BENCHMARK LANDSCAPE BRIEF · JULY 2026

FinRetrieval: No Independent Results Outside Daloopa

Who has run, cited, or reported on Daloopa's FinRetrieval benchmark

As of July 2026, every published FinRetrieval benchmark number comes from Daloopa itself. This brief maps everyone who has run, cited, or reported on the benchmark. It explains why no independent party has reproduced it.

Benchmark Landscape Brief · Compiled July 2026
Primary source: `arxiv.org/abs/2603.04403`

Contents

Audience and Scope

This brief is for anyone weighing FinRetrieval as an external reference point when evaluating financial-document retrieval. It assumes familiarity with retrieval benchmarks but no prior knowledge of Daloopa or FinRetrieval. It reports who has engaged with the benchmark and how. It does not re-run the benchmark or contest its reported scores.

1 The Finding: No Independent Results

Every reported FinRetrieval number traces back to Daloopa. As of July 2026 there is no independent run, no third-party leaderboard entry, and no practitioner reproduction. Everyone who has touched the benchmark falls into three groups: Daloopa and its own channels, press outlets, and two academic papers. Only the first group produces actual numbers.

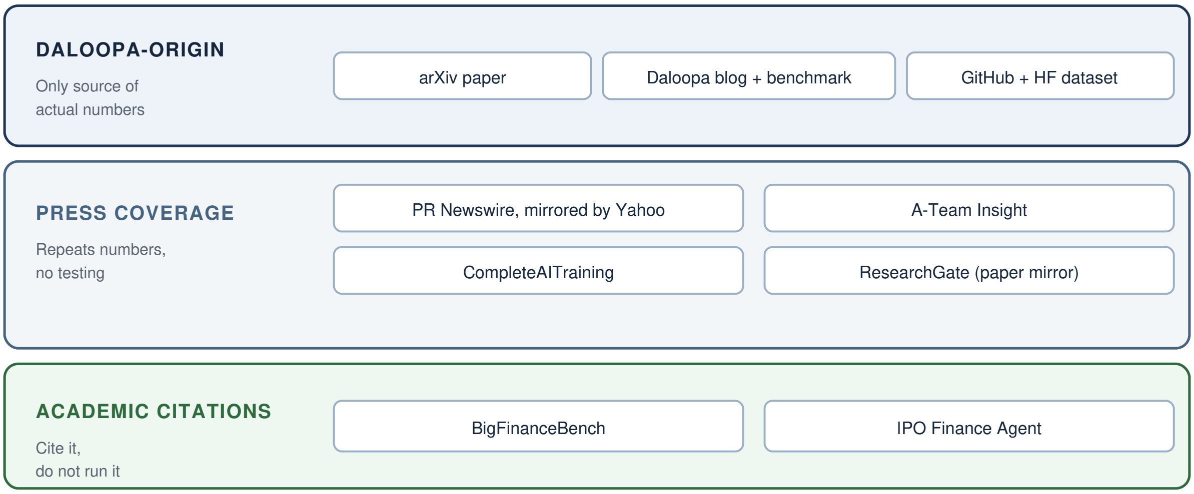

Figure 1. Everyone who has touched FinRetrieval, grouped by what they contribute.

1.1 Daloopa-Origin Sources

Daloopa's own channels are the only place actual FinRetrieval numbers appear:

- The arXiv paper, by Eric Y. Kim and Jie Huang, who are both Daloopa employees (`arxiv.org/abs/2603.04403`).
- Daloopa's own blog and benchmark pages (`daloopa.com/benchmark`).
- The GitHub repository and Hugging Face dataset, both under the `daloopa` organization (`github.com/daloopa/finretrieval`, `huggingface.co/datasets/daloopa/finretrieval`).

1.2 Press Coverage

Press coverage repeats Daloopa's numbers without independent testing:

- A PR Newswire release (`prnewswire.com`), mirrored by Yahoo Finance.

- A-Team Insight (`a-teaminsight.com`) and CompleteAITraining (`completeaitraining.com`), which both restate the 90.8% and +71-point figures from the release.
- ResearchGate, which is a mirror of the same Daloopa paper.

1.3 Academic Citations

Two academic papers mention FinRetrieval but do not run it:

- BigFinanceBench (Wang et al., 2026, `arxiv.org/abs/2606.03829`) cites FinRetrieval only in a related-work comparison table. It runs its own 928-item benchmark and uses no Daloopa data or tools.
- IPO Finance Agent (Benhenda, 2026) cites FinRetrieval, then builds its own IPO-diligence benchmark and reports no FinRetrieval results.

The Semantic Scholar citation graph shows exactly these two citing papers. Neither reports its own FinRetrieval scores.

2 Why No One Else Has Attempted It

Three reasons explain the absence of outside runs.

The high-scoring configuration is not independently runnable. The 88–91% results come from the MCP configuration. That configuration requires a bearer token for Daloopa's own API. Only Daloopa and its paying customers can reproduce those numbers. An outsider can only run the web-only configuration, which scores a noisy 20–70%. There is little incentive to run the benchmark as a neutral comparison.

Configuration	Reported score	Access required	Who can run it
MCP	88–91%	Bearer token for Daloopa's API	Daloopa and its paying customers
Web-only	20–70% (noisy)	None	Any outside party

The design is self-referential, and the authors say so. Daloopa is the data source, the question generator, the retrieval target, and the author. The paper lists this in limitation §7.3. Setups like this tend to be cited as prior work rather than adopted. That is the pattern the citation graph shows.

The benchmark is recent. It was published in January 2026, about six months ago. The usual independent-reproduction cycle has not had time to run, even apart from the access barrier.

> **BOTTOM LINE**
>
> No one outside Daloopa has reported FinRetrieval results. The only outside engagement is two papers citing it as related work and press outlets repeating Daloopa's own figures. FinRetrieval currently functions as a vendor demonstration rather than an independently contested benchmark.

Sources

`arxiv.org/abs/2603.04403` · `github.com/daloopa/finretrieval` · `huggingface.co/datasets/daloopa/finretrieval` · `arxiv.org/abs/2606.03829` (BigFinanceBench) · `prnewswire.com` · `a-teaminsight.com` · `completeaitraining.com`