



המכון
לחירות ואחריות

בית ספר לאודר
לממשל, דיפלומטיה
ואסטרטגיה



פיתוח של בינה מלאכותית לזיהוי מקדים של אלימות בשיח פוליטי במרחב הווירטואלי

נייר עמדה לקראת שולחן עגול שיתקיים ב 30/4/2023

ד"ר אורלי אידן, ד"ר יעל בן דוד ורפאלה גויכמן

מטרתו של השולחן העגול היא לדון בהשפעות שיש לשפה ולרטוריקה על תהליכי הסלמה ודה-אסקלציה בשיח פוליטי ברשתות. במסגרת הדיון יוצג כלי תיאורטי המזהה סימנים מוקדמים לאלימות בשיח פוליטי, ומשמעותיו לפיתוח מודל אלגוריתמי מבוסס בינה מלאכותית אשר יוכל לזהות ואף להתערב בתהליכי הסלמה בקבוצות און ליין. המודל פותח מתוך עבודה מחקרית שנעשתה על ידינו בשנים האחרונות בה תיקפנו אינדיקטורים לשיח מסלים, וראינו כיצד התערבות בתהליכי הסלמה בשיח יכולה להועיל להפחתת קיטוב וקידום שיח בקרב משתתפים שאינם מסכימים זה עם זה.

במפגש נציג את המחקר, ונדון באפשרויות הפיתוח של כלים העשויים לאפשר למנהלי פורומים וקבוצות און ליין זיהוי והתערבות מוקדמת בשיח, כמו גם תיעול השיח אל עבר מקומות קונסטרוקטיביים יותר ובכך למנוע את עזיבתם של משתתפים מדיונים אלימים מידי, הוצאה של משתתפים או סגירה מוחלטת של קבוצות שיח. המסמך הנוכחי מהווה בסיס לדיון שתוצריו יהיה ניסוח תובנות והמלצות לפעולה להפחתת שיח מסלים ברשתות באמצעות שימוש בבינה מלאכותית מבוססת ניתוח לינגוויסטי של תהליכי הסלמה.

תהליכים של הסלמה בשיח פוליטי : היבטים פסיכו וסוציו-לינגוויסטיים

השיח בחברה הישראלית מצוי במצב של קיטוב מתמיד, המושפע מבעיות עומק הנוגעות לכיבוש המתמשך, ליחסי דת ומדינה, לאופייה החברתי המרובד של החברה הישראלית ועוד¹. בשנים האחרונות ניתן לזהות בישראל התגברות של תהליכים אלה, כחלק ממגמה עולמית רחבה יותר הקשורה בין היתר להיחלשות גוברת והולכת של הסכמה על סמכויות מקצועיות של בעלי הידע כגון גופי התקשורת, הפוליטיקאים, וברי סמכא בתחומים שונים². התערערות זו של הסמכות האפיסטמית נובעת מכמה סיבות ביניהן: מקורות ידע אלטרנטיביים ו'שטוחים' במרחב האינטרנטי³; ירידה בתקשורת פוליטית בינאישית בלתי אמצעית⁴; עלייתם של מנהיגים פופוליסטיים, ושילוב של כל התהליכים הללו יחד. תהליכים אלה מביאים לבולטות יתר לרטוריקה ולאופן השימוש במילים, שכן הרטוריקה בה נאמרים הדברים קובעת באיזו מידה המסר נתפס כאמיתי ומהימן⁵. לדוגמא, במחקר שהשווה שימוש במילים מופשטות לעומת מילים קונקרטיות בקרב פוליטיקאים, נמצא שהמילים המופשטות הגבירו את תפיסת הסמכות והמהימנות של אותם פוליטיקאים⁶.

הקיטוב בשיח הפוליטי ניכר גם באופן שבו מתנהל שיח פוליטי במרחב הוירטואלי המביא להקצנה של השיח, בין היתר בשל רמת מחויבות בינאישית ירודה⁷ ואי הבנות ופערים בתקשורת⁸. לאור זאת, עולה הצורך להבין לעומק את המופעים השפתיים של תהליכי הסלמה, ומתוך כך לנסות לזהות פרקטיקות אשר

ברטל, ד. ושל, י. (2013). הכובש והנכבש: המקרה הישראלי. בתוך ד. בר-טל, וי. שנל (עורכים), השפעת הכיבוש על החברה הישראלית – מבוא (ע'ע 22-33). מוסד ביאליק.

² Pew Research Center (2019). *Trusting the news media in the Trump era*. Retrieved from www.pewresearch.org.

³ Fandos, N., & Roose, K. (2018). *Facebook identifies an active political influence campaign using fake accounts*. Retrieved from the *New York Times* <https://www.nytimes.com/2018/07/31/us/politics/facebook-politicalcampaign>

⁴ John, N.A., & Dvir-Gvirzman, S. (2015). "I Don't Like You Any More": Facebook Unfriending by Israelis During the Israel-Gaza Conflict of 2014. *Journal of Communication*, 65(6), 953-974.

⁵ Cichocka, A., Bilewicz, M., Jost, J. T., Marrouch, N., & Witkowska, M. (2016). On the grammar of politics—or why conservatives prefer nouns. *Political Psychology*, 37, 799-815.

Chilton, P. (2004). *Analyzing political discourse*. Routledge.

⁶ Wakslak, C. J., Smith, P. K., & Han, A. (2014). Using abstract language signals power. *Journal of Personality and Social Psychology*, 107, 41-55.

⁷ Weinberg, H., 2018. *The paradox of Internet groups: Alone in the presence of virtual others*. Routledge.

⁸ Parks, C. D. (2020). Group dynamics when battling a pandemic. *Group dynamics: Theory, research, and practice*, 24(3), 115.

יאפשרו התמודדות עם תהליכים אלה באופן שיביא להפחתת הקיטוב בין יריבים פוליטיים, ואף לקידום שיח והתמרה של קונפליקט בסוגיות הנדמות בלתי פתירות⁹.

במחקרים שלנו חקרנו את הדינאמיקה של שיח מסלים בקבוצות קטנות שקיימו שיח פוליטי פנים אל פנים. זיהינו מספר עקרונות המתכתבים עם הספרות הפסיכו-חברתית של מצבי קונפליקט והסלמה ומציעים הבנה של התהליכים השפתיים המביאים לכך ששיח מסלים ומקצין. בנוסף על כך, אנחנו מצביעות גם על פרקטיקות המסייעות לשבור תהליך זה של הסלמה בשיח.

בהיבט הקוגניטיבי, ניתן לזהות כי בסיטואציות של קונפליקט ואיום, פרטים נוטים להיות פחות סבלניים למצבי עמימות ואי וודאות והם תרים אחר תשובות ברורות ואמיתות אשר יוצרות וודאות ומפחיתות את החרדה הכרוכה במצבי קונפליקט¹⁰. בהקשר של שיח פוליטי, החוקרת ספייק פיטרסון¹¹ מצביעה על הנטייה של שיח זה להתבסס על הנחות יסוד גבריות, אשר יוצרות הפרדה בין הפרטי לציבורי. בהקשר זה, שיח סובייקטיבי המבוסס על תחושות ורגשות נתפס כלא לגיטימי בתקשורת בינאישית, ומביא את הדוברים לבסס את דבריהם על רטוריקה של עובדות. בהיבט השיחני, ניתן לזהות הסלמה במעבר של השיח מדיבור סובייקטיבי, בו הדוברים מוכנים להכיר בקיומם של נראטיבים שונים משלהם תוך הבנה שהשיח שלהם עצמם הינו סובייקטיבי ולא מייצג אמת מוחלטת, אל עבר שיח של אמיתות מוחלטות, הכולל הדיפה של כל נראטיב שלא מתיישב עם 'אמת' של הדובר. פרקטיקות אלה של שיח מביאות את הדוברים לעסוק בהגנה על הטעוניהם שלהם ולא בהקשבה לזולת ועל כן גורמות לדיון להסלים¹².

⁹ Galtung, J. (1996). *Peace by peaceful means: Peace and conflict, development and civilization*. International Peace Research Institute Oslo: Sage Publications, Inc.

¹⁰ Bar-Tal, D., Soli, V., & Fuxman, S. (2020). Between open-minded critical thinking and closed-minded allegiance: Educational tensions in societies involved in intractable conflict. *Political Psychology*, 42(10), 126-137.

¹¹ Peterson, S.V. (1999). Political identities/nationalism as heterosexism. *International Feminist Journal of Politics*, 1(1), 34-65.

¹² Ben David, Y., & Idan, O. (2019). "We don't have to talk about how I feel": Emotionality as a tool of resistance in political discourse among Israeli students—a gendered socio-linguistic perspective. *International Feminist Journal of Politics*, 21(2), 271-294.

Ben David, Y., & Idan, O. (2021). Eyes wide shut: Political ideology as a tool of discursive avoidance among Israeli-Jewish students in the context of escalating conflict. *Political Psychology*, 42(3), 371-387.

בהיבט הרגשי, קבוצות בקונפליקט נוטות לתפיסה מקוטבת של העצמי למול ה'אחר', כך שביסוס ותיקוף העצמי נעשה על ידי שלילה והאשמה של ה'אחר'¹³. במצבי הסלמה, גם מחוות חיוביות עשויות להתפרש כמהלך מניפולטיבי. בעבודה שלנו מצאנו כי דינאמיקה זו באה לידי ביטוי במעבר משיח המאופיין בהתבוננות עצמית ולקייחת אחריות של הדובר על רגשותיו, לדינמיקה בה הרגשות השליליים (כגון אשמה, בושה ופחד) מושלכים על האחר דרך האשמות ודיבור שלילי.

בהיבט הרטורי, מצבי קונפליקט מעצימים את הנטייה של פרטים לחשוב ולפעול באופן קולקטיבי¹⁴. במילים אחרות, פרטים נוטים לזהות את עצמם יותר עם קבוצות שהם משתייכים אליהן ולחשוב ולדבר כחברי קבוצה. יתרה מכך, מצבי קונפליקט מקדמים חשיבה סטריאוטיפית על האחר, וראייה מונוליטית של העצמי והאחר¹⁵. הנטייה לדבר במונחים מוכללים וסטריאוטיפים באה לידי ביטוי בשיח במעבר מדיבור ביחיד לדיבור ברבים. במחקרים שלנו הראינו כיצד דיבור קולקטיבי במונחים כגון "אנחנו" "הם", ו-"אתם", ובעיקר המעבר מדיבור בלשון יחיד לדיבור בלשון רבים תוך כדי דברים של דובר ספציפי, מהווים אינדיקטורים לתהליכי הסלמה בשיח פוליטי קבוצתי.

לסיכום, ניתן להצביע על שלושה מאפיינים לשיח מקטב, אותם ניתן לזהות כסימנים מקדימים להידרדרות של שיח לאלימות:

1. שיח של אמיתות מוחלטות
2. שיח מאשים שאינו לוקח אחריות על רגשות שליליים אלא מתיק אותם להאשמת האחר
3. שימוש ברטוריקה קולקטיביסטית ומכלילה בהתייחסות לעצמי או לאחרים

¹³ Tajfel, H., & Turner, J. C. (2004). The social identity theory of intergroup behavior. In J. T. Jost & J. Sidanius (Eds.), *Political psychology: Key readings* (pp. 276–293). Psychology Press.

¹⁴ Bar-Tal, D., & Antebi, D. (1992). Siege mentality in Israel. *International Journal of Intercultural Relations*, 16(3), 251-275.

¹⁵ Kelman, H. C. (1999). The Interdependence of Israeli and Palestinian. *Journal of Social Issues*, 55(3), 581-600.

הדיבור המסלים מעורר עניין ותגובות, אבל לרוב אלה יהיו תגובות אשר יקצינו את הדיון ויתרמו לקיטוב ולא להפחתתו. זהו דיבור שמחזק מחנאות ולא באמת פותח פתח לדיאלוג ולחשיבה שיש בה גם למידה אלא חיזוק והתבצרות של כל מחנה בנראטיב שלו, תוך הקפאה ושימור של הנראטיבים.

תהליכי הסלמה ברשתות החברתיות : מבט עדכני

הרבה לפני סבבי הבחירות שהחלו ב-2019, הרשתות החברתיות כבר היו זירת התגוששות שבה אנשים שמעולם לא נפגשו, הרשו לעצמם להתבטא בצורה בה לא היו מדברים לו היו נפגשים במרחב הפיזי. בשונה מדיונים בנושאים כמו 'איזו קבוצת כדורגל יותר טובה' או 'האם את או אתה אנשים של כלבים או חתולים?' – כשזה מגיע לסוגיות בתחום הפוליטי-חברתי, השיח מקצין מיד משני הצדדים הפוליטיים. במהלך מבצע שומר חומות ניתן היה להבחין לראשונה בקריאות לאלימות פיזית – בעיקר באפליקציות המסרים כמו וואטסאפ וטלגרם. לאחר עליית ממשלת השינוי ברשות נפתלי בנט ויאיר לפיד, ניתן היה להבחין בעליית מדרגה נוספת באמצעות שיח שעשה דה-לגיטימציה לממשלה שכיהנה.

בשלושת החודשים האחרונים נרשם שיא מדאיג בשיח ברשת בעקבות החקיקה האנטי-דמוקרטית שמובילה הקואליציה ברשות בנימין נתניהו. ההסלמה נובעת מהיקף החקיקה הדורסנית שמקדמת הלכה למעשה הפיכה משטרית במדינת ישראל. התוצאה היא שיח היפר מוקצן ורווי שנאה וקיטוב חברתי שכמוהו לא נראה בישראל עד כה. יש לכך כמה סיבות: הראשונה היא שהאלגוריתם מעצימים תוכן אלים וחדשות כזב כי הם זוכים למעורבות גבוהה יותר של המשתמשים. עבור הרשתות, שמתפרנסות ממכירת שטחי פרסום, ככל שאנשים מבלים יותר זמן בפלטפורמה הם נחשפים לעוד מודעות וכך ההכנסות שלהן עולות, ולכן אין תמריץ להנמיך את הלהבות. סיבה שנייה היא שהרשתות לא כפופות לרגולציה ולכן כשמופץ בהן תוכן שקרי, מידע לא מבוסס, הסתה לאלימות או גזענות – הן לא מחויבות בהסרתן והן מתנהלות בשיטה של רגולציה עצמית. עבור הפוליטיקאים, הרשתות הן פלטפורמה חלומית – בשונה ממדיות מסורתיות אין שם שאלות קשות של עיתונאים, אם מישהו מציג דעה שונה הם יכולים לחסום אותו והם מדברים ישירות לבייס שלהם – כלומר רווח נקי. התוצאה היא שאחרי חמישה סבבי בחירות בתוך שלוש שנים, שכוללים את הסגרים הארוכים

של מגיפת הקורונה שבה רוב תשומת הלב הופנתה לרשתות - המרחב הדיגיטלי הפך כמעט ברובו מייצג של החברה הישראלית - זירת מלחמה של שמאל נגד ימין, יהודים נגד ערבים, מזרחים נגד אשכנזים, חילונים נגד דתיים וגם אזרחים נגד מוסדות המדינה.

כיצד ניתן לפעול על מנת להפחית הסלמה בשיח הפוליטי?

באופן מעניין, גילינו כי אותם עקרונות המדרדרים שיח לאלימות, מקדמים באופן הפוך תהליכים של דה-אסקלציה. ריבוי קולות ודעות, הזמנה לשתף בסיפורים אישיים ובעמדה רגשית אשר נמצאת בבסיס העמדה אותה מציג האדם, יכולים לשבור דינאמיקות שיח מתנצחות בין שני משתתפים, ולהוציא את השיח מדיון באמיתות מוחלטות של שחור ולבן, ולהגביר סיכויים להקשבה הדדית. דיבור מתקף רגשית, המכיר גם ברגשותיו של האחר בשיחה תוך לקיחת אחריות על הרגשות האישיים, יכול להביא להשתנות של השיח ולמעבר משיח של מתקפה/מגננה אל עבר שיח מתבונן הנותן תוקף למגוון של רגשות שנושא קונפליקטואלי זה או אחר מעלים. לבסוף, דיבור אישי, המדבר בגוף ראשון יחיד, או פונה לאחר בגוף שני יחיד, יכול להביא להתייחסויות אישיות יותר בקרב משתתפים נוספים ולהשתנות של שיח מכליל וסטריאוטיפי.

בשולחן העגול המוצע להלן, נבקש להעלות לדיון את השאלות הבאות, בשני חלקים:

חלק א': דיון קונספטואלי

1. האם ניתן לקדם שיח פוליטי מבלי להיגרר להסלמה?
2. מהי האחריות וכיצד היא צריכה להתחלק בין גורמים שונים (הרגולטור, התקשורת, הפוליטיקאים, הציבור, האקדמיה)?
3. כיצד ניתן לעשות שימוש מושכל באחריות ברשת כדי לקדם שיח פוליטי בונה ולא מקטב?

חלק ב': דיון אופרטיבי

1. האם ניתן לפתח אלגוריתם שיכול לסייע בניטור או במניעה של הסלמה?
2. מהן המשמעויות שיש לפיתוחים של בינה מלאכותית כמו היוזמה הנוכחית המוצעת על המרחב הפוליטי הדיגיטלי, ובהשלכות האתיות של שימוש במודלים מסוג זה?