוב או היתוך של שני סוגי "מידע" מובנים הוא כבר מלאכה קלה יותר, וזו מתאפשרת בין השאר בבסיסי נתונים סמנטיים ובטכניקות מבוססות רשתות סמנטיות (על כך ראו בהמשך). בנקודה זו אנו נפגשים עם המושג "ביג דאטה", שכן אנו נדרשים לאגור את כמויות הענק של המידע מן הסוגים השונים בבסיסי נתונים מסוג חדש, בעלי יכולת אגירה עצומה, ויותר מזה, עם יכולת אינדוקס מעולם חדש.

אינדוקס פירושו שלכל חלקיק מידע שנכנס לבסיס הנתונים, יש כתובת שבה הוא "נמצא", וידיעתה מאפשרת לשולפו על ידי שאילתה שמוגדרת על ידי חוקר או על ידי מערכת אוטומטית, שמבקשת להשתמש בחלקיק המידע הרלוונטי על פי "זמן ומקום" כדי להתיך אותו עם מרכיב/ חלקיק מידע מסוג אחר באותה נקודת זמן ובאותו המקום.

עד לאחרונה הייתה היכולת לשלוף מבסיס הנתונים הענק את כל המידע הרלוונטי במהירות ולהתיכו, מוגבלת. הפתרונות הטכנולוגיים למימוש היכולות הללו במסגרת הנקראת ביג דאטה נוצרו רק בשנים האחרונות, וגם הפתרונות הללו היו חלקיים. פריצות הדרך הגדולות טמונות בשנים האחרונות בשיפור יכולת האינדוקס להבדיל מיכולת האגירה. לדוגמה, שיטת ה-Elastic Search או כלי שפיתחה חברת Attivio.

אבל בזאת לא הסתיימו האתגרים שנולדו בגין צרכים אלה, שכן איך "מתיכים" מידע סיגינטי עם מידע אופטי או טקסטואלי? מה פירוש להתיך? איך מכונה אוטומטית (כלומר, תוכנה) יודעת לקשר מרכיב מידע מסוג אחד לאחר, וכיצד מכונה אוטומטית יכולה להסיק מה שחוקר מודיעין אמור להסיק, אם בכלל? לפני שאנסה להבהיר כיצד מכונה אוטומטית אמורה לבצע את ההיתוך, אדגיש שרק מכונה אוטומטית יכולה להתמודד עם האוקיינוסים העצומים של המידע שנאגר במהירות פנטסטית, בבסיסי הנתונים שהצגתי לעיל.

היתוך

היתוך פרושו הפקת מידע שנוצר משילוב של לפחות שני מרכיבי מידע על אותו אובייקט (אנשים, מקומות, שטחים, מבנים וכיוצא באלה). היתוך בין תוצרים איסופיים שהם "סיגנלים אלקטרומגנטיים" מבוסס על זמן ומקום. כלומר, אם, לדוגמה, אדם נקלט על ידי מכשיר קומינטי במקום מסוים ובנקודת זמן מוגדרת, וכמו כן, מכ"ם כלשהו גילה את המכונית שבה הוא שוהה או נוהג באותו מקום ובאותו הזמן, ניתן בחישוב לא מסובך להגדיר שהאיש שדיבר בטלפון הסלולרי הוא זה שנוסע במכונית הספציפית הזאת. כאשר ההיתוך הוא בין חומר שמוצג בטקסטים לבין מידע אלקטרוני, מבוסס סיגנלים, ההיתוך הופך מסובך יותר, והוא מחייב "שפה משותפת".

בעשרים השנים האחרונות פותחו "גשרים" סטנדרטיים המאפשרים "הבנה" בין מערכות ממוחשבות שונות, כגון ה-XML כסטנדרט שנוצרו ממנו נגזרות שונות. אבל השפה המשותפת אינה מספיקה. היא מחייבת גם הגדרה של תחום העניין שבו מתקיים "השיח" בין שתי מערכות שונות.

אונטולוגיה

שפה משותפת מחייבת הגדרת תחום העניין שבו מתקיים ההיתוך. למשל, אם התחום שבו אנו עוסקים הוא ימי, אוסף הישיות הרלוונטיות בים שונה מאוסף הישיות שאנו מכירים במרחב האווירי. בים אנו עוסקים בישיות כמו ספינות, צוללות, מצבי ים, דגים, זרמים וכיוצא באלה, ואילו במרחב האווירי אנו עוסקים במטוסים, מסוקים, מל"טים, בעלי כנף וכיוצא באלה. לעיתים גם הגיאוגרפיה היא מרכיב בהגדרת תחום העניין, שכן באזור מסוים – יבשתי – הישיות הרלוונטיות הן ואדיות, הרים ומערות, ובמרחב אחר הישיות הרלוונטיות הן מקורות מים מדבריים, נאות מדבר וכיוצא באלה. "יצירת אונטולוגיה" היא הגדרת הישיות והפעילויות הרלוונטיות לנושא שבו אנו רוצים לבצע היתוך או יותר מזה, תובנות מצטברות כתוצאה של "התהוות דפוסים" שאומנם פרטיהם נמצאים בתוך בסיסי הנתונים הענקיים, אבל הם בבחינת מחט בערמת שחת.

מהנאמר עד כאן ברור שהנושא הקרוי ביג דאטה אינו מתמקד רק באגירת כמויות גדולות של מידע. הוא מחייב קיומו של אינדקס המאפשר את שליפת הנתונים ויכולת חיפוש בים הנתונים המאפשרת זיהוי דפוסים שונים של התנהגות של ישויות, או התפתחות של תהליכים פיזיקליים, כימיים ואחרים. הפרטים המעידים של קיום דפוסים כאלה נמצאים בבסיסי הנתונים, אבל הם מעטים ולעיתים אינם מופיעים בדיוק כפי שרצינו לאתרם. והקשה מכול – היכולת לחבר נתון שאותר עם נתון אחר שאותר ולבנות מהם "תוצר מורכב יותר" הקרוי דפוס. הקושי הזה טמון גם בהקשרים מתמטיים של היכולת לזהות דפוס וגם בהקשרים של יכולת שליפה מבסיסי מידע גדולים מאוד. הקושי הזה מגיע לשיאו (המשברי) כאשר עוסקים בבסיסי נתונים המבוססים על גרפים סמנטיים, שיוצגו בהמשך.

דפוסים

דפוס הוא, לדוגמה, תהליך שבו הגברת כהן, שהיא אם לשני ילדים, יוצאת מביתה כל בוקר, עם שני ילדיה, מכניסה אותם למכוניתה, נוסעת לכן של הילד הצעיר, במסלול מוגדר ובמשך חמש דקות, וכשהיא מגיעה, יוצאת מהמכונית, פותחת את הדלת לבנה הצעיר ומאפשרת לו לצאת,

מנשקת את מצחו והילד רץ אל הגן ונעלם. הגברת כהן נכנסת למכוניתה וממשיכה לבית הספר, מורידה באותו נוהל את בנה הגדול יותר ואחר כך ממשיכה אל בית הקפה "אפרסק" שבקצה העיר. הנתונים לגבי הגברת כהן נאגרים ממקורות איסוף שונים, ומוכנסים אל בסיס נתונים אחד, על פי אותה שפה משותפת שהצגתי לעיל. המקורות הם טקסטואליים – דיווח של תצפיתן שראה את הגברת כהן במקומות שונים, שיחות הסלולר של הגברת כהן, שמצביעות על מקום הימצאה בנקודות זמן שונות לאורך מסלול נסיעתה, וכן הידע המוקדם על מיקומם של הגן ובית הספר של ילדיה ובית הקפה "אפרסק", ונתונים על אותה גברת. נתונים אלה מוגדרים כתכונות של אותן ישויות (הגברת כהן, שני הילדים, בית הספר, הגן ובית הקפה הם ישויות).

הקושי טמון בעיקר בשליפת כל הנתונים הרלוונטיים, ביכולת לקשר ביניהם, ובעיקר בכך שהנתונים שנאגרו הם "לא מדויקים". כלומר, אינם בדיוק אותו "סיגנל או מילה" שאנו מחפשים. כי הרי אנחנו לא יודעים מהו דפוס הנסיעה של הגברת כהן. אנחנו יודעים כמה פרטים מראש, אבל לא את המסלול שחוזר על עצמו ומוצג באופן חלקי יחד עם עוד מיליארדי נתונים על מכוניות דומות, וגברות דומות לגברת כהן.

"אי־דיוקים" או נתונים שאינם מעידים באופן ישיר ומיידי על קיום הדפוס של הגברת כהן, לדוגמה: תצפיתן זיהה את סוג מכונית הסובארו שהגברת כהן נוהגת בדרך כלל באמצע הדרך לגן הילדים, אבל אין לו שום עדות לכך שהיא זאת שנהגה במכונית. סנסור סלולרי שהגברת כהן הפעילה נקלט כשיצאה מבית הספר במרחק של מאה מטר ממנו, ותצפיתן נוסף מעיד שראה את מכונית הסובארו שתי דקות אחרי הרגע שבו דיברה בסלולר. מנתוני מכ"ם שצפה על העיירה מטעם העירייה, התברר שמכונית "מסוימת" אכן עברה במקום שבו היה איכון של הסלולר. בבדיקה מהירה שמבצעת התוכנה מתברר שמסלול הנסיעה של המכונית הזאת עבר דרך בית הספר וגן הילדים, וכן, נקודת הזמן שבה המכונית נצפתה על ידי התצפיתן היא "בערך" אותה נקודת הזמן שמסלול המכ"ם מציג.

מכל הנתונים הללו, על המערכת לקבל החלטה שייתכן מאוד שזאת הגברת כהן. אבל זה עדיין אינו דפוס. דפוס הוא תהליך שחוזר על עצמו. כלומר, נדרש מן המערכת לזהות חזרות של התהליך הזה של הגברת כהן. ואכן, למוחרת שוב התקבלו נתונים, קרובים לקודמים – אבל הסלולרי של הגברת כהן לא פעל. כלומר, עדיין קשה לומר שנוצר דפוס וכך, במשך כמה ימים המשיכו הסנסורים הללו "למלא" את בסיס הנתונים.

הקושי טמון בעובדה שחולפות עוד מכוניות על אותו מסלול, עוצרות עוד מכוניות ליד גן הילדים ובית הספר, ויש הרבה מאוד מכוניות שנצפו בדרך לבית הקפה, וכן הלאה. כך גם לגבי שיחות הסלולר - מדובר על מספרים עצומים של אימהות שמשוחחות בזמן הזה בסלולר, והשאלה היא איך המערכת יכולה לאתר את הדפוס של הגברת כהן. הקושי מגיע לשיאו כאשר איננו יודעים שהגברת כהן היא הישות שמבצעת את הדפוס הזה. אנחנו יודעים על קיומו "המעורפל" של דפוס כזה, אך איננו יודעים מי הוא זה שמבצע אותו. הטכנולוגיה העוסקת בנייתוח המידע לאיתור דפוסים בבסיסי נתונים גדולים התפתחה מאוד בשנים האחרונות והיא מאפשרת לנו להתמודד עם השאלות האלה. אומנם לא בקלות, אבל בגופים מתקדמים מאוד, היא מתמודדת בהצלחה מסוימת.

מכונה לומדת

אחד הצרכים החשובים למודיעין מסוג כלשהו הוא "חיזוי". כלומר, איתור דפוסי פעולה בבסיסי נתונים גדולים מאוד (טריליונים של ישויות) מהווה צעד גדול קדימה, אבל החיזוי שמבוסס על הנתונים הנמצאים בבסיסי הנתונים הוא קפיצה גדולה עוד יותר.

חיזוי מבוסס על לימוד תהליכים ואירועים שאירעו בהיסטוריה, או על דפוס שחוזר על עצמו מספר משמעותי של פעמים. מכונה לומדת מבוססת על שני מרכיבים: מרכיב האימון (training) ומרכיב הסיווג (קלסיפיקציה). המכונה הלומדת אמורה להחליט כשמוגיע אליה טקסט חדש, לאיזו קטגוריה (נושא עניין) הוא שייך: לספורט? למדע? לפוליטיקה? וכן הלאה. הלימוד פירושו – המכונה הלומדת "מתאמנת" על "דוגמאות". הדוגמאות אמורות לייצג את מה "שהמכונה מחפשת". כלומר, בעולם טקסטואלי, למשל, יידרשו כמה עשרות טקסטים המייצגים נושא עניין. המכונה תתאמן על הטקסטים הללו, ותיצור "פיצ'ר וקטור" המייצג את הקטגוריה הזאת.

"פיצ'ר וקטור" הוא אוסף של מילים או שילובי מילים שמייצגים את הנושא המדובר. למשל, מילים כמו שופט, כדור, קהל, שחקנים וכיוצא באלה, מעידות על טקסט הקשור למשחק כמו כדורגל או כדורסל. כדי להפריד בין השניים (כדורגל או כדורסל), על האימון ללמוד מטקסטים על כדורסל וכדורגל. בתהליך הלימוד ייווצר וקטור המבוסס על עוגנים (מילים והרכבי מילים) אופייניים לכדורגל או כדורסל. למשל: בעיטת 11 קיימת רק בכדורגל, ואילו בכדורסל ישנם פיצ'רים אופייניים אחרים כגון time-out או זריקות עונשין שאינם קיימים בכדורגל.

ככלל, כאשר יגיע מסמך טקסטואלי חדש למכונה לומדת, המכונה "תקרא" אותו, תשלוף ממנו את אוסף העוגנים הרלוונטי, ותשווה אותו אל אוסף העוגנים (פיצ'ר וקטור) המייצג את הקטגוריה. כך גם לגבי זיהוי תמונות, תהליכים או דפוסים. זיהוי תהליך פירושו היכולת להפיק חיזוי. לדוגמה, אם נתייחס לגברת כהן שמובילה את ילדיה כל בוקר לגן ולבית הספר, המערכת יכולה להגדיר את הצעד "הבא" של הגברת כהן בכל אחד משלבי התהליך שהיא מבצעת. כלומר, כאשר הגברת כהן יוצאת מביתה, המערכת

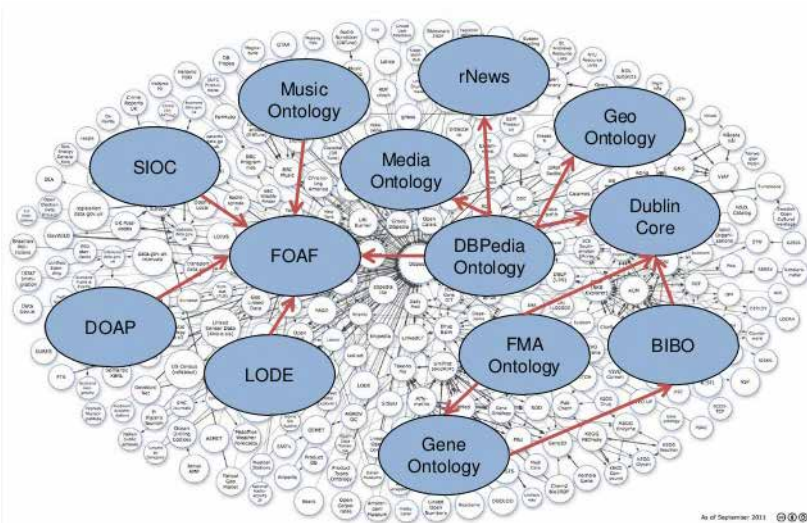
**כאשר יגיע מסמך
טקסטואלי חדש
למכונה לומדת, המכונה
"תקרא" אותו, תשלוף
ממנו את אוסף העוגנים
הרלוונטי, ותשווה אותו
אל אוסף העוגנים
המייצג את הקטגוריה.
כך גם לגבי זיהוי
תמונות, תהליכים או
דפוסים.**

תוכל להעריך בהסתברות גבוהה שהיא תגיע לבית הקפה לאחר זמן מוגדר.

הדוגמה של הגברת כהן פשוטה, אבל מקופלים בתוכה תהליכים רבים אחרים, מורכבים מאוד. למשל, המערכת זיהתה שבכל פעם שיורד שלג בקוריאה והבורסה היפנית יורדת, ונשיא חדש נבחר בארצות הברית, צפון קוריאה מבצעת ניסוי טילים חדש לאחר שנה (מבחירתו של הנשיא האמריקאי). כדי לזהות את הדפוס הזה, המערכת נאלצה לחקור מאות מיליארדים של נתונים, ואולי טריליונים, ולהגיע למסקנה שאכן הדפוס הזה חוזר על עצמו, והוא אכן דפוס. בפעם הבאה

כשירד שלג בקוריאה, והבורסה היפנית תיפול ואם זה יהיה רק כמה חודשים לאחר היבחרו של נשיא אמריקאי, תוכל המערכת לנבא, שבסיכוי גבוה יבצעו הצפון קוריאנים ניסוי שיגור טילים. היא תוכל גם לנבא שאחרי השלג בדרום קוריאה והיבחרו של הנשיא האמריקאי בחודשים האחרונים, צפון קוריאה עלולה לעשות ניסוי טילים בסבירות יותר נמוכה מן המקרה הקודם, אבל בסבירות מסוימת שמחייבת ערנות.

הקושי הגדול בפרוצדורה שהוצגה לעיל טמון ביכולת לזהות את הדפוס הזה (שלג בקוריאה וכולי) בתוך אוקיינוס עצום של חלקיקי מידע, ולאתר דפוס מסוג כלשהו שחוזר על עצמו מספיק פעמים, אפילו בווריאציות שונות של "שמות הישיות" או "שמות האירועים". החיזוי עצמו, בהינתן זיהוי הדפוסים, אינו מורכב במיוחד. המהפכה טמונה בזיהוי הדפוסים הללו בין הררי הנתונים שלא חייב להיות כל קשר הגיוני או אחר ביניהם.



SEMANTIC WEB

ביג דאטה הוא נושא כללי מאוד. הוא מכיל שני מרכיבים עיקריים: אופן אגירת כמויות עצומות של מידע בבסיס נתונים, וכן היכולת לשלוף אותם באופן מהיר ולקשר ביניהם כדי ליצור "תובנה או היסק". אחת הדרכים לממש זאת טמונה ביכולת לקשר בין הנתונים עוד בתוך בסיסי הנתונים על ידי ייצוג אחיד של המידע. לדוגמה, ייצוג של ישויות באופן אחיד, ייצוג של פעולות באופן אחיד וכן הלאה. הגדרת "הייצוגים" היא עיקרה של ה"אונטולוגיה" במובנה ההנדסי. כלומר, מעין עורך (שפותח על ידי חברות רבות) המאפשר לייצג ישויות ופעולות ביניהן ואת תכונותיהן בגרף אחד. יכולת זאת מנסה להפיק טכנולוגיית ה-3 WEB המבוססת על בסיסי נתונים סמנטיים או מה שמכונה גרפים סמנטיים המאפשרים יצירת קשרים בין ישויות על פי התכונות המשותפות שלהן,

ועל פי הקשרים שלהן עם ישויות אחרות (linkification), וכן על פי מרכיבים רבים אחרים כפי שהמשתמש יהיה מסוגל להגדיר. הגרפים הסמנטיים מהווים מהפכה גדולה בעולם עיבוד המידע, אבל המערכות שפותחו עדיין מתקשות להתמודד על כמויות עצומות של מידע (עד 50 מיליארד ישויות), אם כי גם סוגיה זאת נמצאת על סף פתרונות יצירתיים ביותר, כולל של חברות ישראליות.

סיכום

ביג דאטה הוא מושג המכיל את שני המרכיבים: יכולת אגירה שכוללת את טכנולוגיית האינדוקס, המאפשרת לשלוף במהירות כל חלקיק מידע מבסיסי נתונים עצומים, וכן היכולת לבצע מניפולציות בין טריליוני הנתונים במהירות (על סף זמן אמיתי) וכך לחשוף דפוסים וקשרים, וליצור היסקים שיצור אנושי לא יכול לבצע, בגין העומס העצום של הנתונים. תחום ההיסקים עדיין בחיתוליו, אבל כבר כיום קיימת התפתחות מהותית במרחב הטכנולוגי, שמתמקדת באלגוריתמים ובתהליכי חיזוי מבוססי מכונות לומדות ו"לימוד עמוק", שהוא למעשה שכלול של מכונה לומדת שמפעילה רשתות נוירונים המאפשרות משיכת אינפורמציה מכמה שכבות מידע (להבדיל ממכונה לומדת קלאסית). טכנולוגיית הלמידה העמוקה מאפשרת פריצות דרך משמעותיות בנושא כמו ניתוח תמונה. כמו כן, תהליכי ניתוח מבוססי אלגוריתמים בייסיאניים¹ רשתיים ודינמיים, המאפשרים "קבלת החלטות" על ידי רובוטים גם בהיעדר מידע היסטורי. לעזרת היכולת הזאת פותחה טכנולוגיה המכונה reinforcement learning שפירושה לימוד מחוזק. כלומר, המערכת "מתקנת" את מרכיבי הלימוד תוך כדי תנועה – שוב – כאשר מפעילים מערכת רובוטית ללא מידע היסטורי. כל היכולות הללו אמורות להפיק היסקים מתוך ים הנתונים שאנו חיים בתוכם, ולהערכתנו, היכולות הללו עוד ישתכללו וילכו, וייצרו יכולות מסוג חדש גם לחיינו האזרחיים וגם לצבאיים. כבר היום אפשר לחזות את הצעד הבא של המרחב הקיברנטי. אם עד היום הופץ על ידי המרחב הזה מידע, אם ברשתות החברתיות ואם במערכות נוסח גוגל ויאנדקס, בעתיד הלא רחוק נפגוש גם את הפצת "הידע". כלומר, ידע רלוונטי לכל אחד, שמופק משילוב של מידעים מכמה מקורות, ומופק בטכניקות של "למידה עמוקה" ומודולים אחרים, על פי הצרכים והאינטרסים של כל אחד מאיתנו. האינטרסים הללו יאותרו על ידי מערכת גלובלית (נוסח פייסבוק) והיא זאת שתציע "ידע" רלוונטי על פי צרכינו בדחיפה (push).