



Press Release
Paris, May 22, 2025

LightOn Unlocks Agentic RAG with new SOTA Model Reason-ModernColBERT



After the recent release of [GTE-ModernColBERT](#) exhibiting state-of-the-art results on long-context retrieval, LightOn announces a major leap forward in AI-driven knowledge discovery: **Reason-ModernColBERT**, an open-source multi-vector model purpose-built for Deep Research applications.

Redefining Retrieval: From Matching to Reasoning

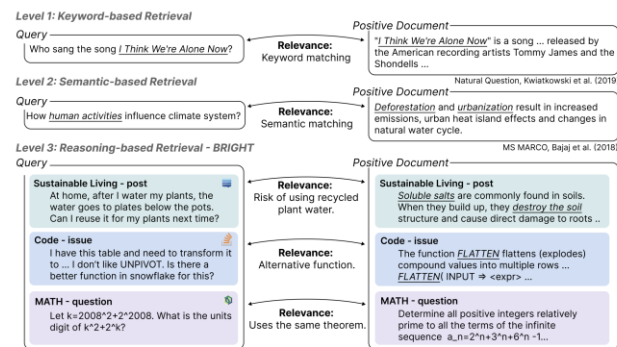


Figure 1: Existing retrieval benchmarks focus on keyword-based retrieval (level 1), or semantic-based retrieval (level 2), e.g., NQ, MS MARCO datasets (Kwiatkowski et al., 2019; Bajaj et al., 2018). **BRIGHT introduces level 3 retrieval, where the relevance between queries and documents requires intensive reasoning to determine.** Our data consists of natural user queries from diverse domains (e.g., economics, math, earth sciences, etc.). BRIGHT corpora also span across web data, such as blogs, syntax documentation, and STEM problem-solutions.

Source: [BRIGHT: a realistic and challenging benchmark for reasoning-intensive retrieval](#)

With the rise of Deep Research, organizations are demanding more than simple lexical or semantic matching. Today's cutting-edge enterprise insights require reasoning, the ability to connect, synthesize, and uncover knowledge beyond what's explicitly stated.

Recent advancements in Large Language Models have sparked a boom in reasoning, reshaping what's possible in AI. Yet, until now, information retrieval systems have lagged behind, lacking the reasoning capabilities needed to fully support this new paradigm. That gap has finally been bridged.

Reason-ModernColBERT rises to this challenge thanks to LightOn's long-standing commitment to late-interaction architectures. Its results have been made possible thanks to an entire ecosystem purposefully built over time, from pioneering the [PyLate library](#), to developing [ModernBERT](#), and setting new standards with [GTE-ModernColBERT](#). This sustained investment enables us today to unlock game-changing performance in reasoning-driven retrieval with remarkable simplicity. The result: a new model for complex, reasoning-intensive search, powered by an infrastructure designed from the ground up for exactly this purpose.

Breaking New Ground in Reasoning-Intensive Retrieval

- **Small Model, Big Results:** Despite being just 150M parameters (over 45 times smaller than certain competitors), Reason-ModernColBERT outperforms all models up to 7B parameters on BRIGHT, the gold-standard benchmark for reasoning-intensive retrieval. It even outperforms ReasonIR-8B by over 2.5 NDCG@10 on Stack Exchange queries.

- **Lightning-Fast and Streamlined Training:** Built using LightOn's powerful PyLate library, Reason-ModernColBERT has been trained in less than two hours with fewer than 100 lines of code.
- **Late-Interaction Advantage:** Direct comparisons with dense single-vector models, trained on identical data, highlight the consistent, striking lead enabled by late-interaction architecture.

Unlocking the Next Frontier of Research

Reason-ModernColBERT is built to drive advanced knowledge exploration, addressing cases in which questions are nuanced and relevance is often subtle or implicit.

As agentic RAG, advanced document understanding, and domain-specific research become central to enterprise AI, LightOn's new model provides:

- Enhanced retrieval for subtle, implicit, or reasoning-based queries
- Drastically reduced inference latency relative to massive LLMs
- Easy reproducibility and transparency via open-source release

Reaffirming LightOn's Commitment to Open Research

As with our previous models, LightOn is making Reason-ModernColBERT, its training code, and the relevant datasets publicly available. Anyone can freely access, extend, and build upon it, leveraging PyLate to drive the next generation of multi-vector retrieval innovation.

Get Started Today

Reason-ModernColBERT is available now for use and experimentation on Hugging Face through PyLate, with comprehensive documentation and code for easy fine-tuning and deployment. Whether for knowledge management teams, AI developers, or scientific researchers,

Reason-ModernColBERT opens new horizons in the age of Deep Research.

 [Try Reason-ModernColBERT on Hugging Face](#)

About LightOn

Founded in 2016 in Paris and the first European generative AI company listed on Euronext Growth, LightOn is a pioneering player in the field of sovereign GenAI. Its Paradigm platform enables organizations to deploy large-scale AI while ensuring the confidentiality of their data. LightOn's technology ensures essential strategic independence by offering tailored solutions. This technological mastery is accompanied by the ability to process large volumes of data for industrial uses, with applications in various sectors such as finance, industry, health, defense, and public services.

LightOn is listed on Euronext Growth® Paris (ISIN: FR0013230950, ticker: ALTAI-FR). The company qualifies for PEA and PEA PME investment plans and is recognized as an "Innovative Company" by Bpifrance. To learn more : <https://www.lighton.ai>

Contacts

LIGHTON invest@lighton.ai	SEITOSEI•ACTIFIN Investor Relations Benjamin LEHARI lighton@seitosei-actifin.com
KALAMARI Media Relations Camille Bernisson - +33 7 64 44 14 49 Maroua Derdega - +33 7 63 77 73 20 lighton@kalamari.agency	SEITOSEI•ACTIFIN Financial Media Relations Jennifer JULLIA - +33 6 47 97 54 87 jennifer.jullia@seitosei-actifin.com