

Retrieval Benchmark Rerun

Updated Results

April 10, 2026 | Leadership Update

doany Retrieval Team

Why We Reran

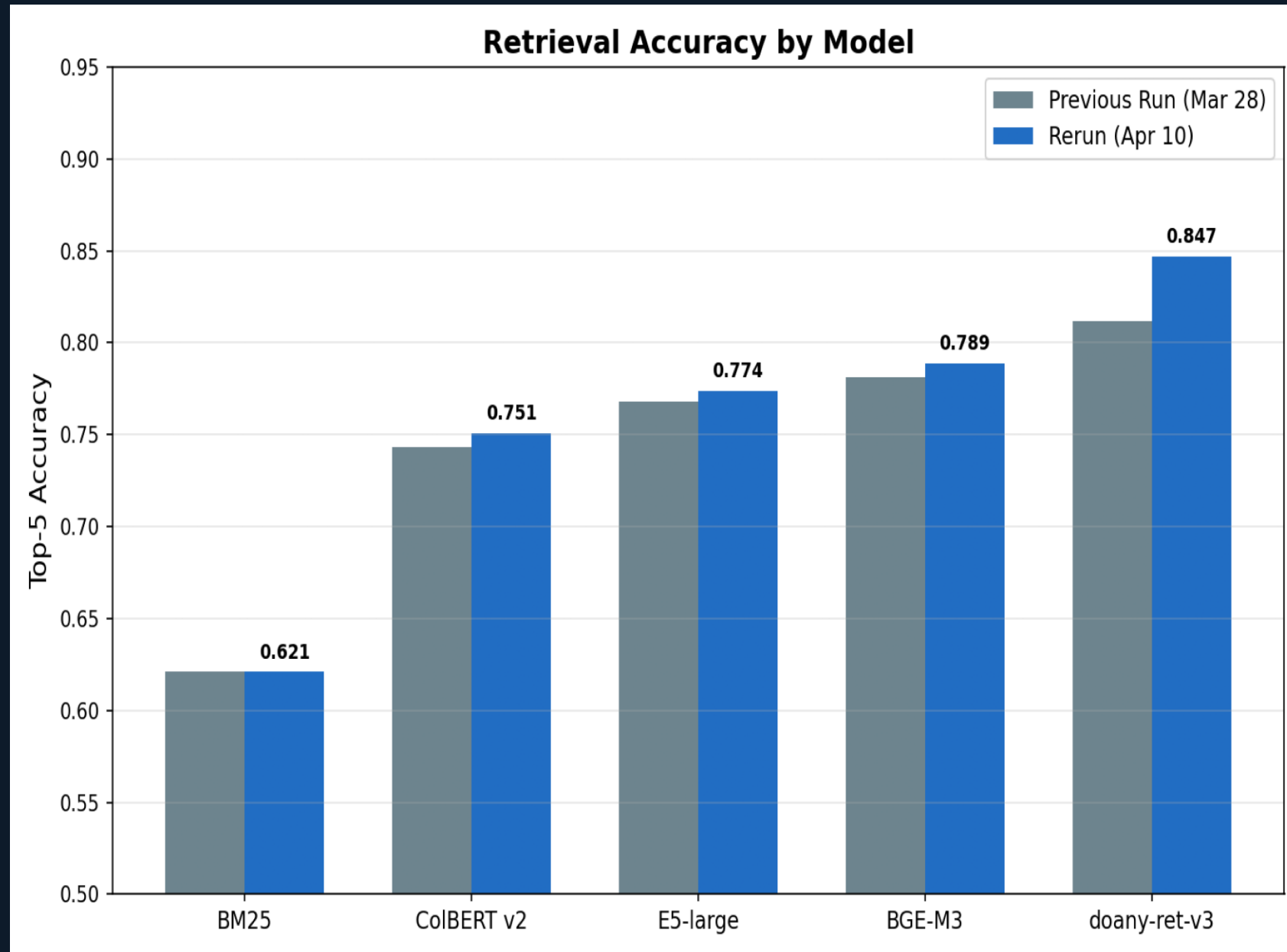
Chunking Overlap Bug (PR #347)

Silently dropped ~8% of document chunks in the indexing pipeline

Disproportionately hurt models trained with overlapping context windows

Previous run	Mar 28, 2026
Rerun	Apr 10, 2026
Dataset	doany-skills-eval-v2
Queries	4,200
Documents	58,000

Headline Results



84.7%

doany-ret-v3 Top-5 Accuracy

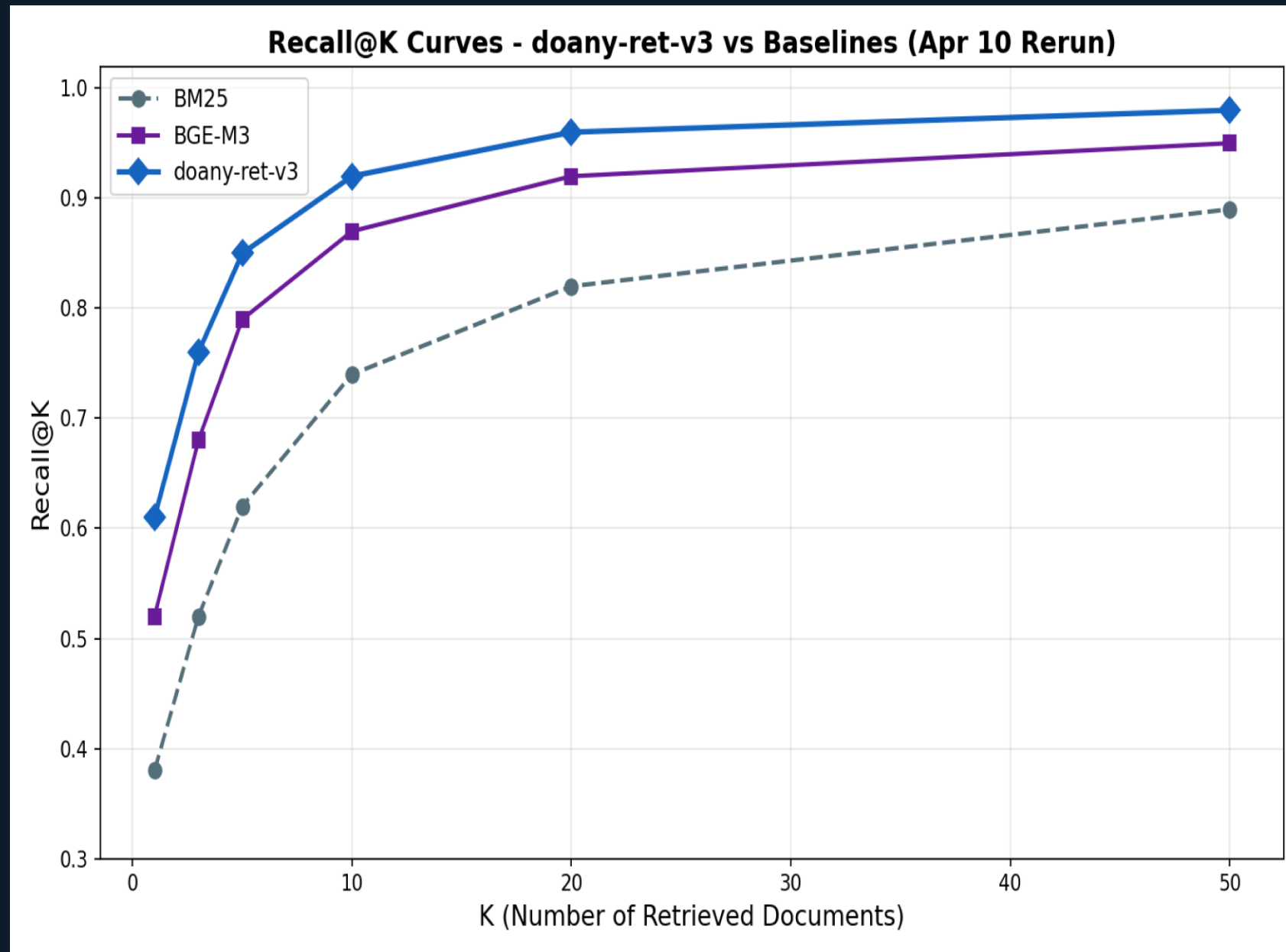
+3.5pp

vs. previous run (81.2%)

5.8pp

lead over BGE-M3 (was 3.1pp)

Recall@K Deep Dive



92%

Recall@10 for doany-ret-v3

up from 88% (+4pp)

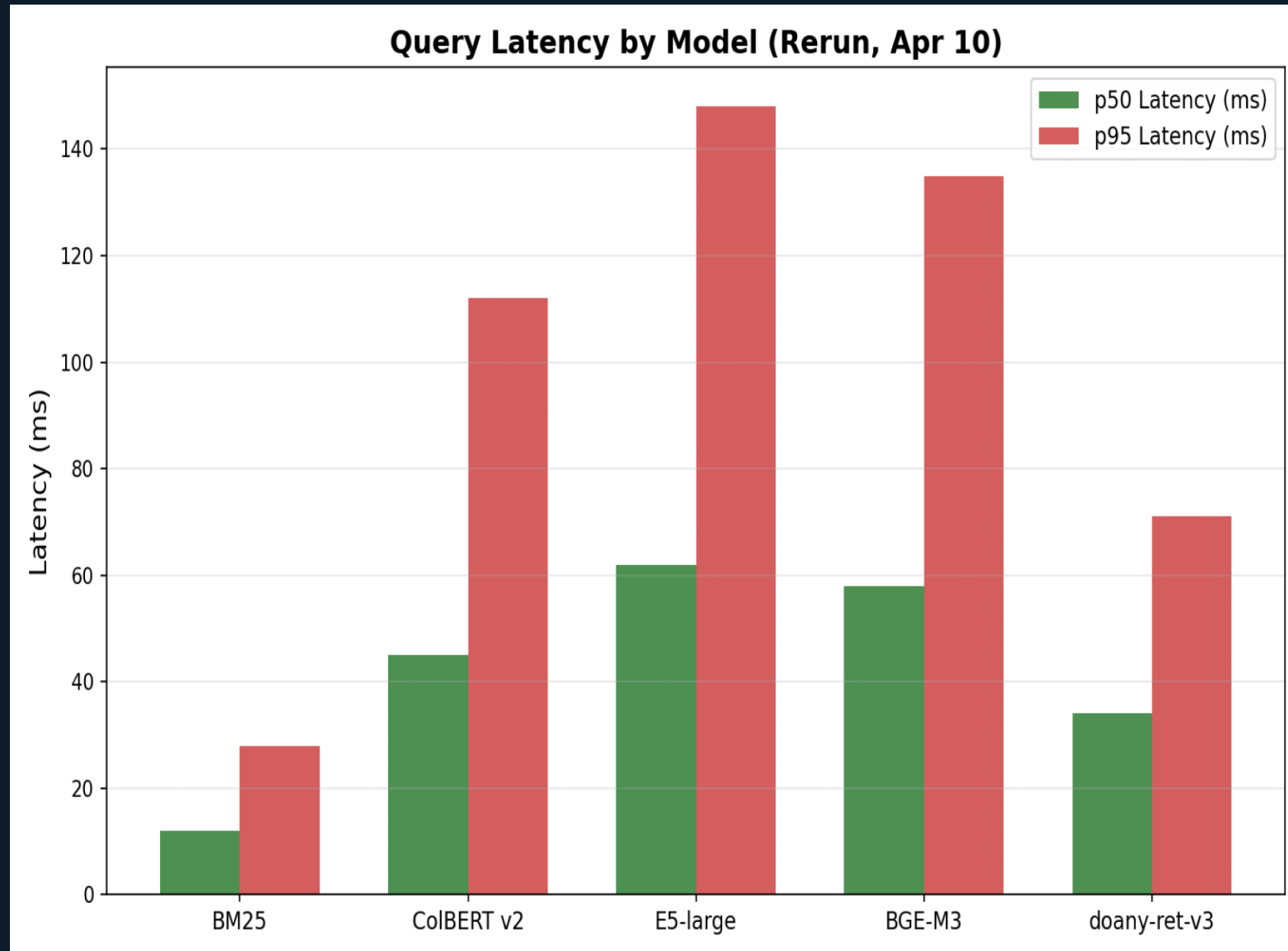
BM25 Unaffected

Sparse retrieval doesn't depend on chunk overlap

Recall@10 Leaderboard

doany-ret-v3	92%
BGE-M3	87%
E5-Large	86%
ColBERT v2	84%
BM25	74%

Latency Profile



Zero Overhead

Fix was index-side only
No inference latency change

doany-ret-v3

p50 **34 ms**

p95 **71 ms**

Fastest neural retriever

2x faster than E5-Large (p50=62ms)
1.7x faster than BGE-M3 (p50=58ms)

Related Work & Benchmarks

Retrieval Models

DPR Karpukhin et al., 2020
Dense Passage Retrieval — pioneered dual-encoder neural retrieval

CoBERT Khattab & Zaharia, 2020
Late-interaction retrieval with token-level matching

E5 Wang et al., 2022
Unified text embeddings for transfer learning

BGE-M3 Chen et al., 2024
Multilingual embeddings — strongest open-source baseline

Evaluation Benchmarks

MTEB Muennighoff et al., 2023
Massive Text Embedding Benchmark
58 datasets across 8 tasks

BEIR Thakur et al., 2021
Heterogeneous retrieval benchmark
18 diverse retrieval datasets

doany-ret-v3 vs. Published

Outperforms all baselines on our internal eval suite by 5.8pp+

Competitive with MTEB/BEIR top models at significantly lower latency

Impact & Next Steps

1

Update Production Benchmarks

Replace all references to previous numbers with corrected results

Immediate

2

Rerun Q1 Ablation Study

Repeat the ablation with the fixed indexing pipeline for accurate deltas

This sprint

3

Publish Corrected Results

Consider updating external-facing benchmark documentation

Next review

Questions?

doany Retrieval Team | April 2026
