



**Communiqué de presse**

Paris, le 7 mai 2025

## **LightOn redéfinit les standards de la recherche d'information complexe (RAG) avec GTE-ModernColBERT.**

**Paris, le 7 mai 2025** — Levier d'innovation clé pour les entreprises, les IA génératives sont confrontées à la complexité et la spécificité des données internes aux organisations. Aujourd'hui LightOn annonce la sortie de GTE-ModernColBERT, un modèle dédié à la recherche d'information redéfinissant les standards de l'industrie par sa gestion des documents longs et sa capacité de généralisation sur données complexes inconnues. Ce nouveau modèle surclasse ceux des acteurs de l'écosystème (Alibaba, Cohere, Snowflake, BAAI, JinaAI) sur le benchmark de référence LongEmbed.

### **Une innovation majeure pour la recherche documentaire.**

GTE-ModernColBERT, est le premier modèle d'« interaction tardive », multi-vecteurs de pointe, entièrement développé grâce à l'architecture ModernBERT et la bibliothèque PyLate.

Jusqu'ici, la plupart des solutions de génération par recherche augmentée (Retrieval Augmented Generation - RAG) reposaient sur des modèles à vecteur unique. Modèles de références pour un usage courant, ils tendent à montrer des limites dès qu'il s'agit de traiter des contextes longs ou des bases documentaires spécialisées, comme c'est le cas des entreprises et établissements publics qui manipulent des volumes massifs et hétérogènes de données nouvelles.

### **Une capacité de généralisation hors norme pour les longs documents.**

GTE-ModernColBERT établit une nouvelle référence (SOTA — State of the Art) sur la généralisation avec contextes longs. Il surpasse de 10 points les meilleurs modèles (benchmark

LongEmbed) sur des documents allant jusqu'à 32 000 tokens, soit des textes de plusieurs dizaines de pages, alors même qu'il n'a été initialement entraîné que sur des extraits de 300 tokens issus du jeu de données MS MARCO. Ces premiers tests montrent que GTE-ModernColBERT pourrait étendre ses capacités de manière encore plus ambitieuse, offrant d'excellentes performances même au-delà de cette fenêtre contextuelle.

- **Gestion avancée du contexte** : capacité à traiter des documents atteignant 32 000 tokens,
- **Adaptation aux contenus spécialisés** : capacité de généralisation supérieure pour des contenus spécialisés, confidentiels, et des corpus peu standardisés,
- **Performances révolutionnaires** : GTE-ModernColBERT est le premier modèle à surpasser ColBERT-small sur le benchmark de référence BEIR.
- **Efficacité architecturale** : grâce à ModernBERT, le modèle allie puissance et optimisation des ressources, répondant aux exigences opérationnelles.

## Démocratiser le déploiement du Retrieval Augmented Generation (RAG) en entreprise.

GTE-ModernColBERT transforme en profondeur la manière dont les organisations peuvent déployer la génération augmentée par recherche d'information (Retrieval-Augmented Generation, RAG) en :

- Améliorant la qualité de la recherche au sein des bases de connaissances propriétaires ;
- Maintenant des performances élevées même sur des contenus hautement spécialisés ;
- Permettant un traitement massif de documents à l'échelle de l'entreprise ;
- Améliorant significativement la pertinence des réponses générées par IA ;

## Un impact concret.

Les équipes de gestion des bases documentaires et les développeurs de solutions IA disposent désormais d'un socle robuste pour bâtir des services d'information de nouvelle génération. GTE-ModernColBERT se révèle particulièrement puissant pour :

- L'analyse automatisée de documents juridiques.
- L'extraction intelligente dans des corpus scientifiques.
- L'accès facilité à la documentation technique.
- L'optimisation des bases de connaissances pour le support client
- Le pilotage efficient de l'information stratégique en entreprise.

## Savoir-faire technologique de LightOn.

LightOn a développé GTE-ModernColBERT en construisant une solution unique à partir d'éléments clés :



1. **Encodeur moderne** : LightOn a conçu ModernBERT afin de permettre la création de modèles de recherche puissants et à la pointe. GTE-ModernColBERT est la suite directe de cette première version pour exploiter le potentiel très prometteur de l'approche multi-vecteurs.
2. **Bibliothèque PyLate** : Nous avons développé un framework permettant de faciliter la mise en œuvre, l'expérimentation et l'entraînement de modèles de recherche multi-vecteurs. Seules 80 lignes de code sont nécessaires pour reproduire le processus d'entraînement de ce modèle.
3. **Distillation des connaissances** : En entraînant notre modèle sur MS MARCO via la distillation de connaissances, nous avons obtenu un modèle léger mais très performant, sans compromis sur l'efficacité.
4. **Focus sur la compatibilité** : La plupart des bases de données vectorielles majeures telles que QDrant, LanceDB, Weaviate et Vespa prennent désormais en charge l'indexation multi-vecteurs, facilitant ainsi l'adoption en entreprise.

## Engagement pour l'open source et la souveraineté technologique

Après la sortie de ModernBERT et ModernBERT-embed, la publication de GTE-ModernColBERT en open source illustre l'engagement continu de LightOn pour l'avancement de l'IA, tout en permettant à des organisations de toutes tailles de bénéficier des technologies de recherche de pointe et en favorisant la recherche grâce à la mise à disposition libre de PyLate.

Poursuivant sa politique d'innovation ouverte, LightOn publie GTE-ModernColBERT sous licence open source Apache 2.0, dans la lignée de ModernBERT et ModernBERT-embed. Cette démarche s'inscrit dans l'engagement continu de LightOn pour le développement libre de l'IA, tout en permettant à des organisations de toutes tailles de bénéficier des technologies de recherche de pointe et en favorisant la recherche grâce à la mise à disposition libre de PyLate.

Pour les organisations souhaitant rester à la pointe de la gestion de la connaissance et du RAG, GTE-ModernColBERT est désormais disponible. Testez-le et (re)découvrez la valeur cachée de vos documents !

 [Essayez-le dès aujourd'hui sur HuggingFace](#)  
 [Pour commencer : Documentation PyLance](#)

## 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

|   |  |
|---|--|
| <b>LIGHTON</b><br><a href="mailto:invest@lighton.ai">invest@lighton.ai</a>  | <b>SEITOSEI•ACTIFIN</b><br><b>Relations investisseurs</b><br>Benjamin LEHARI<br><a href="mailto:lighton@seitosei-actifin.com">lighton@seitosei-actifin.com</a>   |
| <b>KALAMARI</b><br><b>Relations médias</b><br>Camille Bernisson - +33 7 64 44 14 49<br>Maroua Derdega - +33 7 63 77 73 20<br><a href="mailto:lighton@kalamari.agency">lighton@kalamari.agency</a> | <b>SEITOSEI•ACTIFIN</b><br><b>Relations presse financière</b><br>Jennifer JULLIA - +33 6 47 97 54 87<br><a href="mailto:jennifer.jullia@seitosei-actifin.com">jennifer.jullia@seitosei-actifin.com</a> |