Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 292 (2026) 1845–1853

Procedia Computer Science

www.elsevier.com/locate/procedia

30th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2026)

Enhancing Multilingual Fact-Checking Retrieval with Ensemble Bi-Encoders and Architecture-Aware Hard Negative Reranking

Van-Truong Le[a,b], Duy-Minh Nguyen-Tran[a,b,c], Tung Le[a,b], Huy Tien Nguyen[a,b,*]

[a] *Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam*
[b] *Vietnam National University, Ho Chi Minh City, Vietnam*
[c] *Japan Advanced Institute of Science and Technology, Ishikawa, Japan*

Abstract

Multilingual fact-checking requires retrieving relevant verified claims across languages, a task where dense retrievers often struggle to precisely rank top candidates, especially in cross-lingual settings. In this work, we propose a two-stage retrieve-then-rerank pipeline, combining an ensemble of Bi-Encoders for candidate generation with a Cross-Encoder reranker trained using semantically mined hard negatives. Experiments on the SemEval-2025 Task 7 dataset show that reranking significantly improves performance, yielding an 8.0% absolute gain in cross-lingual retrieval. Our analysis further shows that semantically aligned hard negatives are critical for effective reranking, whereas lexically derived negatives can substantially degrade performance.

© 2026 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the KES International Association

Keywords: Fact-Checking Retrieval; Cross-lingual Information Retrieval; Bi-Encoder; Cross-Encoder; Hard Negative Mining; Ensemble Learning; Reranker Architecture

1. Introduction

Multilingual fact-checking requires retrieving previously verified claims across languages, a task that is particularly challenging in cross-lingual settings where queries and relevant evidence may not share lexical overlap. Recent multilingual encoders such as XLM-R and E5 [1, 2] enable semantic matching through dense representations. However, dense retrievers often struggle to precisely rank top candidates, especially when semantically similar but factually different claims exist. This limitation is critical in fact-checking, where users typically rely on only a few top-ranked results.

A common solution in information retrieval is the retrieve-then-rerank paradigm, where a Cross-Encoder refines the top candidates retrieved by a Bi-Encoder. Despite its effectiveness in other domains, the role of reranking in multilingual fact-checking remains underexplored. In particular, the impact of negative sampling strategies on reranker

* Corresponding Author: Huy Tien Nguyen
E-mail address: ntienhuy@fit.hcmus.edu.vn (Huy Tien Nguyen).

1877-0509 © 2026 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the KES International Association

performance has not been systematically studied, even though it directly affects the model's ability to distinguish between semantically similar but irrelevant claims.

In this work, we adopt a two-stage retrieve-then-rerank pipeline for multilingual fact-checking. We use an ensemble of Bi-Encoders to improve retrieval robustness and a Cross-Encoder reranker trained with semantically mined hard negatives to refine top-ranked candidates. Through experiments on the SemEval-2025 Task 7 dataset, we show that reranking significantly improves performance, especially in cross-lingual settings.

Our contributions are summarized as follows:

1. We demonstrate that a Cross-Encoder reranker substantially improves multilingual fact-checking retrieval, with notable gains in cross-lingual scenarios.
2. We show that negative sampling plays a critical role in reranker effectiveness, where semantically mined hard negatives outperform both random and BM25-based negatives.
3. We provide an empirical analysis of reranker architectures and training strategies for multilingual fact-checking.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents our approach. Section 4 describes the experimental setup. Section 5 reports results and analysis. Section 6 concludes the paper.

2. Related Work

Multilingual Fact-Checking Retrieval. Early approaches to claim retrieval relied primarily on lexical matching techniques such as BM25 [3] and TF-IDF. While effective for exact term matches within a single language, these methods suffer from the "lexical gap" problem—the inability to match semantically equivalent but lexically different expressions. This limitation is further exacerbated in cross-lingual settings, where translation equivalents may share few surface features and cultural references may not directly align.

The advent of pre-trained multilingual language models has significantly advanced cross-lingual retrieval. Models such as mBERT [4] and XLM-RoBERTa [1] learn cross-lingual representations through large-scale masked language modeling on multilingual corpora. More recent approaches, including Multilingual-E5 [2] and BGE-M3 [5], are explicitly designed for embedding alignment across languages, enabling queries and documents in different languages to be compared in a shared semantic space. Despite their strong performance, these models often struggle with fine-grained ranking in fact-checking scenarios, where semantically similar claims may differ in factual correctness.

Retrieval Architectures. Modern neural retrieval systems typically adopt either Bi-Encoder or Cross-Encoder architectures. *Bi-Encoders* independently encode queries and documents into dense representations, enabling efficient large-scale retrieval using approximate nearest neighbor search libraries such as FAISS [6]. However, this independence limits their ability to capture fine-grained interactions between query and document tokens [7].

In contrast, *Cross-Encoders* jointly encode query-document pairs using cross-attention, allowing them to model detailed semantic interactions at the cost of significantly higher computational complexity. To balance efficiency and effectiveness, a two-stage *retrieve-then-rerank* paradigm is widely adopted [8], where a Bi-Encoder first retrieves a set of candidate documents, and a Cross-Encoder subsequently refines their ranking. While this paradigm has proven effective in general information retrieval, its behavior in multilingual fact-checking—particularly under semantic ambiguity—remains insufficiently explored.

Hard Negative Mining Strategies. The effectiveness of neural retrieval models critically depends on the quality of negative samples used during training. Random negatives are computationally efficient but often too easy, resulting in weak supervision signals. To address this, prior work has focused on *hard negatives*—non-relevant documents that are semantically similar to the query. Dense Passage Retrieval (DPR) [9] leverages in-batch and BM25-mined negatives, while Approximate Nearest Neighbor Negative Contrastive Estimation (ANCE) [10] iteratively mines hard negatives from evolving retrieval models. RocketQA [11] further improves training with cross-batch negatives and denoised hard negative selection.

Despite these advances, most prior studies focus on improving retrieval effectiveness in general-domain settings, with limited analysis of how different negative sampling strategies affect reranker behavior. In multilingual fact-checking, where semantically similar yet factually incorrect claims frequently occur, the choice of negative sam-

ples can critically influence model performance. However, this aspect remains underexplored, particularly for Cross-Encoder reranking.

3. Methodology

We propose a two-stage retrieval framework for multilingual fact-checking that explicitly addresses the challenge of distinguishing semantically similar yet factually different claims. While the retrieve-then-rerank paradigm is widely used, its effectiveness in this setting critically depends on the quality of the representations and the design of the training signals. Our approach introduces improvements at both stages: (1) a robust dense retriever with enhanced pooling and ensembling, and (2) a Cross-Encoder reranker trained with semantically aligned hard negatives.

Formally, given a collection of fact-checked claims $C = \{c_1, \ldots, c_N\}$ and a query q, the goal is to learn a ranking function $f(q, c_i)$ that assigns higher scores to relevant claims.

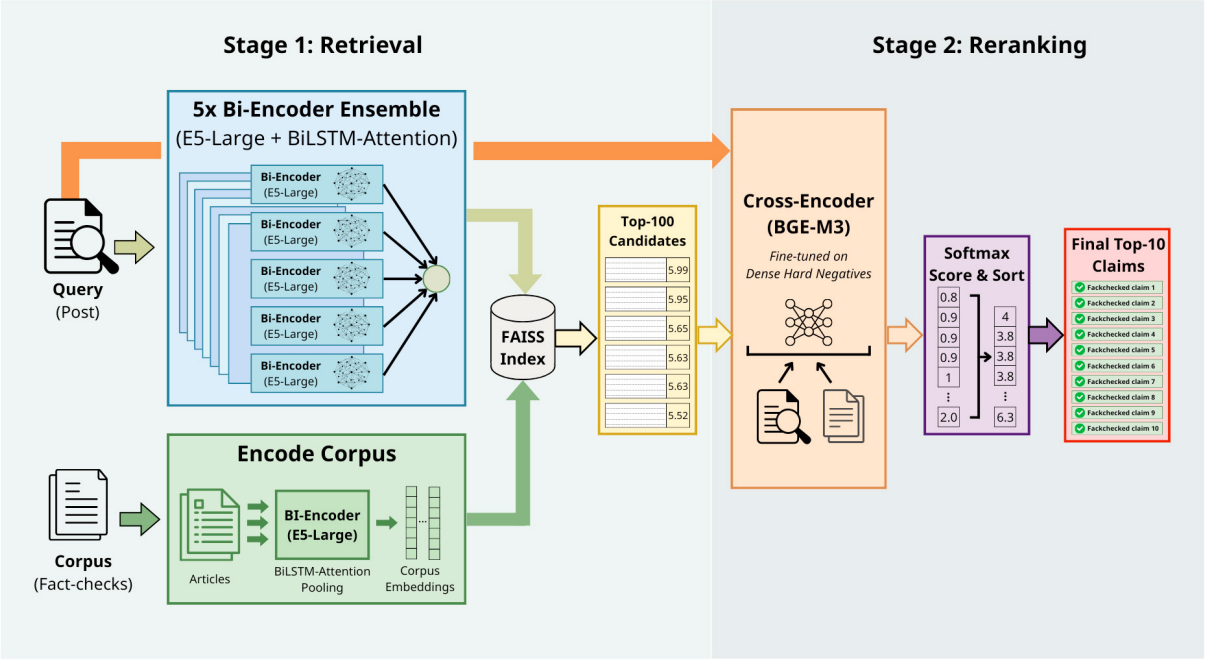


Fig. 1: Overview of the proposed two-stage framework.

3.1. Stage 1: Dense Retrieval with BiLSTM-Attention Ensemble

We build our retriever on `multilingual-e5-large-instruct` [2], a multilingual embedding model trained with instruction tuning. This choice enables strong cross-lingual alignment, which is essential for handling both multilingual and cross-lingual retrieval scenarios.

BiLSTM-Attention Pooling. Standard pooling strategies such as mean pooling or `[CLS]` representations often assume that semantic information is uniformly distributed across tokens. However, in fact-checking tasks, relevant evidence is frequently sparse and dispersed across multiple parts of the text. This can lead to diluted sentence representations and suboptimal retrieval performance. To address this limitation, we introduce a BiLSTM-Attention pooling layer that captures both sequential dependencies and token-level importance. Given token embeddings $e_1, \ldots, e_T$ from the encoder, we compute contextualized token representations and aggregate them into a single sentence embedding as follows:

$$
\begin{aligned}
s &= \sum_{t=1}^{T} \alpha_t h_t. \\
&\textit{with} \\
h_t &= [\overrightarrow{\mathrm{LSTM}}(e_t); \overleftarrow{\mathrm{LSTM}}(e_t)], \\
\alpha_t &= \frac{\exp(\tanh(W_a h_t + b_a)^\top v_a)}{\sum_{j=1}^{T} \exp(\tanh(W_a h_j + b_a)^\top v_a)}
\end{aligned}
\tag{1}
$$

This mechanism allows the model to emphasize semantically salient tokens while preserving contextual information across the sequence. As a result, the representation better captures fine-grained evidence patterns typical in fact-checking content.

Ensemble Retrieval. A single dense retriever can produce unstable rankings due to sensitivity to initialization and training dynamics. To improve robustness and generalization, we train an ensemble of five models, each with a different random seed and data shuffling. At inference time, we aggregate their outputs by averaging the similarity scores, producing a more stable and reliable ranking. This ensemble not only improves recall but also provides higher-quality candidate sets for downstream reranking. The final query and document embeddings are obtained by averaging the outputs of the five models:

$$
s_{\text{ensemble}} = \frac{1}{5} \sum_{k=1}^{5} s^{(k)}. \tag{2}
$$

3.2. Stage 2: Cross-Encoder Reranking with Hard Negatives

While the dense retriever efficiently narrows down the search space, it encodes queries and documents independently, limiting its ability to capture fine-grained interactions. To overcome this, we apply a Cross-Encoder that jointly processes each query–candidate pair and produces a refined relevance score.

Reranking Strategy. For each query, we take the top-K candidates from Stage 1 and feed them into the Cross-Encoder using the standard input format `[CLS] query [SEP] document [SEP]`. The model evaluates each pair holistically, enabling it to capture subtle semantic mismatches and factual inconsistencies that are not visible to Bi-Encoders.

The final ranking is obtained by sorting candidates according to their predicted relevance scores.

Dense Hard Negative Mining. A critical factor in training an effective reranker is the choice of negative examples. Random negatives are often too easy and fail to provide meaningful learning signals, while lexical negatives like BM25-based tend to emphasize surface-level overlap rather than semantic similarity.

To address this, we adopt a dense hard negative mining strategy based on the Stage 1 retriever. For each query, we retrieve the top-K candidates and remove all ground-truth positives. The highest-ranked remaining candidates are then selected as hard negatives.

These negatives are semantically close to the query but factually irrelevant, forcing the model to learn fine-grained distinctions. Importantly, since the negatives are drawn from the retriever's output distribution, the training process is well aligned with the inference scenario.

4. Experimental Setup

4.1. Dataset

We conduct experiments on the SemEval-2025 Task 7 dataset [12], a multilingual benchmark for fact-checked claim retrieval. The dataset consists of social media posts (queries) paired with fact-checked claims (documents). Each post includes a main textual component and may additionally contain OCR-extracted text from images, reflecting the multimodal nature of real-world misinformation. Metadata such as source and verdict are also provided, but are not directly used in our retrieval models. Each claim corresponds to a verified statement published by professional fact-checking organizations, typically consisting of a concise claim title and a reference URL. An illustrative example is shown in Table 1.

Table 1: Example of a social media post and a relevant fact-checked claim (German, with English translations).

	Original (German)	**English Translation**
Post	*Wer an Wahlen glaubt, ist auch der Meinung, dass das Ordnungsamt die Küche putzt. Wussten Sie eigentlich das Das Besatzerstatut besagt: dass die Fremdverwaltung allein Durch die Wahlen vom Wahler akzeptiert wird ** Bei unter 50% Wahlbeteiligung wäre dies hinfällig Denn es gabe dann ein rechtliches Problem Das Besatzerstatut wäre ungültig...*	*Anyone who believes in elections also believes that the regulatory office cleans the kitchen. Did you actually know that? The occupier statute states: that the foreign administration is only accepted by the voters through the elections ** If the turnout is less than 50%, this would be obsolete Because then there was a legal problem The occupier statute would be invalid...*
Claim	Deutschland ist ein besetztes Land.	Germany is an occupied country.

Language Coverage and Distribution. As summarized in Table 2, the dataset spans eight languages in the training set (Arabic, English, French, German, Malay, Portuguese, Spanish, Thai), along with two unseen languages (Polish and Turkish) in the test set. This setting introduces a realistic zero-shot cross-lingual generalization challenge. A notable characteristic of the dataset is the severe imbalance in data distribution across languages. High-resource languages such as English and Spanish contain significantly more training pairs compared to low-resource languages like Thai or Arabic. This imbalance poses additional challenges for multilingual representation learning and may bias models toward dominant languages.

The test set contains 4,959 queries, providing a robust basis for evaluation. Following the shared task setup, we report performance on both monolingual and cross-lingual retrieval scenarios. This dual evaluation allows us to analyze not only overall retrieval effectiveness but also the generalization ability of models across languages.

Table 2: Dataset Statistics for the Monolingual Track (Number of SMP–Claim Pairs)

Language	**ISO Code**	**Train Pairs**	**Dev Pairs**	**Test Pairs**
Arabic	ara	680	78	514
English	eng	5,446	627	574
French	fra	1,667	200	740
German	deu	830	101	549
Malay	msa	1,169	116	96
Portuguese	por	3,386	403	613
Spanish	spa	6,313	692	576
Thai	tha	465	42	183
Polish	pol	–	–	564
Turkish	tur	–	–	550
Total		**19,956**	**2,259**	**4,959**

4.2. Evaluation Metrics

We adopt Success@K (S@K) as the primary evaluation metric, which measures whether at least one relevant claim is retrieved within the top K results. Formally, for a set of queries Q,

$$\text{S@K} = \frac{1}{|Q|} \sum_{q \in Q} \mathbf{1}\{\exists c \in \text{Relevant}(q) \text{ s.t. } c \in \text{Top}_K(q)\},$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. We report S@1, S@5, and S@10 to capture both precision at the top ranks and recall-oriented performance.

S@K is particularly suitable for fact-checking retrieval, where the goal is not to retrieve all relevant claims but to ensure that at least one correct verification appears among the top results. In practical applications, downstream systems (or human fact-checkers) typically inspect only a small number of top-ranked candidates, making early precision critical.

We evaluate performance on three subsets. *Monolingual* includes queries whose relevant claims are in the same language, measuring within-language retrieval performance. *Cross-lingual* includes queries paired with claims in different languages, directly assessing the model's ability to bridge linguistic gaps. *Overall* aggregates all queries regardless of language pairing, providing a holistic performance measure.

4.3. Implementation Details

All models are implemented using the SentenceTransformers library [7]. Training is conducted on NVIDIA T4 GPUs with 16GB VRAM using mixed precision (FP16) for efficiency. We use the AdamW optimizer [13] with weight decay of 0.005 and gradient clipping at 1.0. The complete hyperparameter configuration for both stages is detailed in Table 3.

Table 3: Hyperparameters for Model Training

Component	Parameter	Value
Bi-Encoder	Base Model	`intfloat/multilingual-e5-large-instruct`
	Pooling Layer	Custom BiLSTM-Attention (hidden dim 512)
	Loss Function	MultipleNegativesRankingLoss
	Learning Rate	5×10^{-6}
	Batch Size	24
	Epochs	4
	Ensemble Size	5 models
	Gradient Checkpointing	Enabled
	Temperature (cosine)	0.05
Cross-Encoder	Primary Model	`BAAI/bge-reranker-v2-m3`
	Negative Strategy	Dense Hard Mining (Top 10)
	Learning Rate	1×10^{-5}
	Batch Size	8
	Epochs	4
	Max Sequence Length	512

4.4. Comparative Baselines

We compare our approach against a diverse set of baselines spanning lexical, dense, and competitive systems. As a traditional sparse retrieval method, BM25 [3] is included to highlight the limitations of lexical matching in multilingual

and cross-lingual settings. For dense retrieval, we use Multilingual-E5-Large [12] as a strong off-the-shelf baseline without our proposed BiLSTM-Attention pooling or ensembling, and Dense Retrieval Only, which corresponds to our Bi-Encoder ensemble used without reranking, to isolate the contribution of the first-stage retriever. Finally, we compare against top-performing systems from the SemEval-2025 Task 7 leaderboard, including PINGAN AI (Rank 1), RACAI, and TIFIN India [12], providing a strong real-world benchmark against state-of-the-art approaches.

5. Results and Analysis

We evaluate our approach on the SemEval-2025 Task 7 benchmark, focusing on overall retrieval effectiveness and cross-lingual generalization.

5.1. Overall Performance

Table 4 presents the performance of our final system compared to the SemEval-2025 baselines and top-performing teams. Our pipeline, combining a Bi-Encoder ensemble with a BGE-M3 reranker trained on dense hard negatives, achieves an overall S@10 of 0.9163.

Compared to the organizers' dense baseline (Multilingual-E5-Large), our approach significantly improves cross-lingual performance (0.8833 vs. 0.6372), demonstrating the effectiveness of reranking with semantically informed negatives. More importantly, our system achieves the best cross-lingual performance among all compared methods, surpassing the competition winner (PINGAN AI) by 2.46% absolute.

While our monolingual performance (0.9215) is slightly lower than the top system (0.9601), the strong gains in cross-lingual retrieval indicate that our approach is more robust in multilingual scenarios where lexical overlap is limited.

Table 4: Overall Performance of the Final Pipeline vs. SemEval-2025 Task 7 Systems

Method	**Monolingual S@10**	**Cross-lingual S@10**	**Overall S@10**
SemEval-2025 Organizers' Baselines			
BM25	0.8123	0.6027	–
Multilingual-E5-Large	0.8589	0.6372	–
SemEval-2025 Top Performing Teams			
PINGAN AI (Rank 1)	**0.9601**	0.8587	–
RACAI (BGE-M3 adapted)	0.9377	0.8245	–
TIFIN India	0.9383	0.8102	–
Our Proposed Methods			
Dense Retrieval (BiLSTM-Attn Ensemble)	0.8973	0.8033	0.8845
Final Pipeline (BGE-M3 + Dense Hard Negatives)	0.9215	**0.8833**	**0.9163**

5.2. Ablation Study

Reranker Architecture. We first analyze the impact of reranker architecture while fixing the negative sampling strategy (dense hard negatives). As shown in Table 5, lightweight models such as MiniLM-L12 perform poorly (0.3479 S@10), indicating insufficient capacity for fine-grained semantic reasoning. XLM-RoBERTa improves performance (0.5402 S@10) but remains significantly below optimal.

In contrast, BGE-M3 achieves 0.9104 S@10, substantially outperforming all alternatives. This highlights the importance of specialized reranker architectures for fact-checking, where distinguishing between semantically similar yet factually different claims is critical.

Table 5: Comparison of Reranker Architectures (Trained with Dense Hard Negatives)

Reranker Model	Params (M)	Overall S@10	Mono S@10	Cross S@10
Dense Baseline (No Reranker)	560	0.8845	0.8973	0.8033
MiniLM-L12	117	0.3479	0.3481	0.3467
XLM-RoBERTa	278	0.5402	0.5482	0.4900
BGE-M3	568	**0.9104**	**0.9189**	**0.8567**

Negative Sampling.. We further evaluate the impact of negative sampling strategies using the BGE-M3 reranker. As shown in Table 6, dense-mined hard negatives achieve the best performance (0.9163 S@10), while BM25-mined negatives lead to a severe degradation (0.2015 S@10).

This result indicates that lexical hard negatives are misaligned with semantic reranking objectives, whereas dense hard negatives provide more informative supervision by capturing semantically similar but non-relevant claims.

Table 6: Ablation Study: Negative Sampling Strategies with BGE-M3 Reranker

Negative Sampling Strategy	S@1	S@5	S@10
Random	0.4789	0.7362	0.8049
BM25-mined	0.0396	0.1155	0.2015
Dense-mined (Ours)	–	–	**0.9163**

5.3. Cross-Lingual Analysis

Our approach yields the largest improvement in the cross-lingual setting, where S@10 increases from 0.8033 to 0.8833—an absolute gain of 8.0%. This reduces the performance gap between monolingual and cross-lingual retrieval from 9.40% to 3.82%, indicating substantially improved alignment across languages.

This gain can be explained by the limitation of Bi-Encoder retrieval in cross-lingual scenarios. Since queries and claims are encoded independently, similarity relies on coarse semantic alignment, often leading to high scores for topically related but factually incorrect claims.

The Cross-Encoder mitigates this issue by modeling query–claim interactions at the token level, enabling finer semantic discrimination beyond surface similarity. This effect is further strengthened by dense hard negatives, which expose the model to semantically similar but non-relevant examples during training, improving its ability to resolve cross-lingual ambiguities.

6. Conclusion

We presented a two-stage retrieval pipeline for multilingual fact-checking, achieving state-of-the-art performance on the SemEval-2025 Task 7 benchmark, particularly in cross-lingual retrieval. Our results demonstrate that combining a scalable Bi-Encoder with a Cross-Encoder reranker trained on dense hard negatives is critical for handling semantically ambiguous, cross-lingual claims.

These findings highlight the importance of training signal design and fine-grained interaction modeling in multilingual retrieval systems. Future work will explore multimodal reasoning, temporal awareness, and explainable retrieval to further improve robustness and transparency.

Acknowledgement

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA2386-24-1-4034.

References

[1] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8440–8451.

[2] L. Wang, N. Yang, X. Huang, L. Yang, R. Majumder, F. Wei, Multilingual e5 text embeddings: A technical report, arXiv preprint arXiv:2402.05672 (2024).

[3] S. Robertson, H. Zaragoza, The probabilistic relevance framework: Bm25 and beyond, Vol. 3, Now Publishers Inc., 2009, pp. 333–389.

[4] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2019).

[5] S. Xiao, Z. Liu, P. Zhang, P. Liu, Z. Zhuang, Y. Shen, Z. Zhao, B. Hou, Y. Lin, J. Lin, et al., Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation, arXiv preprint arXiv:2402.03216 (2024).

[6] J. Johnson, M. Douze, H. Jégou, Billion-scale similarity search with gpus, IEEE Transactions on Big Data 7 (3) (2019) 535–547.

[7] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3982–3992.

[8] R. Nogueira, K. Cho, Multi-stage document ranking with bert, 2019.

[9] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, W.-t. Yih, Dense passage retrieval for open-domain question answering, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 6769–6781.

[10] L. Xiong, A. Gupta, J. Du, N. Thakur, C. Dyer, V. O. Cárdenas, Approximate nearest neighbor negative contrastive learning for dense text retrieval, in: International Conference on Learning Representations, 2021.

[11] Y. Qu, Y. Ding, J. Liu, K. Liu, R. Ren, W. Zhao, D. Dong, H. Wu, H. Wang, Rocketqa: An optimized training approach to dense passage retrieval for open-domain question answering, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 5835–5847.

[12] Q. Peng, R. Moro, M. Gregor, I. Srba, S. Ostermann, M. Simko, J. Podroužek, M. Mesarcik, J. Kopačn, A. Søgaard, Semeval-2025 task 7: Multilingual and crosslingual fact-checked claim retrieval, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025.

[13] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101 (2017).