Endnotes

1 Abdulkader Tayob, “Sermons as Practical and Linguistic Performances: Insights from Theory and History,” *Journal of Religion in Africa* 47, no. 1 (2017): 132–44.

2 Constantine Boussalis, Travis G Coan, and Mirya R Holman, “Political Speech in Religious Sermons,” *Politics and Religion* 14, no. 2 (2021): 141.

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, no. 2 (2002): 596–609; Melissa Deckman, Sue ES Crawford, and Laura R Olson, “The Politics of Gay Rights and the Gender Gap: A Perspective on the Clergy,” *Politics and Religion* 1, no. 3 (2008): 384–410; Rebecca A Glazier, “Bridging Religion and Politics: The Impact of Providential Religious Beliefs on Political Activity,” *Politics and Religion* 8, no. 3 (2015): 458–87.

4 Boussalis, Coan, and Holman, “Political Speech in Religious Sermons.”

5 A key distinction between this study and previous research lies in the methodological approach. Earlier studies relied on classical machine learning techniques, such as Latent Dirichlet Allocation (LDA), which analyze sermons by identifying statistical patterns in word co-occurrence. While these methods are effective for detecting broad trends, they are inherently limited in their ability to capture deeper semantic meaning or the nuanced ways in which political discourse is embedded within religious rhetoric. By treating texts as mere collections of words without considering their contextual relationships, LDA struggles to grasp the complexities of meaning, tone, and implicit messaging that characterize sermons. In contrast, our study employs large language models (LLMs), specifically ChatGPT-4o, which offer a more sophisticated means of textual analysis. LLMs understand text contextually, allowing for a richer interpretation of themes, sentiment, and rhetorical strategies. This enables us to move beyond simple keyword frequency and uncover more complex patterns of political discourse within sermons.

6 For previous studies on the history of Jewish sermons consider: Robert V Friedenberg, “Hear O Israel’: The History of American Jewish Preaching, 1654-1970,” (No Title), 1989; Marc Saperstein, *Jewish Preaching in Times of War,* 1800-2001 (Liverpool University Press, 2012).

7 A full description of the data collection, computational methods, and validation procedures can be found in the Appendix: Methodology.

8 Based on sample analyses of 20 sermons, the ChatGPT model demonstrated an accuracy rate of approximately 92.5% in correctly identifying political content within sermons.

9 The analysis achieved a precision rate of 90% on a sample of 10 sermons. Additionally, it attained a recall rate of 100% based on another sample of 10 sermons.

10 The analysis conducted by human readers of a sample of sermons confirmed that it is inherently challenging to precisely quantify the proportion of political content within a sermon. However, their evaluation indicated that sermons classified as political contained a substantial volume of political discourse across all three denominations. This finding aligns with the results generated by ChatGPT, further reinforcing the conclusion that political themes constitute a significant component of sermons identified as politically oriented. The analysis of both the volume of political content and the structure of sermons was conducted across all 4302 sermons, specifically focusing on those categorized as political, regardless of the specific time period. In subsequent analyses, only sermons from the three main periods outlined above will be considered.

11 The result was of 100% accuracy on a random sample of 20 sermons.

12 Based on a sample of 20 sermons, an accuracy rate of 92.5% was achieved.

13 The analysis achieved a precision rate of 93.33% in identifying whether a sermon included criticism, based on a sample of 15 sermons. This means that out of 15 sermons classified as containing criticism, 14 were correctly identified, demonstrating a high level of accuracy in detecting critical discourse within the dataset. Additionally, to evaluate recall, we selected a separate sample of 15 sermons where we already knew which ones contained explicit criticism. We tested whether ChatGPT would miss any of these, and it achieved a recall rate of 100%, correctly identifying all sermons with explicit criticism in this sample

14 To assess the model's ability to retrieve all relevant cases within the dataset, the **recall metric** was employed. Two samples were analyzed: one consisting of 15 sermons and another of 20 sermons, both selected from a dataset previously identified as containing critical content. The primary objective was to evaluate the extent to which ChatGPT accurately and comprehensively identified and analyzed nuanced criticism across various topics. For general criticism, recall was assessed using the 15-sermon sample, which focused on themes such as settler violence, Benjamin Netanyahu, and related issues. The 20-sermon sample, in contrast, was used to evaluate recall concerning criticism directed at the government, judicial reform, and the Haredi community.

15 The analysis achieved a precision rate of 100% and Recall of 100%.

16 The analysis achieved a precision rate of 100% and Recall of 100%.

17 The analysis achieved a precision rate of 100% and Recall of 100%

18 The analysis achieved a precision rate of 86%.

19 The analysis achieved a precision rate of 100% and Recall 100%.

20 The analysis achieved a precision rate of 100%. And Recall 100%.

21 The analysis achieved a precision rate of 100% and Recall 100%.

22 The analysis achieved a precision rate of 83% and Recall 100%.

23 The analysis achieved a precision rate of 100% and Recall 100%.

24 For all topic analyses, we took a sample of 10 sermons and achieved an accuracy rate of 100%, meaning that every sermon identified as containing a certain topic was correctly classified.

25 For the task of topic classification we got accuracy of 100%.

26 The analysis based on a sample of 20 sermons achieved a precision rate of 85% and Recall of 100%.

27 On Few-shot learning consider: 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," 2024, 187–92.

28 See:Jiuhai Chen et al., "When Do You Need Chain-of-Thought Prompting for Chatgpt?," arXiv Preprint arXiv:2304.03262, 2023; Zhipeng Chen et al., "Chatcot: Tool-Augmented Chain-of-Thought Reasoning on Chat-Based Large Language Models," *arXiv Preprint arXiv:2305*.14323, 2023.

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

30 On interpreting topic modeling consider: Emil Rijcken et al., "Toward Interpreting Topic Models with ChatGPT," 2023.

31 On the Human in the Loop approach consider: 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, no. 2 (2025): 31; Carl Orge Retzlaff et al., “Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities,” *Journal of Artificial Intelligence Research* 79 (2024): 359–415; Xingjiao Wu et al., “A Survey of Human-in-the-Loop for Machine Learning,” *Future Generation Computer Systems* 135 (2022): 364–81.\\uc0\\u8221{} {\\i{}Eng} 6, no. 2 (2025