

LightOn met à niveau la recherche dans le RAG agentique avec son nouveau modèle de pointe Reason-ModernCoBERT

Après la récente sortie de GTE-ModernCoBERT, un modèle dédié à la recherche d'information redéfinissant les standards de l'industrie par sa gestion des documents longs, LightOn annonce une nouvelle avancée dans la recherche d'information pilotée par l'IA : **Reason-ModernCoBERT**, un modèle open-source multi-vecteurs spécialement conçu pour les applications de recherche approfondie (Deep Research).

Redéfinir la recherche d'information : de l'association au raisonnement

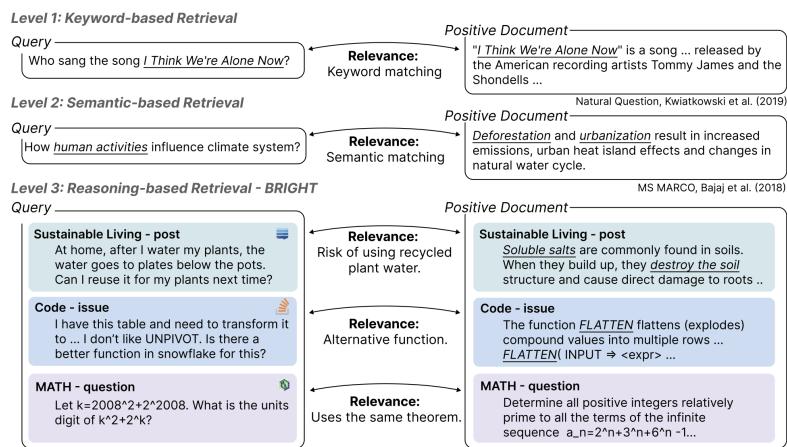


Figure 1: Existing retrieval benchmarks focus on keyword-based retrieval (level 1), or semantic-based retrieval (level 2), e.g., NQ, MS MARCO datasets (Kwiatkowski et al., 2019; Bajaj et al., 2018). **BRIGHT** introduces level 3 retrieval, where the relevance between queries and documents requires intensive 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](#)

Avec l'essor de la Recherche approfondie, de nouveaux cas d'usage émergent et ils exigent plus qu'une simple association lexicale ou sémantique. Pour produire un réponse réellement précis, il faut disposer d'une capacité de raisonnement, de pouvoir connecter, synthétiser et retrouver les données nécessaires, au-delà de ce que peut proposer une requête explicite.

Les avancées récentes des grands modèles de langage (Large Language Model - LLM) ont déclenché un essor des capacités de raisonnement en inférence, redéfinissant ce qui est possible en IA. Pourtant, jusqu'à présent, les systèmes de recherche d'information accusaient un retard, manquant des capacités de raisonnement nécessaires pour accompagner cette révolution. Cet écart a enfin été comblé.

Reason-ModernColBERT relève le défi grâce au travail engagé de longue date par LightOn sur les architectures d'interaction tardive. Ses résultats ont été rendus possibles grâce à tout un écosystème construit au fil du temps, depuis le lancement de la bibliothèque [PyLate](#), de [ModernBERT](#) et plus récemment [GTE-ModernColBERT](#). Cet investissement nous permet aujourd'hui de débloquer des performances révolutionnaires dans la recherche d'information nécessitant du raisonnement, avec une simplicité remarquable. Le résultat : un nouveau modèle pour des recherches complexes et nécessitant du raisonnement, propulsé par une infrastructure conçue précisément pour cet objectif, dès le départ.

Repousser les frontières de la récupération d'information par raisonnement intensif

- **Petit Modèle, Grands Résultats** : Bien qu'il ne compte que 150 millions de paramètres (plus de 45 fois moins que certains concurrents), Reason-ModernColBERT surpasse tous les modèles allant jusqu'à 7 milliards de paramètres sur BRIGHT, benchmark de référence pour la recherche d'information nécessitant du raisonnement (Reasoning-Intensive Retrieval). Il surpasse même ReasonIR-8B de plus de 2,5 NDCG@10 sur les requêtes de Stack Exchange.
- **Entraînement Ultra-Rapide et Simplifié** : Construit à l'aide de la puissante bibliothèque PyLate de LightOn, Reason-ModernColBERT a été entraîné en moins de deux heures avec moins de 100 lignes de code.

- **Avantages de l'Interaction Tardive** : Des comparaisons directes avec des modèles à vecteur unique dense, entraînés sur des données identiques, mettent en évidence l'avantage frappant et constant que permet l'architecture d'interaction tardive.

Ouvrir une nouvelle ère dans la recherche d'informations

Reason-ModernColBERT a été conçu pour parcourir et retrouver des bases de connaissances complexes, apportant des réponses précises aux questions nuancées et quand la pertinence sous-jacente est subtile ou implicite.

Alors que le RAG agentique, la compréhension avancée des documents et la recherche de données spécifiques deviennent centrales pour l'IA d'entreprise, le nouveau modèle de LightOn offre :

- L'amélioration de la recherche pour les requêtes subtils, implicites ou basées sur le raisonnement
- Une réduction drastique de la latence d'inférence par rapport aux LLMs massifs
- Une reproductibilité facilitée et transparente grâce sa disponibilité en open source

LightOn s'engage pour l'open source

Comme pour les précédents modèles, LightOn partage Reason-ModernColBERT, son code d'entraînement et les jeux de données pertinents en accès libre. Tout le monde peut y accéder, l'étendre et le développer, en utilisant PyLate pour propulser la nouvelle génération d'innovation dans la recherche d'information multi-vecteurs.

Commencez dès Aujourd'hui

Reason-ModernColBERT est désormais disponible pour l'utilisation et l'expérimentation sur Hugging Face via PyLate, avec une documentation complète et un code facilitant le fine-tuning et le déploiement. Que ce soit pour les équipes Data, les développeurs IA ou les chercheurs,

Reason-ModernColBERT ouvre de nouveaux horizons à l'ère du Deep Research

 [Essayez Reason-ModernColBERT sur Hugging Face](#)

 [Documentation PyLate](#)

À propos de LightOn

Fondée en 2016 à Paris et première société européenne de l'IA Générative cotée sur Euronext Growth, LightOn est un acteur pionnier dans le domaine de la GenAI souveraine. Sa plateforme Paradigm permet aux organisations de déployer des IA à grande échelle, tout en garantissant la confidentialité de leurs données. La technologie de LightOn assure une indépendance stratégique essentielle, offrant ainsi des solutions sur mesure. Cette maîtrise technologique s'accompagne d'une capacité à traiter d'importants volumes de données pour des usages industriels, avec des applications dans des secteurs variés tels que la finance, l'industrie, la santé, la défense et les services publics.

LightOn est cotée sur Euronext Growth® Paris (ISIN : FR0013230950, mnémonique : ALTAI-FR). La société est éligible au PEA et au PEA PME, et est qualifiée « Entreprise innovante » par Bpifrance.

Pour en savoir plus : <https://www.lighton.ai/fr>

Contacts

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