

BM25 Returns, Benchmarks Break, Labs Get Instituted

2026-09-16 / 00:06:50

“BM25 hasn't changed. It's the same scoring function, but the user has become very good at search.”

— Seln Oriax, today's narration

Two talks from AI Engineer's stage today converge on the same point: retrieval is becoming the bottleneck again. Will Bryk of Exa lays out a world where AI searches exceed human searches in 2026 and reach a thousand-fold multiplier within a few years — built on hybrid embedding-keyword infrastructure that serves Cursor, HubSpot, and financial agents. Jo Bergam of Hornet Dev argues BM25 is having its moment because modern large language models are now capable query generators with parametric knowledge, making exact lexical matching more valuable than embeddings for specific lookups.

Beyond search: Ethan Mollick reports that public AI benchmarks are dire — the easy ones are maxed out, and the remaining ones are riddled with errors. Demis Hassabis and Shane Legg announce the DeepMind Institute for interdisciplinary AGI research. Naval proposes holding labs fully liable for their models' behavior. And Arcee.ai closes a Series B at over \$1 billion for security-focused AI infrastructure. Paul Dix's team also used AI to rewrite Influx DB from C++ to Rust in 11 days for \$165K.

00:00:04 The Search Infrastructure Shift

00:03:16 Benchmarks and Institutions