

# בינה מלאכותית, זהויות דיגיטליות וחופש העיתונות

## נעם למלשטריך לטר ודיוויד נורדפורס

### הקדמה

בינה מבוססת מחשב (computer-assisted), כלומר הסיוע בתוכנת מחשב בקבלת החלטות, היא חלק מחיי היום-יום שלנו. אנו מדברים על "אינטליגנציה מורחבת" של פרטים המשתמשים במחשב האישי<sup>1</sup>, ועל "אינטליגנציה קולקטיבית" של קבוצות המתחברות לרשתות תקשורת. למושגים אלה יש להוסיף את המושג "בינה מלאכותית" (AI — Artificial Intelligence), המתייחס למחשבים הפועלים באופן תבוני בלא לתקשר עם בני אנוש, תוך חיקוי האינטליגנציה האנושית.<sup>2</sup>

האינטליגנציות השונות הללו נמצאות בתהליך של התמזגות והתלכדות. האינטליגנציה המורחבת של הפרט, האינטליגנציה הקולקטיבית והבינה המלאכותית מתפתחות בתהליך מקביל ומושפעות אהדדי. האינטרנט הופך חלק ממוחנו, והמוחות שלנו נעשים חלק מהאינטרנט. המחשב האישי נכנס לעולם העיתונות בשנות השמונים, האינטרנט חדר לעיתונות בשנות התשעים, וכיום אנו עדים לחדירה מאסיבית של אינטראקציה חברתית לעולם העיתונות, החל בתגובות הקוראים וכלה במיקור חוץ להמונים (crowdsourcing), כאשר המוני המשתמשים באינטרנט יכולים לייצר תכנים בעצמם.

יחסי הגומלין המתקיימים בין התקשורת לבין האינטרנט והאינטראקציה החברתית מסייעים לנו להבין את יכולתה של העיתונות לשפר את האינטליגנציה הקולקטיבית ומושכים תשומת לב מחקרית רצינית ומעמיקה. במאמר זה אנו מתמקדים בתרומתה של הבינה המלאכותית לקידום האינטליגנציה הקולקטיבית באמצעות אלגוריתמים הפועלים באופן אוטומטי לתיוג חדשות, למדרוגם (קביעת רייטינג) ולניתוח הקונטקסט שבו נקלט התוכן.

מערכות המדרוג הן כלי לפיקוח על מערכות חברתיות והן משפיעות על ההתפתחות החברתית, שכן בני האדם תמיד ינסו לשפר את הדירוג שלהם. מערכת מדרוג שמשפיעה על חייהם של אנשים היא נושא מורכב וקשה למחקר, שכן גם דירוג מתסכל, לא מועיל או מטעה משפיע על המערכת, מאחר שההנחה היא שהצרכנים מושפעים ממנו. לעיתונות אין ברירה אלא לחקור את מערכות המדרוג במטרה להציב להן אלטרנטיבות. דוגמה לשימוש באלגוריתם לדירוג חדשות הוא האתר TechMeme<sup>3</sup>, החותר לשתף את הקוראים בידיעות החשובות ביותר בנושאי טכנולוגיה. מודל כגון זה יתמרץ אתרי עיתונות להשביח ידיעות במטרה להשיג דירוגים גבוהים יותר.

אלגוריתמים המבוססים על בינה מלאכותית משנים את פניה של העיתונות המקצועית ושל תחומי המחקר האקדמי הנלווים אליה. הבינה המלאכותית קונה דריסת רגל במאחזים המסורתיים של העיתונות: תוכן חדשותי (באמצעות ניתוח תוכן ותיוג באופן אוטומטי לפי נושאי הכתבה בכל סוגי המדיה) ופרסום (על ידי מדידה של תשומת לב הצרכנים על בסיס פרטני וקוגניטיבי והתאמת פרסומות ל"זהות הדיגיטלית" של הגולש או ל"אישיותו" בהתאם לדפוסי התנהגותו ברשת). הן התוכן והן הפרסום עוברים כיום שינויים דרמטיים.

סוכני השינוי הם המדיה החדשה והבינה המלאכותית, המבוססות על כוח חישובי הולך וגובר. המדיה החדשה והאינטראקטיבית מאפשרת לראשונה למדוד את תשומת הלב שכל משתמש מקדיש לתוכן תקשורתי. מפרסמים עתידיים לדרוש תיקוף (הוכחה) מלא של דירוגי גולשים. כלי המדידה שבשימוש היום ייעלמו. התחרות על תשומת לב המפרסמים תגבר, ומדד החזר על ההשקעה (ROI — Return On Investment) יקבע את עתידם של ערוצי הפרסום, של העיתונאים, ושל סוכנויות חדשות בכל הפלטפורמות (סוגי הצגים – הטלוויזיה, הסלולר, האיפד) התלויות בהכנסות מפרסום.

תכני התקשורת עוברים שינויים דרמטיים בהשפעת פלטפורמות אינטראקטיביות ובהשפעת העובדה שהתכנים זמינים כיום מכל מקום בכל עת. לפני זמן לא רב, אמצעי התקשורת היו בשליטתן של החברות שהפיקו את התכנים. המודל העסקי המסורתי של תאגידי החדשות והבידור היה לשלוט בו זמנית במדיום התקשורת ובתוכן. עם בוא האינטרנט הופיע דור חדש של שחקני מדיה: חברות כגון טוויטר, פייסבוק וגוגל נמנעות במכוון מלהפיק תוכן. הן מסתפקות במתן גישה לתכני תקשורת.

העיתונות נמצאת בתהליך של היפרדות מתשתיות התקשורת.<sup>4</sup> ספקי תוכן מהדור החדש מתמקדים בהפקת תוכן, בפרסום מקוון ובהפצת התכנים שלהם על גבי התשתיות הפתוחות והניטרליות החדשות (היכן שהחברות בעלות התשתית אינן מייצרות תוכן). שני עיתוני רשת חשובים ומוכרים, *Huffington Post* ו-*TechCrunch*, התחילו כבלוגים וזכו להצלחה רבה בלי לשלוט בתשתיות המידע.

ארגוני תקשורת מהדור הקודם משקיעים הון עתק במדידת גודל קהל הקוראים, במדידת תשומת לבו ובחישוב סטטיסטיקות והסתברויות שייצעו לאינטרסים של המפרסמים. ברשת, המדיה החדשה מציעה למפיקי התכנים ולמפרסמים מידע מדויק במקום

תכנים לפי הקשר הצריכה (Contextual Targeting). המאמר מתאר "מנוע" המבוסס על בינה מלאכותית, שמסוגל לסנן ולהתאים תוכן תקשורתי לזהות הדיגיטלית של הצרכן במטרה למקסם את ההחזר של חברות על כל שקל שמושקע בפרסום.

בסיכום אנו מפנים את הזרקור לכמה שאלות על המורכבות וזהויות דיגיטליות, התאמת תכנים אוטומטית ובינה מלאכותית מביאים לעולם העיתונות, ועל הצורך לעדכן את הקוד האתי שמדריך את העיתונות. אנו מדגישים את הצורך הדוחק להעמיד קוד אתי חדש בשימוש בזהויות דיגיטליות בתחום העיתונות, ומציעים כמה עקרונות לקוד זה. לבסוף אנו מציגים שאלות מחקר חדשות שתובלנה להעמקת המחקר בשאלת ההשלכות האפשריות של אלגוריתמים המבוססים על בינה מלאכותית על מקצוע העיתונות ועל תפקידה של התקשורת בחברה דמוקרטית.

### התקשורת בעידן הדיגיטלי: הגדרה מחודשת

עד היום שני המונחים "תקשורת" ו"אמצעי תקשורת" היו בעצם נרדפים. התקשורת הייתה נושאת הדגל של התשתיות שסיפקו את אמצעי התקשורת להמונים, ולהפך. הקריאה "עצרו את מכבשי הדפוס" הייתה זהה להודעה "מבזק חדשות". הארגונים ששלטו בתשתיות התקשורת להמונים שלטו גם בתכנים ששודרו על גבי אותן תשתיות. המציאות הזאת משתקפת בהגדרות המילונאיות למושג "תקשורת", כפי שאפשר לראות, לדוגמה, במילון *Compact Oxford English Dictionary* (המילון זמין ברשת דרך האתר [AskOxford.com](http://AskOxford.com)).

**עיתונאי:** אדם שכותב ידיעות חדשות עבור עיתונים או כתבי עת או מפיץ חדשות לשידור ברדיו או בטלוויזיה.<sup>8</sup> האירוניה היא שהמילון המקוון אינו מאזכר את רשת האינטרנט ברשימת אמצעי התקשורת. בכל מקרה, גם אילו היה המילון מציין את האינטרנט, ההגדרה הייתה נשארת הגדרה מתה שאבד עליה הכלת. היום כל אחד יכול לשדר חדשות או להפיץ ידיעות באינטרנט, אבל זה לא הופך אותו לעיתונאי.

עד לעת האחרונה תשתיות התקשורת אפשרו תקשורת משני סוגים. הן אפשרו ליחיד להתקשר עם הרבים – תשתיות אלה נקראו "אמצעי התקשורת ההמוניים" – ועם פרט אחר, למשל באמצעות הטלפון. תשתיות התקשורת שקישרו בין היחיד לרבים נחשבו לאמצעי התקשורת בה' הידיעה וסיפקו בעיקר תוכן חדשותי ובידורי. במסגרת אותם אמצעי התקשורת נשאו מערכות העיתונים וההפקה באחריות לשדר את התכנים ואילו הצרכנים היו פטורים מאחריות כלשהי. הם היו חופשיים להחליט אם לקלוט את השידורים או לא. אמצעי התקשורת להתקשרות בין יחידים, ובראשן רשת הטלפוניה, לא נחשבו אמצעי תקשורת אלא רק כלי להתקשרות בין-אישית.

היום נעלם ההבדל בין התשתיות המשמשות לתקשורת בין יחידים לבין אלה המשמשות לתקשורת בין היחיד להמון. יתרה מזאת, האינטרנט תומך בתקשורת בין המונים להמונים (many-to-many) וכל אחד יכול לפרסם. הדור הראשון של האינטרנט, Web 1.0, הרחיב את אפשרויות ההתקשרות בין יחיד לרבים. כיום, בדור השני, Web 2.0, האפשרויות לפתיחת ערוצי תקשורת בין המונים להמונים רבות יותר. ציבור של אנשים יכול לפרסם יחד ודבר זה לא מרתיע את חברות התקשורת החדשות – החברות שלא מפקות את התכנים שהן מספקות.

טבלאות עם הסבריות. היא מספקת מספרים מדויקים: איזה משתמש הסתכל על תוכן מסוים, ולמשך כמה זמן. מפרסמים יכולים לדעת אם קורא הקליק על הפרסומת שלהם ובאיזה הקשר. בערוצי הפרסום הישנים הפרסומות נשלחות לכל הצרכנים ללא הבחנה ותג המחיר שלהן קבוע ואינו מושפע ממספר האנשים שרואים את הפרסומת או ממספר האנשים שהפרסומת מניעה לפעולה. האינטרנט, לעומת זאת, תומך בפרסום תלוי הקשר, שבו הפרסומות המשוגרות לכל גולש מותאמות לתכנים שמוצגים לו.

הכלים למעקב אחר משתמשים ולהתאמת תכנים ופרסומות ליחידים מחוללים מהפכה בתחומי התוכן, המדיה והשיווק. בעולם הדיגיטלי-האינטראקטיבי תקציבים נדרשים לתת דין וחשבון להוצאות על פרסום באמצעי התקשורת. חישוב מדד ההחזר על ההשקעה בתחום ההתאמה האישית של הפרסום והתוכן לגולש הפרטי הוא תחום מחקר מדעי שאחראי לשינויים נרחבים בתחומי הפרסום, המדיה והתוכן. בשנת 2007 ההוצאה הגלובלית על פרסום נאמדה ב־385 מיליארד דולר,<sup>9</sup> סכום שהשתווה לתמ"ג של מדינה שדורגה במקום ה־26 בעולם בגודל הכלכלה לשנת 2008.<sup>6</sup>

כדי להתאים את התכנים לזהות הדיגיטלית של המשתמש נבנים מנועים המבוססים על בינה מלאכותית, שיהיו מסוגלים לנתח תוכן רב ממדי, לתייגו ולהשוותו למאפייניה של זהות הצרכן. פירוש הדבר שאותם מנועים יידרשו לזהות קשרים בין הקוד הגנטי של אנשים, לבין ה"קוד הגנטי" של התוכן וההקשר (נעשו כמה ניסיונות מוצלחים לזהות את ה"קוד הגנטי" של יצירות מוסיקה וספרות פופולריות). מחקרים בתחומי הביולוגיה, הגנטיקה והפסיכולוגיה שניסו לזהות קשרים בין הקוד הגנטי של אנשים לבין תכונותיהם הקוגניטיביות ומאפייני התנהגות חברתיים או אנטי-חברתיים, מצטרפים לתוכנות לכריית מידע על פעילות בני האדם שמורצות ברשתות חברתיות ואשר מטרתן להרכיב פרופילים עבור הצרכנים. נראה שהזהויות הדיגיטליות עתידות למזג מידע על הקוד הגנטי הפרטי של אנשים עם מידע על הקישורים שהם מקליקים באינטרנט.

יש כיום אלגוריתמים חדשים המבוססים על בינה מלאכותית, שמנתחים תוכן – טקסט כתוב, וידאו, שמע, ותמונות – באופן אוטומטי. בעקבות זאת מתקיימים כיום מאמצים בין-לאומיים לנסח תקן אחיד לתיגו זהויות דיגיטליות של גולשים יחידים. מנועים המבוססים על בינה מלאכותית משמשים לאיתור, לקידוד ולתיגו של תוכן באופן אוטומטי ולהתאמתו לזהויות הדיגיטליות של הגולשים. תהליכים אלה עתידים להשפיע על תכנים עיתונאיים וייתכן שהם אף יגדירו מחדש את העיתונות ואת תחום המחקר האקדמי של העיתונות. התקשורת נאלצת להסתגל למצב החדש ולחפש מודלים עסקיים חדשים כדי – כפי שנסביר בהמשך – להגן על החופש העיתונאי.<sup>7</sup>

במאמר זה אנו מתארים זהויות דיגיטליות ותכנים בין-לאומיים חדשים לניהול זהויות דיגיטליות, דנים בתרומתם של הרשתות החברתיות, של המחקר הגנטי ושל העולמות הווירטואליים ליצירתן של זהויות דיגיטליות ומציגים את המחקר העדכני בתחום הבינה המלאכותית ושימוש בה לצורך התאמת תוכן לזהויות דיגיטליות. חוקרים כיום מסוגלים לתרגם תוכן תקשורתי לנוסחות מתמטיות ("חתימות") במטרה לחקור תוכן והקשר. אנו דנים במושג "מעורבות צרכנית בתכני תקשורת" (Media Engagement) ובנגזרותיו – פרסום מפולח לפי אישיות הצרכן (Behavioral Targeting) והתאמת

אלא מהם המודלים העסקיים החדשים שיכולים לקיים את עקרונות העיתונות בכלכלת היזמות?

### זיהוי ה"קוד הגנטי" של תכני תקשורת

בשנות התשעים הוקדשו מאמצים מחקרניים בין-תחומיים לפיתוח שיטות יעילות לאחזור (שליפת מידע ממאגר) אוטומטי של מידע ושל ידע ממידע חדשותי. המטרה העיקרית הייתה לאפשר לצרכנים למצוא את המידע שהם מחפשים במהירות וביעילות. כיום מנועי החיפוש המובילים, כגון גוגל, יאהו ואחרים, מחזירים מיליוני קישורים על כל שאילתה ואינם יכולים לתת מענה לשאלות פשוטות שמכילות כמה מילות מפתח פשוטות. קהילת החוקרים שעוסקת בתחום זה של חקר אחזור מידע מולטימדיה (MIR — Multimedia Information Retrieval) מכסה גם את תחומי האינטראקציה בין אנשים למחשבים (HCI — Human Computer Interaction), תאוריית המידע (Information Theory) — סטטיסטיקה, זיהוי דפוסים, פסיכולוגיה ולאחרונה גם מדעי החברה. אפשר לראות שמאמרים מדעיים המתפרסמים בתקופה האחרונה בכל התחומים האלה מאמצים שיטות מחקר ומצטטים מפרסומים בתחומים האחרים.

נקודת המוצא של רבים מהמחקרים הבין-תחומיים בתחום אחזור ידע ממידע חדשותי היא חיפושים אחר מידע וידע שנעשים ביוזמת הצרכן. אולם מטרתנו היא לחקור את ההשלכות הצפויות על עתיד העיתונות של תופעה חדשה שבה תכנים מחפשים באופן אוטומטי אחר צרכנים בהתאם לזהויות דיגיטליות שלהם. בשלב הראשון אנו מתארים את הידע העדכני בתחום המחקר של אחזור תוכן אוטומטי מתכני מולטימדיה חדשותיים. רוב קהילות המחקר מחלקות את התוכן ליחידות תוכן דיגיטליות קטנות, מנתחות את יחידות המשנה ומתייגות אותן, ולבסוף מקיימות ניתוח אינטגרטיבי שמטרתו לתת כותרת (לתייג) ליחידת התוכן השלמה באופן שיהיה מובן לצרכנים. חלק מהחוקרים ממירים תוכן ויזואלי בנוסחאות מתמטיות בעזרת אלגוריתמים של בינה מלאכותית.<sup>13</sup>

### אחזור מבוסס על תמונות (Computer Based Information Retrieval CBIR —

השיטה העיקרית לחיפוש במאגרי תמונות, לאחזור תמונות או לתיוג אוטומטי של תוכן ויזואלי היא בחלוקת המסגרות הוויזואליות לאזורים קטנים יותר, שנקראים "בלובים" (blobs). טכנולוגיה זאת נעזרת בכלים סטטיסטיים שונים, ובהם חלוקה לאשכולות (clustering). לכל בלוב מודבקת תווית מילולית. התיאור המילולי של כל תמונה ויזואלית כולל תיאורים של צבע, של טקסטורה, של צורות גאומטריות ושל ארגון פנימי. בעזרת תאוריות סטטיסטיות כל תחום של התמונה מקושר לתוויות מילוליות. בשלב השני נערכת השוואה בין התוויות שהניבו הכלים הסטטיסטיים לתוויות שנתנו בני אדם באופן ידני לתמונות דומות.<sup>14</sup> כיום, בעקבות הצעתם של פינאר דויגולו ואחרים,<sup>15</sup> ישנו ניסיון לתאר תמונות באמצעות לקסיקון מוגדר ותחום. הצעה אחרת, שהציעו יואן ועמיתיו, היא לפתח שיטה שהיא "סדרת הדרכה של תמונות מתויגות שתעמיד [...] מודל לרלוונטיות של תמונות בכל הפלטפורמות".<sup>16</sup> חוקרים בתחום ה-CBIR שוקדים כיום על פיתוח תיאורים מתמטיים של תמונות, המכונות "חתימות". החתימות הן תיאורים מתמטיים שמאפשרים לחוקרים לחשב במדויק

חברות התקשורת ששייכות למודל הישן מתאמצות לשמר את דפוסי הפעולה שלהן ואת זהותן ונטות להיאחז בטכנולוגיות מיושנות לערוצי תקשורת בין היחיד לרבים. המודלים העסקיים שלהן מבוססים על שליטה באמצעי התקשורת ובתכנים, ולפיכך הן נתקלות בקשיים במעבר לרשת האינטרנט. בכל המודלים העסקיים החדשים לפרסום שזכו להצלחה, ובהם גוגל, eBay ו-Craigslist, מכירת פרסומות איננה חלק מדפוס עבודה רחב יותר ששייך לעולם העיתונות. מהותה של הפרקטיקה העיתונאית מתומצתת ברשימת עקרונות עיתונות שהציעו ביל קובץ' ותום רוזנסטיל בספרם *The Elements of Journalism* מטעם המרכז על שם פיו לחקר מצוינות עיתונאית ב-2007 (Pew Research Center's Project for Excellence in Journalism).<sup>9</sup>

1. חובתה העליונה של העיתונות היא לאמת.
2. מחויבותה העליונה של העיתונות היא לציבור האזרחים.
3. העיתונות, בעיקרה, עוסקת באימות עובדות.
4. העוסקים בה חייבים לשמור על עצמאות ממושאי הסיקור שלהם.
5. תפקידה של העיתונות לשמש גוף פיקוח חיצוני.
6. עליה לספק במה לביקורת ציבורית ולתמיכה ציבורית.
7. עליה לשאוף להציג מה שראוי לסיקור באופן מעניין ורלוונטי.
8. עליה לסקר את החדשות באופן מלא ומאוזן.
9. יש להבטיח את חופש הדת והפולחן של העוסקים בסיקור תקשורתי.

העקרונות הללו היו ועודם העקרונות המחייבים את התקשורת גם היום, כשכבר איננו יודעים מי או מה נכלל תחת הכותרת הזאת. הבה נאמץ לרגע הגדרה חדשה, תמציתית, למושג "עיתונות", שמפרידה בין עיתונות לתקשורת, אך נשמכת על עקרונות עיתונאיים המושתתים על מערכת היחסים שבין העיתונות לקהל במקום על היחסים בין העיתונות למדיום התקשורת (זהו מקור הבלבול השורר כיום). להלן דוגמה להגדרה כזאת: "העיתונות היא הפעולה של הפקת חדשות וידיעות עומק שמביאות לידיעת הציבור נושאים ועניינים שנמצאים בתחום עניינו. המנדט של העיתונות ניתן לה על ידי הקהל שלה".<sup>10</sup>

העיתונות מחויבת לפעול למען הקהל ולא למען מקורותיה או המפרסמים. אחד המודלים העסקיים השכיחים ביותר כיום הוא של עיתונות החותרת לעורר עניין ולמכור שטח פרסום ושמנדט הקיום שלה מגיע מהקהל (attention work).<sup>11</sup> חלק גדול מהעיתונות כיום שייך לקטגוריה הזאת. כאשר כלי תקשורת מתבסס על הכנסות ממקורות המידע שהם תכנים יחצ"ניים ושיווקיים הוא איננו עיתונות. העיתונות, בתפקידה כמי שקובעת את סדר היום הציבורי (לפי מקסוול מקקומבס ודונלד שאו),<sup>12</sup> תלויה ביכולתה למקד את העניין הציבורי בנושאים המעניינים את הציבור, זאת סיבת הקיום של העיתונות. כלומר, העיתונות זקוקה למודל עסקי שישכיל למשוך תשומת לב ציבורית ולמקדה. מודלים עסקיים עיתונאיים המושתתים על מכירת מידע ולא על תשומת לב, כגון עלונים, לא בהכרח נותנים תמריץ להשגת עניין ציבורי רחב, וייתכן שהם אף נותנים תמריץ שלילי. מי יהיה מוכן לשלם מחיר גבוה עבור עלון שמכיל מידע שכבר ידוע לציבור ושכביר להניח גם זמין באינטרנט?

השאלה המרכזית שתקבע את עתידה של העיתונות כתחום עצמאי ובלתי-תלוי איננה כיצד העיתונות תתאים את עצמה לאינטרנט,

הצרפתי ז'ורז' פולטי, שנולד ב-1868, יצירות מהספרות היוונית הקלאסית ומהספרות הצרפתית ומנה את היסודות שמרכיבים כל יצירה ספרותית נחשבת, ובמילים אחרות את ה"קוד הגנטי" של כל ספרות טובה. פולטי הרכיב רשימה של 36 סיטואציות דרמטיות שמצויות בספרות דרמטית טובה. אלה כוללות פנייה לכוחות על-טבעיים, פשע שמונע מרגשות נקם, שכול או איחוד מחדש של משפחות, אסון, רגשות אשמה, התקוממות נגד רודן עריץ, אניגמה ועוד. הרשימה של פולטי פופולרית גם בימינו וסופרים רבים נעזרים בה כדי לפתח עלילה. טרי רוסיו, שכתב את התסריט של שרק, סיפר שהוא נעזר ברשימה של פולטי כדי לצאת ממבוי סתום בעלילת הסרט. את המאמץ לחשוף את ה"קוד הגנטי" של היצירה הדרמטית המשיכו אחרים אחרי מותו של פולטי.

הניסיונות לזהות את מרכיבי היסוד של סיפור טוב ושל דרמה משכנעת אמנם לא נכתבו למטרות אחזור מידע, אבל הם מספקים את ה"בלובים" הספרותיים שיכולים לאפשר לחוקרים לפרק סיפור עיתונאי למרכיבי התוכן שלו. אפשר לתרגם את מרכיבי התוכן המצויים בטקסטים לנוסחות מתמטיות שתמוכנה בניתוח תוכן ממוחשב. המטרה היא לאפשר לאנשים לאחזור מידע לפי 36 הסיטואציות של פולטי או לפי רשימה אחרת של סיטואציות. רשימה זאת תורכב מניתוח תוכן של ידיעות עיתונאיות, ותתמוך בניתוח משווה בין ידיעות עיתונאיות או בשיווקן לצרכנים, מותאמות לזהויותיהם הדיגיטליות.

### תוכן תקשורתי ומעורבות צרכנית

#### מעורבות צרכנית בתכני מדיה

עד לאחרונה, הבסיס הכלכלי של העיתונות ושל ערוצי התקשורת שאינם חלק מהשירות הציבורי היה ההכנסות מפרסום. ספקי חדשות שפעלו בתחום העיתונות המודפסת, המשודרת או הרדיו, מכרו תשומת לב צרכנית למפרסמים. המדרוג, למרות חוסר הדיוק, שימש כמדד העיקרי וככלי התמחור העיקרי. לפני האינטרנט לא הייתה שיטה למדידת העניין האמיתי שצרכן מסוים מגלה ביחידת תוכן נתונה. עם חדירתן של פלטפורמות מדיה חדשות, ובמיוחד טכנולוגיות אינטראקטיביות חדשות, גברה התחרות על תשומת לב הצרכן ובד בבד היא גם מורכבת יותר. על תעשיית העיתונות כיום לפתח שיטות חדשות למדוד עניין צרכני בכמה פרמטרים, ובהם פרמטר הפרופיל הקוגניטיבי, הפרופיל ההתנהגותי וההקשר. ההיבט האינטראקטיבי של פלטפורמות המדיה החדשה פותח אפשרויות חדשות למדידה מדעית של תשומת לב צרכנית בהתבסס על מאפייניו האישיים של כל צרכן. התחרות על תשומת הלב של הצרכן מתוארת כיום בעזרת המונח "מעורבות צרכנית" (consumer engagement). זהו מונח חדש יחסית שמתאר את מערכת היחסים החדשה בין הצרכנים לתוכן העיתונאי. החוקרים מקס קילגר ואלן רומר מבהירים את המושג: "לעורר רגשות חיוביים כלפי מותג מסוים בעזרת פרסום שמתאם לאווירה ולהקשר של התכנים"<sup>21</sup>. החוקרים מייחסים חשיבות להקשר שבו מועברים תכני מדיה בכל הפלטפורמות. קילגר ורומר זיהו שלושה מכניזמים שמעלים את רמת מעורבות הצרכן בהקשר של תוכן עיתונאי:

מכניזם קוגניטיבי (הרלוונטיות של תכני המידע והפרסום לצרכן)

את הדמיון בין התכנים של בלובים שונים. החתימות ששיתות סטטיסטיות שונות, כגון חלוקה לאשכולות וסיווג לתכנים, נותנות לתכנים ויזואליים, תומכות בחישוב ממוחשב ואוטומטי של הדמיון בין התמונות. החישובים האלה נעשים על ידי סיווג התמונות על פי מאפיינים כגון גוני צבע, הבדלים צורניים, הבדלים בטקסטורה ועוד שלל מאפיינים בולטים אחרים.

#### אחזור תכני וידאו

היו חוקרים שניסו לפתח שיטות שאינן תלויות בניתוח אנושי סובייקטיבי לאחזור אוטומטי של תכני וידאו. לשם כך היה צורך לפתח תוכנות חדשות להבחין בין היסטוגרמות צבע שונות שמקבילות לפריימים רצופים של וידאו.<sup>17</sup> פליקנר וצוותו מדווחים על מנוע חיפוש בשם ImageSpace המאפשר למשתמשים להזין שאילתות על עצמים חזותיים כגון שמים, עצים, מים וכדומה. תוכנות עזר מסוג זה משמשות בין היתר לזיהוי אוטומטי של תכנים פורנוגרפיים.<sup>18</sup>

#### ניתוח תוכן ממוקד אנשים

לסיפוק העניין האנושי בחיפוש אחר ידע ומידע בתכני מולטימדיה יש לשלב את הממד הרציונלי עם הממד הרגשי. החיפוש באינטרנט של הצרכן הוא תלוי הקשר ומושפע ממצבו הרגשי, ולכן תוצאות חיפוש מעוררות תגובות שונות בקרב אותו אדם בהתאם למצבו הרגשי בזמן החיפוש. רקע אישי, השכלה, וערכים – כולם משפיעים על האופן שבו אנשים מגיבים על תוצאות החיפוש ועל מידת הסיפוק שלהם מהן.

זווית חשובה למחקר בתחום ניתוח תוכן ממוקד אנשים נוגעת ברגשות שתכנים מסוימים מעוררים באנשים. ריטנדה דאטה ואחרים, בסקירה רחבה של תחום אחזור תמונות בשימוש בתוכנות בינה מלאכותית, מכנים את הממד הזה אסתטי. "אסתטיקה היא הרגש שעולה באנשים למראה תמונה". רגשות הן תגובות סובייקטיביות ויש לחקור אותן בהתאם. בעולם המחקר קוראים לזה לפעמים "חישוביות מעוררת רגשות" (affective computing), מתוך התמקדות בניסיון להבין את העולם הרגשי של המשתמש וכיצד אותו עולם רגשי משפיע על מידת שביעות הרצון של המשתמש מהמידע המאוחזר ומתוצאות השאילתות.<sup>19</sup>

בקרב חוקרים מתחום ה-CBIR קיימת ההבנה שמעורבות אנושית בתהליך הרב תבניתית לניתוח תוכן אוטומטי (automatic multi-model content analysis), ובמיוחד שילוב של משוב אנושי, חיוניים להפחתת טעויות ולהעלאת שביעות הרצון של המשתמשים. הכיוון החדש הזה במחקר מכונה "חישוביות ממוקדת אנשים" (Human Centered Computing). יש חוקרים המנסים להגדיר תמונות בהתאם לקטגוריות רגשיות.

אנדרו סולוויי ומייק גרהם פיתחו שיטה חדשה לזיהוי המצב הרגשי של דמויות בסרטים, שמקשרת בין המצב הרגשי של הדמויות לאירועים המתרחשים בו.<sup>20</sup>

#### ה"קוד הגנטי" של יצירות ספרותיות

במאה התשע-עשרה, לפני המצאת המחשב (אבל אחרי שכבר הלוגיקה הבוליאנית ותאוריות ההסתברות של בייס הניחו את היסודות המתמטיים למחשב ולא לגורמים המודרניים), ניתח הסופר

כדי להצליח בתקופה שבה שולטים החסכנות והציפיות הגבוהות, אנשי שיווק בתחום הדיגיטלי נדרשים לתת דין וחשבון על כל שקל [...] ההתמקדות בהחזר על ההוצאה (ROI) תכריח סוכנויות המפיקות חדשות [...] להתייעל, ואנו רואים תלות הולכת וגדלה באוטומציה [...] שינויים מהזמן האחרון [להגברת הליברליזציה] בתפיסת פרטיות הגולשים פתחו פתח עבור מפרסמים לעשות שימוש בבינה מלאכותית במטרה להריץ קמפיינים פרסומיים יעילים.<sup>25</sup>

המטרה הסופית של מדידת מעורבות צרכנית, מנקודת המבט של תעשיית הפרסום, היא להתאים פרסומות לצרכנים בהתאם לפרמטרים של הקשר ואישיות כפי שהוגדרו על ידי צוות קילגר: קוגניציה, רגשות והתנהגות. הפעולות האלה נחקרות ומוצאות לפועל כבר היום בערוצי מדיה חדשה, תחת הכותרת "פרסום מפולח" (BT Behavioral Targeting —) בקרב קהילות המחקר, העיתונות ותעשיית הפרסום.

**פרסום מפולח לפי התנהגות הצרכן (BT) ותוכן תקשורתי**  
בסוף שנות התשעים החל לצבור תאוצה תחום שיווק חדש שמשך אליו את תשומת הלב של קהילת המחקר האקדמי ושל עולם התעשייה: פרסום מפולח לפי אישיות הצרכן. החידושים בתחום האינטרנט והאפשרויות האינטראקטיביות שהוא מאפשר, שהפכו את הצרכנים ליצרני התוכן ולספקי התוכן, סללו את הדרך לאפיקים חדשים בתחום הפרסום המפולח ולהתאמת הפרסום להתנהגות הצרכנים ברשת. פרסום מפולח ברשת פירושו "היכולת לספק פרסומות לצרכנים בהתאם להתנהגותם בזמן שהם צופים בעמודי תוכן ברשת, רוכשים מוצרים ושירותים, מזינים מנועי חיפוש במילות מפתח או כל שילוב אפשרי של שלוש הפעולות האלה".<sup>26</sup>

חברות אינטרנט רבות פועלות בתחום הפרסום המפולח, ובהן גוגל, מיקרוסופט ויאהו. מייקל קסנר סקר את השימוש של גוגל בפרסום מפולח, דבר שגוגל מאשרת באתר הרשמי שלה.<sup>27</sup> למטרה הזאת גוגל משתמשת בשתי מערכות נפרדות: Adwords – מערכת AdSense. מערכת AdSense מילת מפתח ומשגרת פרסומות בהתאם לנושא החיפוש שהוזן; מערכת AdSense מתאימה את הפרסומות לתוכן העמוד הנצפה. "לשם דוגמה, כאשר אדם מבקר באתר בנושא גינון, הפרסומות באותו אתר יהיו קשורים בדרך זו או אחרת לגינון".<sup>28</sup> למערכת AdSense נוספה גם היכולת לחפש תמונות וסרטונים מתויגים ביוטיוב. לדברי קסנר, "גוגל מנסה לשכלל את היכולת שלה להציג פרסומות רלוונטיות באפליקציית ה-Gmail [...] המערכת סורקת את כל תכתובות המייל כדי לסנן דואר זבל [...] ולשדר פרסומות בהתאם למילות מפתח [...] והתהליך נעשה באופן אוטומטי ואיש איננו מתאים את הפרסומות לתכנים ב-Gmail".<sup>29</sup> הרציונל שמניע את גוגל גורס שככל שהפרסומות תהיינה רלוונטיות יותר לצרכנים, כך ההחזר על ההשקעה יהיה גדול יותר. בינתיים שתי המערכות מתבססות עדיין על ניתוח טקסטואלי. לכשתגיע השעה וכלי הניתוח יוכלו לסרוק ולתייג באופן אוטומטי גם תכני וידאו ותמונות, יש להניח שפרסום מפולח על פי התנהגות הצרכן ברשת יופעל על תכנים עיתונאיים בכל הפלטפורמות.

**מכניזם רגשי** (באיזו מידה הצרכן אוהב את תכני המידע והפרסום)  
**מכניזם התנהגותי** (באיזו מידה הצרכן מתעניין בתכני המידע והפרסום).<sup>22</sup>

ההיפותזה המרכזית השלטת בעולם הפרסום גורסת שצרכנים יוצאו יותר כסף על המוצר המפורסם ככל שמעורבותם בתכנים תהיה רבה יותר. התובנה הזאת, שמעורבות הצרכן בתכנים העיתונאיים נמדדת ברמה הקוגניטיבית, הרגשית וההתנהגותית של הצרכן, מעוררת את העניין שלנו בטכנולוגיות חדשות של בינה מלאכותית לאחזור ממוחשב של תכני מידע המשמשות לניתוח תוכן.

מחקר שערך צוות החוקרים בראשות קילגר על הקשר שבין מעורבות צרכנית להסתברות הסטטיסטית שהצרכן ירכוש מוצר שמופיע בפרסומת הצביע על כך שככל שהמדדים למעורבות צרכנית גבוהים יותר, כך גדלה ההסתברות הממוצעת שהמוצרים שפורסמו יירכשו. המחקר הקיף שלוש פלטפורמות מדיה: טלוויזיה, אינטרנט וכתבי עת מודפסים. בכל שלוש הפלטפורמות התוצאות היו דומות. "באינטרנט ובכתבי העת נצפו גרפים דומים ביותר, בעוד שבטלוויזיה נצפה גרף תגובה דומה אך הממוצע של הסתברות רכישה היה מעט נמוך יותר".<sup>23</sup>

הפרמטרים האישיים שקילגר וצוותו בחנו היו הפרמטרים הסטנדרטיים שנחקרים במדעי החברה: מין, גיל, רמת השכלה, רמת הכנסה, קבוצת השתייכות (race), ומעמד אישי. אלה שנמצאו משמעותיים ביותר היו הגיל, רמת הכנסה וקבוצת השתייכות. בבדיקה של טלוויזיה ואינטרנט, אנשים בעלי השכלה נמוכה יותר הביעו יותר אמון בפלטפורמת המדיה מאשר בעלי השכלה גבוהה יותר, ואילו מבוגרים יותר דיווחו על רמת מעורבות נמוכה יותר מאשר צעירים. אין זה צריך להפתיע שמאפיינים אישיים משפיעים על מעורבות בתכני מדיה. לממצא זה יש חשיבות רבה בחקר זהויות דיגיטליות. המידע על זהויות דיגיטליות הוא מידע חשוב למפרסמים, ואין ספק שהם יעשו שימוש במידע כזה ברגע שהוא יהיה זמין לשימושם בהיקף משמעותי, ובוודאי אם הוא יהיה זמין באופן אוטומטי. הדרך להפעלת לחצים על עיתונאים להפיק תכנים בכיוון של הגברת מעורבות צרכנית תהיה קצרה. היקף המחקר של קילגר אפשר לו לחקור רק מספר מוגבל של פרמטרים אישיים.<sup>24</sup>

האינטרנט מציע לא רק עניין צרכני ומעורבות צרכנית אלא גם תגובה צרכנית. המודל הפרסומי של תשלום עבור הקלקה – pay-per-click – הוא דוגמה למודל מסוג זה, שבו מפרסמים משלמים לא על חשיפה כי אם על פעולה של הצרכן. תאורטית אפשר עוד להקציף את המודל הזה. לדוגמה, לחיצה על פרסומת בדרך כלל מביאה את הגולש לאתר מכירות והדבר יכול להוביל להמשך התקשרות בין הצרכן למוכר, לרבות פעולת רכישה. לפיכך אפשר לדמייין מצב שבו פרסומות, ובאמצעותן תוכן עיתונאי, ממומנים על ידי עמלות מכירה. דא עקא, שמודל כזה מצביע בפני עיתונאים תמריצים עסקיים שעלולים לחתור תחת עקרונות מקצוע העיתונות.

כדי להפוך את שדה המחקר על היחסים בין מעורבות בתכנים לרכישות למדע של ממש, יהיה צורך לנתח עוד משתנים רבים, ובהם משתנים הקשורים להקשר. נוסף על כך יהיה צורך להפוך את התהליך לתהליך אוטומטי ולהכניס טכנולוגיות של בינה מלאכותית.

יראו מידע רלוונטי במקורות החדשות והעדכונים שלהם בנושא הרלוונטי".<sup>36</sup> דוויץ מכנה את הפעולה של התאמת ידיעות חדשותיות לתחומי העניין האישיים של החברים – פרויקט Smart Push, ובתרגום לעברית – שיווק מושכל. לדברי אולסן, "SRI בוחנת שימושים מסחריים עבור טכנולוגיית Smart Push שאינם קשורים לפרסום [...] כבר עכשיו אפשר ללמוד יותר על אנשים מ־MySpace ומפייסבוק".<sup>37</sup>

כאשר מכון מחקר מוביל ומשפיע כדוגמת SRI מקדם פיתוחים כמו "סמרט פוש", התוצאה הבלתי נמנעת היא שחברות תקשורת – שאצלן הרייטינג הוא הגורם הקובע – תתאמנה את התכנים העיתונאיים לפרופילים הדיגיטליים של הצרכנים. זה יכול להיעשות על ידי שימוש במנוע המבוסס על בינה מלאכותית, שיסנן או יפח את שוק השירותים ברשת על פי זהויות דיגיטליות.

### בינה מלאכותית: זהויות דיגיטליות ומנועים להתאמת תוכן לפי התנהגות הצרכן

ניהול זהויות דיגיטליות – פיתוח תקן בין־לאומי  
זהות הדיגיטלית היא מרכיב מרכזי בטכנולוגיית פילוח השוק הצרכני באינטרנט, ויש לה השפעה ישירה על סוג השירותים והמידע שהצרכן מקבל. תעשיית המידע הגלובלית משקיעה כיום משאבים אדירים בפיתוח שיטות לניהול זהויות דיגיטליות. התחום מתפתח בקצב מהיר ומכונה בעגה המקצועית Federated Identity Management (זהות משולבת). המושג מתייחס לכמה מרכיבים בפרופילים של המשתמשים המשולבים בזמן שהם גולשים ברשת באתרים שונים ומכונסים יחד לפרופיל אחד בסטנדרטים בין־לאומיים. מושג הזהות המשולבת משמש גם בתהליך הסטנדרטיזציה של זיהוי משתמש בפלטפורמות השונות. כיום התקן המוביל ליצירת זהויות דיגיטליות נקרא SAML2, ראשי תיבות של Security Assertions Markup Language 2.0, ובתרגום לעברית – שפה לתיג הרשאות ביטחוניות, דור 2.<sup>38</sup> תקן ה־SAML2 תומך בהאחדת הזהויות הדיגיטליות של גולשים בפלטפורמות שונות ובניהולן. מעבר לכך, תקן ה־SAML2 מאפשר "לגזור" מרכיבים מהגדרות הזהות הדיגיטלית של גולש, כפי שהיא מוגדרת ברשתות חברתיות שונות, ולמזג אותם יחד לפרופיל וירטואלי אחד. התקן נחל הצלחה בארגונים פיננסיים ואקדמיים, בממשל האמריקני שמעביר את הקשר עם האזרח לזירה האלקטרונית, ועוד.

תהליך הסטנדרטיזציה להגדרת זהויות דיגיטליות בהתאם לתקנים בין־לאומיים אחידים הוא רב חשיבות. הוא יאפשר לחוקרים לעקוב אחרי גולשים בכל אתר ברשת העולמית ולחקור את הקשרים שבין הזהות הדיגיטלית של הגולשים לאישיותם, לתחומי העניין שלהם וליכולותיהם הקוגניטיביות. לכל גולש יש דפוס גלישה אופייני שנגזר מיכולתו לקבל החלטות ומן הזיכרון שלו, נוסף על מאפיינים קוגניטיביים. אלגוריתמים המבוססים על בינה מלאכותית יכולים לנתח את המאפיינים הקוגניטיביים האישיים הללו באופן אוטומטי.

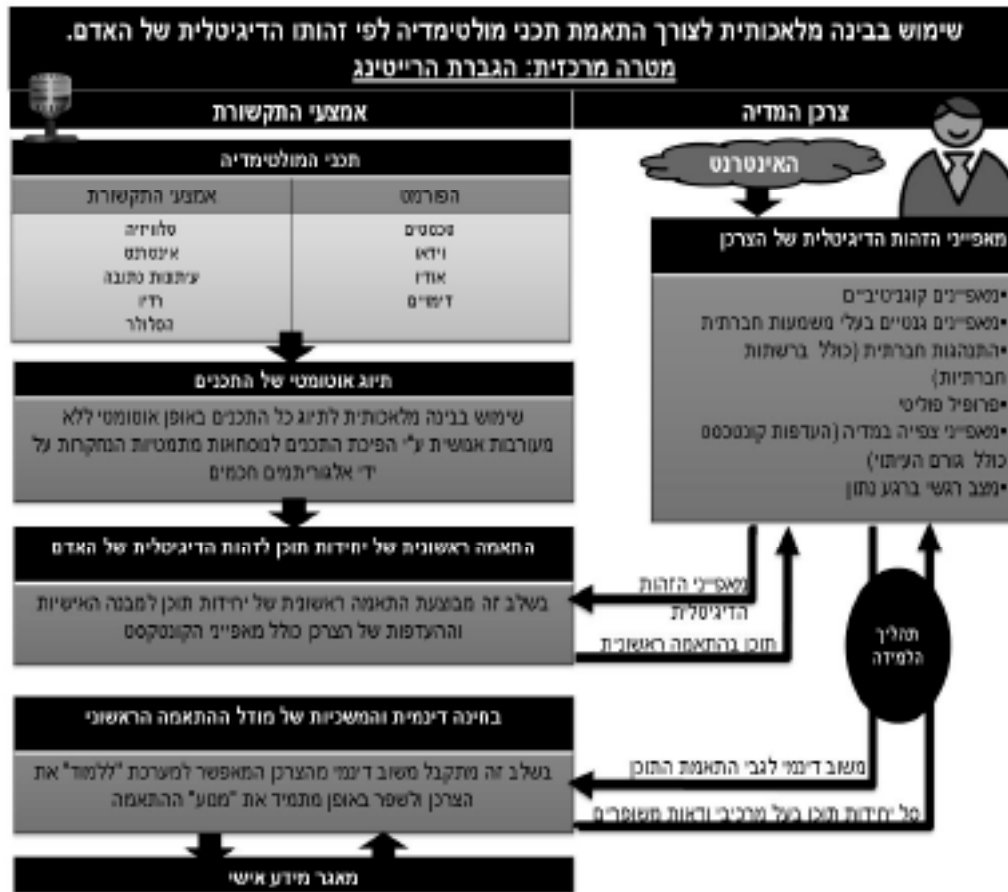
לא רחוק הזמן שבו אלגוריתמים המבוססים על בינה מלאכותית יוכלו להרכיב זהות דיגיטלית אישית לכל משתמש באינטרנט. "רובוטים" לכריית מידע באופן סטטיסטי יוכלו לנתח טקסטים כתובים, תכני וידאו ושמע, ולהמיר אותם למידע גנטי סוציולוגי (SDNA — Sociological DNA) שיתארו את אישיותם.<sup>39</sup> תהליך

**פרסום מפולח לפי התנהגות הצרכן ברשתות חברתיות**  
רשתות חברתיות שהחברים בהן מספקים מרצון מידע אישי בצורת טקסטים כתובים, תמונות וסרטוני וידאו, הן מקור בלתי נדלה לבניית פרופילים של צרכנים ולפרסום מפולח. הפרופילים האישיים של החברים ברשתות החברתיות כוללים רשימות של חברים, של תחביבים, של נתונים דמוגרפיים ושל תחומי עניין נוספים. תחום הפרסום המפולח בהתאם להתנהגות הצרכן ברשת מתרחב ומתפתח במהירות רבה ברשתות החברתיות. חברות הזנק עסוקות בפיתוח טכנולוגיות חדשות לפרסום מפולח על פי התנהגות הצרכן ברשת שמתאמות במיוחד לרשתות חברתיות. סטפני אולסן מביאה דוגמה של חברה ניו יורקית: 33Across.com.<sup>30</sup> האלגוריתמים של החברה עוקבים אחרי דפוסי הפעולה של הצרכנים ברשתות חברתיות, מזהים "סוציוגרמים" בין החברים, כלומר החברים בעלי ההשפעה על עמדות של אחרים והתנהגותם, מנתחים את הדינמיקה של התכתובות בין החברים ומצביעים עבור חברות הפרסום על האנשים בעלי ההשפעה הגדולה ביותר. חברת ההפקות Universal Pictures שכרה את שירותיה של חברת 33Across כדי לברר כיצד אנשים משתפים חברים בתכני האולפנים ובקדימונים לסרטים.<sup>31</sup> חברות שכבר התחילו לעשות שימוש בפילוח צרכנים בהתאם להתנהגותם ברשת למטרות פרסום הן Tacoda Systems, Reverence Science (עכשיו חברת AOL, שרכשה אותה). יאהו השיקה את מערכת SmartAds שמשלבת מידע התנהגותי ומידע דמוגרפי למטרת התאמת פרסום. ההערכה היא שההוצאה על פרסום מפולח בהתאם להתנהגות הצרכן ברשת תגיע למיליארד דולר ב־2010 ותנסוק ל־3.8 מיליארד דולר עד 2011.<sup>32</sup>

טכנולוגיות לפרסום מפולח בהתאם להתנהגות הצרכן ברשת מעוררות שאלות קשות בנושאי פרטיות. שאלות אלה נדונות בהרחבה בחוגים אקדמיים ופוליטיים, אך חורגות מנושא מחקרנו. טים ברנרס־לי, מי שנחשב להוגה הרעיון של רשת אינטרנט עולמית, נשא הרצאה בפני הפרלמנט הבריטי בנושא פרטיות והאינטרנט, שמטרתה לדבריו הייתה "להעלות את רמת המודעות להשלכות הטכניות, המשפטיות, והמוסריות של פעולות הציטוט ויצירת פרופילים של צרכנים שעושות ספקיות אינטרנט בשיתוף עם חברות לפרסום מפולח ברשת".<sup>33</sup> לדבריו, "יש לכך חשיבות עילאית שכאשר אדם מקליק על משהו ברשת, הוא מקליק ללא החשש שצד שלישי יודע על מה הוא לוחץ [...] באתי לכאן במטרה להגן על האינטרנט כמדיום תקשורת".<sup>34</sup> סקרים שערכה חברת TRUSTe (חברת אבטחה) מראים שהציבור "מגלה נכונות [...] להיות נתון למעקב וחשוף לשידור תכנים ממותגים".<sup>35</sup> אין ספק שזהו גילוי מרתק שדורש מחקרים נוספים.

### פרויקט Smart Push (שיווק מושכל)

ג'פרי דוויץ (Davitz) מחברת SRI מיישם טכניקות מעולם "המחשבים הלומדים" כדי לחקור תקשורת ברשתות חברתיות במסגרת פרויקט רחב היקף שזכה למימון בהיקף של עשרות מיליוני דולרים מהסוכנות לחקר קידום ההגנה שבמשרד ההגנה האמריקני (DARPA — Defense Advance Research Project Agency). המטרה המוצהרת של דוויץ הייתה "לעקוב באופן אוטומטי אחר תחומי העניין וההשפעה של אנשים ששייכים לסוכנויות ההגנה [...] לזהות את האנשים המשפיעים ביותר [...] ולהבטיח שאותם האנשים



מנוע המבוסס על בינה מלאכותית לפרסום מפולח לפי אישיות הצרכן שמסתמך על תוכן תקשורתי ועל זהותו הדיגיטלית של הצרכן

כוונתו להתחיל בכריית מידע סטטיסטית של הפרופילים והבלוגים האישיים של כמאה מיליון חבריה הרשומים במטרה להתאים להם פרסומות למוצרים ושירותים. נראה שאנו עומדים בתחילתו של תהליך הקמתה של מערכת סינון רחבה שמטרתה להתאים שירותים ומידע לחברים רשומים בהתאם לזהותם הדיגיטלית.<sup>41</sup> המטרה המוצהרת של הרשתות החברתיות היא לשפר את חוויית הגלישה בהן, או "להביא ערך מוסף לחוויית המשתמש" (בפרפרזה על דברי אלדוס האקסלי בספרו **עולם חדש מופלא**).

כניסתן של רשתות חברתיות לתחום הזהויות הדיגיטליות יוצרת הרחבה משמעותית ומסוכנת של התחום, משום שהן מביאות אליו מיפוי של הקשרים החברתיים והמקצועיים של הגולשים. תכונות המיפוי האלה תצטרפנה לפעולות של הגולשים ויש סיכוי שהן תהפוכנה לפילטר שיסנן את המידע והאפשרויות של הגולשים, וכל זאת ללא ידיעתם.

ההרכבה של הזהות הדיגיטלית הוא תהליך דינמי שמתעדכן כל זמן שאדם פעיל ברשת האינטרנט.

### זהויות דיגיטליות ורשתות חברתיות

אחת מהפעולות העיקריות של אנשים ברשת היא גלישה ברשתות חברתיות. מיליוני אנשים כיום חברים ברשתות חברתיות שמספקות מענה לצרכים חברתיים, כלכליים ופוליטיים. רשת חברתית היא קבוצת אנשים השומרים על קשר ביניהם במטרה לשתף זה את זה בתכנים כתובים, תכני וידאו ושמע, ובתמונות או לכל מטרה חברתית אחרת. כל חבר רשום ברשת חברתית מתחייב לספק פרטים אישיים שיהיו זמינים לחברים הרשומים האחרים ברשת או לחלקם, לפי רצון המשתמש.<sup>42</sup> חלק מהרשתות המובילות שהוקמו כמאגרי מידע לשימוש הגולשים, רואות כיום את מטרתן כספקיות שירותים, מידע ומוצרים בהתאם לזהויות דיגיטליות של החברים הרשומים. בספטמבר 2007 הודיע אתר Myspace למחזיקי המניות שלו על

## גנטיקה חברתית והזהות הדיגיטלית

הנפש והגוף קשורים זה בזה והמדע חותר בהתמדה לשפר את הבנת הקשר הזה. כיום ידוע שהתנהגות חברתית קשורה לגנטיקה. ככל שנעמיק בהבנת הקשר הזה ובאופן פעולתו בהקשר חברתי נשפר את יכולתנו להרכיב זהויות דיגיטליות, וכנראה גם לנתח את הגוף את הנפש ואת המערכת השלמה שבה הם נתונים, הווה אומר החברה. פירוש הדבר שמידע על הקוד הגנטי של אנשים פרטיים יכול לקדם את המחקר בתחום הזהויות הדיגיטליות נוסף על מידע שמספקות רשתות חברתיות.

מיפוי הקוד הגנטי של האדם מתפתח בקצב מהיר ומכוני מחקר מובילים, כמו גם חברות מסחריות גלובליות, שוקדים על פיתוח כלי עבודה ומחקר בתחום. המטרה העיקרית היא לבודד גנים הקשורים למחלות גנטיות ולפתח תרופות המתבססות על שיטות טיפול גנטיות. מאז הושלם פרויקט מיפוי הגנום האנושי בשנת 2001 גברה התחרות בין גופים מסחריים על היכולת לפתח מכשירים למיפוי הקוד הגנטי של האדם. פרויקט המחקר המרכזי בתחום נקרא "פרויקט הגנום האישי" (Personal Genome Project).<sup>42</sup>

הצטברות הידע על הקשר שבין גנים לתכונות אישיות וכניסתן של ענקיות אינטרנט כגוגל ומכוני מחקר מובילים כאוניברסיטאות הרוורד וקורנל לתחום המחקר הגנטי, צפויות להביא לסגירתם של פערי המידע בתחום במהירות רבה יותר. היקף גדול של משתתפים בניסויים, מאגרי מידע נרחבים על הזהויות הדיגיטליות של המשתתפים וטכניקות לכריית מידע על דפוסי ההתנהגות החברתית של אנשים ברשת בצירוף עם אלגוריתמים חכמים יאפשרו לחזות התנהגות חברתית שלילית וחיונית בהתאם למפת הגנום של המין האנושי.

תרשים הזרימה מתאר מנוע מבוסס בינה מלאכותית לפרסום מפולח לפי אישיות הצרכן. התרשים מתאר בקווים כלליים את האלמנטים הבסיסיים המנתחים באופן אוטומטי תוכן עיתונאי בכל הפלטפורמות ומעבירים יחידות תוכן רלוונטיות ופרסומות לצרכנים בהתאם לזהויותיהם הדיגיטליות. התרשים מתאר מודל של מחשב בעל כושר למידה דינמי שמתעדכן באופן רצוף, ו"לומד" את הפרופילים והעדפות התוכן של הצרכנים. גורמים בלתי מוכרים מקבלים ביטוי באמצעות הסתברויות שמתעדכנות במהלך תהליך ה"למידה". המנוע עוקב אחר בחירות הצרכנים בתכנים העיתונאיים ותגובותיהם, ובמקביל גם אחר הפרמטרים של ההקשר ושל מצבם הרגשי בתהליך האינטראקציה על ידי ניתוח תגובות מילוליות ואחרות של הצרכנים. להלן הסבר תמציתי של תרשים הזרימה.

שלב ראשון: אלגוריתמים חכמים המבוססים על בינה מלאכותית מנתחים את כל התכנים העיתונאיים ומצמידים להם באופן אוטומטי תגים. שלב שני: הזהויות הדיגיטליות של הצרכנים והתכנים המתויגים מוזרמים למנוע בדיקה ראשונית – Assessment Rule Engine – שמגדיר באופן ראשוני אילו תכנים יזרמו לצרכנים; הפרסומות המתאימות מוזרמות לצרכנים בהתאם לפרופילים שלהם. שלב שלישי: הצרכנים באים במגע עם התכנים והפרסומות; מידע על תגובות הצרכן ומעורבות הצרכן בתכנים נאסף ונמדד באופן רצוף. שלב רביעי: המנוע הלומד מנתח את המשוב מהצרכנים ומעדכן באופן אוטומטי את ההסתברויות שמתארות התנהגות צרכנית; תכנים חדשים נשלחים לצרכנים. שלב חמישי: המנוע הלומד שולח מידע

מעודכן למאגר המידע האישי שבו מוקמים, נשמרים ומעודכנים ברציפות הפרופילים של הצרכנים.

השילובים הרביעי והחמישי הם ליבת תהליך הלמידה ומאפשרים למנוע המבוסס על בינה מלאכותית להמשיך לחזות ברמת דיוק גבוהה את העדפות התוכן, תחומי העניין והמוצרים המועדפים של הצרכנים בהקשרים משתנים.

## זהויות דיגיטליות ופילוח שוק ברשת (Weblining)

סינון תוכן עיתונאי בהתייחס לפרופיל הצרכן יכול להוביל לאי שוויון חברתי חמור. מרסיסה סטרפנק (Sterpanek) טבעה את המונח Weblining כדי לתאר את התופעה הבאה:

נתאר זאת במונח weblining [פילוח שוק ברשת], הגרסה החדשה של עידן המידע לאותו מנהג דוחה המוכר בשם red lining [סימון באדום] שבו ספקי הלוואות ועסקים אחרים מסמנים שכונות מסוימות במרקר אדום, כמחוז' לתחום. למרחב המקוון אמנם אין גבולות גאוגרפיים, אך העובדה הזאת אינה מונעת מחברות לפלח את השוק ברשת ולסמן צרכנים מסוימים [...] פילוח השוק יכול לסגור אפשרויות בפניך או בפני העסק שלך לתמיד.<sup>43</sup>

מרשל בלונסקי (Blonsky), סוציולוג מאוניברסיטת ניו יורק, מרחיב על המשמעות של המונח: "מרגע שאני מסומן כחסר ערך, לעולם לא אזכה בגישה למוצרים ולשירותים ולהזדמנויות הכלכליות שזורמות לאחרים ברשת".<sup>44</sup>

זהויות דיגיטליות הן רכיב המפתח בפילוח השוק. אף על פי שהדגש של סטרפנק ובלונסקי היה על הממדים הכלכליים של הנהגה הזו, לתופעה זאת יש השלכות גם על דפוסי ההפצה של תכנים עיתונאיים בהתאם לפרופילים האישיים של המשתמשים. אי אפשר לדרוש מנציגי כוחות השוק – המפרסמים וסוכנויות הידיעות – לגלות אלטרואיזם ולהקים מנגנונים להבטיח את זכותנו לגישה שווה לתוכן. המקור העיקרי לדאגה הוא שהלחצים להתאים את התכנים לפרופיל הצרכני של הגולשים עלולים להשפיע על איכותם של התכנים העיתונאיים.

## זהויות דיגיטליות והפרקטיקה העיתונאית

הפילוח הממוקד באישיות הצרכן יאפשר למו"לים של העיתונות ולעיתונאים לערוך דמיות ואומדנים לחישוב ההשפעה של כתבות על קהל היעד ועל גורמים בעלי עניין אחרים עוד בשלב פיתוח הידיעות. הם יוכלו להריץ את הידיעות במבדק מקדים לפני פרסומן, כשם שמפרסמים בודקים כיום מוצרים חדשים לפני שיוקם. האפשרות הזאת תפתח בוודאי אפשרויות מעניינות בפני עולם העיתונות, אך גם תעמיד בפניו סכנות ואתגרים חדשים.

פרסום טקסטואלי פשוט לא חייב לאיים על העיקרון העיתונאי של הפרדה בין תחום הפקת המידע לתחום המכירות, שעניינו מכירת תשומת הלב של הצרכנים למפרסמים (שאוילי תלויים בכתבות העיתונאיות בדרכים נוספות). ככל שהפרסום המבוסס על הקשר ישתכלל הפרסומות תהינה רלוונטיות יותר. המודלים העסקיים הקיימים כיום, המבוססים על תשלום על לחיצות עכבר של צרכנים על הפרסומות – pay-per-click – ייצרו תמריץ למו"לים העיתונאיים להתמקד בסיפורים שמתאימים לפרסומות. אם התחזית הזאת תתממש, התוצאה עלולה לאיים על חופש העיתונות. "ההפרדה



ייתכן על מכירת מעורבות צרכנית למפרסמים. העיתונות תתקשה לשמור על נאמנותה לקהל האזרחים אם העיתונאים ישתכרו יותר אם הם יגלו נאמנות לזהויות הדיגיטליות של המפרסמים.

**העיתונאים חייבים לשמור על מעמדם העצמאי והבלתי תלוי בנושאי הסיקור שלהם.** ייתכן שאפשר יהיה לשלב מודלים התנהגותיים של אלה שמסוקרים בכתבות במסגרת ה"נוסחה" של האלגוריתם. אפשרות כזאת תשפר את יכולתם של העיתונאים לתכנן סדרות של כתבות ולדעת מראש כיצד הכתבות תתחברנה. המידע הזה ייתן לעיתונאים כלי לחזות את השפעת הכתבות על בעלי העניין. המודעות המוגברת הזאת היא חרב מתהפכת שיכולה לסייע לעיתונאים לשמור על עצמאותם בחלק מהמקרים, אך להקשות עליהם לשמור על מעמדם העצמאי במקרים אחרים.

**יש לשמור בקנאות על חופש הפעולה המוסרי של עיתונאים.** אם המודלים העסקיים ומערכת הזהויות הדיגיטליות והפילטרים יאפשרו לחזות מראש את ההכנסות שכתבה מסוימת צפויה להניב עוד בזמן שהיא נכתבת, או כיצד הסיפור ישפיע על בעלי העניין בארגון העיתונאי, קיימת הסתברות גבוהה שהמצפון האישי של העיתונאים יתעמת עם האינטרסים של בעלי העניין האחרים, בעיקר אלה העסקיים. במילים אחרות, העיתונאי יחשוב ש"אם אני כותב את הכתבה באופן שאני רוצה לכתוב אותה, המו"ל שלי יידע שאני בוחר להרוויח פחות כסף" או "אם אני כותב את הכתבה באופן שאני רוצה לכתוב אותה, המו"ל שלי יידע שאני בוחר להגביר את הסיכון שאכניס אותנו לעימות עם המפרסמים".

אלה דוגמאות לסוגי השאלות שיש לתת עליהן את הדעת בתהליך הפיתוח של מערכות זהויות דיגיטליות ומערכות סינון עבור העיתונות.

**קוד אתי להכוונת השימוש בזהויות דיגיטליות בתקשורת** לדעתנו יש לפתח עקרונות מוסכמים שיסייעו בפיקוח על השימוש בזהויות דיגיטליות בתחום העיתונות. הנה כמה עקרונות כאלה שאנו מציעים: (1) הצרכים האנושיים קודמים לצורכי הזהויות הדיגיטליות. הזהויות הדיגיטליות לעולם לא תהיינה זהות לאישיות השלמה של הפרט ויש להביא זאת בחשבון. יתר על כן, האנשים החיים חשובים יותר מהזהויות הדיגיטליות שלהם. לכן יש לחייב את המערכות לשקפות מלאה ולחשוף בפני הצרכן את הזהות הדיגיטלית שהוגדרה עבורו לפי בקשתו. (2) אין לאפשר לשימוש בזהויות דיגיטליות בתחום העיתונות לפגוע בנאמנות העיתונות לקהל קוראיה או במעמדה הבלתי תלוי במקורות המידע שלה. (3) אין לאפשר לשימוש זה לפגוע בחופש של העיתונאים לפעול על פי עקרונות המוסר האישיים שלהם.

**הצורך להמשיך את הדיון בהשתתפות כלל בעלי העניין בחברה הדמוקרטית**

בכנס שארגנה ב-2009 האגודה לקידום בינה מלאכותית (AAAI – Association for the Advancement of Artificial Intelligence), נפגשה קבוצה של חוקרים מתחום מדעי המחשב, בינה מלאכותית ורובוטיקה במרכז הקונגרסים אסילומר שבמפרץ מונטריי, קליפורניה, כדי לדון בשאלה אם ראוי שיוטלו מגבלות על מחקר שיכול להוביל לאובדן השליטה האנושית על מערכות המבוססות על מחשב שנושאות בנתח גדל והולך של פעילות החברה.

המסורתית בין דת ומדינה – מטבע הלשון המשמשת את המו"לים בהשאלה להבחנה בין ידיעות חדשות לפרסום בתשלום – תטושטש. נניה שהקוראים יוכלו להשתמש בזהויותיהם הדיגיטליות בשילוב פילטרים כדי לבחור אילו ידיעות הם רוצים שיובאו לידיעתם. ונניה שאיכות הפילטרים והזהויות הדיגיטליות טובה מספיק כדי להעריך שידיעה מסוימת תמשוך את תשומת לבם ותדרבן אותם לפעול. נדמיין לעצמנו מפרסמים מבוססי הקשר (כלומר כאלה שהם עצמם זהויות דיגיטליות) המשלמים עבור תשומת לב הקוראים ואינטראקציה עם הקוראים, ועיתונאי שיש לו גישה לפילטרים ולזהויות הדיגיטליות האלה, כמו גם למפרסמים מבוססי הקשר בזמן שהוא כותב ידיעה. אותו עיתונאי יכול לבחון את התוכן של הכתבה או הידיעה במהלך כתיבתה על זהויות דיגיטליות הן של הקהל והן של מפרסמים על ידי בחינה של תגובות לתוכן דומה שכתב בעבר. אותו עיתונאי יוכל לאחר מכן להתאים את תוכן הכתבה במטרה להשיג את התוצאות ה"טובות" (כלומר הגדלת ההכנסות למו"ל) ביותר שתוגדרנה כשילוב של מטרות העיתונאי, הקהל והמפרסמים. התנהגות כזאת נוגדת את האתיקה העיתונאית. לפי האתיקה הזאת העיתונאי מקבל את זכות הכתיבה העיתונאית מקהל קוראיו.

נניה שהזהות הדיגיטלית של אותו עיתונאי יכולה להיכלל באינטראקציה. זהות זו מורכבת מפילטרים שעוזרים לו לבחור את הנושאים שהוא מבקש לסקר בהתאם לזהויות הדיגיטליות של הקוראים ושל המפרסמים המקושרים למקורות החדשות. הווה אומר, הציצים בטוויטר, הפרסומים בבלוגוספירה וכל מקורות החדשות האחרים באינטרנט, כגון שוקי ההון, חיישנים שמודדים את מזג האוויר או מתריעים מפני רעידות אדמה וכיוצא באלה. באופן הזה העיתונאי יכול לקבל מידע על אירועים שיש בהם מתאם בין תחומי העניין האישיים שלו לתחומי העניין של הקהל והמפרסמים.

בעיקרו של דבר, לידיעה חדשותית מוצלחת נדרש שילוב של זהות העיתונאי, של הקהל ושל המודל העסקי, המפרסם. שילוב נכון בין זהויות דיגיטליות לפילטרים יוצר תהליך מעגלי והיריסטי של זיהוי תחומי עניין משותפים, ואופטימיזציה במטרה להשיג תוצאה שתמקסם את הערך (יוקרה ופרנסה לעיתונאי ורווח למו"לים ולמפרסמים) לכל אחד מהצדדים. בכל אינטראקציה מספקים המשתמשים האמיתיים שמאחורי הזהויות הדיגיטליות משוב המאשר או מחייב עדכון של הזהויות הדיגיטליות ושל פעולות הפילטרים במטרה לשפר את התוצאות בסבב הבא. האם תהליך כזה יעמוד בערכי האתיקה העיתונאית?

**עקרונות להכוונת הממשק בין התקשורת לזהויות דיגיטליות**

נקודת המפגש שנדונה לעיל בין זהויות דיגיטליות לתכנים היא תחילתו של תהליך רגיש שעלול לגרום תוצאות מסוכנות. מבחינת העקרונות שמנחים את עולם העיתונות כיום, אפשר להצביע על כמה אלמנטים בפרקטיקה העיתונאית הקיימת שעשויים-עלולים להיות מושפעים באופן מרחיק לכת מהתהליך המתרחש.

**הנאמנות העלוונה של העיתונות היא לאזרחים.** ייתכן מצב שבו עיתונאים יהיו נתונים לחצים לגלות נאמנות לזהויות הדיגיטליות של אזרחים במקום לאזרחים עצמם. הבעיות עלולות להתעורר במצב שבו כל ידיעה עיתונאית תשוך ישירות למודל העסקי והמודל העסקי

and Hall, London, UK. Reprinted, Methuen and Company, London, UK, 1964. 2) 'Man-Computer-Symbiosis' Licklider, J.C.R. (1960). "Man-Computer Symbiosis", *IRE Transactions on Human Factors in Electronics*, vol. HFE-1, 4-11.

2 A. Turing (October 1950). "Computing Machinery and Intelligence", *Mind*, 49, pp. 433-460.

3 Techmeme, <http://techmeme.com>, arranges tech journalism story links into a single page. Techmeme works by scraping news websites and blogs, and then compiles a list of links to the most popular technology-related news of the day, which is continuously updated. The stories selected are all chosen by an automated process. <http://en.wikipedia.org/wiki/Techmeme> (11 Jan 2010).

4 D. Nordfors (2008). "Separating Journalism and the Media", *EJC Magazine*, 4 Dec 2008, European Journalism Centre, [http://www.ejc.net/magazine/article/separating\\_journalism\\_and\\_the\\_media/](http://www.ejc.net/magazine/article/separating_journalism_and_the_media/)

5 Wikipedia (29.8.2009). [http://en.wikipedia.org/wiki/Advertising#cite\\_note-2](http://en.wikipedia.org/wiki/Advertising#cite_note-2). "Global Entertainment and Media Outlook: 2006-2010, a report issued by global accounting firm PricewaterhouseCoopers". Pwc.com. Retrieved 2009-04-20.

6 "The World Bank: World Development Indicators database, 1.7.2009. Gross domestic product (2008) <http://siteresources.worldbank.org/DATASTATISTICS/Resources/GDP.pdf>.

7 N. Lemelshtrich Latar & D. Nordfors (2009). "Digital Identities and Journalism Content", *The Innovation Journalism Publication Series*, 6(7), Nov. 11. VINNOVA, Stanford Research Center of Innovation Journalism, Wallenberg Hall, Stanford University. *Compact Oxford English Dictionary*, published on-line by AskOxford.com, retrieved Sep 6 2009

8 B. Kovach & T. Rosenstiel (2001). *The Elements of Journalism: What Newspeople should Know and the Public should Expect*. New York: Three Rivers Press.

9 D. Nordfors (2009). "Innovation Journalism, Attention Work and the Innovation Economy: A Review of the Innovation Journalism Initiative 2003-2009," *Innovation Journalism*, 6 (1), p. 16. <http://www.innovationjournalism.org/archive/injo-6-1.pdf>, Retrieved Sep 9 2009.

10 D. Nordfors (2006). "PR and the Innovation Communication System," *Innovation Journalism*, 3 (5), p. 8. <http://www.innovationjournalism.org/archive/INJO-3-5.pdf>, also published by *Strategic Innovators* (July-Sept 2007, Volume I | Issue 3).

11 M.E. McCombs & D.L. Shaw (1972). "The Agenda-Setting Function of Mass Media," *Public Opinion Quarterly*, 36, pp. 176-187.

12 J. Jeon V. Lavrenko & R. Manmatha (2003). "Automatic Image Annotation and Retrieval Using Cross Media Relevance Models," in J. Callan et al (eds.), *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York: Association for Computing Machinery, pp. 119-126 (Jeon et al, להלן: "Automatic Image").

14 A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta & R. Jain, (2000). "Content Based Image Retrieval at the End of the Early

דאגתם התמקדה בעיקר בחשש שהתפתחויות נוספות במחקר יובילו לבעיות חברתיות ואף להשלכות מסוכנות של ממש, ותאלצנה את בני האדם ללמוד לחיות עם מכונות שמחקות התנהגות אנושית.<sup>45</sup> החוקרים בכנס חששו בעיקר מאובדן מקומות עבודה או מכך שעבריינים ישיגו גישה לכלים האלה. הם לא התייחסו להשלכות ההרסניות על העיתונות בשל השימוש בכלים המבוססים על בינה מלאכותית. מארגן המפגש, ד"ר אריק הורוביץ מחברת מיקרוסופט, אמר שהוא "מאמין שחוקרים מתחום מדעי המחשב חייבים לתת מענה לתחושה שאנו מאבדים שליטה על מכונות 'סופראינטליגנטיות' ועל כלי בינה מלאכותית [...] הפאנל ניסה להתוות דרך למחקר עתידי בתחום במטרה להבטיח שהטכנולוגיה תשפר את החברה במקום להפיל אותה לקטסטרופה טכנולוגית".<sup>46</sup>

הגיעה העת לארגן כנס דומה בהשתתפות מדעני מחשב, מומחים לבינה מלאכותית, חוקרים מתחום אחזור מידע מולטימדי, עיתונאים ומומחי עיתונות, מומחים לתקשורת ולחברה וכלכלנים המתמחים במודלים עסקיים לתעשיית התקשורת, כדי לדון בהשלכות הפוטנציאליות של אלגוריתמי בינה מלאכותית על מקצוע העיתונות ועל תפקיד העיתונות בחברה דמוקרטית. בין השאלות שצריכות להישאל בכנס כזה יש למנות את השאלות הבאות:

1. האם אנשים ינהלו את הזהויות הדיגיטליות שלהם או שהזהויות הדיגיטליות שלהם תנהלנה אותם?
2. מהן ההשלכות של זהויות דיגיטליות על ההגדרה של העיתונות?
3. בעידן האינטרנט, תפקיד העיתונות כבר אינו מוגבל לשידור ידיעות אלא כולל גם ערוץ תקשורת דו-כיווני עם קהל הקוראים וארגון פורומים לניהול שיח ציבורי. אם כן, מהו תפקיד העיתונות בחברה?
4. מה תהיה השפעת המפגש עם זהויות דיגיטליות על עקרונות הפרקטיקה העיתונאית?
5. באילו מודלים עסקיים תומכות זהויות דיגיטליות? באיזו מידה יספקו העיתונאים עניין צרכני וישתכרו ממכירת עניין הקוראים למפרסמים ובאיזו מידה יהיו העיתונאים ספקי מידע שישתכרו מהפצת מידע?
6. יש לקבוע עקרונות שיתאימו למקצוע העיתונות בעולם שבו המפגש עם זהויות דיגיטליות משיק לעשייה העיתונאית, לדרכים שבהם העיתונות מייצרת תמורה עבור אנשים ולאופן שבו העיתונות מייצרת רווחים עבור חברות התקשורת.
7. יש לבחון בניית מודלים עסקיים חדשים לעיתונות שיבטיחו את חופש העיתונאי למלא את תפקידו החשוב בשמירה על הדמוקרטיה.
8. כיצד ליישם את העקרונות העיתונאיים ואת תהליך עדכוןם בסביבה של זהויות דיגיטליות?

## הערות

- 1 The expression 'augmented intelligence' is attributed to D.C. Engelbart (Oct 1962). "Augmenting Human Intellect: A Conceptual Framework", Summary Report AFOSR-3233, Stanford Research Institute, Menlo Park, CA. Related concepts: 1) IA or 'Intelligence Amplification' by Ashby, W.R. (1956), *An Introduction to Cybernetics*, Chapman

- G. Abramovitch (2007). "Myspace has Data Mining Plans," 41  
www.dmnews.com, Sept. 24. www.personalgenomes.org.
- M. Sterpanek (2000). "Weblining," *businessweek on-line*, 42  
April 3.
- J. Markoff (2009). "Scientists Worry Machines May Outsmart 44  
Man," *NYT.com*, July 26.
- 45 ש.מ.
- ### מקורות
- Abramovitch, G. (2007). "Myspace has Data Mining Plans" retrieved  
on 20.4.2009 from: <http://www.dmnews.com/myspace-has-data-mining-plans/article/98564/>.
- Aho Williamson, D. (2005). "White Paper on Behavioral Targeting,"  
*Wall Street Journal and e-Marketer*, May 11.
- Ashby, W.R. (1956). *An Introduction to Cybernetics*. London, UK :  
Chapman and Hall.
- Boyd, D.M. & N.B. Ellison (2007). "Social Network Sites:  
Definition, History and Scholarship" *Journal of Computer  
Mediated Communications*, 13 (1), pp. 210-230
- Duygulu, P., K. Barnard, N. de Freitas & D. Forsyth (2002). "Object  
Recognition as Machine Translation: Learning a Lexicon for  
Fixed Image Vocabulary" *Seventh European Conference on  
Computer Vision*, pp. 97-112.
- Engelbart, D.C. (1962). "Augmenting Human Intellect: A Conceptual  
Framework" *Summary Report AFOSR-3233*, Stanford Research  
Institute: Menlo Park, CA, p. 19.
- Flickner, M. et al. (1995). "Query by Image and Video Content: The  
QBIC System" *Computer Magazine*, 28 (9), pp. 23-32.
- Jeon, J., V. Lavrenko & R. Manmatha (2003). "Automatic Image  
Annotation and Retrieval Using Cross Media Relevance  
Models" in J. Callan et al (eds.), *Proceedings of the 26th  
Annual international ACM SIGIR Conference on Research and  
Development in Information Retrieval*, New York: Association  
for Computing Machinery, pp. 119-126.
- Joshi, D. et al. (2008). "Image Retrieval: Ideas, Influences and  
Trends of the New Age" *ACM Computing Surveys*, 40 (2), p. 46.
- Kassner, M. (2009). "Google Quietly Starts Behavioral Targeting"  
*ZDNetAsia*, April 21. <http://www.zdnetasia.com/google-quietly-starts-behavioral-targeting-62053320.htm>.
- Kilger, M. & E. Romer (2007). "Do Measures of Media Engagement  
Correlate with Product Purchase Likelihood?" *Journal of  
Advertising Research*, 47(3), pp. 313-325.
- Kovach, B. & T. Rosenstiel (2001). *The Elements of Journalism:  
What Newspeople should Know and the Public should Expect*.  
New York: Three Rivers Press.
- Lemelshtich Latar, N. (2004). "Personal Web Social DNA and  
Cybernetic Decision Making" *International Communication  
Association Conference*, Hubert Burda Center for Innovative  
Communications, BGU, Feb 2004.
- Lemelshtich Latar, N. & D. Nordfors (2009). "Digital Identities and  
Journalism Content" *The Innovation Journalism Publication  
Series*, 6(7). available at:
- Years," *IEEE Transactions on Pattern Analysis and Machine  
Intelligence*, 22(12), pp. 1349-1380.
- P. Duygulu K. Barnard, N. de Freitas & D. Forsyth (2002). 15  
"Object Recognition as Machine Translation: Learning a  
Lexicon for Fixed Image Vocabulary", *Seventh European  
Conference on Computer Vision*, pp. 97-112.
- Jeon et al, "Automatic Image" 16  
ההרגום שלנו.
- M. Flickner et al. (1995). "Query by Image and Video 17  
Content: The QBIC System," *Computer*, 28(9), pp. 23-32.  
doi:10.1109/2.410146.
- M. Flickner et al. (1995). "Query by Image and Video Content: 18  
The QBIC System" *Computer Magazine*, 28 (9), pp. 23-32.
- D. Joshi et al. (2008). "Image Retrieval: Ideas, Influences and 19  
Trends of the New Age", *ACM Computing Surveys*, 40 (2), p.  
46.
- A. Salway & M. Graham (2003). "Extracting Information 20  
about Emotions in Films," *Proceedings of the Eleventh  
ACM International Conference on Multimedia, AAAI Spring  
Symposium*, November 2-8, pp. 299-302.
- M. Kilger & E. Romer (2007). "Do Measures of Media 21  
Engagement Correlate with Product Purchase Likelihood?"  
*Journal of Advertising Research*, 47(3), pp. 313.
- 22 ש.מ.
- 23 ש.מ, עמ' 322.
- 24 ש.מ.
- 25 ש.מ, עמ' 325.
- Aho Williamson, D. (2005). "White Paper on Behavioral 26  
Targeting," *Wall Street Journal and eMarketer*, May 11
- M. Kassner (2009). "Google Quietly Starts Behavioral 27  
Targeting," *ZDNetAsia*, April 21 URL:<http://zdnetasia.com/techguide/security/0,39044901,62053320,00/htm>.
- 28 ש.מ.
- 29 ש.מ.
- S. Olsen (2008). "33Across: The Next Generation of Behavioral 30  
ad Targeting," *news.cnet.com*, June 23.
- 31 ש.מ.
- Mills, E. (2007). "AOL buys ads from Tocada," *ZDNetAsia*, July 32  
25.
- Watson, F. (2009). "Behavioral Targeting: Profiling or Projecting 33  
User Experience," *Search Engine Watch*, Mar 13.
- 34 ש.מ.
- 35 ש.מ.
- Olsen, S. (2008). "33Across: The Next Generation of Behavioral 36  
ad Targeting", *news.cnet.com*, June 23.
- 37 ש.מ.
- P. Madsen, "SAML2: The building blocks of federated identity," 38  
Jan 2005, xml.com.
- N. Lemelshtich Latar (2004). "Personal Web Social DNA and 39  
Cybernetic Decision Making," *International Communication  
Association Conference*, Hubert Burda Center for Innovative  
Communications, BGU, Feb 2004.
- D.M. Boyd & N.B. Ellison, "Social Network Sites: Definition, 40  
History and Scholarship," *Journal of Computer Mediated  
Communications*, 13(1), 2007, pp. 2.

- Initiative 2003-2009" *Innovation Journalism*, 6(1), p. 8. Available at: <http://www.innovationjournalism.org/archive/injo-6-1.pdf>.
- Olsen, S. (2008). "33Across: The Next Generation of Behavioral ad Targeting" *news.cnet.com*, June 23. Retrieved on 21.4.2009 from: [http://news.cnet.com/8301-10784\\_3-9974210-7.html](http://news.cnet.com/8301-10784_3-9974210-7.html).
- Salway, A. & M. Graham, (2003). "Extracting Information about Emotions in Films" Proceedings of the Eleventh ACM International Conference on Multimedia, AAAI Spring Symposium November 2-8, pp. 299-302.
- Smeulders, A.W.M. et al (2000). "Content Based Image Retrieval at the End of the Early Years" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), pp. 1349-1380.
- Sterpanek, M. (2000). "Weblining" *businessweek on-line*, April 3. Retrieved on 20.4.2009 from: [http://www.businessweek.com/2000/00\\_14/b3675027.htm](http://www.businessweek.com/2000/00_14/b3675027.htm).
- Turing, A. (1950). "Computing Machinery and Intelligence" *Mind*, 49, pp. 433-460.
- Watson, F. (2009). "Behavioral Targeting: Profiling or Projecting User Experience" *www.SearchEngineWatch.com*, Mar 13. [www.innovationjournalism.org/archive/INJO-6-7.pdf](http://www.innovationjournalism.org/archive/INJO-6-7.pdf).
- Licklider, J.C.R. (1960). "Man-Computer Symbiosis" *IRE Transactions on Human Factors in Electronics*, 1, pp. 4-11.
- Markoff, J. (2009). "Scientists Worry Machines May Outsmart Man" *NYT.com*, July 26, retrieved on 21.4.2009 from: <http://www.nytimes.com/2009/07/26/science/26robot.html>.
- McCombs, M.E. & D.L. Shaw (1972). "The Agenda-Setting Function of Mass Media" *Public Opinion Quarterly*, 36, pp. 176-187.
- Mills, E. (2007). "AOL buys ads from Tocada" *ZDNetAsia*, July 25.
- Nordfors, D. (2006). "PR and the Innovation Communication System" *Innovation Journalism*, 3(5). Available at: <http://www.innovationjournalism.org/archive/INJO-3-5.pdf>. Also published by *Strategic Innovators* (July-Sept 2007), 1 (3), p. 16.
- , (2008). "Separating Journalism and the Media" *EJC Magazine*, 4 Dec 2008, European Journalism Centre. Retrieve on 20.4.2009, from: [http://www.ejc.net/magazine/article/separating\\_journalism\\_and\\_the\\_media/](http://www.ejc.net/magazine/article/separating_journalism_and_the_media/).
- , (2009). "Innovation Journalism, Attention Work and the Innovation Economy: A Review of the Innovation Journalism