



Press release

Paris, 08 October 2026

LightOn releases LightOnOCR-3, a world-leading model family for document intelligence

LightOn introduces LightOnOCR-3, a family of AI models that delivers world-leading benchmark results, ahead of competing US and Chinese models, with state-of-the-art performance on French-language documents and handwriting. Building on LightOnOCR-2, it goes beyond text recognition to describe images, extract data from charts and preserve document structure. Available in 0.8-billion and 4-billion-parameter versions under the Apache 2.0 license, it helps organizations turn complex documents into usable information at a fraction of the cost of competing models.

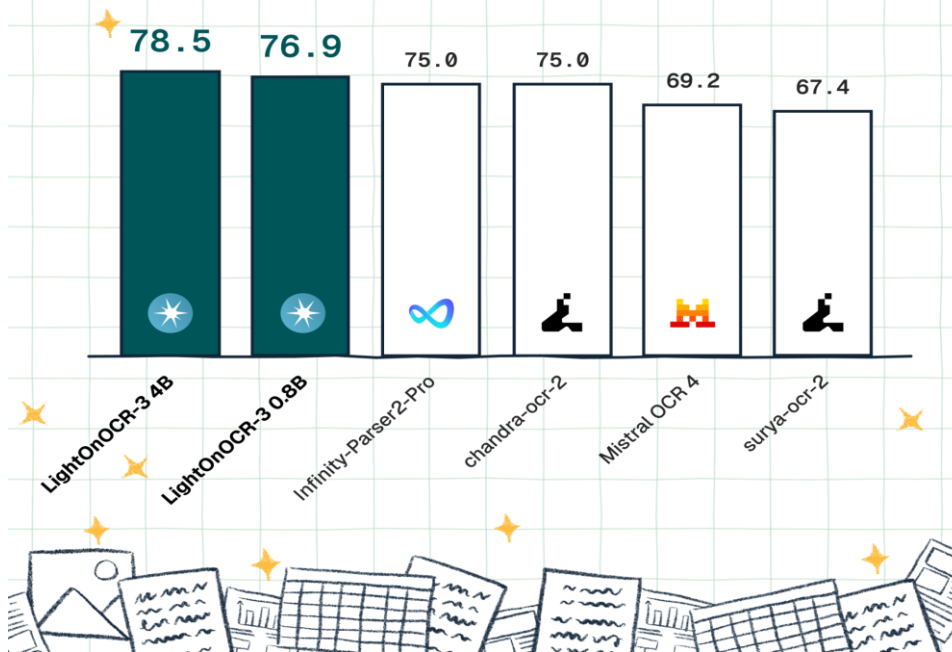
World-leading performance on document recognition benchmarks

Building on the success of LightOnOCR-2, which sees 240,000 monthly downloads on Hugging Face, LightOn today introduces LightOnOCR-3, a new generation that extends its document intelligence capabilities.

LightOnOCR-3 delivers leading results across aggregated benchmarks for text recognition and the interpretation of visual information, ahead of existing OCR models across Europe, the United States and China.

LightOn OCR-3

Average of olmOCR-Bench, ParseBench and fr-bench-pdf2md



On OlmOCR-Bench, the industry-leading OCR benchmark which evaluates document text extraction, it ranks second behind Infinity-Parser2-Pro, a model more than eight times its size, and ahead of Chandra-OCR-2, Mistral OCR 4.1 and dots.mocr. It leads open-weight models on ParseBench, which evaluates document layout, image descriptions and structured data extraction from charts and scientific figures - demonstrating its enhanced vision capabilities.

It also ranks first on FRBench-pdf2md, which focuses on French-language documents, with the largest gains in handwriting recognition.

Detailed scores and comparisons are available in the accompanying technical article.

[Read the technical article](#)

Beyond text recognition: making more document information usable

Financial reports, administrative files and technical publications convey essential information through text, tables, charts, images and page layouts. When document processing captures only text, some of that information is lost, limiting what AI applications can retrieve and use.

LightOnOCR-3 brings these different types of information together in a single processing step. It recognizes text, identifies document elements and their location coordinates on the page, describes images and extracts structured data from charts and scientific figures. All of this information can then be used downstream in a wide variety of contexts.

These capabilities address concrete document processing needs:

- **Financial analysis:** extracting information from text, tables and charts helps teams use more of the content in financial reports within the same workflow.
- **Administrative workflows:** stronger recognition of French-language documents and handwriting helps teams turn handwritten forms and correspondence into searchable content and data for further processing.
- **Technical document search:** preserving sections, page structure and the location of supporting information helps applications retrieve relevant content and connect it to the original document.

The resulting content can be checked, indexed or passed to business applications, reducing manual preparation before it can be used.

Compact models for faster processing and lower costs

With 0.8-billion and 4-billion-parameter versions, LightOnOCR-3 gives organizations a choice of models to suit their computing resources and document processing needs.

With their compact size, LightOnOCR-3 models are among the fastest on the market: they can process documents up to twice as fast as LightOnOCR-2. They are also the cheapest among leading OCR models. When run on an organization's own infrastructure, processing costs can fall below one cent per thousand pages, depending on hardware, usage levels and workload.

Teams can choose straightforward text transcription or a richer output that includes document structure and visual information. Combining these capabilities in a single model reduces the need to assemble separate tools for text recognition, layout analysis and image processing.

Released under the permissive Apache 2.0 license, LightOnOCR-3 is available for commercial use, modification and redistribution. Organizations can deploy it on their own infrastructure, adapt it to their needs and integrate it into their applications.

About LightOn

Founded in Paris in 2016, and the first European AI company listed on Euronext Growth, LightOn develops an enterprise AI platform designed to enable organizations to connect cutting-edge AI to their sensitive data. LightOn offers an integrated architecture built for large-scale production deployment, robust, efficient, and secure, allowing organizations to industrialize use

cases in regulated environments. LightOn's solutions are intended in particular for the finance, industrial, healthcare, defense, and public sectors.

LightOn is listed on Euronext Growth® Paris (ISIN: FR0013230950, ticker: ALTAI-FR). The company is eligible for PEA and PEA-PME investment schemes and has been 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	SEITOSEI•ACTIFIN Financial Media Relations Jennifer JULLIA - +33 6 47 97 54 87 jennifer.jullia@seitosei-actifin.com