# Computational Analysis of Character Development
# in Holocaust Testimonies

**Anonymous ACL submission**

## Abstract

This work presents a computational approach to analyze character development along the narrative timeline. The analysis characterizes the inner and outer changes the protagonist undergoes within a narrative, and the interplay between them. We consider transcripts of Holocaust survivor testimonies as a test case, each telling the story of an individual in first-person terms. We focus on the survivor's religious trajectory, examining the evolution of their disposition toward religious belief and practice along the testimony. Clustering the resulting trajectories in the dataset, we identify common sequences in the data. Our findings highlight multiple common structures of religiosity across the narratives: in terms of belief, most present a constant disposition, while for practice, most present an oscillating structure, serving as valuable material for historical and sociological research. This work demonstrates the potential of natural language processing techniques for analyzing character evolution through thematic trajectories in narratives.

## 1 Introduction

Characters in narratives are shaped by the continuous interplay of internal and external events they experience, each influencing their thoughts, feelings, and behaviors. As narratives progress, these events unfold along various dimensions. Some are common to essentially all narratives, such as changes in location, while others are specific to particular genres. Within the domain of Holocaust testimonies, as people experience unimaginable atrocities, the change in religious dispositions plays a central role (Brenner, 1980). Applying NLP technology for analyzing these testimonies has been advocated (Artstein et al., 2016; Wagner et al., 2022), offering the potential to extract valuable insights from the vast number of testimonies, instead of relying on small-scale, manual studies.

In this work we extract representations of this progression along the narrative timeline. This allows us to analyze religious trajectories across the entire corpus of testimonies, rather than focusing on individual cases. We leverage off-the-shelf LLMs for inferring the character arc of the protagonist, focusing on their development by exploring the progression of their religious trajectory throughout the narrative. We conduct a dual perspective analysis that captures the protagonist's descriptions of their beliefs, and religious practices.

As people change and adapt, their observance of religion and their disposition towards religion constantly evolve. This is particularly relevant when discussing Holocaust testimonies, which follow survivors' experiences in extreme conditions. We expect different patterns of change in religiosity throughout the testimonies, each telling the story of an individual and covering a wide range of times and places in their life.

We define a *religious trajectory* as a series of mentions of religiosity throughout a given storyline of an individual's life course. We view each trajectory as a combination of the evolution of religious practices and beliefs, assuming that they each describe religiosity from a unique perspective.

For extracting the trajectories, we identify the segments within the text that constitute the trajectory, and assess the protagonist's religious valence and intensity in each segment. We then cluster the resulting trajectories, identifying common sequences (Figure 1). Our research contributes to a deeper understanding of character evolution within narratives by highlighting the potential of machine learning techniques to study thematic trajectories.

## 2 Related Work

**Narrative Analysis.** Recent NLP research has underscored the importance of narrative analysis in the study of human behavior and belief systems.
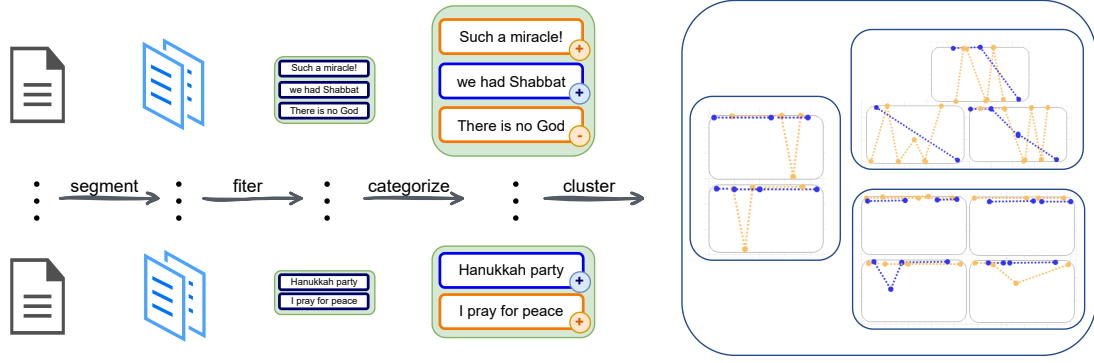
Figure 1: Our pipeline for identifying the religious trajectories from a set of Holocaust survivor testimonies: (1) **Segmentation:** segment the testimonies by question-answer pairs; (2) **Filtration:** train and run a classifier to filter all segments containing religious content; (3) **Determining Valence:** use LLMs to identify the protagonist's valence of religious practice and/or beliefs in a given segment; (4) **Schematization:** cluster the resulting trajectories to identify common patterns of evolution of religiosity in the testimony dataset.

While these studies often center around event and location schemas (Wagner et al., 2023), character personalities, narrator intentions (Piper et al., 2021; Zhu et al., 2023), or narrative structure (Wagner et al., 2022), there remains a gap in tracking the evolution of characters, particularly their development. This gap is critical, as understanding character evolution can provide deeper insight into how narratives convey complex human experiences.

Another aspect studied is character attribute inference, analyzing personas or archetypes, relationships, and emotional trajectories (Chaturvedi et al., 2017). An example of studying thematic trajectories of characters is (Brahman and Chaturvedi, 2020), which models the emotional trajectory of the protagonist by generating and modeling stories that follow various emotional arcs.

This study builds upon these foundations by analyzing character development, focusing on the protagonist's religious practices and their expressed religious belief systems.

**Religious Trajectories.** Our analysis relies on the ability of LLMs to recognize various valences of religious activity and beliefs, including sacred emotions associated with spiritual or religious practices. Plaza-del Arco et al. (2024) explore how religion is represented in LLMs, showing that these models can attribute emotions tied to religion, and display some awareness of sacred emotions, although in an inconsistent and biased manner.

In the social sciences, much work has examined religious trajectories with Ingersoll-Dayton et al. (2002) analyzing patterns of change and stability in religiosity over the lifecourse, and uncovering four distinct patterns of religious trajectories: stable, increasing, decreasing, and curvilinear. Similarly, McCullough et al. (2005) study religious trajectories among immigrant populations. More recently, Exline et al. (2020) explore the religious and spiritual struggles of college students distancing from religion, including conflicts and intellectual questions about God's existence.

In the context of the Holocaust, Lassley (2015) examines religious transformation among Jewish survivors of the Holocaust, focusing on testimonies that describe a shift from being connected with religious traditions to abandoning belief in God completely. Their work discusses how survivor testimonial information can serve as valuable sources regarding faith after the Holocaust.

We leverage recent advances in NLP to analyze religious trajectories on a broader scale using a large dataset of personal testimonies. Rather than focusing on individual cases, our approach aims to uncover common patterns and sequences across many narratives, providing insights into how religious beliefs and practices evolve throughout the life course and under trauma, within this context.

**Trajectory Clustering.** Alqahtani et al. (2021) held a thorough review of time-series data analysis, focusing on deep time-series clustering (DTSC), proposing a clustering approach using deep convolutional auto-encoders (DCAE). Chang et al. (2023) discuss methods for this task, including traditional metrics for trajectories, and deep learning approaches that embed trajectories for similarity measurement. However, these methods typically require labeled data for supervised learning.

## 3 Task Definition

Given a segmented testimony $(x_1, x_2, \ldots, x_n)$, our framework outputs a subsequence of segments representing its religious trajectory $(\langle b_1, p_1, t_1 \rangle, \langle b_2, p_2, t_2 \rangle, \ldots, \langle b_k, p_k, t_k \rangle)$, where $b_j$ are the labels for the belief trajectory, $p_j$ for practice, and $t_j \in (0, 1)$ represents their position within the testimony.

## 4 Data and Annotation

We received 1000 Holocaust survivor testimony transcripts archived by the Shoah Foundation (SF).[1] The interviews were recorded on video in English and transcribed as time-stamped text. The significance of this dataset in the study and remembrance of the Holocaust cannot be overstated. As the number of surviving witnesses diminishes, there is an urgent need to find new approaches to engage with the extensive collection of testimonies.

Due to the lack of a structured format, our first step is to segment the data using natural divisions created by question-answer pairs followed by merging segments with fewer than 10 words and dividing those exceeding 100 words. The method aims to create segments that contain just enough information to determine its relevance to the religious trajectory. Our segmentation method yields a dataset, with the majority length of the segments being between 50-100 words and reasonable paragraph breaks, mostly keeping separate ideas apart from each other. However, some of the samples may contain multiple topics or conflicting religious signals. For example: *"Saturday I would go, and I was encouraged to go to the synagogue. And Sunday, I was encouraged to go to church."*

### 4.1 Annotations

After segmenting the data, we randomly sample and annotate approximately 4000 segments. The trajectories are constructed from all religious content segments from the narrator's perspective. The first annotation assignment is for capturing descriptions of Jewish religious practices and beliefs. This is designed as a binary classification task – the annotators are directed to classify all segments that describe any Jewish religious practices or beliefs of the protagonist or indicate their absence. We provide the complete guidelines in Appendix A.

This complex annotation task involves knowledge of history and the Jewish religion. Many seg-

ments can be ambiguous, rely on context, or require an understanding of the context. Therefore, there may be a lack of consensus among the annotators as to the exact classification of a given segment. All these factors contribute to the complexity of the annotation task and require careful consideration and attention.

To enhance the accuracy of the annotation process and the quality of its results, we recruit multiple annotators, each tasked with annotating over 2500 segments, with an overlap of about 500, for measuring inter-annotator agreement (IAA). The overlap samples are selected randomly and the annotators did not know of the overlap. To evaluate the IAA, we use Krippendorff's alpha, obtaining a score of 0.64, indicating a substantial level of agreement.

The second annotation task is multi-label and consists of identifying the class of each sample and determining the speaker's valence toward the specific class. Each segment belonging to the trajectory is categorized by its content – practice, belief, or both. The segments are further categorized by their valence toward the specific practice or belief they describe. Specifically, we map each sample to one or more of the classes or mark it as falsely recognized religious content.

We distinguish between religious practices and beliefs, as this is discussed in the Sociology of Religion. Chaves (2010) describes common incongruities among religious observers. He states: *"Ideas and practices exist as bits and pieces that come and go as situations change, producing many inconsistencies... Religious ideas, values, and practices generally are not congruent."*

The classification schema is as follows: a segment describing a practice can be annotated with one of the classes – *Active*, *Inactive* or *Other*. The first two classes identify descriptions of participating or violating religious practices while *Other* represents descriptions that do not meet the criteria of *Active* or *Inactive*, or match both of the classes simultaneously. For belief, the classification schema is similar, with the classes being *Positive*, *Negative*, and *Other*. An example of the annotation process is illustrated in Figure 2, and the annotation guidelines are detailed in Appendix B.

To construct the dataset for this task, we first run the religious content classifier to select positive predictions. Next, we use GPT-4o[2] to identify

---

[1] https://sfi.usc.edu/

[2] https://platform.openai.com/docs/models/

3

**Text sample:**

And you know, and sometimes how I feel, you know, I-- I believe there is somebody there, you know. But then, you know, you-- you take these-- like when the Germans came in in 1939, these pious Jews, their whole life was God, whole life was the praying and all that, and what-- and how they-- the things that they did to them, how they laughed at that, they ridiculed them. It-- it's-- I didn't think about it when I was younger, but as-- the older I get, you know, I question.
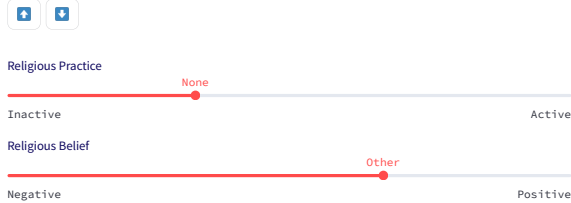
Religious Practice

None

Inactive                                                    Active

Religious Belief

Other

Negative                                                  Positive

Figure 2: An annotation example from the platform we provided the annotators with to identify the survivor's valence of religious practice and belief in each segment.

samples that belong to the minority classes, specifically the Belief classes and *Other-Practice*. The prompts used with GPT-4o include definitions of these minority classes. We then incorporate into the dataset all samples that GPT-4o classified as belonging to these minority classes. This approach helps balance the dataset by increasing the number of examples in the underrepresented classes, ensuring both positive and negative instances are well-represented. The resulting dataset is the one provided to the annotators.

The Krippendorff's alpha score for the second task is 0.44, indicating moderate agreement and suggesting that this classification task is challenging even for human annotators. The Belief dataset contains 141 *Positive* samples, 76 *Negative*, and 42 *Other*. For Practice predictions, the class distribution is 351 *Active*, 76 *Inactive*, and 44 *Other*. We divide the data into three splits with equal portions of each class.

## 5 Modeling Trajectories

### 5.1 Experimental setup

Extracting the religious trajectory of a given segmented testimony consists of two annotation tasks followed by prompting and training adapters for LLM classification. The main steps are: (1) filter religious content, (2) identify practices and beliefs, and (3) measure the valence of the protagonist's expressed attitude toward religion.

The annotation process yields approximately 700 positive samples with religious content and the rest negative. We fine-tune a RoBERTa-large (Liu

et al., 2019) classifier on this task, using a balanced dataset of approximately 1400 segments, divided into three splits: train, validation, and test, by a ratio of 0.8, 0.1 and 0.1, respectively.[3] We evaluate the model on the test set, which yields accuracy and F1 scores of 95%. Additionally, cross-validation over 10 random splits produces an average accuracy of 92%. Given that the goal of this assignment is to identify segments that form a node in the trajectory – i.e., contain relevant practice and belief descriptions, these results demonstrate that the RoBERTa classifier is a reliable tool for this purpose.

For modeling the trajectories, we fine-tune LoRA adapters for Mistral-7B-instruct on data from the second annotation task and compare their performance with GPT-4o. Model selection was based on the average F1 score. For training, we use Mistralai's mistral-finetune repository[4] with their default hyperparameters and test on multiple seeds.[5] For the four practice classes, the average F1 scores are 0.51 for GPT-4o and 0.55 for Mistral-7B. For the parallel belief classes, the scores are 0.57 and 0.56, respectively. In a zero-shot setting, the average F1 scores for practice are 0.45 for GPT-4o and 0.33 for Mistral-7B. For the parallel belief classes, the scores are 0.56 and 0.23, respectively.

Both models perform comparably on our test data, leading us to select Mistral-7B for generating the trajectories. The prompts we run fit the format of Self-Consistency (Wang et al., 2022) and are carefully designed with guidance from Anthropic's prompt generator.[6] Additionally, we tested GPT-4[7] on a subset of the dataset, but it did not respond in the specified format. Considering the complicated annotation process and IAA, together with the models' bias toward stereotypes in the context of religion (Plaza-del Arco et al., 2024), these moderate F1 scores are unsurprising.

While memorization is a valid concern when applying pre-trained LLMs, the specific combination of Holocaust testimony data with our labels is unlikely to have been seen in this exact form during pre-training and therefore is not expected to limit

---

[3]Hyperparameters: 3 epochs, batch size=8, learning rate=1e-5, seed=33

[4]https://github.com/mistralai/mistral-finetune

[5]Hyperparameters: Lora Rank=64, sequence length=16,384, batch size=1, learning rate=1e-4 number of micro-batches=1. Number of training steps for belief=100; for practice=400, weight decay=0.1, pct-start=0.05. Seed for belief=1; for practice=5

[6]https://console.anthropic.com/login

[7]version: gpt-4-turbo-2024-04-09
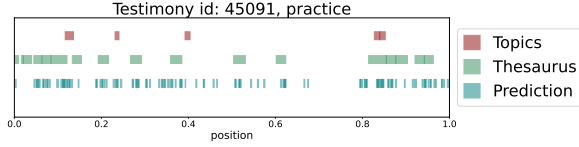
gpt-4o version gpt-4o-2024-05-13

Figure 3: Alignment of predicted trajectories with reference trajectories for testimony ID: 45091, illustrating the differences and overlaps between them. The colored rectangle widths correspond to the segment lengths. The x-axis represents the normalized position within the testimony transcript.

| Baseline | Definition |
|----------|------------|
| Equal | Scatter the points evenly between 0 and 1. |
| Original | Randomly select points from the original distribution of the predicted label. |
| Edges and middle | Randomly select points from three equal splits of (0, 1); each third contains the same portion of points that it has in the predicted distribution of the label. |
| G-Edges and middle | Same as Edges&middle except we sample the point from the normal distribution. |
| 2-Gaussian | All of the points sampled normally from the first half, same for the second half. |
| Normal-original | Sample from the Gaussian with the variance and mean values of the predictions. |

Table 1: The different baseline definitions for evaluating the predicted trajectories.

the generality of our results.

## 5.2 Evaluation

As we are not aware of any directly comparable reference for the trajectories. We therefore evaluate them against two reliable proxies, albeit not directly comparable ones: a topic-modeling-based and a thesaurus-based.

**Topic modeling based reference.** Ifergan et al. (2024) conducted topic modeling using BERTopic, on the same set of testimony transcripts. Their results suggest multiple topics that align with our definitions of religious practices and beliefs. Using this prior knowledge, we extract the labeling by these topics as reference trajectories. Their segmentation method creates segments about three times larger than ours, leading to relatively sparse sequences. The topics we address, and their matching categories appear in Appendix E.

**Thesaurus-based reference.** The testimony transcripts we received from the SF were divided into segments according to their duration. Each segment was indexed with labels describing several terms from the SF thesaurus, containing $\sim 8000$ unique labels.[8] We leverage the terms related to religious practices and beliefs to extract reference trajectories. The full list of addressed terms can be found in Appendix F.

Both of these references have several limitations. First, the segments in the reference trajectories are larger than ours, leading them to contain information irrelevant to the practice and belief labels. Second, the reference sequences are sparse compared to the predicted ones. For the topics, since our analysis for the topic model selects a single topic for each segment, rather than all the top relevant

topics. For the thesaurus, despite the large number of terms, many segments lack the labeling of their matching terms. This may be a result of the exceedingly large label set ($\sim$ 8K labels), which may have resulted in recall issues in the SF annotation. In addition, the label set of the references is partial, meaning that the majority of the topics and terms that we address do not have a specific valence. Considering these points, the false positive rate of the predictions is unknown, making the references suitable only for measuring recall. An example illustrating the alignment of the references and predictions is shown in Figure 3.

We run the fine-tuned Mistral-7B model on the full dataset, producing 1000 trajectories. In this setting, relying on our fine-tuned classifier, only religious content predictions are prompted, and not every single one of the original dataset segments. For each of the reference and predicted sequences, we quantify their distance. For a given non-empty trajectory $T$ and a reference path $R$:

$$min\_sum\_dist(T, R) = \sum_{r \in R} min_{t \in T}(|t - r|)$$

If $T$ is empty, return the number of segments in $R$, and if $R$ is empty, return 0.

We compare the sum of the distances to baseline trajectories. These baselines are artificial sequences designed by using our prior knowledge about the distribution (Figure 5) of practice and belief segments in the full dataset; each is the same length as the predicted sequence. The different baselines for a given class are defined in Table 1.
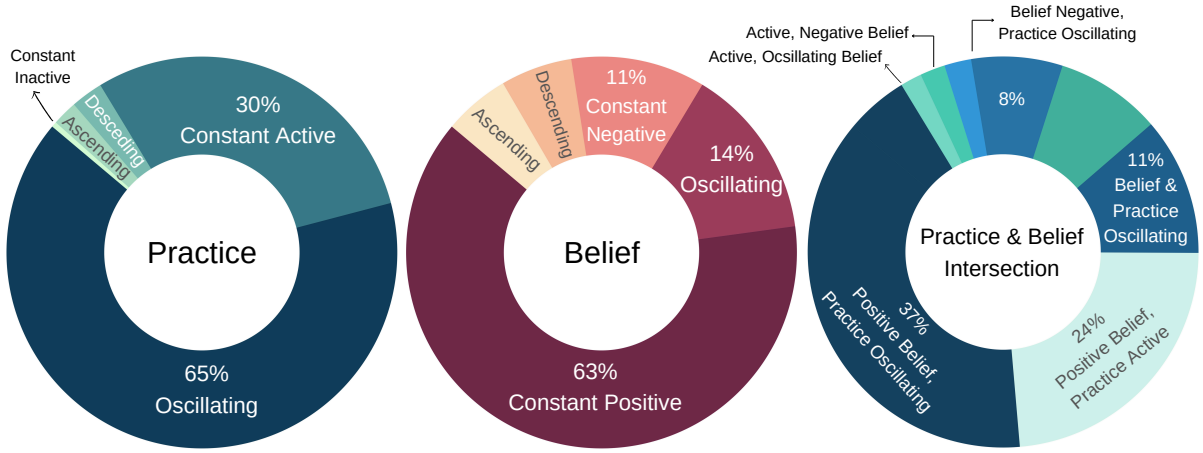
---

[8] sfi.usc.edu/content/keyword-thesaurus

Figure 4: Religious Trajectory structure distributions, from left to right: 65% of the practice trajectories have an oscillating structure and 30% are constant-active. The belief trajectories have a similar distribution, with the majority sharing a constant-positive structure (63%) and the rest distributed among oscillating (14%), constant-negative (11%), descending (6%), and ascending (5%). For the intersection of the two aspects, the two large groups cover 61% of the intersection, all have a constant-positive belief structure, while the practice valence varies: 37% oscillate and 24% have a constant-active structure.

## 6 Trajectory Clustering

After extracting the trajectories, we cluster them in two ways: according to a predefined taxonomy and using unsupervised methods.

We expect to identify a few main clusters, each representing a different trajectory type that can be identified by generating a path that is more similar to each of the sequences within its cluster than to any of the other sequences.

### 6.1 Predefined Taxonomy

We divide the trajectories into classes based on their valence and structure: *constant positive/negative, ascending/descending, oscillating*, and the degree of coverage, relative to the testimony storyline: $t_{max} - t_{min}$ (Table 2). For this evaluation, each trajectory is filtered by removing any neutral values it contains and then shrunk by replacing a sequence of constant consecutive values with a single value. For example, [-1,1,0,1,1] was shrunk to [-1,1].

| Structure | Filtered&shrunk series |
|---|---|
| Constant-Negative/Inactive | [-1] |
| Constant-Positive/Active | [1] |
| Ascending | [-1, 1] |
| Descending | [1, -1] |
| Oscillating | $\{1, -1\}^n$ |
| **Coverage Level Definitions** | |
| Low | $t_k - t_1 \leq 0.33$ |
| Medium | $0.33 \leq t_k - t_1 \leq 0.33$ |
| High | $0.67 < t_1 - t_k$ |

Table 2: Definitions of the structures and coverage levels for the predefined taxonomy.

### 6.2 Unsupervised Hierarchical Clustering

For hierarchical clustering, we use Dynamic Time Warping (Berndt and Clifford, 1994) with multiple window sizes to measure the distances between the trajectories, and then run Agglomerative (Zepeda-Mendoza and Resendis-Antonio, 2013) and HDB-SCAN clustering (McInnes et al., 2017). Further details are in Appendix H.

## 7 Results

We evaluate the predicted trajectories against the references and present the results in Table 3. Almost every reference point aligns with a predicted one, validating the predictions' recall.

Besides the evaluation of the full dataset, we randomly sampled 101 testimonies and compared their predicted trajectories in three settings: prompting GPT-4o and the trained Mistral-7B on religious content predictions, and separately prompting Mistral on the full set of segments. Evaluation results for this subset are provided in Appendix G. We observe that prompting on all segments, as opposed to just religious content, leads to a significant over-prediction rate for each class. Specifically, when using the non-filtered data, the proportion of positive classes does not align with the proportion of religious content in the full dataset, as predicted by a reliable classifier trained on this data.

| | Topic | | | Thesaurus | | | | |
|---|---|---|---|---|---|---|---|---|
| | **B** | **P** | **P⁻** | **B** | **P** | **P⁺** | **B⁺** | **B⁻** |
| Predicted | **12** | **12** | **41** | **7** | **9** | **10** | **13** | **127** |
| Original scatter | 22 | 104 | 48 | 16 | 71 | 54 | 14 | 139 |
| Edges & middle | 27 | 33 | 57 | 20 | 30 | 31 | 20 | 139 |
| Equal scatter | 35 | 24 | 53 | 23 | 18 | 25 | 14 | 180 |
| Normal-original | 24 | 62 | 51 | 21 | 52 | 60 | 20 | 141 |
| # Reference paths | 335 | 905 | 187 | 282 | 787 | 761 | 98 | 217 |
| # predicted paths | 874 | 996 | 665 | 874 | 996 | 988 | 659 | 278 |
| # Reference points | 456 | 2,768 | 301 | 439 | 2,434 | 2,214 | 171 | 253 |
| # predicted points | 4,500 | 21,193 | 1,719 | 4,500 | 21,193 | 17,462 | 2,339 | 427 |

Table 3: Sum of $min\_sum\_dist$ for fine-tuned Mistral-7B predictions and baseline trajectories on the full dataset. **Original:** Sample from the distribution of the predictions. **Edges&middle:** Random sample from three equal splits of $(0, 1)$; according to the predicted distribution of the label. **Equal:** Even scatter. **Normal-original:** Sample from the Gaussian with the variance and mean values of the predictions.
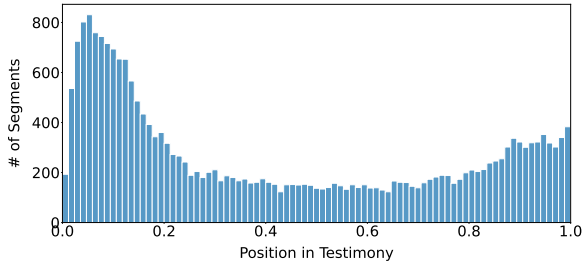


Figure 5: Distribution of all religious content, the distributions that we based the baselines on are according to each label separately.

### 7.1 Clustering

Among the trajectories extracted by the pipeline, 68% contain at least two points for both practice and belief. The taxonomy reveals that, for the belief aspect, more than half (63%) share a constant positive shape, 14% oscillate, and the rest are distributed among constant negative (11%), descending (6%), and ascending (5%). For the positive trajectories, 80% cover a substantial portion of the plot, whereas negative trajectories are sparser, 59% covering up to two-thirds of the narrative. Practice trajectories have two large structures, 65% oscillate and 30% follow a constant-active pattern. 96% of the two main structures have high coverage levels. Examples for clusters are given in Appendix H.

When comparing the trends according to the two aspects, the two trends that we recognize, covering 61% of the intersection, all have a constant-positive belief structure, while the practice valence varies: 37% oscillate and 24% have a constant-active structure. This discrepancy may reflect external circumstances the survivors went through that prevented them from practicing Judaism, or social changes. These distributions are shown in Figure 4.

### 7.2 Clustering Evaluation

We evaluate the clusters against human judgment, using a subset of trajectories with a length of at least 4 and a span greater than 0.75 of the testimony's length.

**IAA.** Annotators reviewed text segments from 118 trajectories and performed two tasks – classifying them according to the taxonomy and identifying the more similar pairs from a list of triplets. IAA levels for classification were substantial for belief ($\alpha = 0.66$) and moderate for practice ($\alpha = 0.56$), while for the triplet task, IAA was moderate ($\alpha = 0.58$ for belief, $\alpha = 0.32$ for practice).

**Comparing human annotation with DTW.** The DTW distances correlate with human judgment for similarity, validating its use for identifying similar trajectories. For similar pairs, the average distances are 0.46 (belief, $\sigma = 0.19$) and 0.41 (practice, $\sigma = 0.16$). For dissimilar pairs, these distances increase to 0.57 (belief, $\sigma = 0.18$) and 0.47 (practice, $\sigma = 0.20$) on average. A t-test reveals statistical significance for belief between similar and dissimilar pairs ($p < 0.01$) but not for practice, suggesting that additional annotation is needed to validate these results.

Taxonomy annotation shows that the DTW distance also correlates with the taxonomy's distinctions. The average distances are 0.29 (belief, $\sigma = 0.12$) and 0.30 (practice, $\sigma = 0.13$ for pairs of trajectories from the same class; while for pairs
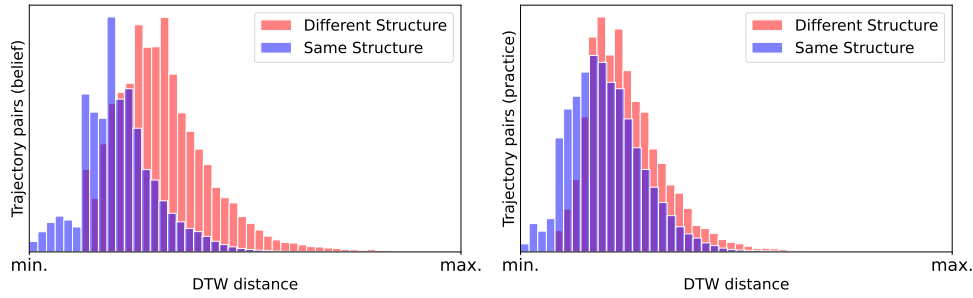
7

Figure 6: DTW distance distributions of trajectory pairs: the left plot represents belief and the right practice. Blue indicates pairs with the same structures, and red indicates pairs with different structures. The distribution for identical structures is shifted toward smaller DTW distance values.

with different structures, the average distance is 0.32 for both aspects (belief, $\sigma = 0.12$, practice, $\sigma = 0.13$), with t-tests confirming statistical significance ($p < 10^{-7}$ in both cases). This provides further validation for DTW by showing that part of the distance it captures is related to fundamental characteristics of the trajectory.

**Manual and automatic taxonomy classification.** Comparing the automatic and manual classification of trajectories into taxonomy classes, we observe F1 scores of 0.45 for belief and 0.43 for practice, showing that the task is challenging, even for humans. Some discrepancies arise from the assumption that the predicted trajectories depend on the religious trajectory progressing along the narrative timeline; however, it does not hold for each case.

As with manual taxonomy classification, automatic taxonomy prediction results correlate with DTW distances (Figure 6). Pairs with the same structure exhibit smaller distances, with average distances of 0.10 (belief, $\sigma = 0.04$) and 0.08 (practice, $\sigma = 0.04$) for structures of the same class, compared to 0.16 (belief, $\sigma = 0.05$) and 0.12 (practice, $\sigma = 0.05$) for other pairs, reinforcing the link between structural alignment and DTW metrics.

## 8 Discussion

Our analysis identifies common structures of religious trajectories, grouping those with similar structures. For belief, the majority presents a constant positive structure, as to practice, while for practice most oscillate.

The distribution of religious content throughout the testimonies demonstrates a common narrative sequence within the dataset that justifies comparison of trajectories on the narrative timeline (Figure 5). This can be caused by the SF interviewers'

guidelines regarding the themes and order of the testimony, including religious belief and activity.

Other directions we aim to cover include examining the testimonies that share structural similarities, exploring additional commonalities, and how these may correlate with metadata such as age, gender, pre/post-war hometowns, and experiences during the war. Any analysis will need to consider the general distribution of the data, as most survivors in our sample were born in Poland, with most interviews conducted in the United States and Canada.

Another perspective expected to contribute to a more sophisticated picture of the patterns is including the non-positive-or-negative parts of the trajectory in the structure taxonomy analysis. The importance of incorporating these segments arises from the unsupervised clustering that identifies a separation between trajectories with multiple non-positive or negative segments, suggesting that these play a separate role in the trajectories.

Additionally, inspecting the relationship between the practice and belief trajectories, investigating whether one follows the other's pattern or is a shift of the other, and whether one can be predicted based on the other is an idea to explore.

## 9 Conclusion

This paper explores character development in narrative text, by analyzing religious trajectories from a dual perspective. We develop a framework to extract and cluster trajectories from 1,000 Holocaust testimony narratives. These findings encourage further research on utilizing LLMs to capture character development and thematic trajectories and offer valuable material for historical and sociological research. We believe that alignment with relevant literature can lend insights to the modeling.

8

## Ethical Considerations

According to the guidelines given by the SF archive, although the testimonies were not given anonymously, no identifying details are included in our analysis. Our codebase and scripts will be released on request, but they will not contain any data from the archives. The data and trained models used in our work will not be shared with third parties without the archives' consent. Permission can be requested from the SF archive to browse and research the testimonies.

## Limitations

As for the framework's limitations, it is important to consider the existence of a reporting bias. Our data is limited to interview transcripts, which capture the survivors' descriptions of their beliefs without offering a broader understanding of their theological positions. Although the SF recorded testimonies in several languages and countries, it is important to note that our investigation only covered testimonies given in English.

We examine the trajectories without accounting for differences in their lengths and frequencies. However, these could provide a more nuanced understanding of the religious, social, and theological aspects reflected in the trajectories. Omitting the blank trajectories from our study might overlook a significant aspect. The absence of a religious trajectory may result from either a low level of religiosity or religion not being mentioned during the interview, which conveys a story, though the narrative behind it remains unknown.

Additionally, much of this work relies on human annotators. Despite their shared professional backgrounds, each annotator brings different prior knowledge, which can affect agreement levels. For the cluster similarity evaluation, our analysis is restricted to the subset of testimonies that the annotators reviewed and analyzed.

## References

Ali Alqahtani, Mohammed Ali, Xianghua Xie, and Mark W. Jones. 2021. Deep time-series clustering: A review. *Electronics*.

Ron Artstein, Alesia Gainer, Kallirroi Georgila, Anton Leuski, Ari Shapiro, and David R. Traum. 2016. New dimensions in testimony demonstration. In *North American Chapter of the Association for Computational Linguistics*.

Donald J. Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series. In *KDD Workshop*.

Faeze Brahman and Snigdha Chaturvedi. 2020. Modeling protagonist emotions for emotion-aware storytelling. *ArXiv*, abs/2010.06822.

Reeve Robert Brenner. 1980. The faith and doubt of holocaust survivors. In *The faith and doubt of Holocaust survivors*.

Yanchuan Chang, Egemen Tanin, Gao Cong, Christian S. Jensen, and Jianzhong Qi. 2023. Trajectory similarity measurement: An efficiency perspective. *Proc. VLDB Endow.*, 17:2293–2306.

Snigdha Chaturvedi, Haoruo Peng, and Dan Roth. 2017. Story comprehension for predicting what happens next. In *Conference on Empirical Methods in Natural Language Processing*.

Mark Chaves. 2010. Sssr presidential address rain dances in the dry season: Overcoming the religious congruence fallacy. *Journal for the Scientific Study of Religion*, 49:1–14.

Julie J. Exline, Daryl R. Van Tongeren, David F. Bradley, Joshua Adam Wilt, Nick Stauner, Kenneth I. Pargament, and C Nathan DeWall. 2020. Pulling away from religion: Religious/spiritual struggles and religious disengagement among college students. *Psychology of Religion and Spirituality*.

Maxim Ifergan, Omri Abend, Renana Keydar, and Amit Pinchevski. 2024. Identifying narrative patterns and outliers in holocaust testimonies using topic modeling. In *Proceedings of the First Workshop on Holocaust Testimonies as Language Resources (HTRes) @ LREC-COLING 2024*, pages 44–52, Torino, Italia. ELRA and ICCL.

Berit Ingersoll-Dayton, Neal Krause, and David L. Morgan. 2002. Religious trajectories and transitions over the life course. *The International Journal of Aging and Human Development*, 55:51 – 70.

Jennifer Lassley. 2015. A defective covenant: Abandonment of faith among jewish survivors of the holocaust. *International Social Science Review*, 90(2):1–17.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692.

Michael E. McCullough, Craig K. Enders, Sharon Brion, and Andrea R. Jain. 2005. The varieties of religious development in adulthood: a longitudinal investigation of religion and rational choice. *Journal of personality and social psychology*, 89 1:78–89.

Leland McInnes, John Healy, and S. Astels. 2017. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2:205.

Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative theory for computational narrative understanding. In *Conference on Empirical Methods in Natural Language Processing*.

Flor Miriam Plaza-del Arco, Amanda Cercas Curry, Susanna Paoli, Alba Cercas Curry, and Dirk Hovy. 2024. Divine LLaMAs: Bias, stereotypes, stigmatization, and emotion representation of religion in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4346–4366, Miami, Florida, USA. Association for Computational Linguistics.

Eitan Wagner, Renana Keydar, and Omri Abend. 2023. Event-location tracking in narratives: A case study on holocaust testimonies. In *Conference on Empirical Methods in Natural Language Processing*.

Eitan Wagner, Renana Keydar, Amit Pinchevski, and Omri Abend. 2022. Topical segmentation of spoken narratives: A test case on holocaust survivor testimonies. In *Conference on Empirical Methods in Natural Language Processing*.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Huai hsin Chi, and Denny Zhou. 2022. Self-consistency improves chain of thought reasoning in language models. *ArXiv*, abs/2203.11171.

Marie Lisandra Zepeda-Mendoza and Osbaldo Resendis-Antonio. 2013. *Hierarchical Agglomerative Clustering*, pages 886–887. Springer New York, New York, NY.

Lixing Zhu, Runcong Zhao, Lin Gui, and Yulan He. 2023. Are nlp models good at tracing thoughts: An overview of narrative understanding. In *Conference on Empirical Methods in Natural Language Processing*.

## A   Religious Content Annotation Guidelines

We are working on exploring changes in religious beliefs as reflected in testimony transcripts of Holocaust survivors; through looking at Jewish practices and beliefs described in the text. We want to capture all data segments from the testimonies that hold to the definition of religious content, we will label "TRUE" all segments that describe any Jewish religious practices or beliefs of the interviewee, **as well as segments that indicate the absence of religious practices or beliefs.** In detail:

- Religious practices and beliefs: Content that describes *"Who we were, what we did, and what we believed in"*. For instance phrases in the style of:

    - *"We were orthodox" / "We were not religious"*

    - *"My family belonged to a synagogue" / "We didn't keep the Sabbath"*

    - *"We hosted Pesach Seder" / "I wasn't familiar with the Jewish holidays"*

- Descriptions of the interviewee's theological inner life, giving the reader information about their faith in God. For example:

    - *"I looked up to the sky, there must be a God above"*

    - *"I didn't believe that God can look at this cruelty and all these dreadful things"*

- If the sample is written in first person (We, us, I), it is telling the story of the interviewee or their family and not stories about others, and its content meets our description of religious practices and beliefs then it will be labeled True.

- We will consider samples written in the third person, only if they reflect on the interviewees' beliefs, or describe the environment that he grew up in.

    - *"My parents explained that we were Jewish, secular Jews, nonbelievers"* - True

    - *"My elder brother Rabbi Yosef Meyer was with me"* - False

    - *"I think my brother probably was affected because he changed his religion, he became a Roman Catholic."* - True, since implies that the speaker wasn't affected.

- The phrases "Baruch Hashem " and "Thank God " are religious phrases and, therefore will be labeled True.

- Zionism descriptions will be considered False; since the Zionist movement saw the land of Israel and the Hebrew language as integral parts of the Jewish national heritage, and not necessarily of religious significance.

- Descriptions of Jewish Identity will be labeled False. We are focusing on **religious Identity and not Jewish Identity**, therefore we want to capture only segments that contain religious practices and beliefs, not signs of Jewish culture.

- Jewish food or music, or speaking Yiddish - Not considered religious content

10

- Hnnukah candles, Bris, Bat Mitzvah celebrations, and Friday night dinner - are considered positive religious content.

- Getting a Jewish education

  - Secular Jewish education, such as attending Hebrew School, or a Tarbut institution, or participating in secular youth movements - Not considered religious content

  - Studying at Yeshivas or in any Jewish religious school, Bnei Akiva or Mizrahi - Positive religious content.

- Mentions of miracles

  - If reflecting a belief will be labeled True, and not when used as a phrase.

  - It is suggested to view the wider context of segments that mention miracles.

- Description of friendship relations with the non-Jewish population will be considered False, unless it indicates a religious approach.

  - *"We were friendly with our gentile neighbors, though didn't eat at their houses"* - True

**Some Examples**

1. *"He gave me a Brucha, a blessing. I was very sick, a very sick child. And he told, don't worry. She'll grow. She'll be well, God will help her. "*

2. *"he was able to survive, which it was a, a miracle. It was simply a miracle. Because, by rights, he was dying. He was on death's door"*

3. *"I got my presents under the Christmas tree, and then my girlfriend came to Hanukkah to my place. She got a present."*

4. *"I send them to the best yeshivas. This is the whole future of my life. And this was– baruch HaShem– always could be better. But I'm satisfied– baruch HaShem. I accomplished."*

5. *"He will always tell you, if you ask him, what's your religion, I am a Jew. And the boys really have neither been circumcised nor baptized. And it doesn't seem to have done them any harm so far."*

- Although it uses the word religion, The first part alone would not be considered Religious Content: "If you ask him, what's your religion, I am a Jew" since it is about Jewish identity, not Religion.

6. *"I don't think God was absent, but I really don't understand yet."*

7. *"A lot of kids thought that I am just a Catholic like they are. But it came a time, that when we came to school we had to stand up to say the prayers. And, of course, the Jewish children, we stood up, but we didn't say, and we didn't cross ourselves."*

## B Practice and Belief Identification and Rating - Annotation Guidelines

We are working on exploring changes in religious beliefs as reflected in testimony transcripts of Holocaust survivors; through looking at Jewish practices and beliefs described in the text. This task has two parts: Identifying the class of each sample and then determining the speaker's approach toward the specific class.

**Part 1 - Practice and Belief Identification**  Practice and Belief Identification is a multi-label classification task. The dataset for this task was randomly sampled from data classified as religious content by a machine learning classifier with high accuracy. The task is to map each sample to one or more of the classes or mark it as falsely recognized as religious content.

**Label Definitions**

- Practice - A ritual, action, or activity motivated by the Jewish religion. Also descriptions reflecting an absence of Jewish religious practices will be mapped to this label.

  - Not including descriptions of Jewish communities or culture.

  - Examples: Saying prayers, going to synagogue, keeping Kosher.

- Belief - Inner life descriptions of ideas or thoughts related to God or religion. Including philosophy and feelings about God, and descriptions of miracles.

  - Excluding descriptions of Jewish identity or Zionism.

11

- Other - Samples that were falsely recognized as religious content

A general note: We will consider samples written in the third person, only if they reflect on the interviewees' beliefs, or describe the environment that he grew up in.

**Some Examples   Practice**

1. *"He gave me a brucha, a blessing. I was very sick. He gave me a blessing. I was usually a very sick child– wild, spoiled, and sick. And he told, don't worry. She'll grow. She'll be well, – God will help her. "*

   - *Note: Although phrased in the third person this example is relevant since the narrator of the text went to get the Bracha. It is also considered Belief, since it contains the phrase "God will help her".*

2. *" I got my presents under the Christmas tree, and then my girlfriend came to Hanukkah to– to my place. She got a present."*

**Belief**

1. *"He gave me a brucha, a blessing. I was very sick. He gave me a blessing. I was usually a very sick child– wild, spoiled, and sick. And he told, don't worry. She'll grow. She'll be well, – God will help her. "*

2. *"he was able to survive, which it was a, a miracle. It was simply a miracle. Because, by rights, he was dying. He was on death's door"*

   - Note: While the text is written in the third person point of view, it reflects the narrators faith.

3. *"I had the will, strong willpower to live. Every night I would take my prayer book. Every night somebody was in charge from the twins to look out. Each bunk beds, everybody is asleep, it's quiet. Then came my turn. One night, I would take my prayer book and cry. Say Psalms. It's so good. I felt close and came–"*

   - Note: this example is also considered Practice, since it describes a daily ritual.

**Other - Neither a practice or belief descriptions**

1. *"Was there any particular reason for going to a Jewish school for secondary school? The reason was that I didn't want to sit in exam to go to grammar school. I was too frightened. It was a grave mistake."*

**Part 2 - Rating Jewish Practice and Belief Text Samples**   After identifying the classes of a text sample, our task is to rate the narrator's approach to the specific class on a scale from -1 to 1.

**Rating Definitions:**

- Active - Actively practicing a Jewish religious ritual

- Inactive - Violating Jewish religious practices, descriptions of not observing, actively not practicing a Jewish religious ritual, or practicing a different religion.

- Neither active or inactive valence of Jewish religious practice, or both simultaneously.

**Practice examples   Active**

1. *"How would you describe your family's religious life? Orthodox."*

2. *"When did you have a Bar mitzvah? When I was 13. In the– in the main synagogue. "*

**Inactive**

1. *"And did you– did you observe Shabbat in any way, light candles, go to synagogue? No, no, no candles, no Shabbat."*

2. *"The church became a very important part of my life. Although we didn't go often, any time that I went with my Catholic family, the Leszkowicz or Maciejewski family, I felt safe. And this cross that you're looking at was given to me by my Catholic grandmother."*

**Other**

1. *"Was there davening on the train? Yeah. Sure, they were davening."*

2. *"We were sleeping in an attic, in a cellar, here. If we were caught, that would be the end. Meanwhile, Rosh Hashanah– and– and don't forget."*

   - We don't know whether the speaker celebrated Rosh Hashana or is just mentioning the time od the year, therefor it's valence is "other".

3. *"You said your family was not religious. So in what way did you identify with Judaism? I would say– ethnically and community feeling. my grandmother and also my maternal grand-parents, they went also only I think mainly Rosh Hashanah and Yom Kippur, not more than that."*

- On the one hand, going to synagogue on the high holidays - Active.
  On the other, not a religious family - In-active.
  → Altogether the signal is "other".

**Belief examples   Positive**

1. *"And my mother kept on saying, this is going to be the Passover of Passovers, because we were liberated. Moses took us out of Egypt. And now, the world is going to take us out of Egypt from Hitler. But it didn't happen. The Jews fought for a month there. They fought most– the strongest army in the world. They, of course, at the end, lost."*

- Note: This is a practice example as well.

2. *"What did you do when you saw that being done to the children? You think I could go up and say don't do it? Nothing. I just prayed. I just prayed. But I had to look. When I look at this now, that's what I saw. That's what had happened. And still Germany is so strong. And Germany is a country."*

3. *"But I still belong– believe in God. And if he's good, or if he can close his eyes, and then see us, and not to see and not to hear, but it comes a time when he takes his hands away, and from his eyes and ears, and he sees. And I hope that I will see, too. If I don't see, as my children, grandchildren will see and hear that he took his hands away from his eyes and ears, and he can help them."*

**Negative**

1. *"Why not? How can I believe? How is it possible? I just keep the tradition for my children and grandchildren sake and my friends who don't even keep the tradition. So they come. And we have nice dinner. And we try to be in good mood, but nothing Jewish in us anymore, no."*

2. *"How were you mentally at this stage, Leah, after all your experience? Very angry, very bitter. I didn't believe there was a god. I didn't want to know such a thing as a god. Very guilty."*

3. *"And that was our religion. But we didn't keep Shabbat. We didn't keep kosher. Nothing. And during the time in Auschwitz, I didn't even be-lieve in religion. I didn't believe that the God can look at this cruelty and at all these dread-ful things. I didn't believe in God. Maybe it was just a shock to the– I don't know."*

- Note: This is a practice example as well.

**Other**

1. *"What do you think kept you going? What kept me going? Good question that, obstinacy– I didn't want to die. And however much I didn't want to die, the decision didn't rest with me. This– this is the interesting– it didn't rest with me at all. Maybe God helped– helped me to– to somehow that Dr. Mengele didn't pick me every time he went through these selections. I don't know. I don't know. It certainly wasn't me."*

2. *"And what about your beliefs? Do– do you practice Judaism? Do you believe in a deity? And is that in any way impacted by your expe-riences one way or the other? Well, I believe very much in the importance of what I would call tradition. And I think that our tradition is– is something to be very proud of. I think we have a very rich tradition, we have very rich history, a very unique history in the world. And I think it is– it is very much part of us."*

- The speaker doesn't deny or assure faith in God.

3. *"What has– have you inherited that's helped you survive it? Well, I have more questions, I mean, about God than my father does, I'll tell you that. But I do respect very much my father's faith in God and feel that, if he can live though that and believe in God, then I certainly don't have a right to question that. So I've learned that people can go through hell, still believe in God, and go on in, in most ways in their lives, go on and enjoy life, and enjoy children. My father loves young*

*children. He loves children. His whole life is children."*

## C  GPT-4o Prompt for enlarging the portion of the minority classes

**Belief Recognition**

Your task is to determine if the text contains a phrase describing a religious belief. Any length of selection is considered a phrase!

Class definitions: BELIEF = One or more of the following: Spirituality, inner-life descriptions, beliefs about God, relationship with God, principles, identity, philosophy about God, feelings about God, Grace to God, thoughts about God, emotional experiences that come from beliefs about God, mentions of miracles. God == The Lord == Hashem.

Other = descriptions of anything else.

The answer must contain a single class, Do not add any words! Do not provide any explanations! The only labels allowed are: BELIEF OTHER

Examples:

Text: "My mother was a very religious person and a righteous person, and she felt that perhaps it would be best that she go to live with her aunt. Her aunt kind of appealed to her religious fair being, I guess, if you will." Label: OTHER

Text: "And God should be with you. So my father said, OK, you're going to see. They're going to kill us from the back. From here, we're not going out alive. I said, look it, whatever God wants, this is what's going to be." Label: BELIEF

Text: "With my education that I had on the Zionist movement, I realized that they taught me what it means to be a Jew. And my Hebrew school taught me. And the Bible taught me. But then I realized what it meant to be a Jew." Label: BELIEF

Text: "And my mother kept on saying, this is going to be the Passover of Passovers, because we were liberated. Moses took us out of Egypt. And now, the world is going to take us out of Egypt from Hitler. But it didn't happen. The Jews fought for a month there. They, of course, at the end, lost." Label: BELIEF

Text: "And that's the way my mother went to Palestine as a tourist, my brother with the Youth Aliyah. We were big Zionists. " Label: OTHER

Text: "And you're proud today to be a Jew? Oh, my goodness. And my friend upstairs, I made a deal with him a long time ago. And he kept his side of the bargain, I must keep my side of the bargain, because he's given me miracles, and he's shown me he is there." Label: BELIEF

Text: "I want to believe that there is a purpose for all this. Sometimes things are very hard to accept, but I still believe." Label: BELIEF

Text: "It was the time between Pesach and, and, and Lag B'Omer. And we were supposed to get together for Lag B'Omer." Label: OTHER

Text: "My parents explained that we were Jewish, secular Jews, nonbelievers" Label: BELIEF

Text: "My father's family was a fairly large traditional Jewish family. Orthodox. I remember my mother wearing a wig." Label: OTHER

Text: "I have more questions about God than my father does. But I do respect very much my father's faith in God and feel that, if he can live through that and believe in God, then I certainly don't have a right to question that. So I've learned that people can go through hell, still believe in God." Label: BELIEF

Text: "This is a prayer book that was my grandmother's. She carried it with her all through Europe, into Israel, and then she brought it to this country." Label: OTHER

Text: "I took her death like a punishment from God, Why am I being punished? I pray every morning. I was so hurt. I felt that God did it, because they said God does it. I was very angry at God." Label: BELIEF

14

Text: "They observe the Jewish holiday, but they weren't real Orthodox. They were very Conservative. To the shul they used to go only on the holidays." Label: OTHER

Text: "Simchas Torah they went over the roof dancing around. And like all the Hasidim doing." Label: OTHER

Text: "After all what my father went through, we could still believe in HaShem, God, and that we have to believe in Him and the continuity of the Jewish people." Label: BELIEF

Text: "How can I believe? How is it possible? I just keep the tradition for my children and grandchildren sake and my friends who don't even keep the tradition. but nothing Jewish in us anymore, no." Label: BELIEF

Text: "We went to church every day." Label: OTHER

Text: "It was like a miracle happened" Label: BELIEF

Text: "Continue on the Jewish life. I'm very proud when I see my grandchildren say the Kiddush. I know my parents they must kvell, in heaven, to see this." Label: BELIEF

Text: "I had a two-tier program, one secular education, the other one is Jewish or Hebrew studies, which first included learning how to read, then Mishnah, Chumash, and Talmud." Label: OTHER

**Practice Recognition**

Your task is to determine if the text contains a phrase describing a religious practice. Any length of selection is considered a phrase!

Class definitions: PRACTICE = An action or activity connected to religion, Jewish community, or Jewish culture. Examples: Being orthodox, saying religious texts, going places connected to the Jewish religion, keeping certain restrictions, speaking Yiddish.

OTHER = descriptions of anything else.

The answer must contain a single class, Do not add any words! Do not provide any explanations! The only labels allowed are: PRACTICE OTHER

Examples:

Text: "My mother was a very religious person and a righteous person, and she felt that perhaps it would be best that she go to live with her aunt. Her aunt kind of appealed to her religious fair being, I guess, if you will." Label: PRACTICE

Text: "I had a few friends that came back. We resumed life, normally, planning to eventually to go to Palestine or to Israel. The community grew to about 300 people after the war. " Label: OTHER

Text: "And God should be with you. So my father said, OK, you're going to see. They're going to kill us from the back. From here, we're not going out alive. I said, look it, whatever God wants, this is what's going to be." Label: OTHER

Text: "With my education that I had on the Zionist movement, I realized that they taught me what it means to be a Jew. And my Hebrew school taught me. And the Bible taught me. But then I realized what it meant to be a Jew." Label: PRACTICE

Text: "And my mother kept on saying, this is going to be the Passover of Passovers, because we were liberated. Moses took us out of Egypt. And now, the world is going to take us out of Egypt from Hitler. But it didn't happen. The Jews fought for a month there. They, of course, at the end, lost." Label: PRACTICE

Text: "And that's the way my mother went to Palestine as a tourist, my brother with the Youth Aliyah. We were big Zionists. " Label: PRACTICE

Text: "And you're proud today to be a Jew? Oh, my goodness. And my friend upstairs, I made a deal with him a long time ago. And he kept his side of the bargain, I must keep my side of the bargain, because he's given me miracles, and he's shown me he is there." Label: OTHER

Text: "I want to believe that there is a purpose for all this. Sometimes things

15

are very hard to accept, but I still believe."
Label: OTHER

Text: "It was the time between Pesach and, and, and Lag B'Omer. And we were supposed to get together for Lag B'Omer." Label: PRACTICE

Text: "My parents explained that we were Jewish, secular Jews, nonbelievers" Label: PRACTICE

Text: "My father's family was a fairly large traditional Jewish family. Orthodox. I remember my mother wearing a wig." Label: PRACTICE

Text: "I have more questions about God than my father does. But I do respect very much my father's faith in God and feel that, if he can live through that and believe in God, then I certainly don't have a right to question that. So I've learned that people can go through hell, still believe in God." Label: OTHER

Text: "This is a prayer book that was my grandmother's." Label: PRACTICE

Text: "I took her death like a punishment from God, Why am I being punished? I pray every morning. I was so hurt. I felt that God did it, because they said God does it. I was very angry at God." Label: PRACTICE

Text: "They observe the Jewish holiday, but they weren't real Orthodox. They were very Conservative. To the shul they used to go only on the holidays." Label: PRACTICE

Text: "Simchas Torah they went over the roof dancing around. And like all the Hasidim doing." Label: PRACTICE

Text: "After all what my father went through, we could still believe in HaShem, God, and that we have to believe in Him and the continuity of the Jewish people." Label: OTHER

Text: "How can I believe? How is it possible? I just keep the tradition for my children and grandchildren sake and my friends who don't even keep the tradition. but nothing Jewish in us anymore, no." Label: PRACTICE

Text: "We went to church every day." Label: PRACTICE

Text: "It was like a miracle happened" Label: OTHER

Text: "Continue on the Jewish life. I'm very proud when I see my grandchildren say the Kiddush. I know my parents they must kvell, in heaven, to see this." Label: PRACTICE

Text: "I had a two-tier program, one secular education, the other one is Jewish or Hebrew studies, which first included learning how to read, then Mishnah, Chumash, and Talmud." Label: PRACTICE

## D Instruction Prompts

The Instruction prompts we ran for GPT4o and Fine-tuning Mistral:

**Prompt for Belief Classification**

Your task is to carefully read this text and determine the speaker's valence of Jewish religious belief in God, based on the following classification system: POSITIVE: The text expresses the narrator's belief in God according to the Jewish religion, or his existing relationship with God.

NEGATIVE: The text expresses the narrator's lack of belief in God according to the Jewish religion, or a rejection of religious beliefs.

AMBIGUOUS: The text expresses a relationship with God that does not meet the criteria of the classes POSITIVE or NEGATIVE. This includes questioning Gods while believing in his existence.

NONE: The text does not directly imply the speaker's belief in God and religion, or their lack. This includes texts written in the third person that do not describe the speaker's personal beliefs or family environment.

First, write out your reasoning for classifying the text inside <reasoning> tags. Consider the content and tone of the text, and how it aligns with the definitions provided above.

After writing your reasoning, output your final classification as a single word (POSITIVE, NEGATIVE, AMBIGUOUS, or NONE) inside <classification> tags.
Use HTML tags in your response.
Do not add any words after </classification>.

**Prompt for Practice Classification**

Your task is to carefully read this text and determine the speaker's valence of Jewish religious practice described in the text, if any, based on the following classification system:

1. ACTIVE = The text expresses the narrator actively practicing a Jewish religious ritual.
2. INACTIVE = The text expresses the narrator violating Jewish religious practices or not observing/actively not practicing a Jewish religious ritual.
3. AMBIGUOUS = The narrator of the text expresses a Jewish religious practice, that does not meet the criteria of the classes ACTIVE or INACTIVE, or the text matches both of the classes at the same time.
4. NONE = The text does not directly discuss the speaker participating in a religious practice or violating one. This includes texts written in the third person that do not describe the speaker's personal valence of practicing religion or family environment.

First, write out your reasoning for classifying the text inside <reasoning> tags. Consider the content and tone of the text, and how it aligns with the definitions provided above.
After writing your reasoning, output your final classification as a single word (ACTIVE, INACTIVE, AMBIGUOUS, or NONE) inside <classification> tags.
Use HTML tags in your response.
Do not add any words after </classification>.

## E  Topic Model Based Evaluation

The topics we address, and their matching categories are in Table 4.

| Topic | Class | Sub-Class |
|-------|-------|-----------|
| Synagogue, holiday(s), Shabbos, religious, Shabbat, Shul, Friday, Passover, Jewish | practice | - |
| Bar, Mitzvah, Torah, Synagogue, Mitzvahed, Shul, Rabbi, religious | practice | - |
| God, believe, religion/religious, faith, question | belief | - |
| Catholic, church, priest, baptized, communion, religion, Catholicism, convert, prayers | practice | inactive |

Table 4: Practice and belief-related topics from (Ifergan et al., 2024), and a partial list of the words they contain.

## F  Thesaurus Based Evaluation

List of terms from the thesaurus that align with our definitions for religious content:

- **Active Practice**
  ritual circumcision (bio) mohelim Jewish religious observances Jewish schools synagogue attendance mikva'ot Yahrzeit Jewish dietary laws ghetto Jewish religious observances camp Jewish religious observances refugee camp Jewish religious observances hiding-related Jewish religious observances Islamic prayers yizkor Kaddish prison Jewish religious observances Jewish mourning customs yeshivot forced labor battalion Jewish religious observances Hasidic rebbes (bio) synagogue organizations (bio) Baalei Keriah (bio) Baalei Tefillah (bio) synagogues' sisterhood (bio) synagogues' men's clubs deportation Jewish religious observances Jewish Theological Seminary of America transfer Jewish religious observances Beth Jacob schools b'nai mitzvah b'nai mitzvah (stills) Borerim Jewish religious observances (stills) Institute of Jewish Studies shamas Jewish Institute of Religion Mitnagdim observant/practicing Ramah Camping Movement Shema Yisrael

- **Inactive Practice**
  Christian religious observances church attendance religious identity communions (stills) confirmations (stills) Jehovah's Witness missionary activities Jehovah's Witness religious observances Mormon missionary activities Jehovah's Witness religious beliefs camp Jehovah's Witness religious observances baptisms forced labor battalion Jehovah's Witness religious observances camp Christian religious observances Eucharist Islamic religious observances baptisms (stills) prison Jehovah's Witness religious observances ghetto Christian religious observances Christian religious observances (stills) confirmations Islamic dietary laws Seventh-Day Adventist missionary activities non-observant/non-practicing Buddhist religious observances Buddhist lunar days karma Christian missionary activities Christian prayers

- **Practice - Other**
  rabbis

- **Positive Belief**
  prayers Jewish prayers camp Jewish prayers Jewish religious beliefs prison Jewish prayers forced labor battalion Jewish prayers camp prayers ghetto prayers deportation Jewish prayers hiding-related prayers forced march Jewish prayers

- **Negative Belief**
  Christian religious beliefs camp Jehovah's Witness prayers Jehovah's Witness prayers Islamic identity Buddhist religious beliefs Islamic religious beliefs Armenian Genocide faith issues Bosnian War and Genocide faith issues Guatemalan Genocide faith issues Holocaust faith issues Rwandan Tutsi Genocide faith issues

## G   Evaluation on a Subset of Trajectories

Evaluating a subset of the predicted trajectories presents similar trends to the evaluation in the complete dataset, results in Tables 5, 6 and 7.

## H   Hierarchical Clustering

**Belief Clustering Example**   An example from running HDBSCAN on the belief trajectories: 16 We first truncated the values of the position points of the trajectories to two decimal points, then calulated their distance matrix using dtw_ndim.distance_matrix from the *dtaidistance* Python module, with window=7 and the default values for the rest of the hyper-parapeters; followed by calling HDBSCAN from the *hdbscan* module, with: min_cluster_size=30, min_samples=1, cluster_selection_epsilon=1 and alpha=1.

**Practice Clustering Example**   An example from running HDBSCAN on the practice trajectories: 10 Hyper-parameters:                window=6, min_cluster_size=30,    min_samples=1,    cluster_selection_epsilon=1 and alpha=0.95.

|  | **Topic** | | | **Thesaurus** | | | | |
|---|---|---|---|---|---|---|---|---|
|  | **B** | **P** | **P⁻** | **B** | **P** | **P⁺** | **B⁺** | **B⁻** |
| Predicted | **3.92** | **32.53** | **18.46** | **6.74** | **21.87** | **25.64** | **2.74** | **8.06** |
| Original scatter | 5.60 | 33.33 | *17.99* | 8.22 | 23.39 | 30.81 | 4.80 | 8.92 |
| Edges & middle | 5.89 | 33.33 | 19.64 | 8.90 | 23.72 | 26.82 | 5.26 | 9.69 |
| Gaussian | 6.62 | 41.80 | 19.53 | 8.89 | 29.57 | 34.95 | 3.19 | 9.71 |
| G- Edges & middle | 7.07 | 41.30 | 20.47 | 9.65 | 29.05 | 33.70 | 5.17 | 9.54 |
| Equal scatter | 7.31 | 33.15 | 19.81 | 10.47 | 23.07 | 27.89 | 3.06 | 14.0 |
| Normal-original | 8.40 | 33.50 | 17.92 | 10.58 | 23.88 | 25.39 | 2.89 | 13.08 |
| # Reference paths | 38 | 91 | 26 | 31 | 79 | 76 | 8 | 22 |
| # predicted paths | 71 | 88 | 60 | 71 | 88 | 81 | 56 | 35 |
| # Reference points | 50 | 264 | 40 | 48 | 233 | 206 | 19 | 27 |
| # predicted points | 368 | 1279 | 157 | 368 | 1279 | 639 | 189 | 80 |

Table 5: Sum of $min\_sum\_dist$ for GPT-4o on a subset of trajectories

|  | **Topic** | | | **Thesaurus** | | | | |
|---|---|---|---|---|---|---|---|---|
|  | **B** | **P** | **P⁻** | **B** | **P** | **P⁺** | **B⁺** | **B⁻** |
| Predicted | **2.81** | **0.45** | **1.87** | **0.93** | **0.34** | **0.68** | **2.23** | **10.35** |
| Original scatter | 5.27 | 1.22 | *3.38* | 3.71 | 2.33 | 1.84 | 2.36 | 13.00 |
| Edges & middle | 5.00 | 2.16 | 2.77 | 2.53 | 2.65 | 2.32 | 4.67 | 12.31 |
| Gaussian | 4.13 | 7.07 | 3.03 | 3.02 | 5.81 | 6.10 | 2.38 | 13.38 |
| G- Edges & middle | 5.96 | 7.47 | 4.25 | 4.53 | 6.07 | 6.56 | 4.71 | 12.44 |
| Equal scatter | 3.07 | 0.90 | 3.56 | 2.36 | 0.95 | 1.36 | 1.98 | 16.41 |
| Normal-original | 3.89 | 3.80 | 2.89 | 2.68 | 3.43 | 3.95 | 3.89 | 12.81 |
| # Reference paths | 38 | 91 | 26 | 31 | 79 | 76 | 8 | 22 |
| # predicted paths | 93 | 101 | 86 | 93 | 101 | 101 | 63 | 29 |
| # Reference points | 50 | 264 | 40 | 48 | 233 | 206 | 19 | 27 |
| # predicted points | 507 | 2967 | 328 | 507 | 2967 | 2385 | 206 | 54 |

Table 6: Sum of $min\_sum\_dist$ of fine-tuned Mistral7B on a subset of trajectories, prompting all content

|  | **Topic** | | | **Thesaurus** | | | | |
|---|---|---|---|---|---|---|---|---|
|  | **B** | **P** | **P⁻** | **B** | **P** | **P⁺** | **B⁺** | **B⁻** |
| Predicted | **3.62** | **30.09** | **17.38** | **5.68** | **19.62** | **15.78** | **3.66** | **12.35** |
| Original scatter | 5.20 | 33.23 | *18.34* | 9.20 | 23.99 | 18.53 | **3.39** | 15.70 |
| Edges & middle | 4.86 | 31.78 | 18.47 | 7.33 | 22.37 | 18.47 | 6.02 | 14.26 |
| Gaussian | 5.65 | 36.84 | 19.29 | 8.40 | 25.58 | 21.22 | 4.14 | 14.83 |
| G- Edges & middle | 6.16 | 38.46 | 19.67 | 9.16 | 27.13 | 23.73 | 6.09 | 14.02 |
| Equal scatter | 4.04 | 31.45 | 18.93 | 8.22 | 21.14 | 17.43 | 3.75 | 17.99 |
| Normal-original | 4.23 | 33.91 | 19.26 | 8.13 | 23.51 |  | 5.53 | 14.65 |
| # Reference paths | 38 | 91 | 26 | 31 | 79 | 76 | 8 | 22 |
| # predicted paths | 77 | 88 | 69 | 77 | 88 | 86 | 51 | 26 |
| # Reference points | 50 | 264 | 40 | 48 | 233 | 206 | 19 | 27 |
| # predicted points | 391 | 1816 | 199 | 391 | 1816 | 1468 | 198 | 46 |

Table 7: Sum of $min\_sum\_dist$ of fine-tuned Mistral7B on a subset of trajectories, prompting filtered content
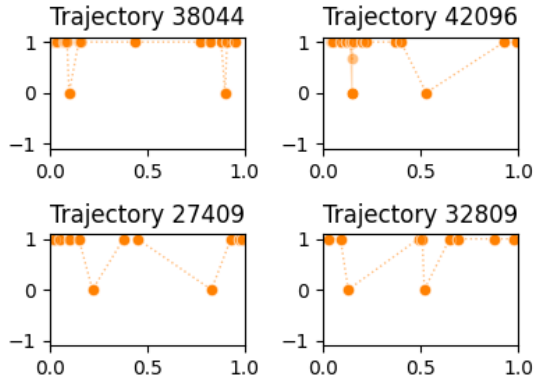
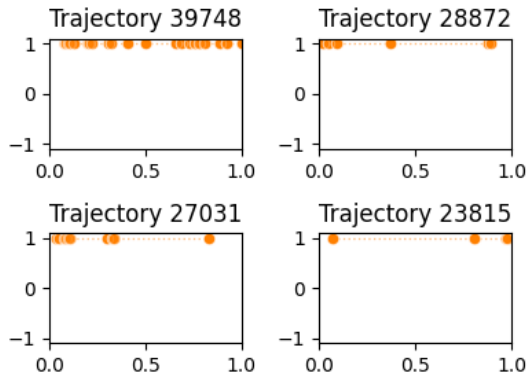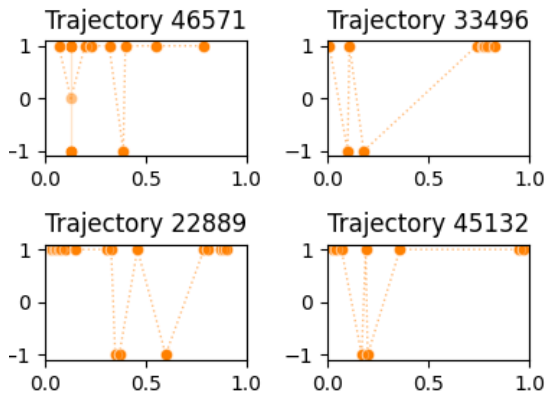Figure 7: (a) Active - Med



Figure 8: (b) Active



Figure 9: (d) Oscillating

Figure 10: Overall figure caption
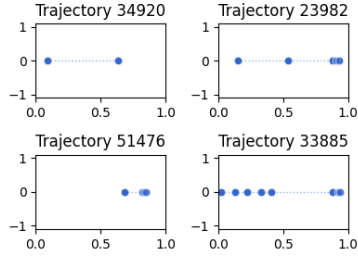
Figure 11: (a) Neutral



Figure 12: (b) Ascending from neutral
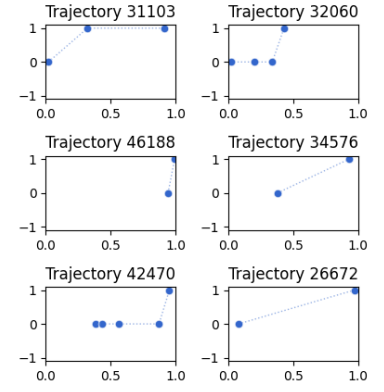


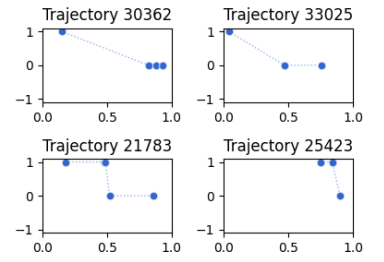Figure 13: (d) Positive



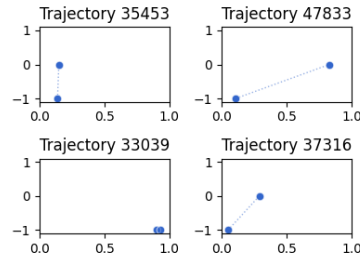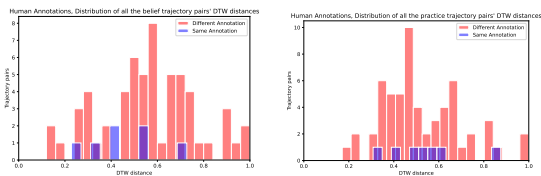Figure 14: (e) Descending Positive Neutral



Figure 15: (g) Negative to Neutral

Figure 16: Belief cluster examples



Figure 17: DTW distributions