



Communiqué de presse

Paris, le 8 octobre 2026

LightOn dévoile LightOnOCR-3, une famille de modèles de référence mondiale en intelligence documentaire

LightOn présente LightOnOCR-3, une famille de modèles d'IA qui affiche des résultats parmi les meilleurs au monde sur les benchmarks, devant les modèles concurrents américains et chinois, avec des performances à l'état de l'art sur les documents en français et la reconnaissance de l'écriture manuscrite. Dans la continuité de LightOnOCR-2, ces modèles vont au-delà de la reconnaissance de texte pour décrire les images, extraire les données des graphiques et préserver la structure des documents. Disponibles en versions de 0,8 milliard et 4 milliards de paramètres sous licence Apache 2.0, ils permettent aux organisations de transformer des documents complexes en informations exploitables, pour une fraction du coût des modèles concurrents.

Des performances parmi les meilleures au monde sur les benchmarks de reconnaissance documentaire

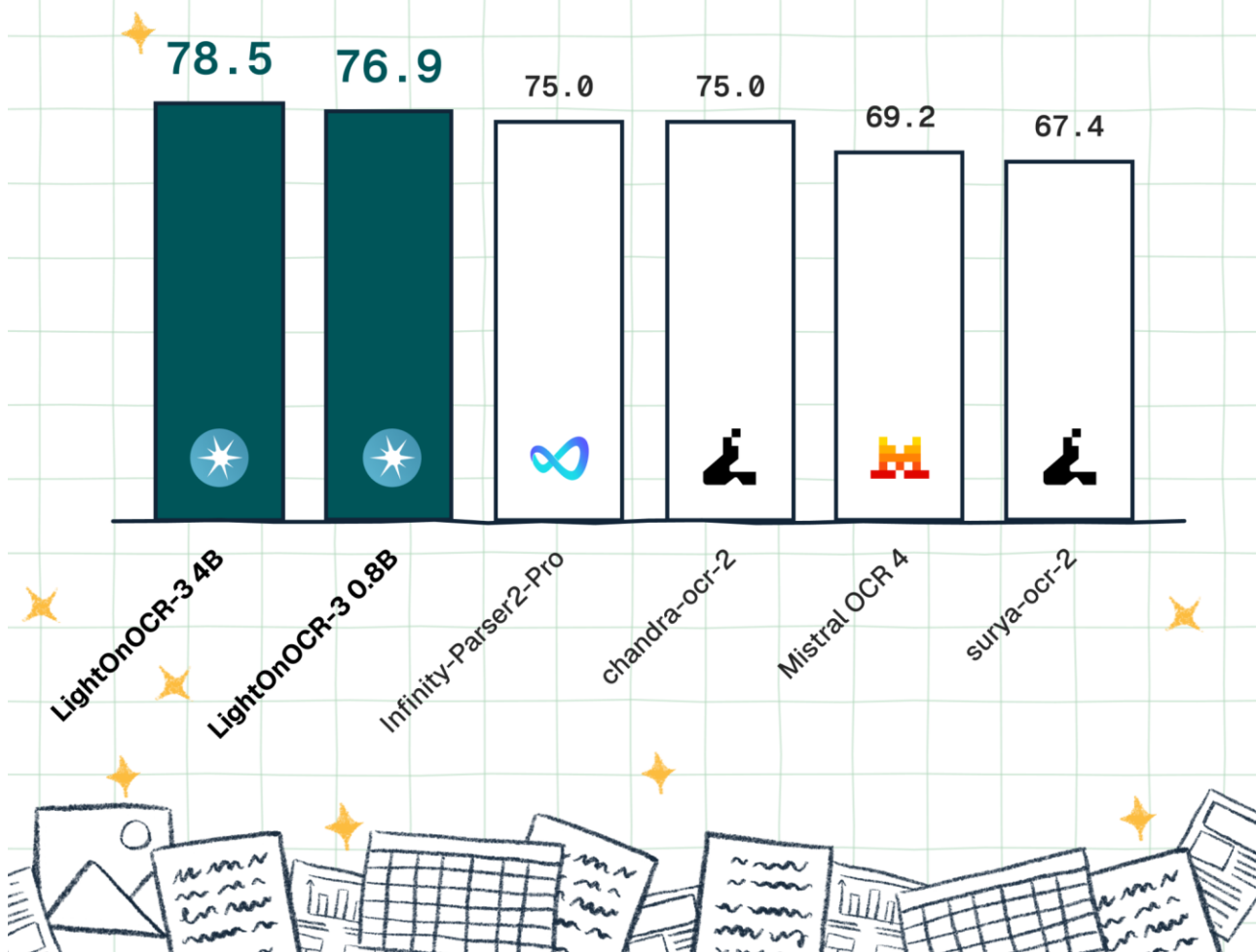
Fort du succès de LightOnOCR-2, qui enregistre 240 000 téléchargements par mois sur Hugging Face, LightOn présente aujourd'hui LightOnOCR-3, une nouvelle génération qui enrichit ses capacités d'intelligence documentaire.

LightOnOCR-3 se classe en première position sur un ensemble de trois benchmarks ouverts évaluant la reconnaissance de texte et l'interprétation des informations visuelles, devant les modèles OCR existants en Europe, aux États-Unis et en Chine.



LightOn OCR-3

Moyenne sur olmOCR-Bench, ParseBench et fr-bench-pdf2md



Sur OlmOCR-Bench, le benchmark de référence du secteur pour l'extraction de texte documentaire, il se classe deuxième derrière Infinity-Parser2-Pro, un modèle plus de huit fois plus volumineux, et devant Chandra-OCR-2, Mistral OCR 4.1 et dots.mocr. Il arrive en tête des modèles à poids ouverts sur ParseBench, qui évalue la restitution de la mise en page, la description d'images et l'extraction de données structurées à partir de graphiques et de schémas scientifiques, démontrant ainsi ses capacités renforcées en vision.

Il se classe également premier sur FRBench-pdf2md, consacré aux documents en français, avec les gains les plus importants en reconnaissance de l'écriture manuscrite.



Les scores détaillés et les comparaisons sont disponibles dans l'article technique associé.

[Lire l'article technique](#)

Au-delà de la reconnaissance de texte : rendre davantage d'informations documentaires exploitables

Les rapports financiers, les dossiers administratifs et les publications techniques transmettent des informations essentielles à travers le texte, les tableaux, les graphiques, les images et la mise en page. Lorsque le traitement documentaire ne capture que le texte, une partie du contenu est perdue, limitant ce que les applications d'IA peuvent retrouver et exploiter.

LightOnOCR-3 traite ces contenus disparates en une seule étape. Il transcrit le texte, identifie la nature des éléments du document et leurs positions sur la page, décrit les images et extrait des données structurées à partir de graphiques et de schémas scientifiques. Il produit ainsi un contenu structuré, prêt à être exploité pour la recherche, l'analyse et l'automatisation documentaire.

Ces capacités répondent à des besoins concrets de traitement documentaire :

- **Analyse financière** : l'extraction de données à partir du texte, des tableaux et des graphiques permet aux équipes d'exploiter davantage d'informations issues des rapports financiers au sein d'un même processus.
- **Processus administratifs** : grâce à des capacités de transcription renforcées par rapport aux générations précédentes, notamment pour les documents en français et l'écriture manuscrite, LightOnOCR-3 facilite la conversion de formulaires et de courriers difficiles à traiter par les OCR traditionnels en contenus directement exploitables
- **Recherche dans les documents techniques** : ces modèles restituent la structure des documents, car ils identifient les sections, la structure des pages et l'emplacement des informations. Ces nouvelles capacités facilitent la recherche de contenus pertinents et permettent de les relier précisément à leur emplacement dans le document d'origine.

Les contenus obtenus peuvent être vérifiés, indexés ou transmis à des applications métier, réduisant ainsi le travail de traitement nécessaire à leur utilisation.

Des modèles compacts pour un traitement plus rapide et moins cher



Communiqué de Presse

Proposant deux versions de 0,8 milliard et 4 milliards de paramètres, LightOnOCR-3 permet aux organisations de choisir le modèle adapté à leurs ressources de calcul et à leurs besoins de traitement documentaire.

Les modèles LightOnOCR-3 figurent parmi les plus rapides du marché: ils peuvent traiter les documents jusqu'à deux fois plus vite que LightOnOCR-2. Ils présentent également les coûts d'utilisation les plus faibles parmi les principaux modèles OCR. Lorsqu'ils sont déployés sur l'infrastructure propre d'une organisation, ces coûts peuvent descendre sous la barre du centime pour mille pages, selon le matériel, le niveau d'utilisation et la charge de travail.

Les équipes peuvent opter pour une simple transcription du texte ou pour une restitution plus riche, intégrant la structure du document et les informations visuelles. La combinaison de ces capacités dans un seul modèle réduit le besoin d'assembler des outils distincts pour la reconnaissance de texte, l'analyse de la mise en page et le traitement des images.

Distribué sous la licence permissive Apache 2.0, LightOnOCR-3 peut être utilisé à des fins commerciales, modifié et redistribué. Les organisations peuvent le déployer sur leur propre infrastructure, l'adapter à leurs besoins et l'intégrer à leurs applications.

Pour en savoir plus : [Lire l'article technique](#)

À propos de LightOn

Fondée en 2016 à Paris et première société européenne de l'IA cotée sur Euronext Growth, LightOn développe une plateforme d'IA d'entreprise conçue pour permettre aux organisations de connecter une IA de pointe sur leurs données sensibles. LightOn propose une architecture intégrée, pensée pour le passage en production à grande échelle, robuste, frugale et sécurisée, permettant d'industrialiser des cas d'usage dans des environnements régulés. Les solutions LightOn s'adressent notamment aux secteurs de la finance, de l'industrie, de la santé, de la défense et du secteur public.

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/>





Communiqué de Presse

Contacts

LIGHTON invest@lighton.ai	
SEITOSEI•ACTIFIN Relations investisseurs Benjamin LEHARI lighton@seitosei-actifin.com	SEITOSEI•ACTIFIN Relations presse financière Jennifer JULLIA - +33 6 47 97 54 87 jennifer.jullia@seitosei-actifin.com