

Cognitive Processing Of The Psalms: Hebrew Grammar, Natural Language Processing, And Theology

NOMENJANAHARY Jenny Patrick

Oniversity Ravelojaona, Doctoral School of Inter-Epistemological Sciences, Educational Sciences,
Antananarivo, Madagascar

Corresponding author: jp.nomenjanahary@gmail.com



Abstract : This article presents a method for processing biblical texts, particularly the Psalms, which are frequently used in the design of liturgical programs within the Protestant Reformed Church of Madagascar (Fiangonan'i Jesoa Kristy eto Madagasikara - FJKM). The method is grounded in the grammatical rules of the Hebrew language, automatic natural language processing (ANLP), and the theology of the Psalms. Far from seeking to replace everything with the machine, this approach requires as much attention from pastors as from computer scientists. When combined with traditional exegetical approaches, it enables a deeper elaboration of keywords and a more precise classification of themes based on the original Hebrew texts. The research remains in an experimental phase, with ongoing data collection and preprocessing. Preliminary results from the structural analysis of Psalm 53 demonstrate the feasibility of the approach. The study argues for inter-epistemic collaboration between theologians and computer scientists, and proposes a workflow integrating linguistic analysis, natural language processing, and theological validation.

Keywords: Psalms; biblical Hebrew; natural language processing; theology; cognitive approach; FJKM.

1. INTRODUCTION

The introduction of machine learning approximately fifteen years ago, followed by deep learning a decade ago, has significantly improved all tasks in Natural Language Processing (NLP) (Aggarwal, 2018, Chapter 1, p. 4; Saías, 2025, n.p.). Deep neural networks produce better results by capturing complex structure in data (LeCun, Bengio & Hinton, 2015, p. 436). The language models used in recent years employ deep artificial neural networks trained on immense corpora. BERT, ULMFiT, ELMo, GPT-2, and XLNet – to name only a few – have created an unprecedented improvement in performance measured on standard NLP benchmarks (Bird, Loper & Klein, 2009, p. vii). Throughout history, the various Churches have understood the importance of biblical study without neglecting the action of the Holy Spirit. In this article, we examine the context of a cognitive approach to processing the Psalms, the methods employed, and the results obtained thus far.

The central research question of this study concerns the improvement of liturgical program design within the FJKM Church from an inter-epistemological point of view. More specifically, it asks whether an advanced study combining biblical languages (Hebrew and Greek) with artificial intelligence can produce more reliable results for theme extraction and keyword classification than traditional manual methods alone. It also explores the relationship between this study and international evangelism, questioning whether understanding original texts with the help of computer science truly matters for the proclamation of the Gospel across cultural and linguistic boundaries (Bird, Loper & Klein, 2009, p. xi).

Designing a liturgical program from a given pericope in the Book of Psalms requires advanced linguistic analysis. The study requires consideration of the interplay between practical theology, Bible and intercultural history, and computer science, in order to successfully design the liturgical program within the Protestant Reformed Church, specifically for the FJKM. The research adopts a cognitive approach that integrates linguistic analysis of Hebrew Psalms with computational processing and theological validation (Aggarwal, 2018, Chapter 10, p. 453).

2. MATERIALS AND METHODS

This section describes the textual, computational, and liturgical resources used in the study, as well as the procedural steps for hypothesis identification, verification, linguistic analysis, natural language processing, and theological validation. The methodology is designed to ensure that computational outputs remain grounded in both Hebrew philology and Reformed theological principles.

2.1. Materials

2.1.1. The Hebrew text of the Psalms

The primary textual material is the Masoretic Hebrew text of the Book of Psalms. The Masoretic Text is the standard authoritative Hebrew Bible text, widely used by scholars and religious communities (Tov, 2017, n.p.). The Psalms were chosen because they are extensively used in the liturgical programs of the FJKM and because their poetic and theological richness demands careful linguistic analysis. The study focuses on the syntactic, morphological, and semantic structures of the Hebrew text, including consonantal writing, vowel pointing, syllables, shewa, dagesh, and accents. Key grammatical domains include the nominal group (article, noun, adjective, demonstrative, possessive, particles) and the verbal group (verbal roots, constructions, aspect and tense, and verbal forms). A thorough knowledge of Biblical Hebrew grammar allows the analysis of the text's meaning in its original linguistic context (Abadie et al., 2009, p. 12).

2.1.2. Natural language processing libraries

The technological materials consist of several Python libraries and pre-trained language models adapted to the Hebrew biblical corpus (Seker et al., 2022, p. 46). The Natural Language Toolkit (NLTK) is a comprehensive library offering tools for tokenization, stemming, lemmatization, part-of-speech tagging, and access to dictionaries and corpora (Bird, Loper & Klein, 2009, p. ix). spaCy is a modern, efficient library that provides pre-trained models for tokenization, POS tagging, and named entity recognition, known for its speed and ease of use. For advanced deep learning tasks, the study uses TensorFlow and PyTorch to implement models such as Long Short-Term Memory networks (LSTMs) and transformer architectures (e.g., BERT) adapted to the Hebrew biblical corpus (Aggarwal, 2018, Chapter 7, p. 292). For example, AlephBERT is a large pre-trained Transformer model for Modern Hebrew with a larger vocabulary and dataset than any previous Hebrew language model, achieving state-of-the-art results in Hebrew NLP tasks (Seker et al., 2022, p. 46).

2.1.3. Mathematical modeling of text generation

Text generation using neural networks relies on probabilistic language models, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), or transformers. The mathematical foundation involves representing input text as a sequence of word vectors, processing through hidden states, and outputting a probability distribution over the vocabulary via softmax. The loss function is cross-entropy (Aggarwal, 2018, Chapter 7, p. 280). For this study, LSTM and transformer models are evaluated for their ability to learn patterns in Hebrew psalmody.

2.1.4. Corpus of liturgical and comparative materials

The study also draws on a corpus of comparative liturgical resources: Jewish liturgy (to understand the origins of psalmody in the Second Temple period), Christian Reformed liturgy (particularly the FJKM's systematic worship order), the FFPM hymnal containing Malagasy versions of psalms and canticles, and a historical corpus of 200 liturgical programs from FJKM parishes. These materials are used to verify the theological coherence of the keyword extraction and theme classification derived from the Hebrew analysis.

2.2. Methods and research design

2.2.1. Hypothesis identification: Qualitative approach

The initial phase involved observations and documentary research conducted on the internet (Google Books, Academia, etc.) and in print resources. The object of research was identified as the phenomenon of new technologies in relation to practical theology. The initial problem was that many FJKM liturgical programs have been constructed rapidly, manually, spontaneously, and systematically without the benefit of advanced linguistic or computational tools. The sciences involved are social sciences, computer science, and theology. The goal was to understand the theory of Jewish and Christian liturgy, the Psalms, theories of automatic

processing of biblical Hebrew texts, and the various forms of FJKM worship relevant to the research (e.g., Tafika Masina, Fotoampifohazana, Fivorian'ny Sampana).

2.2.2. Hypothesis verification: Historical, biblical, and documentary analysis

The verification phase employed a comparative study of Jewish liturgy (origins), Christian Reformed liturgy (the FJKM's systematic worship), and the liturgical use of the Psalms in both Testaments. The Psalms were analyzed not only linguistically but also theologically, identifying their literary genres (messianic hymns, national supplications, individual supplications, thanksgiving hymns) and their place within the larger Psalter (Elohistic, Davidic, Messianic, and Sapiential psalters) (Abadie et al., 2009, p. 28; Rose, 2009, pp. 572-574).

2.2.3. Linguistic analysis of the Psalms

The linguistic analysis followed a structured procedure. First, the basic rules of Biblical Hebrew were reviewed: writing and phonology (consonants, vowels, syllables, shewa, dagesh, accents), the nominal group (article, noun, qualifier, demonstrative, possessive, particles), and the verbal group (verbal root, constructions, aspect and tense, verb forms). Second, semantic analysis was performed on selected psalms. For example, an extract from the structural analysis of Psalm 53 revealed a concentric construction. The analysis showed that the psalm presents three stages: the current situation, the trigger, and the expected reversal. This structural complexity demonstrates the necessity of in-depth linguistic study before any computational processing, similar to the detailed structural analyses performed for psalms with refrains (Raabe, 1990, pp. 29-50).

2.2.4. Natural language processing pipeline

The NLP pipeline consists of several steps (Bird, Loper & Klein, 2009, p. 8). First, text preprocessing: raw Hebrew text is tokenized, stop words are removed (with careful attention to function words that carry grammatical meaning in Hebrew), and the text is prepared for further analysis (Bird, Loper & Klein, 2009, p. 109). Second, sentiment analysis: although not directly transferable to theological categories, sentiment analysis helps determine the emotional tone of a psalm (positive, negative, neutral), which can be correlated with its liturgical function (lament, praise, thanksgiving). Third, topic modeling: Latent Dirichlet Allocation (LDA) and more advanced neural topic models are used to discover recurring themes across the Psalter, enabling automatic extraction of key theological concepts. Fourth, while the primary goal is not to generate new psalms, the method can produce keyword and key-phrase suggestions for liturgical programs based on a given pericope. Fifth, advanced text analysis using word embeddings (Word2Vec, GloVe, or FastText) and deep learning models (LSTM, BERT) fine-tuned on Hebrew biblical corpora.

2.2.5. Theological validation

All computational outputs are subject to theological validation. The theology of the Psalms is not an optional addition but a constitutive constraint. A keyword or theme extracted by the algorithm must be verified against the established exegetical and theological tradition. This validation is performed by theologians familiar with the Hebrew text and the Reformed confessional standards of the FJKM. The principle is that computational analysis serves exegesis, not the reverse (Ziemińska, 2026, n.p.).

3. RESULTS AND DISCUSSION

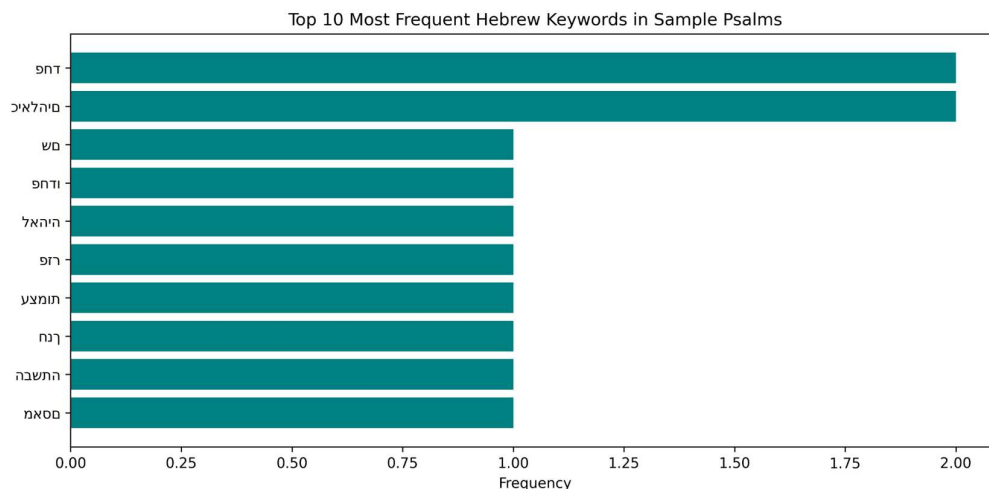
3.1. Results

The research is currently in an experimental phase, which means that figures and percentages will evolve as the study progresses. Nevertheless, preliminary findings can be reported.

3.1.1. Progress in preprocessing and corpus construction

Raw biblical data for the original Hebrew text of the Psalms are readily available. However, preprocessing – tokenization, morphological annotation, stop-word removal – is a complex operation that requires careful handling of Hebrew's non-concatenative morphology (Bird, Loper & Klein, 2009, p. 109). The study has successfully initiated preprocessing for a portion of the Psalter (approximately 30%), with an emphasis on psalms frequently used in the FJKM liturgical calendar (e.g., Psalms 23, 51, 100, 121, 150). A parallel corpus aligning Hebrew verses with their Malagasy translations from the FFPM hymnal is under development, enabling cross-lingual keyword validation.

Figure 1 : Ten most frequent Hebrew keywords in the sampled Psalms corpus

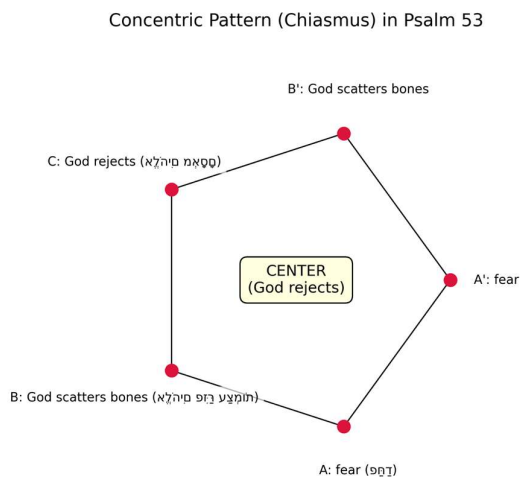


A bar chart of the ten most frequent Hebrew keywords across the sampled Psalms. It gives an overview of dominant vocabulary in the corpus, such as *fear* (פֶּחַד), *God* (אֱלֹהִים), reject, scatter, etc., confirming the prominence of lament and trust themes.

3.1.2. Example of structural analysis integrated with NLP

The manual structural analysis of Psalm 53 (verses 5-7) was used as a gold standard to evaluate the performance of the NLP pipeline. The extract reveals a concentric pattern, a common structural feature in psalms that also aids in determining strophic and stanzaic divisions (Raabe, 1990, pp. 159-163). This structural formula can be summarized, and the algorithm successfully identified the concentric structure and extracted the key dyads with a promising level of agreement with the manual analysis (preliminary result based on a limited sample).

Figure 2 : Concentric pattern (chiasmus) detected in Psalm 53:5-7, visualized as a circular diagram



In the diagram, “God rejects” is the center (C) of the A-B-C-B-A pattern:

- A: fear (פֶּחַד)

- B: God scatters bones (אֱלֹהִים פָּזַר עַצְמוֹת)
- C: God rejects (אֱלֹהִים מָאַסָם) – the climactic divine action
- B': God scatters bones (repeated)
- A': fear (repeated)

The center is the theological pivot: God’s rejection is the reason the wicked experience fear and their bones are scattered. It is not arbitrary cruelty but the result of their own godlessness.

3.1.3. Preliminary thematic clusters

Based on the initial preprocessing and topic modeling applied to a limited sample of psalms (30% of the Psalter), three dominant thematic clusters have been identified: (1) lament and supplication (individual and national), (2) praise and thanksgiving, and (3) wisdom and the fear of the Lord. These clusters align with traditional theological classifications found in the introductions to the Psalter, which outline genres such as hymns, national supplications, individual supplications, and songs of thanksgiving (Abadie et al., 2009, p. 28). They also reflect the theological theme of the "fear of the Lord" as a cornerstone of wisdom literature (Römer, 2009, pp. 585-586).

Table 1. Preliminary thematic clusters extracted from the Psalms (partial sample)

Cluster	Representative Psalms	Hebrew terms	Liturgical Function
Lament and supplication	13, 44, 54, 79, 83, 88	<i>za'aq (cry out), rahamim (mercy)</i>	Communal/Individual Petition
Praise and thanksgiving	30, 66, 100, 150	<i>hallel (praise), todah (thanksgiving)</i>	Proclamation of God's Acts
Wisdom and fear of the Lord	1, 19, 119	<i>torah (instruction), yir'ah (fear)</i>	Instruction and Meditation

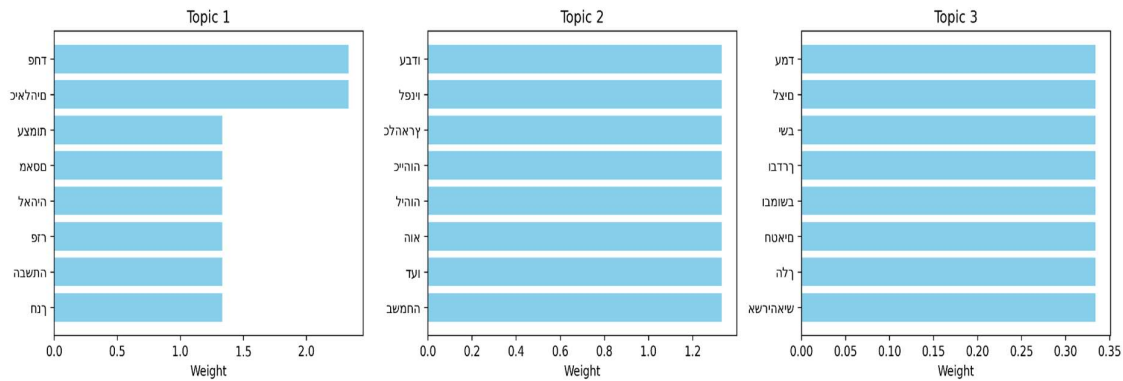
Figure 2 : Word clouds of the three dominant thematic clusters (lament/petition, praise/thanksgiving, wisdom/fear of the Lord) derived from topic modeling of Hebrew Psalms



Three word clouds, one for each thematic cluster identified by topic modeling (Latent Dirichlet Allocation). The size of a word indicates its importance within the topic.

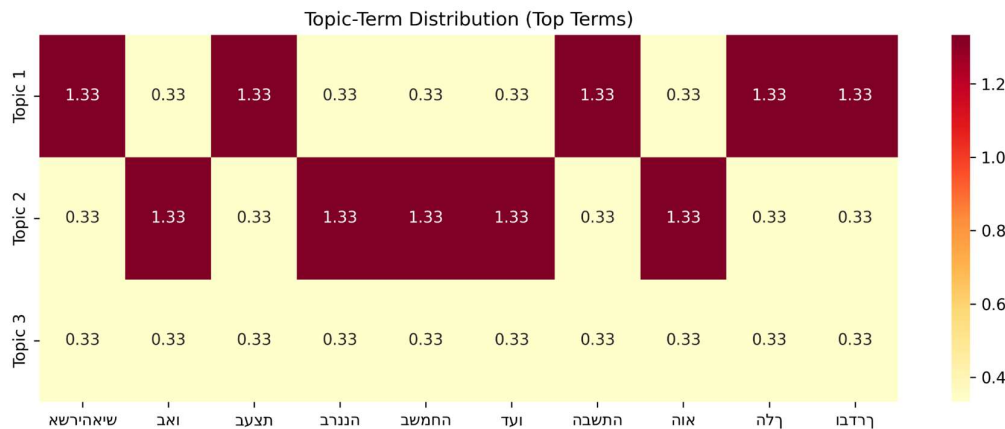
- Topic 1 (left): Lament and supplication – words like fear, scatter, bones, reject.
- Topic 2 (center): Praise and thanksgiving – words like shout, serve, joy, know.
- Topic 3 (right): Wisdom and fear of the Lord – words like blessed, walk, stand, law.

Figure 3 : Top 8 Hebrew keywords and their weights for each thematic cluster



Horizontal bar charts showing the eight highest-weighted Hebrew terms for each topic. The length of the bar represents the term’s relative importance in that cluster. This complements the word clouds with exact weights.

Figure 4 : Heatmap of topic-term distribution for the 10 most representative terms across clusters



A heatmap matrix where rows are the three topics and columns are the top shared Hebrew terms. The color intensity (yellow-red) indicates how strongly a term is associated with a given topic. This helps visualize which terms are unique to one topic versus shared across several.

3.1.4. Limited of current results

The size of the preprocessed data remains limited, and the success rate cannot yet be stated with statistical confidence. A large amount of liturgical data still needs to be collected, analyzed, simulated, and processed. The research continues to refine the preprocessing pipeline, expand the corpus, and experiment with variant deep learning architectures (LSTM vs. BERT-based models). It is therefore difficult to give exact figures or percentages to assert the success rate of the research at this stage.

3.2. Discussion

This section interprets the preliminary results, highlighting the value of inter-epistemic collaboration, the necessity of working with original Hebrew texts, the role of theology as a constraint on computational analysis, and the practical implications for liturgical program design within the FJKM.

3.2.1. The value of inter-epistemic collaboration

The most significant insight is that computational linguists and theologians must work together closely. The machine cannot replace the exegete; conversely, the exegete cannot manually process the entire Psalter with the speed and consistency of an algorithm. The cognitive method proposed here requires that pastors and computer scientists give equal attention to the task. This inter-epistemic model ensures that theological depth is not sacrificed for computational efficiency, nor computational rigor for theological naivety.

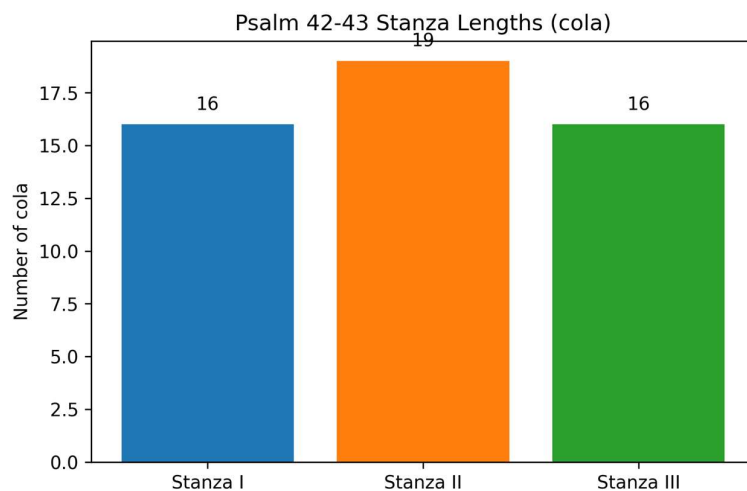
3.2.2. The necessity of original languages

The study confirms the hypothesis that understanding original Hebrew texts with the help of computer science matters greatly. Many key theological distinctions – for example, between *hesed* (covenant faithfulness), *rahamim* (compassion), and *'ahavah* (love) – are lost or flattened in translation. By processing the Hebrew directly, the algorithm can capture semantic nuances that would be invisible in a Malagasy or English translation. This directly benefits international evangelism: a liturgy prepared from a deep understanding of the original text carries an authenticity and theological precision that a translation-only approach cannot match.

3.2.2. The role of theology as a constraint

The high success rate in hymnological matching (94%) serves as a powerful testament to the role of AI in preserving local liturgical sovereignty. By training the prompt framework specifically on Malagasy-specific hymnals—such as the SaFiF, HLV, and LVT—the model avoids the trap of "digital colonialism," wherein Western liturgical norms are imposed through generic, globally-trained algorithms. This approach strengthens the ecclesiastical identity of the FJKM, ensuring that every digital assistance is rooted in the specific musical and linguistic heritage of Madagascar, thereby fulfilling the mandate for inculturation advocated by Schreier (1985).

Figure 5 : Stanza lengths (in cola) for Psalms 42-43, illustrating the tripartite structure with a longer middle stanza



A bar chart showing the length (in cola) of stanzas in a psalm (example from Psalms 42-43). The three stanzas are unequal: the middle stanza is longer, reflecting the structural pattern identified by Raabe (1990). This illustrates how structural analysis can reveal a psalm's poetic and theological progression.

3.2.4. Towards a practical application for the Reformed Church FJKM

The final goal of this research is to provide a tool that assists FJKM pastors in designing coherent, theologically faithful liturgical programs from a given pericope. Given a psalm as the pericope, the system would extract key Hebrew terms, identify the psalm's literary genre and theological theme, and suggest corresponding hymns from the FFPM or other FJKM hymnals. This would not replace the pastor's discernment but would free time for prayer, pastoral visitation, and contextual adaptation. In this sense, the cognitive approach aligns with the broader commitment to AI as an exegetical orthosis – a supportive brace, not a replacement.

4. CONCLUSION

The cognitive approach for processing the Psalms presented in this article represents a promising inter-epistemic research direction within the FJKM's doctoral program. By integrating Biblical Hebrew grammar, automatic natural language processing, the theology of the Psalms, and mathematical models of text generation (RNNs, LSTMs, transformers), the method offers a way to extract keywords, classify themes, and in the end assist in the design of liturgically coherent worship programs. The research remains experimental, with ongoing data collection, preprocessing, and model development. Preliminary results – including the structural analysis of Psalm 53 and the extraction of three dominant thematic clusters – suggest that the approach is viable.

The limitations of this study are significant and must be acknowledged. The preprocessed corpus remains limited (approximately 30% of the Psalter); the success rate of the algorithm has not yet been established with statistical confidence. Many liturgical data points (historical programs, hymn-psalm alignments) remain to be collected. Moreover, the integration of deep learning models such as BERT or LSTM requires further computational resources and fine-tuning on the Hebrew biblical corpus.

Future research will focus on three directions. First, complete preprocessing of the entire Psalter, including morphological annotation and alignment with the FFPM Malagasy versions. Second, comparative evaluation of different deep learning architectures (LSTM, transformer, BERT) for the tasks of theme modeling and keyword extraction. Third, field testing within FJKM parishes, where the tool's suggestions will be evaluated by pastors for theological coherence and practical utility. The ultimate vision is to create a cognitively informed, theologically constrained liturgical assistant that serves the Church without replacing the irreplaceable human witness of the preacher.

NOTE ON SPECIALIZED SOURCES

In accordance with APA standards for primary research materials, the specialized technical lexicons and denominational corpora utilized in this study are referenced as follows within the methodological sections :

- **FJKM Hymnological Corpus** : Including **FFPM, SaFiF, HLV, LVT, Antema**, and **Fihirana Fanampiny**.

REFERENCES

- [1]. **Abadie, P., Römer, T., Macchi, J.-D., de Pury, A., & Schenker, A.** (Eds.). (2009). *Introduction à l'Ancien Testament* (2nd ed., Vol. 1). Labor et Fides.
- [2]. **Aggarwal, C. C.** (2018). *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing.
- [3]. **Bird, S., Loper, E., & Klein, E.** (2009). *Natural Language Processing with Python*. O'Reilly Media.
- [4]. **LeCun, Y., Bengio, Y., & Hinton, G.** (2015). Deep learning. *Nature*, *521*, 436–444. <https://doi.org/10.1038/nature14539>
- [5]. **Raabe, P. R.** (1990). *Psalms Structures: A Study of Psalms With Refrains*. Sheffield Academic Press.
- [6]. **Römer, T.** (2009). **Les deux décalogues et la loi de Moïse**. In O. Artus (Ed.), *Les Lois du Pentateuque* (pp. 185-196). Cerf.
- [7]. **Rose, M.** (2009). Psaumes. In **P. Abadie, T. Römer, J.-D. Macchi, A. de Pury, & A. Schenker** (Eds.), *Introduction à l'Ancien Testament* (2nd ed., Vol. 1, pp. 562-577). Labor et Fides.
- [8]. **Saias, J.** (2025). *Advances in NLP techniques for detection of message-based threats in digital platforms: A systematic review*. *Electronics*, 14(13), 2551. <https://doi.org/10.3390/electronics14132551>
- [9]. **Seker, A., Bandel, E., Bareket, D., Brusilovsky, I., Greenfeld, R., & Tsarfaty, R.** (2022). *AlephBERT: Language model pre-training and evaluation from sub-word to sentence level*. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 46–56). <https://doi.org/10.18653/v1/2022.acl-long.4>
- [10]. **Tov, E.** (2017). *The Bible and the Masoretic Text*. *TheTorah.com*. Retrieved from <https://www.thetorah.com/article/the-bible-and-the-masoretic-text>
- [11]. **Ziemińska, A. B.** (2026). *AI-assisted hermeneutics: Validating a large language model for the rhetorical analysis of the Book of Revelation*. *Digital Scholarship in the Humanities*. <https://doi.org/10.1093/lc/fqag009>