ALBERTIR: A BERT-Based Pretraining for Indonesian Religious Texts Using Qur'an and Hadith Translations

Irwan Darmawan
Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia
irwan.darmawan.2305349@students.um.ac.id

Hakkun Elmunsyah
Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia
hakkun@um.ac.id (corresponding author)

Didik Dwi Prasetya
Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia
didikdwi@um.ac.id

Received: 27 June 2025 | Revised: 12 July 2025, 21 July 2025, and 24 July 2025 | Accepted: 27 July 2025

 DOI: https://doi.org/10.48084/etasr.12977

ABSTRACT

This study introduces Al-Qur'an BERT for Indonesian Religious Texts (ALBERTIR), a domain-adaptive Bidirectional Encoder Representations from Transformers (BERT) model pretrained on Indonesian religious texts, including official Qur'an and Hadith translations. The corpus comprises over 1.2 million tokens sourced from verified government publications and optimized for Masked Language Modeling (MLM). ALBERTIR features weighted MLM, sacred term preservation, and factorized embeddings to enhance understanding of religious semantics and maintain doctrinal integrity. Training was conducted on Google Colab Pro with TPU v3-8, where ALBERTIR outperformed BERT-base and A Lite BERT for Indonesian (ALBERT-ID), improving religious term prediction by 10.9% and reducing training time by more than 40%. Across downstream tasks such as religious question answering, sentiment analysis, and text classification, it achieved up to 8% higher F1-scores. Ablation studies confirmed the effectiveness of its core components, demonstrating advantages in semantic accuracy, contextual sensitivity, and reliability in religious Natural Language Processing (NLP) applications. Unlike general-purpose models like Indonesian BERT (IndoBERT) and multilingual BERT (mBERT), the proposed model is specifically optimized for theological language, thereby reducing vague or contextually inappropriate outputs. This makes it especially suitable for applications such as fatwa retrieval, Islamic education tools, and religious chatbot systems. Cross-lingual evaluations further showed that ALBERTIR surpasses mBERT by +13.3 Bilingual Evaluation Understudy (BLEU)-4 points in religious Questioning-Answering (QA) tasks, while maintaining competitive performance in general benchmarks. Ablation results identified sacred term preservation as the most critical contributor to accuracy gains, underscoring the importance of domain-specific features. Overall, ALBERTIR demonstrates strong capabilities in capturing linguistic precision and theological nuance, establishing a robust foundation for future religious NLP research and applications.

Keywords-Bidirectional Encoder Representations from Transformers (BERT); domain adaptation; religious Natural Language Processing (NLP); Qur'an translation; Hadith; Indonesian language

I. INTRODUCTION

Application of Artificial Intelligence (AI) in processing and preserving religious texts has recently gained scientific attention. Author in [1], for example, demonstrated how AI techniques can preserve sentiment and semantic precision in Qur'an translation, a critical requirement for ensuring accurate religious interpretation. Similarly, authors in [2] observed that Large Language Models (LLMs) often fail to accommodate Arabic dialects, underscoring the need for greater linguistic and geographical diversity.

One of the most significant breakthroughs in Natural Language Processing (NLP) is the Bidirectional Encoder Representations from Transformers (BERT) model, which leverages bidirectional context to generate deep semantic understanding, and its success across diverse tasks, including sentiment analysis, temporal relation classification, and processing morphologically rich or low-resource languages, demonstrates both robustness and adaptability. For instance, fine-tuned BERT and ChatGPT demonstrate complementary strengths in intelligent system development [3]. Other works include enhanced Arabic sentiment analysis through a hybrid ArabBERT-LSTM model [4] and reviews of BERT and LLMs for suicide detection, illustrating their applicability in sensitive domains [5]. Beyond medicine, BERT has proven valuable in educational and academic contexts. Authors in [4] demonstrated improved scholarly knowledge retrieval through BERT-based expert discovery, while authors in [5] employed transformer models to detect AI-generated content in academic writing, an increasingly relevant challenge in the age of generative models. Furthermore, it has supported COVID-19 conversation analysis on Twitter [6], improved sentiment analysis performance across several benchmarks [7], and aided structured data extraction from medical narratives [8]. Transformer models have also been used for subject indexing of large text collections, Arabic text classification [9], and religious text analysis, including CL-AraBERT for Qur'an understanding [10]. Further applications include aspect-based sentiment classification in Arabic [11], fake news detection in Indonesian social media [12], pharmaceutical document classification [13], topic modeling [14], scientific article classification [15], and document classification [16]. These examples emphasize BERT's role as a versatile model across both applied and scientific contexts.

Additionally, BERT has been especially useful to tackle linguistically and culturally challenging tasks in the Indonesian NLP ecosystem. Authors in [17, 18] showed significant improvements over traditional methods in emotion recognition using BERT-based models on Indonesian datasets. Authors in [19] applied hybrid approaches with BERT embeddings for sentiment analysis, achieving excellent classification performance, while authors in [20] fine-tuned BERT to classify exam questions according to Bloom's Taxonomy, a task requiring both instructional and linguistic knowledge. Moreover, authors in [21] extended BERT to toxic comment detection on Indonesian social media, demonstrating its ability to interpret informal language and culturally specific idioms.

Recent advances in Indonesian NLP show strong potential for developing Transformer-based models tailored to specific domains, including religious texts. Authors in [22] emphasized the importance of adapting models to local languages and limited domains, while authors in [23] demonstrated the effectiveness of transformers in handling complex cultural contexts. Studies in [24-26] highlight the relevance of Long Short-Term Memory (LSTM) approaches, transfer learning, and named entity optimization for working with limited-resource corpora. Authors in [27] further stressed the need for precise semantic modeling, which directly aligns with the requirements of religious text processing. Despite these developments, very few BERT models have been specifically trained on Indonesian Islamic texts. To address this gap, Al-Qur'an BERT for Indonesian Religious Texts (ALBERTIR) was pretrained using Masked Language Modeling (MLM) on official Qur'an and Hadith translations, supporting downstream tasks such as classification, sentiment analysis, and Question-Answering (QA).

II. STUDY OBJECTIVES

This paper aims to develop and pretrain a BERT model specifically tailored for Indonesian religious texts, particularly Qur'an and Hadith translations, using transfer learning. By leveraging a corpus enriched with religious vocabulary and the unique linguistic features of Indonesian Islamic texts, the model is expected to capture richer semantic and syntactic contexts than generic BERT models. The study also aims to evaluate the pretrained ALBERTIR model on NLP tasks, including religious text classification, sentiment analysis, and question answering related to the Qur'an and Hadith. Beyond improving accuracy in religious language processing, ALBERTIR is envisioned as a foundation for AI tools supporting education and religious applications in Indonesia.

III. RESEARCH METHOD

A. Corpus Collection and Preparation

The primary dataset used in this study consists of publicly available Indonesian translations of the Qur'an and Hadith. The Qur'an corpus is based on a standardized Indonesian translation, while Hadith texts are drawn from authentic collections translated by qualified religious authorities. These translations were selected for their semantic richness, theological relevance, and widespread use in Islamic education in Indonesia, particularly in aligning Friday sermons with core religious texts. The dataset includes 6,236 Qur'anic verses from the Ministry of Religious Affairs and 7,410 Hadith narrations from verified Indonesian translations of Sahih Bukhari and Sahih Muslim, totaling 13,646 entries and over 4.2 million tokens. Only neutral, non-sectarian texts were selected.

Both corpora are preprocessed to remove non-linguistic elements such as verse numbers, punctuation artifacts, and metadata. Text normalization, tokenization, and sentence segmentation are implemented using the IndoNLP toolkit and HuggingFace's tokenizer library to ensure compatibility with BERT tokenization [23]. The collection process was automated with manual quality checks, and the corpus was normalized into plain text for unsupervised pretraining using the MLM task.

B. Model Architecture

ALBERTIR is preinitialized from the Indonesian BERT (IndoBERT) base model [28], which had previously been pretrained over a large dataset of Indonesian texts. This transfer learning is applied to reduce training time and leverage general Indonesian language features for religious contexts. The model follows the BERT-base configuration with 12 transformer layers, 768 hidden units, and 12 self-attention heads.

C. Pretraining Strategy

The pretraining task applies the MLM objective, masking 15% of general tokens or 20% of domain-specific religious terms and requiring the model to predict them from the surrounding context. This strategy is designed to encourage the model to better capture the semantics of key religious vocabulary without overfitting to general tokens. It also enables deep contextual representation learning within Indonesian religious discourse.

Training is conducted with HuggingFace Transformers and Trainer Application Programming Interface (API) in Google Colab Tensor Processing Unit (TPU) v3-8 runtime. It runs for 10 epochs with a batch size of 256, learning rate of $5 \cdot 10^{-5}$, AdamW optimizer, and linear warmup schedule. The process is logged with Weights & Biases for reproducibility.

D. Standard Baseline BERT and the ALBERTIR Model

1) Baseline (Standard BERT)

In the baseline BERT model, 15% of tokens (subwords or words) are randomly masked with uniform probability. The loss function is given by:

$$L_{MLM} = -\sum_{i \in m} \log P(x_i | x \backslash m) \quad (1)$$

where m denotes the positions of masked tokens, x_i denotes the target token at position i, and $P(x_i | x \backslash m)$ denotes the probability of predicting the target token given the unmasked context.

2) ALBERTIR Modifications

Unlike standard BERT, which may mask only part of a word at the subword level (e.g., "shalat" → "sha[MASK]"), ALBERTIR applies whole-word masking to religious vocabulary (e.g., "shalat" → "[MASK]"). This preserves the semantic integrity of sacred expressions. To emphasize accuracy on religious terms, ALBERTIR introduces term-specific weights w_i:

$$L_{MLM-religious} = -\sum_{i \in m} w_i \log P(x_i | x \backslash m)$$

$$w_i = \begin{cases} 1.5 & \text{if } x_i \in V_{religious} \\ 1.0 & \text{otherwise} \end{cases} \quad (2)$$

where $V_{religious}$ is a curated lexicon (e.g., "iman", "zakat").

3) Visual Comparison

As shown in Table I, ALBERTIR prioritizes masking sacred terms through dynamic whole-word masking, unlike standard BERT's uniform subword approach. This strategy preserves semantic consistency in religious phrases.

TABLE I. MASKING STRATEGIES IN BERT PRETRAINING

Strategy	Example Input	Masked Output
Standard BERT	"Dirikanlah shalat dan tunaikan zakat"	"Dirikanlah [MASK] dan [MASK] zakat"
	Translation	Translation
	"Establish prayer and give alms"	"Establish [MASK] and [MASK] alms"
ALBERTIR	"Dirikanlah shalat dan tunaikan zakat"	"[MASK] shalat dan tunaikan [MASK]" (*whole-word priority*)
	Translation	Translation
	"Establish prayer and give alms"	"[MASK] prayer and give [MASK]"

4) Embedding Architecture

In standard BERT, token embeddings of dimension E are mapped directly into a hidden space of dimension H, with E=H=768. Each token is thus represented as a 768-dimensional vector fed into the transformer layers. However, ALBERTIR adopts an ALBERT-style factorized embedding. The embedding dimension E is decoupled from the hidden dimension H (e.g. E=128, H=768). This reduces embedding parameters by up to 80%, improving training efficiency without sacrificing performance. The reduced parameter space enables pretraining on larger religious corpora within limited computational budgets. Cross-layer parameter sharing further enhances generalization and contextual depth, making the model highly effective for complex theological language.

$$Embedding\ Layer = W_{E*V}$$

$$Projection = W_{H*E}, (E \ll H) \quad (3)$$

where W_{E*V} denotes the embedding matrix of size $E * V$ (E: embedding dimension, V: vocabulary size), and W_{H*E} denotes the projection matrix mapping embeddings to the hidden dimension H.

5) Domain-Specific Masking Strategies

A key limitation of the baseline BERT is its uniform masking strategy, which randomly masks tokens without semantic awareness. This poses a risk in religious texts, where masking sacred terms like "Allah" could lead to inappropriate or blasphemous predictions, undermining both model reliability and cultural sensitivity. Thus, ALBERTIR ensures that sacred tokens V_{sacred} (e.g., "Allah", "Qur'an") are never masked:

$$m' = m \backslash \{i | x_i \in V_{sacred}\} \quad (4)$$

where m' is the adjusted set of masked positions after removing sacred terms.

Furthermore, ALBERTIR introduces verse-aware masking, a strategy designed for paired Qur'anic verses and Hadiths. In this approach, one text from the pair is selectively masked while the other is left intact, encouraging the model to learn cross-textual reasoning. By associating related concepts across both sources, the model develops a deeper understanding of their semantic alignment and contextual interdependence, enhancing its ability to interpret religious texts in a more coherent and meaningful way. An example pseudocode is presented below:

```
FUNCTION religious_masking(text,
sacred_terms, mask_prob = 0.15):
tokens ← tokenize(text)
masked_tokens ← EMPTY LIST
FOR each token IN tokens:
IF token IS IN sacred_terms:
masked_tokens.ADD(token)
ELSE IF random_number_between (0,1) >
mask_prob:
masked_tokens.ADD(token)
ELSE:
masked_tokens.ADD("[MASK]")
RETURN masked_tokens
```

This function tokenizes the input and applies conditional masking. Sacred terms are exempt from masking, preserving doctrinal integrity. Non-sacred tokens are masked with 15% probability, consistent with BERT's MLM framework.

E. Evaluation Tasks

To assess the effectiveness of the pretrained ALBERTIR model, we fine-tuned it on several downstream NLP tasks that reflect religious text understanding:

- Religious QA: adapted from IndoQA [29].
- Religious text classification, using manually curated religious categories
- Religious sentiment analysis, inspired by [17, 19].

Model performance was evaluated using standard metrics: F1-score and accuracy for classification and sentiment tasks, and Bilingual Evaluation Understudy (BLEU) for QA. Results were benchmarked against IndoBERT and multilingual BERT (mBERT), which serve as baselines for domain adaptation

IV. RESULTS AND DISCUSSION

As baselines, ALBERTIR was compared to three top models: BERT-base-ID, A Lite BERT for Indonesian (ALBERT-ID), and mBERT. ALBERT-ID is the Indonesian adaptation of ALBERT, a compact version of BERT that employs parameter sharing and factorized embeddings to improve efficiency. This model has been fine-tuned on Indonesian corpora, enabling a fair comparison with ALBERTIR and BERT-base-ID [30]. These models represent some of the best current architectures for Indonesian and multilingual NLP. All models were trained using the same hardware and settings. For masking, a mixed approach was used: 15% of general tokens were masked randomly, while religious terms were masked with a 20% rate.

ALBERTIR was evaluated using two methods: MLM accuracy to assess its ability to predict masked words, especially religious terms, and real-world tasks like religious text classification and QA using IndoQA-Hadith and IndoQA-Qur'an.

A. MLM Prediction Accuracy

Table II presents the comparison of BERT-base-ID, ALBERT-ID, and ALBERTIR models across four key metrics: general MLM accuracy, religious terms accuracy, number of parameters (in millions), and convergence time (in epochs).

TABLE II. MODEL PERFORMANCE COMPARISON

Model	General MLM Accuracy	Religious Terms Accuracy	Parameters (Million)	Convergence Time (Epochs)
BERT-base-ID	72.1%	65.3%	110	15
ALBERT-ID	70.8%	68.7%	12	12
ALBERTIR	73.5%	76.2%	18	10

ALBERTIR outperforms other models in general MLM accuracy (73.5%) and, more notably, religious terms accuracy (76.2%). ALBERTIR also shows a smaller accuracy gap between religious and general vocabulary (2.7% vs. BERT's 6.8%), proving its balanced grasp of both. This makes ALBERTIR more reliable and adaptable for Indonesian religious NLP tasks. Similar improvements were observed [10], which fine-tuned CL-AraBERT on Qur'anic datasets and reported substantial gains in religious language understanding.

B. Training Efficiency

Tables II and III show that ALBERTIR is the most computationally efficient model compared to BERT-base-ID and ALBERT-ID. ALBERTIR completes training in 28 hours and converges within 10 epochs, faster than the other two models and without changing the batch size.

With 18 million parameters, ALBERTIR shows a deliberately optimized architecture positioned between BERT-base-ID (110M) and ALBERT-ID (12M). This middle-ground technique enables ALBERTIR to still be substantially lighter yet keep much of the representational capability of bigger models. Unlike BERT, which depends on vast embedding matrices and deep layers that need considerable memory and processing capacity, ALBERTIR uses factorized embeddings, a method that splits large matrices into smaller, more effective components. Compared to BERT, this approach lowers memory consumption by as much as 80% without sacrificing understanding of sophisticated linguistic constructions.

TABLE III. PRETRAINING EFFICIENCY COMPARISON BETWEEN BERT VARIANTS

Model	GPU Time (Hours)	Batch Size
BERT-base-ID	48	32
ALBERT-ID	36	32
ALBERTIR	28	32

Graphics Processing Unit (GPU)

C. Downstream Task Evaluation

Table IV presents the performance evaluation results of the BERT-base-ID and ALBERTIR models on several key downstream NLP tasks in Indonesian.

TABLE IV. DOWNSTREAM TASK PERFORMANCE (F1-SCORE)

Task	BERT-base-ID	ALBERTIR	Improvement
Religious QA	0.78	0.85	+7%
Text Classification	0.80	0.88	+8%
Sentiment Analysis	0.75	0.82	+7%

ALBERTIR consistently demonstrates significant improvements over BERT-base-ID, achieving 7-8% gains in F1-score across QA, classification, and sentiment tasks. Similar performance boosts have been reported in [9, 31], which found that domain-adaptive pretraining in Arabic text classification leads to improved semantic sensitivity and generalization.

Additionally, ALBERTIR outperforms BERT-base-ID in generating context-aware religious responses, as shown by its accurate answer "Haram and a great sin" versus BERT's vague "Not advised" when asked about abandoning prayer. This reflects ALBERTIR's superior theological understanding from domain-specific pretraining, enhancing both accuracy and semantic depth in religious QA tasks.

D. Ablation Study

The ablation study, summarized in Table V, evaluates the contribution of each component of the ALBERTIR model to its performance in recognizing and predicting religious terms. The results show that the removal of any component leads to a noticeable decline in accuracy. Specifically, removing sacred term preservation caused the largest accuracy drop (-7.8%), underscoring its critical role in maintaining theological integrity. Excluding weighted loss reduced accuracy by -5.2%, reflecting its importance in addressing class imbalance and improving sensitivity to rare but theologically significant terms. Moreover, omitting whole-word masking led to a smaller but still meaningful reduction (-3.1%), confirming its utility in preserving semantic integrity during pretraining.

TABLE V. ABLATION STUDY OF ALBERTIR COMPONENTS

Configuration	Accuracy	Δ vs. Full
Full ALBERTIR	76.2%	-
Without weighted loss	71.0%	-5.2%
Without sacred term preservation	68.4%	-7.8%
Without whole-word masking	73.1%	-3.1%

E. Comparative Analysis with mBERT

To benchmark ALBERTIR's performance, a comparison was made against mBERT. As shown in Table VI, ALBERTIR substantially outperforms mBERT in religious QA (75.4 vs. 62.1 BLEU-4), demonstrating its effectiveness in modeling domain-specific semantics. However, in general QA, ALBERTIR lags slightly behind mBERT (65.7 vs. 68.3), reflecting a trade-off typical of domain-adaptive pretraining models.

TABLE VI. CROSS-LINGUAL PERFORMANCE (BLEU-4)

Model	ID Religious QA	General ID QA
mBERT	62.1	68.3
ALBERTIR	75.4	65.7

These results are consistent with the study in [14], which found that topic-specific BERT variants often excel within their domain but may underperform in general tasks. The improvements observed for ALBERTIR stem from its specialized pretraining on Qur'an and Hadith corpora, as well as the incorporation of strategies such as weighted loss and sacred term preservation.

F. Comparative Error Analysis via Confusion Matrix

Table VII presents the confusion matrices for BERT-base-ID, ALBERT-ID, and ALBERTIR on the religious QA task, using 1,319 evaluation samples.

TABLE VII. APPROXIMATE CONFUSION MATRIX FOR RELIGIOUS QA (SAMPLE SIZE = 1,319)

Model	TP	FP	FN	TN	F1-score
BERT-base-ID	520	145	145	509	0.78
ALBERT-ID	540	125	125	529	0.81
ALBERTIR	565	99	99	556	0.85

The proposed ALBERTIR achieved the highest number of True Positives (TP) (565) and True Negatives (TN) (556), while it achieved the lowest False Positives (FP) (99) and False Negatives (FN) (99) compared to the baselines.

V. CONCLUSION

This study introduced Al-Qur'an Bidirectional Encoder Representations from Transformers for Indonesian Religious Texts (ALBERTIR), the first BERT-based model pretrained specifically on Indonesian Islamic texts, including Qur'an and Hadith translations. The model incorporates weighted loss, sacred term preservation, and whole-word masking, enabling both computational efficiency and theological integrity.

Evaluations across three downstream tasks, including religious Questioning-Answering (QA), sentiment analysis, and text classification, demonstrated consistent improvements of up to +8% F1-score over strong baselines such as Indonesian BERT (IndoBERT) and multilingual BERT (mBERT). For religious term prediction, ALBERTIR delivered a 10.9% accuracy improvement, while maintaining a lightweight architecture. Unlike general-purpose models, ALBERTIR is tailored for domain-specific semantics, making it particularly suitable for applications such as fatwa retrieval, religious QA systems, and educational tools. By reducing misclassifications and ensuring contextually grounded outputs, ALBERTIR advances the state of Indonesian religious Natural Language Processing (NLP) with both technical and ethical considerations.

REFERENCES

[1] K. Gaanoun and M. Alsuhaibani, "Sentiment preservation in Quran translation with artificial intelligence approach: study in reputable English translation of the Quran," Humanities and Social Sciences Communications, vol. 12, no. 1, Feb. 2025, Art. no. 222, https://doi.org/10.1057/s41599-024-04181-0.

[2] F. Qarah and T. Alsanoosy, "Evaluation of Arabic Large Language Models on Moroccan Dialect," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 22478–22485, Jun. 2025, https://doi.org/10.48084/etasr.10331.

[3] Y. Qiu and Y. Jin, "ChatGPT and finetuned BERT: A comparative study for developing intelligent design support systems," *Intelligent Systems with Applications*, vol. 21, Mar. 2024, Art. no. 200308, https://doi.org/10.1016/j.iswa.2023.200308.

[4] I. A. Mannix and E. Yulianti, "Academic expert finding using BERT pre-trained language model," *International Journal of Advances in Intelligent Informatics*, vol. 10, no. 2, May 2024, Art. no. 280, https://doi.org/10.26555/ijain.v10i2.1497.

[5] J. Campino, "Unleashing the transformers: NLP models detect AI writing in education," *Journal of Computers in Education*, vol. 12, no. 2, pp. 645–673, Jun. 2025, https://doi.org/10.1007/s40692-024-00325-y.

[6] M. Müller, M. Salathé, and P. E. Kummervold, "COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter," *Frontiers in Artificial Intelligence*, vol. 6, Mar. 2023, Art. no. 1023281, https://doi.org/10.3389/frai.2023.1023281.

[7] M. S. Sayeed, V. Mohan, and K. S. Muthu, "BERT: A Review of Applications in Sentiment Analysis," *HighTech and Innovation Journal*, vol. 4, no. 2, pp. 453–462, Jun. 2023, https://doi.org/10.28991/HIJ-2023-04-02-015.

[8] Y. Kim *et al.*, "A pre-trained BERT for Korean medical natural language processing," *Scientific Reports*, vol. 12, no. 1, Aug. 2022, Art. no. 13847, https://doi.org/10.1038/s41598-022-17806-8.

[9] A. S. Alammary, "BERT Models for Arabic Text Classification: A Systematic Review," *Applied Sciences*, vol. 12, no. 11, Jun. 2022, Art. no. 5720, https://doi.org/10.3390/app12115720.

[10] R. Malhas and T. Elsayed, "Arabic machine reading comprehension on the Holy Qur'an using CL-AraBERT," *Information Processing & Management*, vol. 59, no. 6, Nov. 2022, Art. no. 103068, https://doi.org/10.1016/j.ipm.2022.103068.

[11] M. M. Abdelgwad, T. H. A. Soliman, and A. I. Taloba, "Arabic aspect sentiment polarity classification using BERT," *Journal of Big Data*, vol. 9, no. 1, Dec. 2022, Art. no. 115, https://doi.org/10.1186/s40537-022-00656-6.

[12] S. M. Isa, G. Nico, and M. Permana, "IndoBERT for Indonesian Fake News Detection," *ICIC Express Letters*, vol. 16, no. 3, Mar. 2022, https://doi.org/10.24507/icicel.16.03.289.

[13] Y. Wu, Z. Liu, L. Wu, M. Chen, and W. Tong, "BERT-Based Natural Language Processing of Drug Labeling Documents: A Case Study for Classifying Drug-Induced Liver Injury Risk," *Frontiers in Artificial Intelligence*, vol. 4, Dec. 2021, Art. no. 729834, https://doi.org/10.3389/frai.2021.729834.

[14] A. Abuzayed and H. Al-Khalifa, "BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique," *Procedia Computer Science*, vol. 189, pp. 191–194, 2021, https://doi.org/10.1016/j.procs.2021.05.096.

[15] A. K. Ambalavanan and M. V. Devarakonda, "Using the contextual language model BERT for multi-criteria classification of scientific articles," *Journal of Biomedical Informatics*, vol. 112, Dec. 2020, Art. no. 103578, https://doi.org/10.1016/j.jbi.2020.103578.

[16] A. Adhikari, A. Ram, R. Tang, and J. Lin, "DocBERT: BERT for Document Classification." arXiv, 2019, https://doi.org/10.48550/ARXIV.1904.08398.

[17] R. Pramana, M. Jonathan, H. S. Yani, and R. Sutoyo, "A Comparison of BiLSTM, BERT, and Ensemble Method for Emotion Recognition on Indonesian Product Reviews," *Procedia Computer Science*, vol. 245, pp. 399–408, 2024, https://doi.org/10.1016/j.procs.2024.10.266.

[18] D. Endalie, "Fine-Tuning BERT Models for Multiclass Amharic News Document Categorization," *Complexity*, vol. 2025, no. 1, Jan. 2025, Art. no. 1884264, https://doi.org/10.1155/cplx/1884264.

[19] C. H. Lin and U. Nuha, "Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy," *Journal of Big Data*, vol. 10, no. 1, May 2023, Art. no. 88, https://doi.org/10.1186/s40537-023-00782-9.

[20] F. Baharuddin and M. F. Naufal, "Fine-Tuning IndoBERT for Indonesian Exam Question Classification Based on Bloom's Taxonomy," *Journal of Information Systems Engineering and Business Intelligence*, vol. 9, no. 2, pp. 253–263, Nov. 2023, https://doi.org/10.20473/jisebi.9.2.253-263.

[21] G. Z. Nabiilah, S. Y. Prasetyo, Z. N. Izdihar, and A. S. Girsang, "BERT base model for toxic comment analysis on Indonesian social media," *Procedia Computer Science*, vol. 216, pp. 714–721, 2023, https://doi.org/10.1016/j.procs.2022.12.188.

[22] Supriyono, A. P. Wibawa, Suyono, and F. Kurniawan, "Advancements in natural language processing: Implications, challenges, and future directions," *Telematics and Informatics Reports*, vol. 16, Dec. 2024, Art. no. 100173, https://doi.org/10.1016/j.teler.2024.100173.

[23] T. L. M. Suryanto, A. P. Wibawa, H. Hariyono, and A. Nafalski, "Comparative Performance of Transformer Models for Cultural Heritage in NLP Tasks," *Advance Sustainable Science Engineering and Technology*, vol. 7, no. 1, Jan. 2025, Art. no. 0250115, https://doi.org/10.26877/asset.v7i1.1211.

[24] S. Supriyono, "Analyzing Audience Sentiments in Digital Comedy: A Study of YouTube Comments Using LSTM Models," *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 1877–1889, Dec. 2024, https://doi.org/10.47738/jads.v5i4.393.

[25] D. A. Sulistyo, A. P. Wibawa, D. D. Prasetya, and F. A. Ahda, "An enhanced pivot-based neural machine translation for low-resource languages," *International Journal of Advances in Intelligent Informatics*, vol. 11, no. 2, May 2025, Art. no. 258, https://doi.org/10.26555/ijain.v11i2.2115.

[26] D. A. Sulistyo, D. D. Prasetya, F. A. Ahda, and A. P. Wibawa, "Pivoted Low Resource Multilingual Translation with NER Optimization," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 24, no. 5, pp. 1–16, May 2025, https://doi.org/10.1145/3727876.

[27] Sucipto, D. D. Prasetya, and T. Widiyaningtyas, "A Supervised Hybrid Weighting Scheme for Bloom's Taxonomy Questions using Category Space Density-based Weighting," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 22102–22108, Apr. 2025, https://doi.org/10.48084/etasr.10226.

[28] A. Turchin, S. Masharsky, and M. Zitnik, "Comparison of BERT implementations for natural language processing of narrative medical documents," *Informatics in Medicine Unlocked*, vol. 36, 2023, Art. no. 101139, https://doi.org/10.1016/j.imu.2022.101139.

[29] D. Wardani, H. Abdurrahman Syah, A. Wijayanto, and B. Harjito, "Flexible Semantic Qur'an Question Answering Using Graph-Based Summarization and KNN," *JOIV : International Journal on Informatics Visualization*, vol. 8, no. 4, Dec. 2024, Art. no. 2155, https://doi.org/10.62527/joiv.8.4.1907.

[30] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations." arXiv, 2019, https://doi.org/10.48550/ARXIV.1909.11942.

[31] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, "BERT applications in natural language processing: a review," *Artificial Intelligence Review*, vol. 58, no. 6, Mar. 2025, Art. no. 166, https://doi.org/10.1007/s10462-025-11162-5.