Catastrophic Forgetting on Custom Italian RAG Benchmark

NDCG@10

- Base Models
- Fine-tuned Models
- Baseline (BM25)

Qwen3-Emb-8B: 0.535
ml-e5-small (base): 0.488
ml-e5-small (finetuned): 0.434, -11.0%
e5-small-v2 (base): 0.408
BM25: 0.361
e5-small-v2 (finetuned): 0.209, -48.8%

BM25 baseline