AI for business. In-House. Private. Secure.

About Integral Business Intelligence

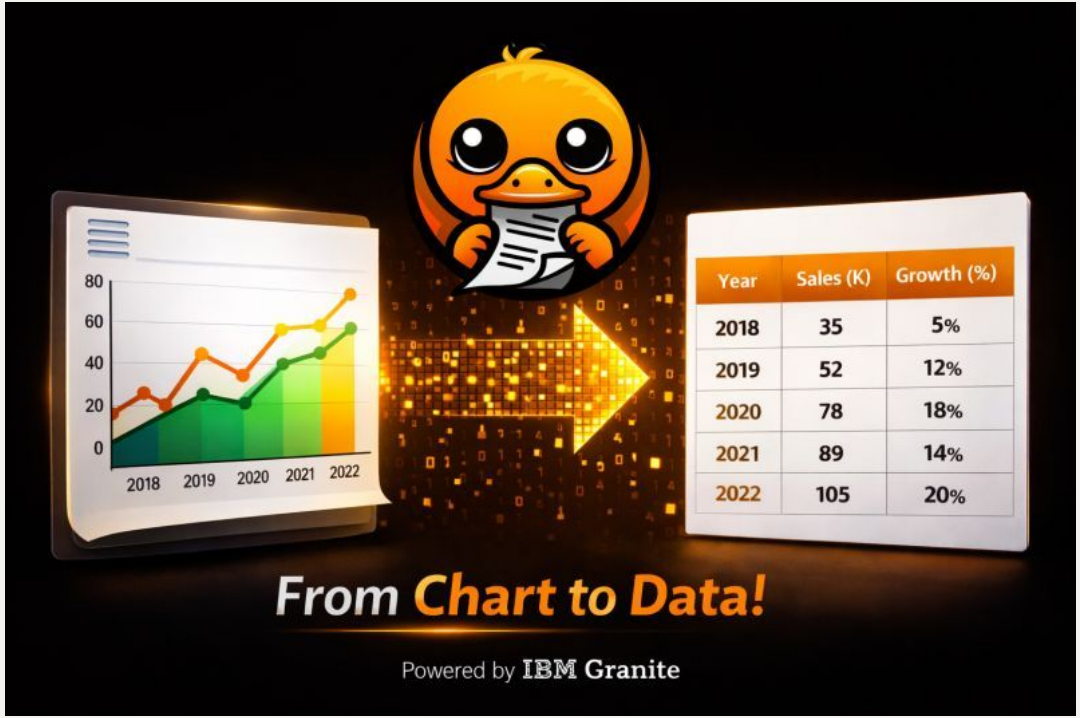
Tony Sclafani
Co-Founder / Partner

https://integralbi.ai

- ✅ AI development & systems integration
- ✅ Ideal customers: small and medium businesses with sensitive data
- ✅ Extract, transform, load (ETL) sensitive data into in-house AI platforms
  - ✅ Paper-to-Digital
  - ✅ Digital-to-Database

In-House AI for Client-Confidential Data

- Text
- Document structure
- Tables
- Charts
- Images



In-House AI for Client-Confidential Data

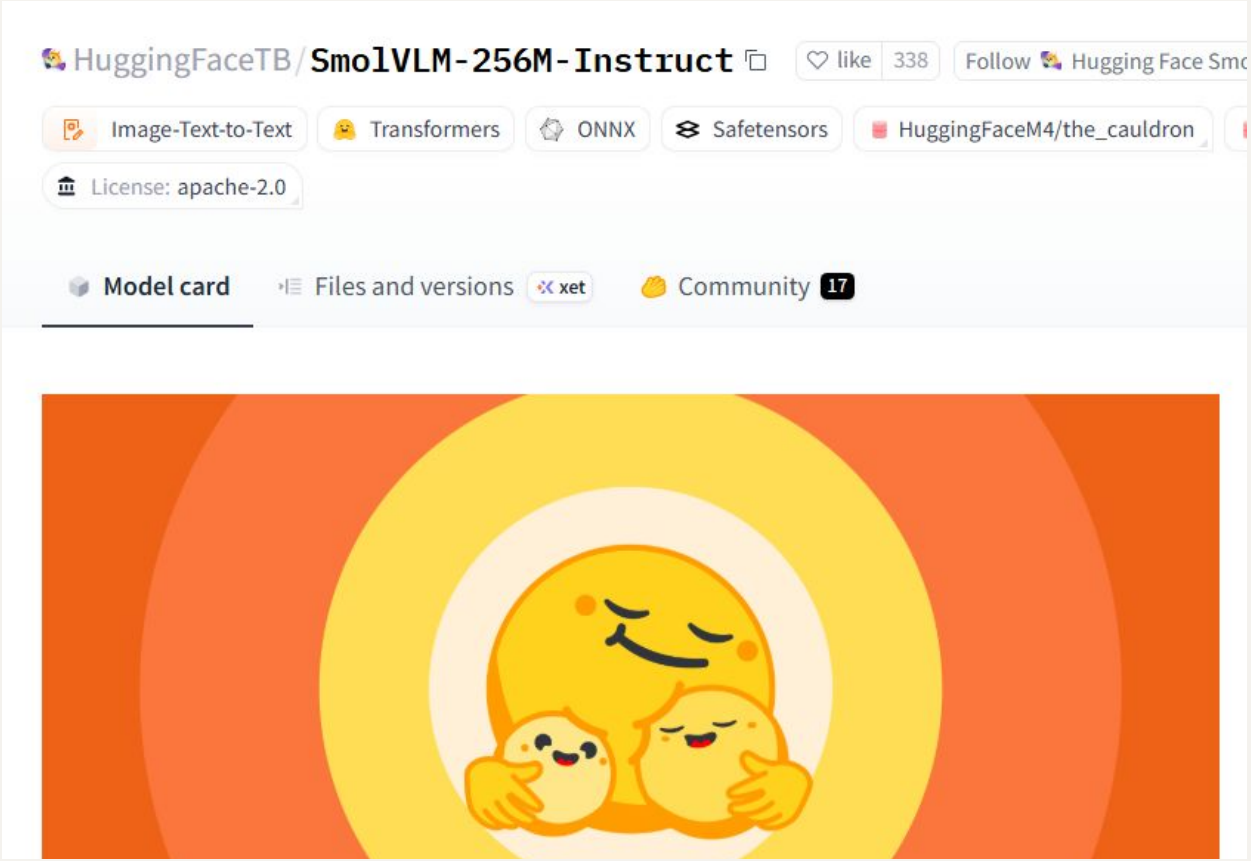
Marker
by datalab.to

Docling
by IBM

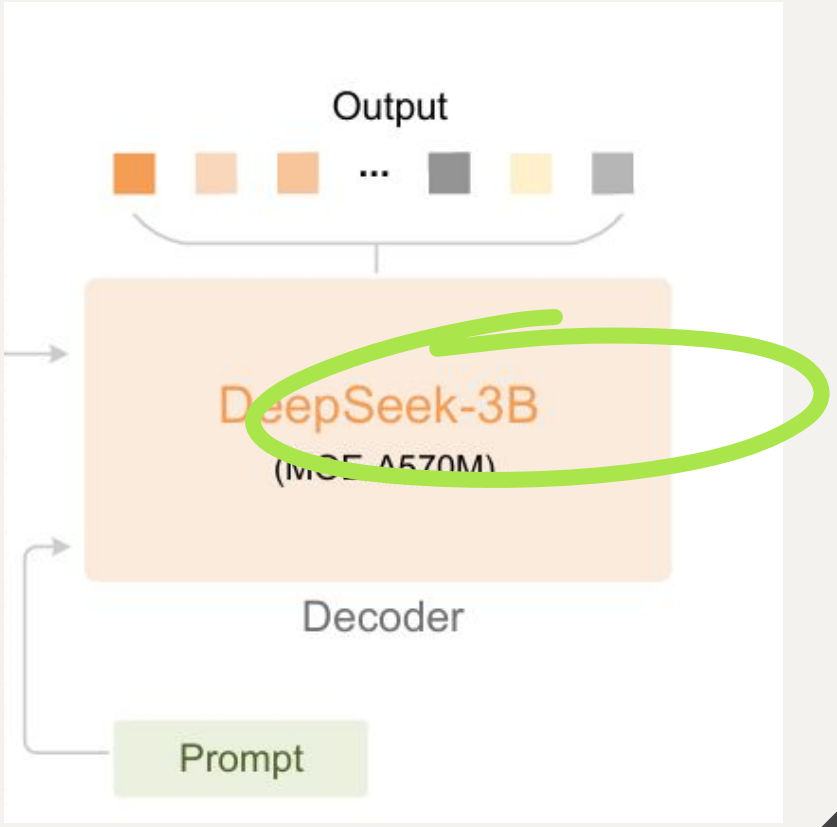
DeepSeek-OCR
by DeepSeek

Kreuzberg

Limitations



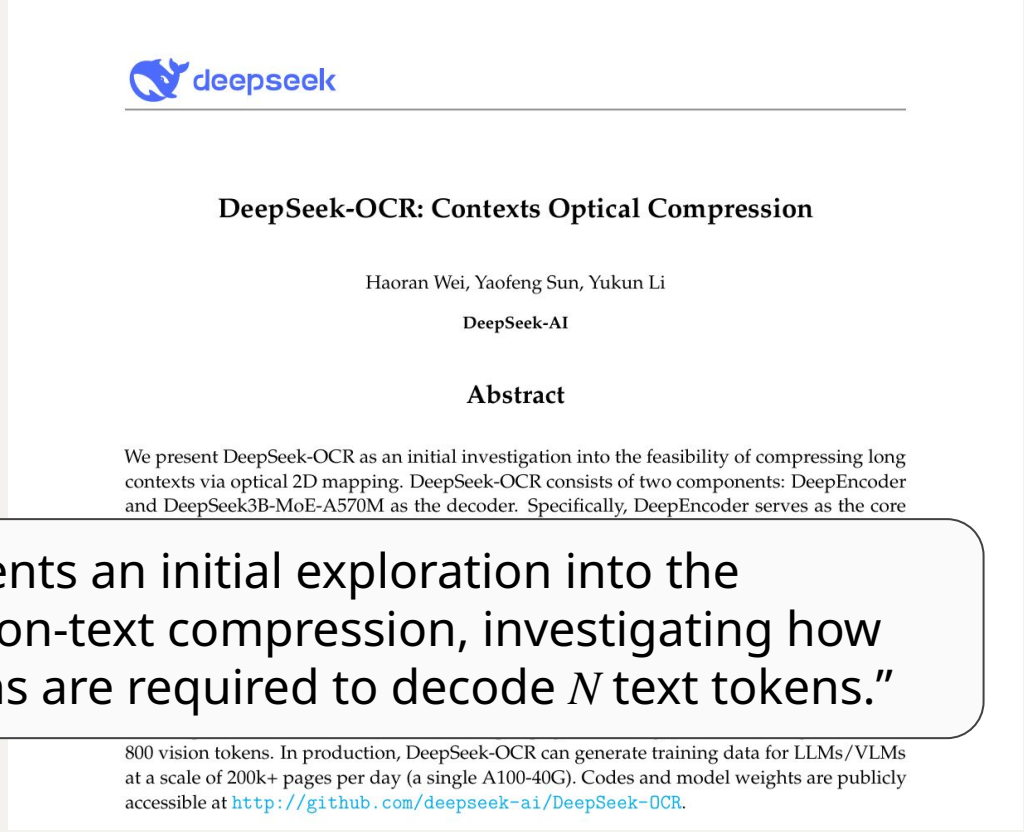
Overcoming Limitations



New Developments

Model	Release	Research Purpose
DeepSeek-OCR	October 21, 2025	Compression
DeepSeek-OCR-2	January 28, 2026	Reading Order

DeepSeek-OCR



deepseek

DeepSeek-OCR: Contexts Optical Compression

Haoran Wei, Yaofeng Sun, Yukun Li

DeepSeek-AI

Abstract

We present DeepSeek-OCR as an initial investigation into the feasibility of compressing long contexts via optical 2D mapping. DeepSeek-OCR consists of two components: DeepEncoder and DeepSeek3B-MoE-A570M as the decoder. Specifically, DeepEncoder serves as the core

800 vision tokens. In production, DeepSeek-OCR can generate training data for LLMs/VLMs at a scale of 200k+ pages per day (a single A100-40G). Codes and model weights are publicly accessible at http://github.com/deepseek-ai/DeepSeek-OCR.

"Our work represents an initial exploration into the boundaries of vision-text compression, investigating how many vision tokens are required to decode N text tokens."

DeepSeek-OCR

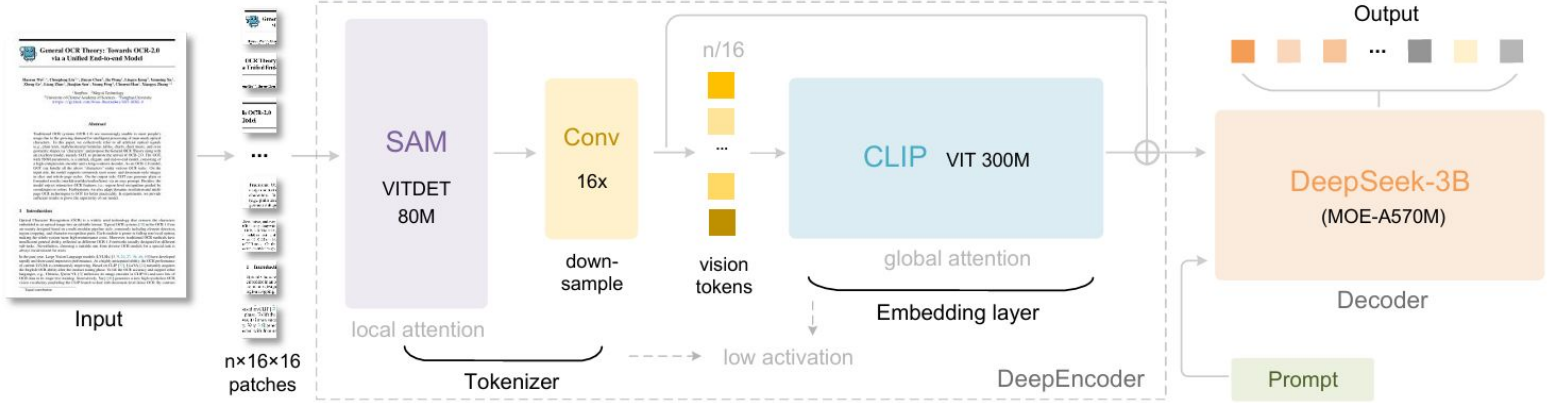
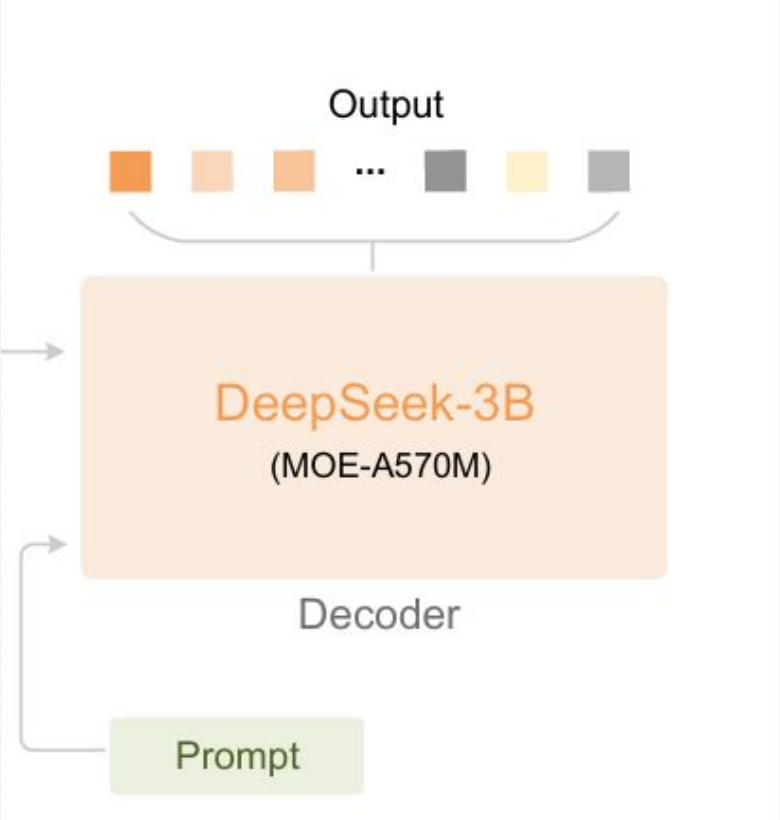


Figure 3 | The architecture of DeepSeek-OCR. DeepSeek-OCR consists of a DeepEncoder and a DeepSeek-3B-MoE decoder. DeepEncoder is the core of DeepSeek-OCR, comprising three components: a SAM [17] for perception dominated by window attention, a CLIP [29] for knowledge with dense global attention, and a 16× token compressor that bridges between them.

2 Stage Training - DeepSeek-OCR

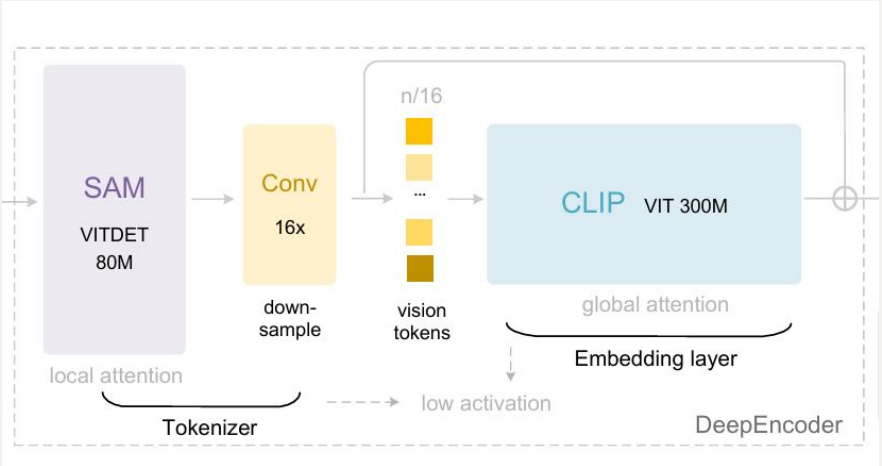
- Stage 1 - DeepEncoder
- Stage 2 - “Full Model” / Full Pipeline

Decoder



DeepEncoder

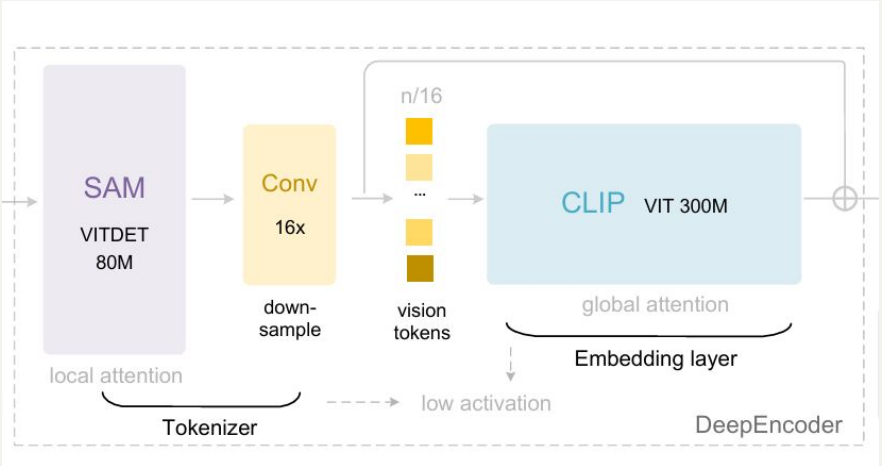
- SAM
  - Perception ("where are things?")
  - Image-to-embeddings
  - 80M params
  - Local attention
  - 1024×1024 image yields 4096 patch tokens
- 16× Convolutional Compressor
  - Token reduction
  - Hard-coded 16x
  - Embeddings-to-less-embeddings
- CLIP
  - Knowledge enrichment ("what do things mean?")
  - Global attention
  - Embeddings-to-embeddings



Compression

- SAM
  - 1024×1024 image
  - 16px tiles
  - 64 patches x 64 patches
  - 1 token per patch
  - Total 4096 patch tokens

- 16× Convolutional Compressor
  - Token reduction
  - Hard-coded 16x
  - Compress 4096 tokens to 256



1. Are the output tokens any good?
2. Input resolution is user-adjustable

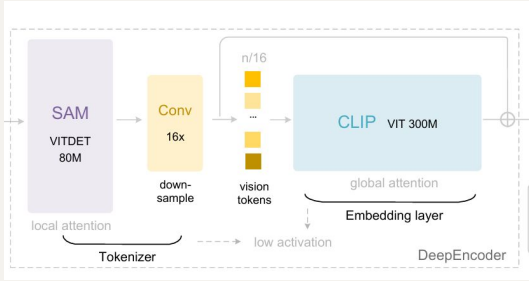
Are the output tokens any good?

Table 2 | We test DeepSeek-OCR's vision-text compression ratio using all English documents with 600-1300 tokens from the Fox [21] benchmarks. Text tokens represent the number of tokens after tokenizing the ground truth text using DeepSeek-OCR's tokenizer. Vision Tokens=64 or 100 respectively represent the number of vision tokens output by DeepEncoder after resizing input images to 512×512 and 640×640.

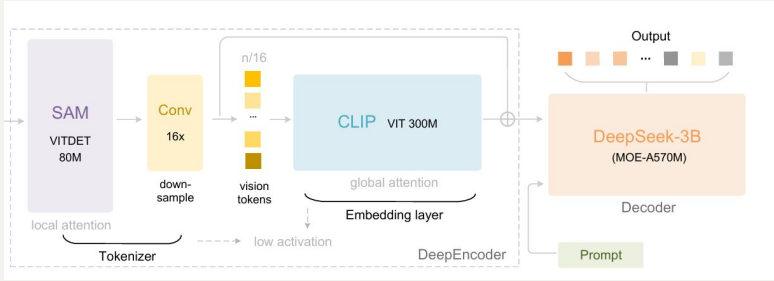
Text Tokens	Vision Tokens =64		Vision Tokens=100		
	Precision	Compression	Precision	Compression	Pages
600-700	96.5%	10.5×	98.5%	6.7×	7
700-800	93.8%	11.8×	97.3%	7.5×	28
800-900	83.8%	13.2×	96.8%	8.5×	28
900-1000	85.9%	15.1×	96.8%	9.7×	14
1000-1100	79.3%	16.5×	91.5%	10.6×	11
1100-1200	76.4%	17.7×	89.8%	11.3×	8
1200-1300	59.1%	19.7×	87.1%	12.6×	4

2 Stage Training - DeepSeek-OCR

- Stage 1 - DeepEncoder
  - ✅ SAM: Training
  - ✅ Compressor: Training
  - ✅ CLIP: Training
  - ❌ DeepSeek-3B-MoE decoder: Omitted



- Stage 2 - “Full Model” / Full Pipeline
  - ❌ SAM: Frozen
  - ❌ Compressor: Frozen
  - ✅ CLIP: Training
  - ✅ DeepSeek-3B-MoE decoder: Training



Research Conclusions

- …with a 10× compression ratio, the model’s decoding precision can reach approximately 97%

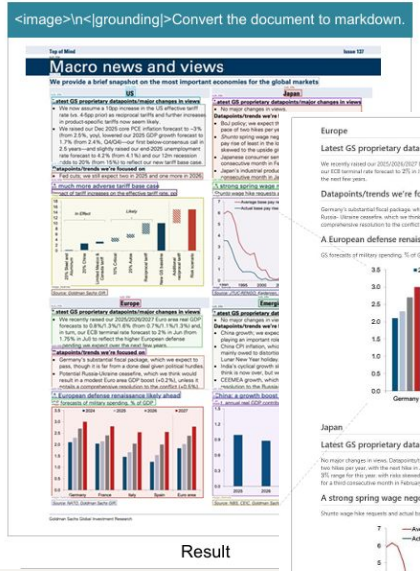
- When compressing tokens by nearly 20×, we find that precision can still approach 60%

Applications - DeepSeek-OCR

- OCR
- LLM/VLM pretraining
- “Deep Parsing”

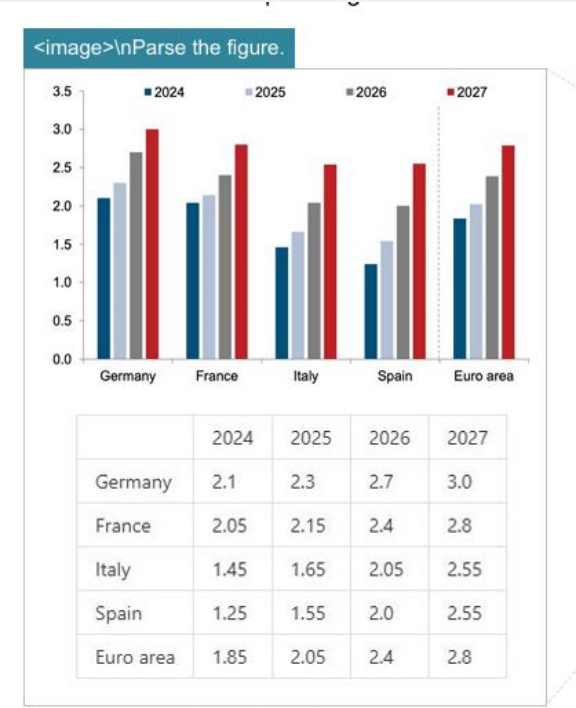
Deep Parsing

Input image



Result

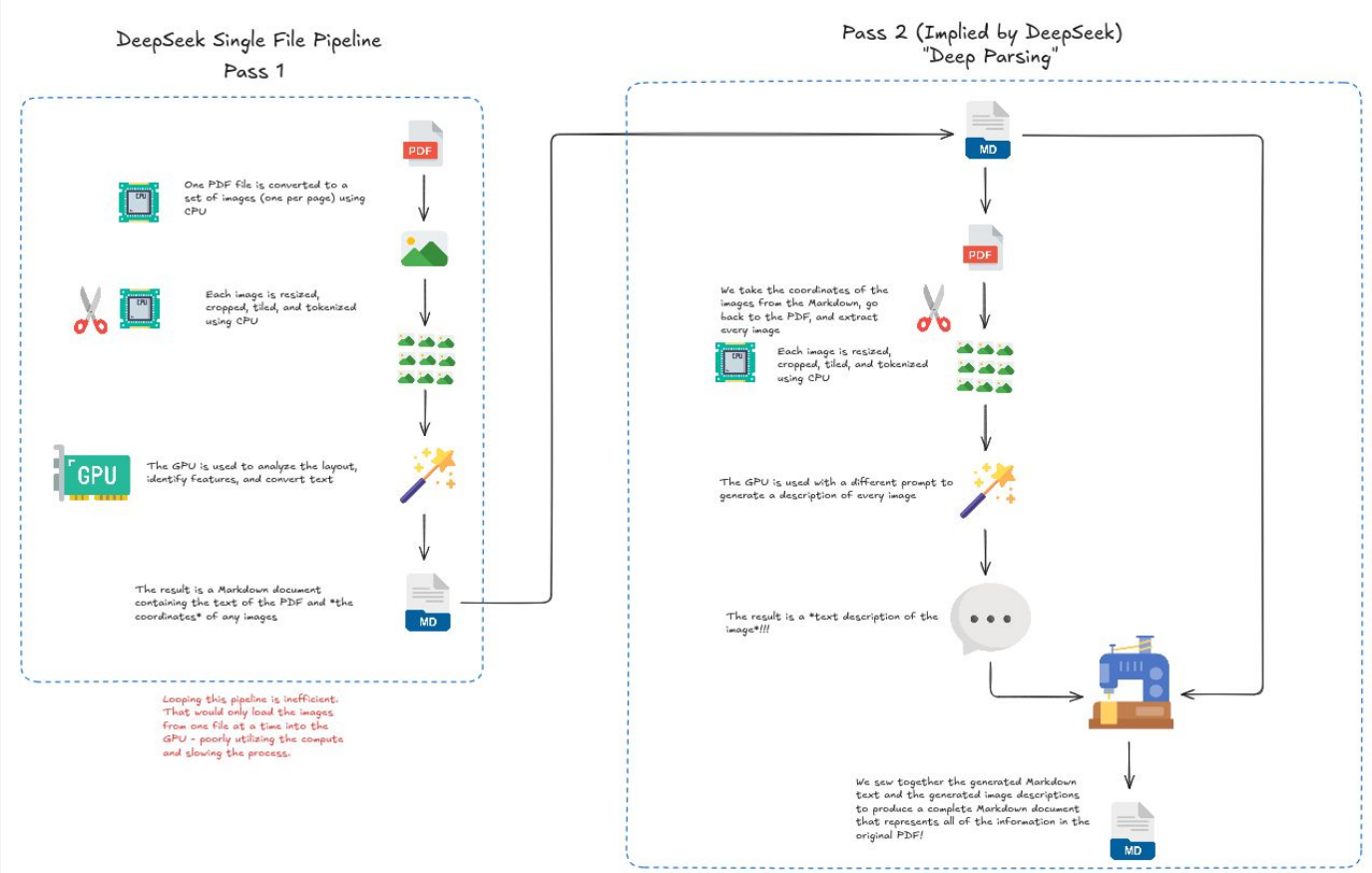
Rendering



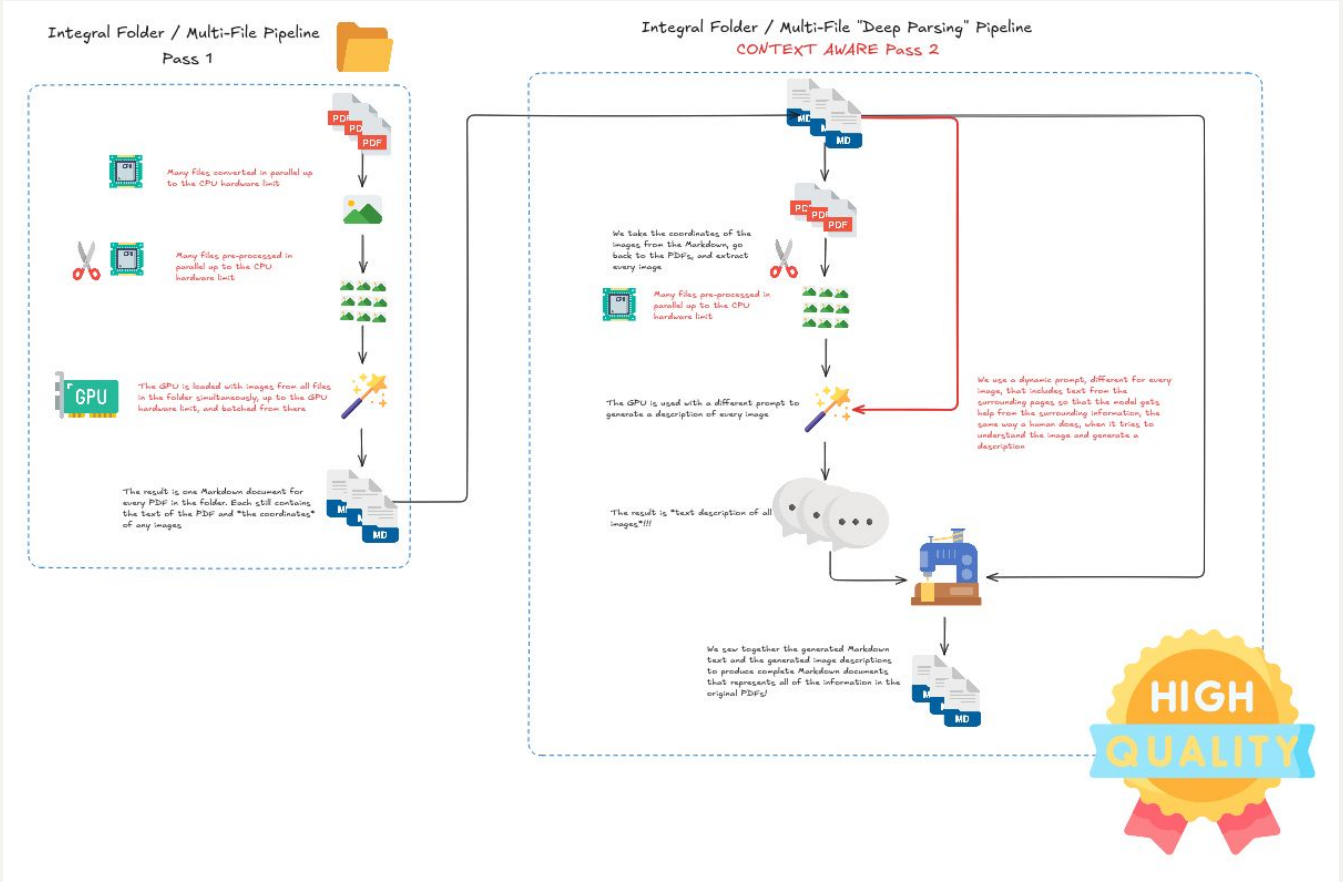
	2024	2025	2026	2027
Germany	2.1	2.3	2.7	3.0
France	2.05	2.15	2.4	2.8
Italy	1.45	1.65	2.05	2.55
Spain	1.25	1.55	2.0	2.55
Euro area	1.85	2.05	2.4	2.8

Deep Parsing

Deep Parsing

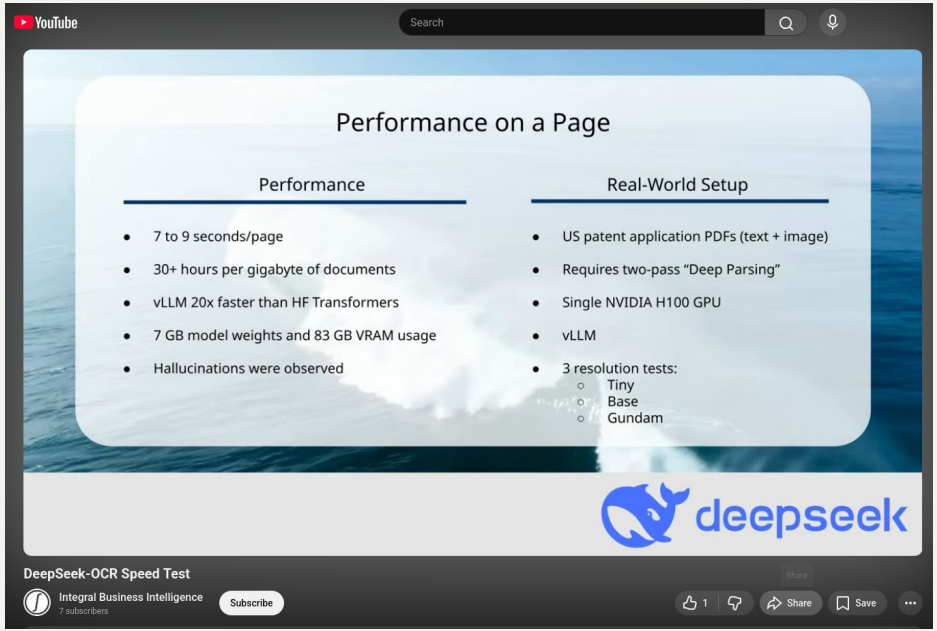


Deep Parsing

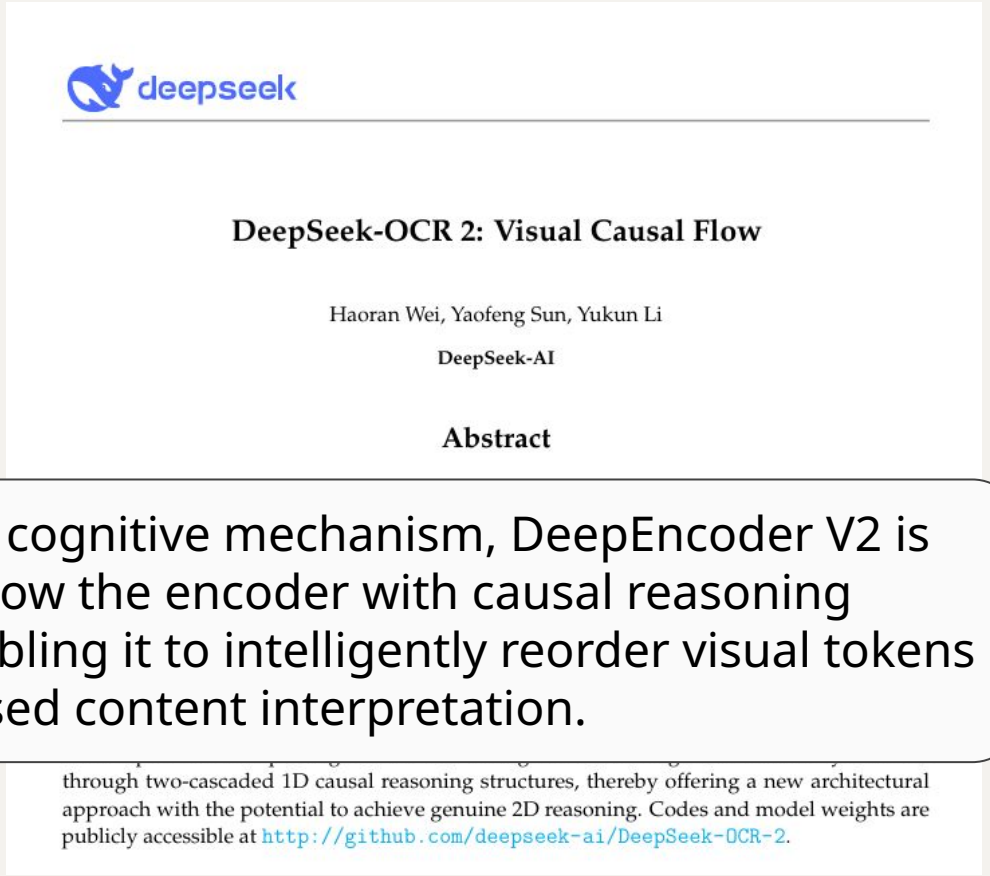


DeepSeek-OCR Companion Docs

https://github.com/integral-business-intelligence/deepseek-ocr-companion



DeepSeek-OCR-2



deepseek

DeepSeek-OCR 2: Visual Causal Flow

Haoran Wei, Yaofeng Sun, Yukun Li

DeepSeek-AI

Abstract

Inspired by this cognitive mechanism, DeepEncoder V2 is designed to endow the encoder with causal reasoning capabilities, enabling it to intelligently reorder visual tokens prior to LLM-based content interpretation.

through two-cascaded 1D causal reasoning structures, thereby offering a new architectural approach with the potential to achieve genuine 2D reasoning. Codes and model weights are publicly accessible at http://github.com/deepseek-ai/DeepSeek-OCR-2.

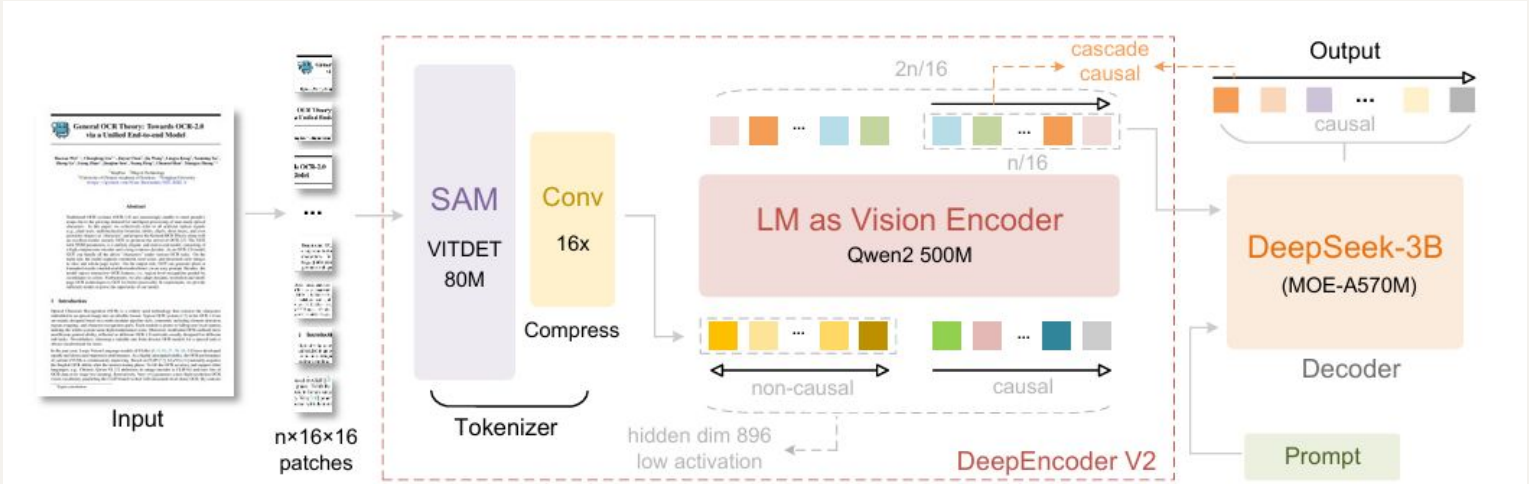
New Developments

Model	Release	Research Purpose
DeepSeek-OCR	October 21, 2025	Compression
DeepSeek-OCR-2	January 28, 2026	Reading Order

Research Hypothesis

- The human visual system … functions as visual tokens, locally sharp yet globally aware.
- …unlike existing encoders that rigidly scan tokens from top-left to bottom-right…
- …visual tokens in models should be selectively processed with ordering highly contingent on visual semantics rather than spatial coordinates
- … fundamentally reconsider the architectural design of vision language models (VLMs), … Directly flattening image patches in a predefined raster-scan order introduces unwarranted inductive bias that ignores semantic relationships.

DeepSeek-OCR-2

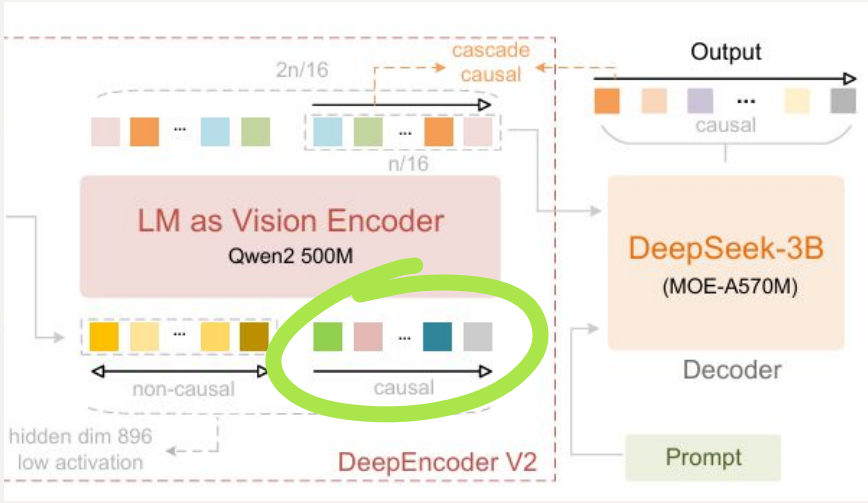


...oken compression mechanism from DeepEncoder,
...sor that reduces visual tokens by a factor of 16.
...coder's CLIP module with a compact language
...on masks, this LM-style vision encoder acquires
...le initiating causal modeling of visual sequences.

Replaced CLIP with Qwen2-0.5B.
Reused 3B decoder model.

3 Stage Training - DeepSeek-OCR-2

- Stage 1 - DeepEncoder v2
  - ✅ SAM: Training
  - ✅ Compressor: Training
  - ✅ Qwen Model: Training
  - ❌ DeepSeek-3B-MoE decoder: Omitted
- Stage 2 - Query Enhancement / Full Pipeline
  - ❌ SAM: Frozen
  - ❌ Compressor: Frozen
  - ✅ Qwen Model: Training
  - ✅ DeepSeek-3B-MoE decoder: Training
- Stage 3 - “Full Model” / Full Pipeline
  - ❌ SAM: Frozen
  - ❌ Compressor: Frozen
  - ❌ Qwen Model: Frozen
  - ✅ DeepSeek-3B-MoE decoder: Training



Called “learnable queries” which can attend to all non-causal/vision tokens + previous queries + itself

Stage 2 learns reading order

DeepSeek-OCR-2 Research Conclusions

- Replacing CLIP with an LLM works
- OmniDocBench score of 91% versus 87% for DeepSeek-OCR

Research Conclusions

- DeepSeek-OCR serves two primary production use cases:
  - An online OCR service that reads image/documents for DeepSeek-LLMs
  - A pretraining data pipeline that performs batch PDF processing

∫ntegral> Business Intelligence

Tony Sclafani
Co-Founder / Partner

https://integralbi.ai