הערות

1 Abdulkader Tayob, Sermons as Practical and Linguistic Performances: Insights from Theory and History," Journal of Religion in Africa 47(1) (2017), pp. 132-144

2 Constantine Boussalis, Travis G. Coan, and Mirya R. Holman, Political Speech in Religious Sermons," Politics and Religion 14(2) (2021), pp. 241-268

3 James L. Guth, The Bully Pulpit: The Politics of Protestant Clergy, University Press of Kansas, 1997; Paul A. Djupe and Christopher P. Gilbert, The Political Voice of Clergy," *Journal of Politics* 64(2) (2002), pp. 596-609; Melissa Deckman, Sue E.S. Crawford, and Laura R. Olson, The Politics of Gay Rights and the Gender Gap: A Perspective on the Clergy," *Politics and Religion* 1(3) (2008), pp. 384-410; Rebecca A. Glazier, Bridging Religion and Politics: The Impact of Providential Religious Beliefs on Political Activity," *Politics and Religion* 8(3) (2015), pp. 458-87

4 Boussalis, Coan, and Holman (לעיל הערה 2).

5 הבדל משמעותי בין המחקר הזה למחקרים קודמים הוא הגישה המתודולוגית. העבודות המוקדמות יותר נשענו על טכניקות טיפוסיות של למידת מכונה, כגון LDA (Latent Dirichlet Allocation), שמנתחת את הדרשות על ידי זיהוי דפוסים סטטיסטיים של מופעים חוזרים של צירופי מילים. שיטות אלה יעילות לאיתור מגמות נרחבות, אך מטבען הן מוגבלות מבחינת היכולת לתפוס את המשמעות הסמנטית העמוקה או את הדקויות המשובצות ברטוריקה הדתית. כאשר שיטה כמו LDA מתייחסת לטקסטים כאל אוסף של מילים ותו לא ואינה מביאה בחשבון את היחסים ההקשריים, היא מתקשה לתפוס את המורכבויות של משמעות הדברים, את נימת הקול ואת המסרים המרומזים המאפיינים דרשות. בניגוד לכך, המחקר שלנו משתמש במודלי שפה גדולים (Large Language Models - LLMs), וספציפית ב־ChatGPT-4o, שמציעים דרך מתוחכמת יותר לניתוח טקסטואלי. מודלים אלה מבינים את הטקסט בתוך ההקשר הנתון ומאפשרים פרשנות עשירה יותר של תמות, עמדות ואסטרטגיות רטוריות. השימוש בהם מאפשר התבוננות מעמיקה יותר מזו של שכיחות ההופעה של מילים בלבד וחושף דפוסים מורכבים יותר של השיח הפוליטי בדרשות.

6 למחקרים קודמים על ההיסטוריה של דרשות יהודיות ראו Robert V. Friedenberg, *"Hear O Israel": The History of American Jewish Preaching, 1654-1970*, University of Alabama Press, 1989; Marc Saperstein, *Jewish Preaching in Times of War, 1800-2001*, Liverpool University Press, 2012

7 לתיאור מלא של איסוף הנתונים, שיטות העבודה והליכי השיקוף ראו נספח 1: "מתודולוגיה".

8 מודל ChatGPT הראה, על סמך ניתוח מדגמי של 20 דרשות, זיהוי נכון של תוכן פוליטי בדרשות בשיעור של 92.5%.

9 בניתוח הושג דיוק בשיעור של 90% במדגם של 10 דרשות. כמו כן הושג שיעור ריקול (recall) של 100% במדגם של 10 דרשות נוספות.

10 לצורך הניתוח קראו קוראים אנושיים מדגם של דרשות. הם קבעו כי מטבע הדברים קשה לכמת במדויק את השיעור היחסי של תוכן פוליטי בתוך דרשה מסוימת. עם זאת הם העריכו כי הדרשות שסווגו כפוליטיות מכילות נפח ניכר של שיח פוליטי בכל שלושת הזרמים. ממצא זה תואם את התוצאות שהופקו ב־ChatGPT, והדבר מאשש את המסקנה שנושאים פוליטיים הם רכיב חשוב בדרשות המזוהות כפוליטיות באופיין. הניתוח הן של נפח התוכן הפוליטי והן של מבנה הדרשות התבצע בכל 4,302 הדרשות, אגב התמקדות מיוחדת בדרשות שסווגו כפוליטיות, ללא התייחסות לפרק זמן ספציפי. בניתוחים הבאים ייבדקו רק דרשות משלושת פרקי הזמן שהוגדרו לעיל.

11 התוצאה הייתה 100% דיוק במדגם מקרי של 20 דרשות.

12 על סמך מדגם של 20 דרשות הושג דיוק בשיעור של 92.5%.

13 בניתוח הושגה, על סמך מדגם של 15 דרשות, רמת דיוק של 93.33% בזיהוי אם הדרשה מכילה ביקורת. פירוש הדבר שמתוך 15 דרשות שסווגו כמכילות ביקורת, 14 זוהו נכונה, כלומר רמת דיוק גבוהה באיתור שיח ביקורתי במאגר הנתונים. בנוסף, לצורך הערכת הריקול בחרנו מדגם נפרד של 15 דרשות שבו כבר ידענו אילו מן הדרשות מכילות ביקורת מפורשת. בדקנו אם ChatGPT עלול להחמיץ חלק מהן, ושיעור הריקול שהושג היה 100%. כלומר, הכלי זיהה נכונה את כל הדרשות במדגם זה שכללו ביקורת מפורשת.

14 לצורך הערכת יכולתו של המודל לשלוף את כל המקרים הרלוונטיים מתוך מאגר הנתונים השתמשנו במדד הריקול. ניתחנו שני מדגמים: אחד של 15 דרשות והשני של 20, ואת שניהם בחרנו מתוך מאגר הדרשות שכבר זוהו כמכילות תוכן ביקורתי. היעד העיקרי היה להעריך עד כמה

ChatGPT מזהה ומנתח במדויק ובאופן מקיף גוני נימה ביקורתית במגוון נושאים. באשר לביקורת כללית – בדקנו את הריקול במדגם של 15 הדרשות שהתמקדו בנושאים כגון אלימות מתנחלים, בנימין נתניהו וסוגיות קשורות. מדגם 20 הדרשות, לעומת זאת, שימש אותנו להערכת הריקול בנוגע לביקורת המופנית ישירות כלפי הממשלה, כלפי הרפורמה המשפטית וכלפי הקהילה החרדית.

15 בניתוח הושגו דיוק בשיעור של 100% וריקול של 100%.

16 בניתוח הושגו דיוק בשיעור של 100% וריקול של 100%.

17 בניתוח הושגו דיוק בשיעור של 100% וריקול של 100%.

18 בניתוח הושגו דיוק בשיעור של 100% וריקול של 100%.

19 בניתוח הושגו דיוק בשיעור של 100% וריקול של 100%.

20 בניתוח הושגו דיוק בשיעור של 100% וריקול של 100%.

21 בניתוח הושגו דיוק בשיעור של 100% וריקול של 100%.

22 בניתוח הושגו דיוק בשיעור של 83% וריקול של 100%.

23 בניתוח הושגו דיוק בשיעור של 100% וריקול של 100%.

24 לצורך כל ניתוחי הנושאים דגמנו 10 דרשות והשגנו דיוק בשיעור של 100%. כלומר, כל דרשה שזוהתה כמכילה נושא מסוים סווגה כְּראוי.

25 במטלת סיווג הנושאים הגענו לרמת דיוק של 100%.

26 בניתוח המבוסס על מדגם של 20 דרשות הושגו דיוק בשיעור של 85% וריקול של 100%.

27 על Few-shot learning ראו Chanathip Pornprasit and Chakkrit Tantithamthavorn, "GPT-3.5 for Code Review Automation: How Do Few-Shot Learning, Prompt Design, and Model Fine-Tuning Impact Their Performance?," *arXiv Preprint arXiv:2402.00905*, 2024; Zhengfei Ren, Annalina Caputo, and Gareth Jones, "A Few-Shot Learning Approach for Lexical Semantic Change Detection Using GPT-4," Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change, Association for Computational Linguistics , 2024 pp. 187-192

28 ראו Jiuhai Chen, Lichang Chen, Heng Huang, and Tianyi Zhou, "When Do You Need Chain-of-Thought Prompting for Chatgpt?," *arXiv Preprint arXiv:2304.03262*, 2023; Zhipeng Chen, Kun Zhou, Beichen Zhang, Zheng Gong, Xin Zhao, and Ji-Rong Wen, "Chatcot: Tool Augmented Chain-of-Thought Reasoning on Chat-Based Large Language Models," *arXiv Preprint arXiv:2305.**14323***, 2023

29 ראו Michael J. Mior, "Large Language Models for JSON Schema Discovery," *arXiv Preprint arXiv:2407.03286*, 2024

30 על הפרשנות של מידול נושאי ראו Emil Rijcken, Floortje Scheepers, Kalliopl Zervanou, Marco Spruit, Pablo Mosteiro Romero, and Uzay Kaymak, "Toward Interpreting Topic Models with ChatGPT," The 20th World Congress of the International Fuzzy Systems Association, Korea, 2023

31 על גישת HITL ראו Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He, "A Survey of Human in-the-Loop for Machine Learning," *Future Generation Computer Systems* 135 (2022), pp. 364–381; Carl Orge Retzlaff, Srijita Das, Christabel Wayllace, Payam Mousavi, Mohammad Afshari, Tianpei Yang, Anna Saranti, Alessa Angerschmid, Matthew E. Taylor, and Andreas Holzinger, "Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities," *Journal of Artificial Intelligence Research* 79 (2024), pp. 359–415; Padma Iyenghar, "Clever Hans in the Loop? A Critical Examination of ChatGPT in a Human-in-the-Loop Framework for Machinery Functional Safety Risk Analysis," *Eng* 6(2) (2025), p. 31