

Reproduction of FlagOpen/FlagEmbedding: BGE-base-en-v1.5 on BEIR SciFact, NFCorpus and FiQA

K-Veritas Team

Submitted by K-Veritas Team

kv:2610.00038v1 · Published 2026-10-11

Independent reproduction · Artifact evaluation · Natural language processing

Abstract

Evaluated the BAAI/bge-base-en-v1.5 embedding model with the repository's BEIR evaluation module (FlagEmbedding.evaluation.beir) at the pinned commit: dense retrieval without a reranker on the test splits of SciFact, NFCorpus and FiQA, top 1000 retrieved, reporting nDCG@10 and Recall@100. Corpus embeddings were computed during the run. The research/ and examples/finetune/ folders are withheld as hash-only entries since this run does not use them. PyTorch 2.11 (CUDA 12.8). One NVIDIA RTX 5060 Ti.

About this record

Title, authors and abstract were provided by the authors and reviewed for completeness, not for scientific soundness. Every number in this record is copied from the signed data of the sealed reports [2], which remain authoritative.

This cover page is not signed. Check it against the SHA-256 listed at <https://kveritas.org/records/2610.00038v1>, and verify each report with its code bundle: `kveritas verify <report.pdf> --bundle <bundle.zip>`.

1 Sealed reports

The experiments were sealed with K-Veritas [1], which binds each result to the code, hardware and time that produced it.

Report	Label	Sealed	Runs	Invocations	Signer
r1 [2]	BGE-base-en-v1.5: BEIR SciF...	2026-10-10	1	1	Signed by the K-Veritas server

2 Results

r1: BGE-base-en-v1.5: BEIR SciFact, NFCorpus, FiQA

Run	Metric	Final value	Step
1	average_ndcg_at_10	49.903	
1	fiqa_ndcg_at_10	39.117	
1	nfcopus_ndcg_at_10	36.799	
1	scifact_ndcg_at_10	73.794	
1	average_recall_at_100	67.768	
1	fiqa_recall_at_100	73.119	
1	nfcopus_recall_at_100	33.184	
1	scifact_recall_at_100	97	

Run	Inline claim	Value	Phase
1	average_ndcg_at_10	49.903	
1	fiqa_ndcg_at_10	39.117	

Run	Inline claim	Value	Phase
1	nfcorpus_ndcg_at_10	36.799	
1	scifact_ndcg_at_10	73.794	
1	average_recall_at_100	67.768	
1	fiqa_recall_at_100	73.119	
1	nfcorpus_recall_at_100	33.184	
1	scifact_recall_at_100	97	

3 Setup

Report	Item	Value
r1	Operating system	linux / amd64
r1	CPU	Intel(R) Xeon(R) W-2145 CPU @ 3.70GHz, 16 cores
r1	GPU	NVIDIA GeForce RTX 5060 Ti, 16311 MiB
r1	Memory	31.0 GB
r1	Total run time	137.2 s

4 Files

Each report and code bundle must hash to these values. A bundle belongs to the report whose signed data names its hash.

File	SHA-256	Size
r1.pdf	1f2f105102ad5c5ed9d7fa0e8f730959b616625723360907c725a809f85ab4ce	2.1 MB
r1.kvbundle.zip	5042560c5889f38755b55d7c6cfe7b4689a2f3fe71b969abf58ab0cd5f56c167	649 KB

5 Links and licence

Paper: <https://arxiv.org/abs/2309.07597>

Reproduction of: <https://github.com/FlagOpen/FlagEmbedding/tree/fd1a2bdf69488ffebe0327999d4400d8c8058a0b>

Record: CC-BY 4.0. Code: MIT, as declared by the authors. Tags: embeddings, retrieval, beir, bge.

Listed, not endorsed. Review checks the artifact, never the science. Reports are never edited; changing them makes a new version. Other fields are corrected in place, and every correction is logged in metadata.yaml.

References

[1] M. K. Keita and C. Homan. Computer Science Conferences Should Require Nonrepudiable Experimental Results. NeurIPS 2026 Position Paper Track. arXiv:2605.08586.

[2] Sealed experiment report r1 (BGE-base-en-v1.5: BEIR SciFact, NFCorpus, FiQA), kv:2610.00038v1/r1. Data hash 6cdf0f4b58cd2acc903e460539b4470551fcb4015ee4c949577760b25e083eeb.
<https://kveritas.org/records/2610.00038v1/r1.pdf>