

האלגוריתמים כובשים את מדעי הרוח

איך מנתחים אוסף של מאות אלפי גלויות של ארץ ישראל מהמאה ה-19, כמה אנשים ידעו קרוא וכתוב בתקופות עתיקות ומה הושמט מעדויות של ניצולי שואה? • ד"ר רננה קידר, שהקימה את המרכז למדעי הרוח הדיגיטליים באוניברסיטה העברית, מאמינה שכלים מתמטיים יכולים לקדם את המחקר

גלי ינבר

בקרוב, לא תצטרכו לדאוג שהילד שהוקסם מהיסטוריה או מפילוסופיה יישאר "כלי מקצוע". גם אנשי הפקולטה למדעי הרוח, מסטודנטים ועד חוקרים, יוכלו להכריז על עצמם כעל "אנשי הייטק" ולהעניק לעצמם תארים סקסיים כמו "מומחי ניתוח תוכן". כן, גם מדעי הרוח הופכים כיום לדיגיטליים.

כבר ב-2010 פורסם מאמר בכתב העת Science, המתאר ניתוח סמני טי של 4% מספרות העולם שהייתה אז במאגר Google Books. המהות קר הוצג כמפגן של יכולות ופוטנציאל לניתוח ספרות דיגיטלי, וחלק ממסגרותיו הוצגו גם בכתבה שפורסמה ב"ניו יורק טיימס" באותה תקופה. לדוגמה, הספרות מזכירה הרבה יותר פעמים את ג'ימי קרטר, נשיא ארה"ב לשעבר, מאשר את מרלין מונרו ומיקי מאלה, לכאורה דמויות הרבה יותר מפורסמות בעולם, ובאנגלית מוזכרת כיכר טיאננמן לעתים קרובות הרבה יותר מאשר בסניגית החוקרים גם הבחינו שהמילה Grill-ing החליפה מסוף שנות ה-90 בהדרגה את המילה Roasting.

"אנחנו רוצים לתת לכל בן 8 כלי לחקור את התרבות האנושית", התגאה אז אחד ממפתחי הכלי הזה, פרופ' ארוז ליברמן איידן, היום חוקר גנטיקה באופן ממוחשב בבית הספר לרפואה של אוניברסיטת ביילור. אלא שהמחקר הזה בוצע על ידי אנשי מחשבים, וגם אם המסקנות שלהם היו נכונות, לא היה להן הקשר, אומרת ד"ר רננה קידר, שיוזמה לאחורונה את הקמתו של המרכז למדעי הרוח הדיגיטליים באוניברסיטה העברית. "לא היה שם חוקר מתחום מדעי הרוח ששאל שאלה מדעיות מעניינות, ולא היה חוקר שיפענח את התשובה". איך נראה מחקר שיש לו הקשר?

"אם, לדוגמה, אנחנו רוצים לדעת אם ספרים נכתבו על ידי נשים בשמות כרויים של גברים", אומרת קידר, "אנחנו יכולים להוין את המערכת בספרים שכתבו גברים ובספרים שאנחנו כבר יודעים שכתבו הספרים שיש לנו במאגר מתקופה מסוימת, והיא תגיד לי 'אני חושבת שאת ספר מס' 501 כתבה אישה'". אבל אחרי שהמערכת הצביעה על הספר "החשוד", קידר עדיין תרצה לגייס פילולוג שיחקור אותו באמצע עים הקלאסיים של התחום.

שיחלפו חוקרים במדעי הרוח. מעד כות לא יודעות כלום אם לא מסבירים להן מה לחפש, ואת המסקנות שלהן מייששו צריך לפענח ולתקף."

"אנשי מדעי הרוח הם הכי מבריקים באוניברסיטה"

קידר, שעסקה בתחום מדעי הרוח הדיגיטליים במסגרת הרוקטורט שלה באוניברסיטת סטנפורד, גייסה לצורך הקמת מרכז רבות חומי ללימוד הכלים האלה מענק הממועצה להשכלה גבוהה.

"זוהינו באוניברסיטה 30 חוקרים במדעי הרוח שעוסקים בפרויקטים שיש להם היטב דיגיטלי", היא מספרת. "הם כרגע ממציאיים כל אחד את המרכז פונה גם לסטודנטים. זה תחו עבוד מחקר של האחר".

המרכז פונה גם לסטודנטים. זה כמה שנים מוצעים לתלמידי התואר הראשון במדעי הרוח קורסי תכנות הרלוונטיים למחקר, אבל בקרוב תחל באוניברסיטה תוכנית חדשה, לסטודנטים לתארים שני ושלישי, שבה יוכלו החוקרים להתמקצע בכלים של מדעי הרוח הדיגיטליים.

"המוטיבציה", אומרת קידר, "היא קודם כול להכשיר חוקרים ולצייד אותם בכלים לביצוע מחקרים חדשים בתחום מדעי הרוח. כדי להיות חוקר ביקורתי בעולם שלנו, חייבים להכיר את הכלים האלה, כי כך יבור צעו לפחות חלק מהמחקרים בעתיד. אני חייבת להכשיר חוקרים שלא ייבהלו מגרפים, מפונקציות. לכן הם לומדים קורס תכנות ומעברת תכנות למדעי הרוח, ושיטות מחקר חישור ביות וקורס מתמטיקה, שאנחנו לא קוראים לו כך, כדי לא להבריח אותם. רק כך הם יוכלו לשמור את האצבע על הרופק של התחום, גם אם לא ירצו לעסוק בסופו של דבר במחקר דיגיטלי בעצמם".

כך תוכלו למושך תלמידים מבריקים קים שאולי חוששים היום להשקיע בתואר מתקדם אבל להישאר בסוף בלי מקצוע מכינים אם לא תימצא להם משרת חוקר בכיר באקדמיה? "אנחנו לא מכוון להכשרת מתמטיים מידע לתעשיית ההייטק, אם זו השאלה. אבל אם כך יותר קל למשוך אותם לתחום, או הרווחנו כולנו". מי שעושים את המעבר הזה בהצ'חה כבר שנים הם הבלשנים. אחוזים לא זניחים מהם, בעיקר אלה ששילבו מלכתחילה את התארים בכלשנות עם תארים בתכנות, מצאו בסופו של דבר עבודה בחברות הייטק שמג'טות ללמד מחשבים לקרוא ולהתבטא בשפה של בני אדם.

אני מודמינת עכשיו דוקטורנט נט בהיסטוריה ופילוסופיה אומר לעצמו, "אלוהים, המתמטיקה רדפה אותי עד כאן". האם אין כבר בכל העולם תחום שבו ניתן להבריק בלי הכישרונים האלה?

"אנשי מדעי הרוח הם הכי מבריקים באוניברסיטה. חוקר שלמד גרמני ניה או יונית עתיקה יודע ללמוד שפות, ותכנות או מתמטיקה כשמלמדים אותם היטב הם פשוט עוד שפה.



צילום: ברונו שרביט, באדיבות אוניברסיטה העברית

תעודת זהות

ד"ר רננה קידר

אישית: בת 44, אמה לשלושה, בעלת תואר ראשון במשפטים ומדעי המדינה מאוניברסיטת תל אביב ודוקטורט בספרות השוואתית מאוניברסיטת סטנפורד

מקצועי: אחרי התואר במשפטים עבדה כפרקליטה כמה שנים במחלקת הבג"צים בפרקליטות המדינה, טרם התקבלה לדוקטורט. היום משמשת ראש המרכז למדעי הרוח הדיגיטליים בפקולטה למדעי הרוח באוניברסיטה העברית, כמינוי משותף של הפקולטות למשפטים ומדעי הרוח

הם לא צריכים להפוך לאלגוריתם מיקאים, אבל בהחלט יכולים ללמוד להשתמש בכלים האלה. היסטוריונים בעבר נלחצו ממעבד תמלילים או מצגת ממוחשבת. הם התגברו. ככל מקרה, מדובר כרגע בתוכנית שפתרחה לשמונה סטודנטים עם מלגות, כך ששום דבר הוא לא חובה".

לזהות את כתב היד על הלוחות העתיקים

בינתיים, אומרת קידר, ההיענות מגיעה דווקא מהחוקרים המבוגרים בפקולטה. "הם מאוד נלהבים לעשות משהו חדש, אחרי שנים שהם חוקרים באותו אופן. זה מרענן מבחינתם, והרעיונות שלהם נהדרים. בנוסף, הם יודעים שאם הם לא יעבירו את האר"כיונים שלהם למאגרים דיגיטליים, יכול להיות שאף אחד לא יתחוק אותם כשהם יפרשו. כדי להצדיק את ההשקעה בריגטיציה של הארכיון, צריך להציג סיבה מחקרית".

כרגע המחקרים משלבים בדרך כלל חוקרים מתחומי המתמטיקה, המחשבים וההנדסה לצד החוקרים מתחומי מדעי הרוח. בעתיד, מקווה קידר, חוקרי מדעי הרוח יוכלו להוביל ולשלוט במחקרים מסוג זה בעצמם, אבל לא על חשבון העבודה הבינתחומית אלא בנוסף לה.

קידר עצמה השתמשה בכלי בינה מלאכותית במחקר שלה על משפט אייכמן (ראו הרחבה במסגרת). לצדו, גויסו למרכז החדש שני פרויקטים נוספים. "פרופ' דני שרירא קיבל תרומה של מאות אלפי, אולי כמה מיליוני גלויות של ארץ ישראל מסוף המאה ה-19 ותחילת המאה ה-20", מספרת קידר, "השאלה היא מה עושים עם אוסף גלויות כזה. הטקסט חשוב, גם התמונה חשובה. הצמר דנו לו עוזר מחקר חישובי ומלווה

עוד זווית

התוכנה שעזרה להבין את התקדים של משפט אייכמן

גה הנאצית, ושימוש בשפה בירוקרטית. אייכמן עצמו השתמש בשפה הזו. ומה בפסק הדין? המחשב הראה שהשופטים השתמשו בשפה הבירוקרטית. הם כן ביקשו משפט צדק, ולא משפט ראווה. בעיניי, בעקבות המחקר נעשה צדק עם משפט אייכמן. השופטים עשו את עבודתם נאמנה".

מה מספר יותר טוב את הסיפור של משפט אייכמן? ענן מילים ורשת מילים, או העדות האחת של קצטניק שנשארה בויתרון לאחר

המחקר של ר"ר רננה קירר עוסק באופן שבו נעשה צדק אחרי מעשים נפשעים רחבי היקף (Mass atrocity). לדוגמה, משפטי נירנברג, ועדות האמת והפיוס בדרום אפריקה ולהבריל – משפט אייכמן, קירר בוחנת לא רק את האירוע המשפטי אלא מגוון מסמכים החריגים שפים את האופן שבו האנושות חשבה על צדק במקרים כאלה, למשל יצירות בריניות כמו הסרט "מזמורים חסרי כבוד" של קוונטין טרנטינו, שמציג נקמה פיקטיבית בנאצים. כך אנתנו יודעים איזה צדק העולם רוצה, אפילו אם הוא לא משיג אותו", היא אומרת.



משפט אייכמן צילום: מילי ג'ון-לע"מ

"משפט אייכמן יכול להיזכר כתקדים בהקשר של משפט המאפסר לעדים להישמע גם לצורך תחושת הצדק האישית שלהם, אבל בלי שהעדויות ימנעו דיון משפטי מהותי"

שהתעלה ברוכז העדים? גם וגם כמוכן, אומרת קירר. "אי אפשר לצלול לתוך מחקר כזה, בלי להבין מה אני רואה. זה התפקיד שלי לפרש את זה. אבל האלגוריתמים מאפשרים לי גם לעבר כמות גדולה של מידע, וגם להשוות את הדעות הקדומות שיש לי. עכשיו המחשב אומר לי – תסתכלי על זה מכיוון אחר. סטטיסטית, דווקא העדויות 78 ו-315 הן הכי מייצגות. וזה מידע שאפשר לקבל תוך כמה דקות."

הדאטה של מלחמת אוקראינה

קירר יוצאת מהמקרה הפרטי של משפט אייכמן לרעיון הכללי שלפיו עדויות במשפט פטים אחרי אירועי וזועה בסדר גודל נרחב משמשות לכמה מטרות במקביל: משפטיות, היסטוריות ופסיכולוגיות.

"לא רחוק היום שבו נראה משפטים דומים באוקראינה, ואני שואלת את עצמי, מה קורה לשופטים אחרי שהם שומעים כל כך הרבה עדויות. זו שאלה רלוונטית לא רק לשופט טים, אלא גם לנו כחברה שמחויבת לזיכרון – איך אנחנו יכולים לתת משמעות ומשקל לכל עדות. ואם לא נשמע אותן בתוך המשפט, כי הן לא נחוצות כדי להוכיח אשמה, אז היכן העולם ישמע אותן?"

"דווקא משפט אייכמן יכול להיזכר כתקדים בהקשר הזה, כמשפט שמאפשר לעדים להישמע גם לצורך תחושת הצדק האישית שלהם, אבל בלי שהעדויות מונעות דיון משפטי מהותי בפסק הדין".

במלחמת רוסיה באוקראינה, היא אומרת, כבר היום מוקדשת מחשבה לעדויות. "זו מלחמה עם המון דאטה. בטוויטר, בפייסבוק, אנשים מספרים את הסיפור שלהם. נפתחה גם תיבת מייל לאיסוף עדויות. כמוכן אין עוד ערך ראיתי. אבל ברמת הביג דאטה יש עוד דפוסים שמתגלים."

"כשהתחלתי את המחקר, היו מוסכמים על חוקרי התחום שני דברים: האחד, שמשפט אייכמן שינה את האופן שבו חושבים על השואה בישראל. לפני כן ניצולים לא דיברו על מה שעברו ואפילו התביישו. למשפט היה תפקיד משמעותי בהבניית זיכרון השואה כמו שאנחנו מכירים אותו היום, על כל מה שנובע מכך, פוליטית וחברתית – הסיפור הציני של משואה לגבורה, שמגיע לנו להיות פה, שארץ ישראל היא הנקמה שלנו."

"אבל היה גם ידוע שמבחינה משפטית אייכמן לא היה תקדים לפסיקות אחרות. אם פט הבינלאומי בהאג, משפט אייכמן שימש שם כתקדים רק לעתים נדירות."

הצדק של משפט אייכמן

קירר שאלה אם גם השופטים עצמם התייחסו למשפט בעיקר כאל אירוע ראווה או שמבחינתם הם ניהלו משפט. "ברור לי שמנקודת המבט של מדינת ישראל, המשפט הזה היה אירוע דירקטי שבו שמענו לראשונה בכל מגוף ראשוני בקולם של הניצולים, מה זו נכללה השנה בכל עדות, ומה האם אפשר לו ריקלימינג כאירוע משפטי?".

לשם כך השתמשה קירר בכלי שנקרא מידול נושאים, שאומר מהם הנושאים העיקריים שעולים בכל עדות, ומה הנושאים העיקריים שעולים במשפט כולו.

היא גילתה שעבור שניצולי השואה דיברו על ההשמדה ועל האנשים שאיבדו, בפסק הדין אין לכך אזור. "היו שם ענני נושאים שלאחר אפשר לקרוא 'השמדה', עם מילים כמו גז, אושוויץ, צריף. זו הייתה השפה שבה השתמשו הניצולים בעדויותיהם. הנושא השני היה בקבלת ההחלטות של אייכמן, עם אוכורים רבים לחברים אחרים בהנהיג

"חוקר שלמד גרמנית או יוונית עתיקה יודע ללמוד שפות, ותכנות או מתמטיקה הם פשוט עוד שפה. היסטוריונים בעבר נלחצו ממעבד תמלילים או מצגת ממוחשבת. הם התגברו"

"מה שאני אוהבת במערכת הממוחשבת היא שלפעמים היא משהה את ההטיות שלנו כבני אדם, ומפתיעה אותנו בהארה של פינות מידע שאנחנו רגילים להזניח"

פס הטיות שהאלגוריתם הממוחשב הכניס לניתוח. הרי בני אדם הם אלה שבונים את האלגוריתם, והוא לומר כיצד לנתח את התוכן החדש על פי ניתוח התוכן הישן, הספוג גם הוא בהטיות.

"מה שאני אוהבת במערכת הממוחשבת היא שלפעמים היא משהה את ההטיות שלנו כבני אדם, ומפתיעה אותנו בהארה של פינות מידע שאנחנו רגילים להזניח", אומרת קירר, אבל, היא אומרת, כדי לוודא שהמערכת נשארת בקשר עם לבין המכונה. דרוש הממשק בין החוקר לסוגיה נוספת נוגעת לתחוק של מאגרי המידע. הרבר דורש תקציב רב, וקיים חשש שזה ירחיב את הפער רים בין חוקרים בעולם, באופן שיע' מיק גם את האי-שוויון בין הקולות המיוצגים באקדמיה. האם היא שוב תייצג אותם קולות שכבר נשמעים זה שנים?

כמוכן, הסיה וחלוקה לא שוויונית של המשאבים קיימות גם במדעי הרוח הקלאסיים. ספריות, תקצירי בי מסע, משרות במסודות הנחשבים – כל אלה אינם מחולקים באופן שווה. אך החשש הוא שתקציבי הענק הדרושים למחשוב מאגרי מידע עלולים להעמיק את הפעיה.

שאלת הקיימות נוגעת לבעיה נוספת – מי יתחוק את מאגר המידע הענק אחרי שהתקציב למחקר ספציפי נגמר? השאלה הזאת רלוונטית גם לארכיונים פיזיים, אולם התקווה היא שהארכיון הדיגיטלי יוכל לקבל כל השנת ובלקלות תוספת של חומרים, עדכונים של החומרים הקיימים ושאלות מחוקרים נוספים, ומישהו צריך לממן זאת. פתרונות אפשריים, מעבר למי מון המחקר המוכר באמצעות מענקים תרומות, הם מימון של המדינה לארכיונים המשמעותיים בעיניה, תמיכה של חברות עם שירותי מחשוב, או תחזוקה של ארכיון על ידי הקהילה שהמידע הרלוונטי בארכיון משמעותי לה.

לכל אחת מהאפשרויות הללו יש השלכות אתיות שיצטרכו להתמודד איתן בעתיד.

מחקר מהמכון, שהחלו לבנות מאגר של הגלויות ולחשוב איך לנתח אותן. האם באמצעות מערכת עיבוד ראייה שתזוהה מקומות או אלמנטים בתמונה (לדוגמה, חמור, איכר, מפעל?)".

הפרויקט השני הוא של פרופ' מיכל מרמורשטיין, שראיינו בעבר על המחקר שלה שבו ניתחה 170 אלף הודעות וואטסאפ. במחקר חדש, היא מעוניינת לנתח שיחות טלפון. וזה כבר עולם אחר מבחינת השקעת הזמן. אם ניתן להשתמש במערכת ממוחשבת שתעזור לה להצביע על אלמנטים מעניינים במיוחד להקשיבה, הזמן יכול להתקצר משמעותית.

דוגמה נוספת למחקר היסטוריה דיגיטלי הוא זה שערך המתמטיקאי ר"ר ברק סובר, הוא חוקר את רמת האוריינות בתקופת המקרא. איך זה מסתדר? סובר השתמש בכלים דיגיטליים כדי לשפר את היכולת לקרוא ולחזות עתיקים, ואחר כך פיתח אלגוריתם שמנתח את תמונת כתב היד, כדי לזהות אם שני לוחות נכתבו על ידי אותו כותב או על ידי כותבים שונים. ככל שמוזהים יותר כותבים, המסקנה היא שהכתיבה והקריאה היו נפוצות יותר. ממציא המחקר אכן מעדיף שרדיעת קרוא וכתוב הייתה נפוצה בתקופה המקראית יותר מכפי שנהוג לחשוב, לפחות בשדרת הארץ מיניסטריצה הצבאית ובקרבת הקצונה הוותיקה בממלכת יהודה. כימים אלה הוא עוסק במיפוי ההתפתחות של האוריינות בימי בית ראשון ושני.

קירר מספרת על מחקר נוסף שהיא עובדת ובו היא מבקשת ממערכת ממוחשבת לנתח עדויות של ניצור לי שואה, כדי לגלות עדויות שניאה שיש בהן השמטות, כלומר, תקופות מסימות או היבטים מסימים שהעדים נמנעים מלדבר עליהם. "ברגע שארע איך נראית עדות שואה, אוכל לזהות את העדויות שחורגות מהנרטיב, וזה מעניין. אולי נגלה, למשל, שמתתרת שם תקיפה מינית. כמעט אין עדויות על תקיפות מיניות בשואה, ואנחנו מניחים שהיו כאלה, כי באירועי זועה המינית כמעט תמיד יש, ואולי כך נבין היכן זה מסתתר".

קירר מפליגה על כנפי הרמיון כשהיא מדברת על האפשרויות המחקריות בעתיד. "אנחנו יכולים להרחיב במכונה את כל הספרות העברית, ולראות מה הנושאים שעולים לדיון בזמנים שונים, או על ידי סופרים מסר גים שונים, או איך נבנות בדרך כלל עלילות בתקופה מסוימת. מחקר כזה יכול לכלול גם ספרים שהם לא קאנון ואולי לא רבים קראו. נוכל גם לראות מה באמת כתוב בספרים שהעבר לעור מת מה שנדרמה לנו שאנחנו זוכרים מהם. במקרה כזה, דווקא לא אבקש מהאלגוריתם לתת משקל רב יותר לספרים החשובים, כי אותם חקרנו ממילא בכלים הקלאסיים".

שאלות על תקציב ועל אתיקה

עם התפתחות התחום עלו וצפו גם שאלות אתיות וסביבתיות. אחת מהן היא השאלה אם ריחוק של החוקר מהטקסטים עלול לגרום לו לפסי