

sirris innovation
forward

sirris

The innovation
companion of the
technology industry



Lokale AI

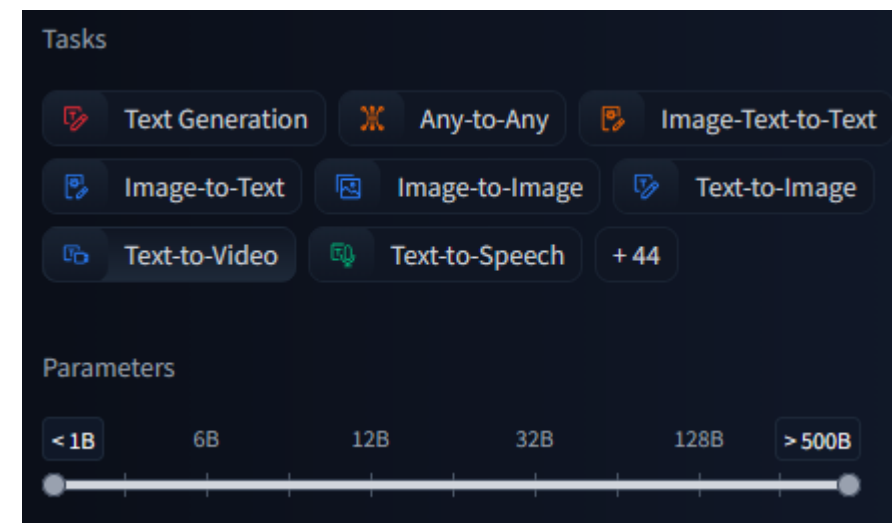
Lokale AI

OPEN WEIGHT

- Open weight
 - Modelparameters (weights) gratis ter beschikking
=> Downloaden en zelf draaien
 - Gemaakt door bedrijven, onderzoekers, non-profits, ...
 - **USA:** GPT-OSS, Gemma, Llama
 - **China:** Qwen, Deepseek, GLM, Kimi, ...
 - **EU:** Mistral
 - Gepubliceerd op HuggingFace
 - > 3.000.000 modellen



Hugging Face



Lokale AI

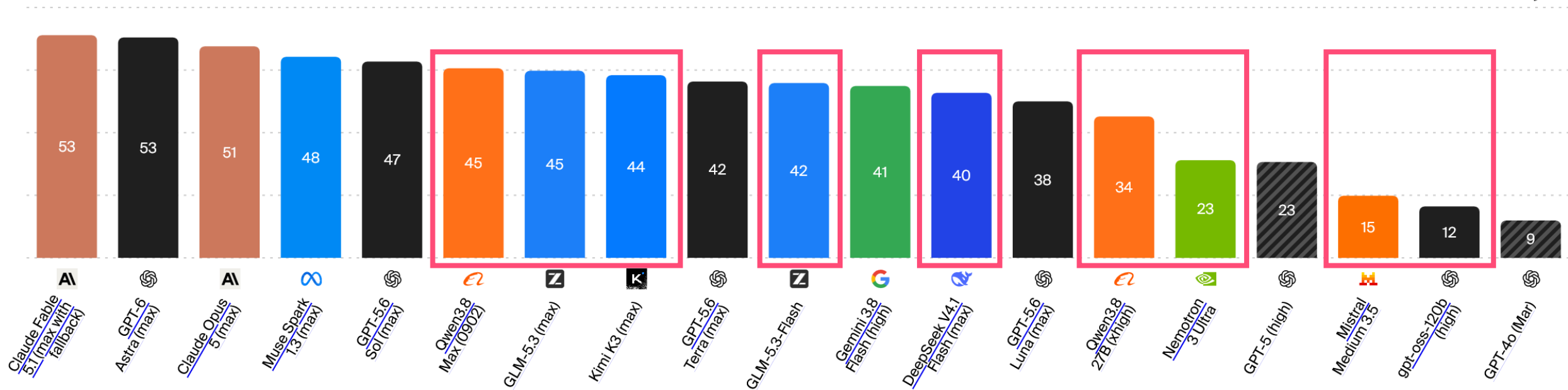
OPEN WEIGHT Artificial Analysis Intelligence Index ↗

Artificial Analysis Intelligence Index v4.3 incorporates 10 evaluations: AA-Briefcase, GDPval-AA v2, AutomationBench-AA, Terminal-Bench 4.0, SciCode, Humanity's Last Exam, GDP.pdf, CritPt, AA-Omniscience, AA-LCR v1.1

■ Estimate (independent evaluation forthcoming)

Elke maand
verouderd

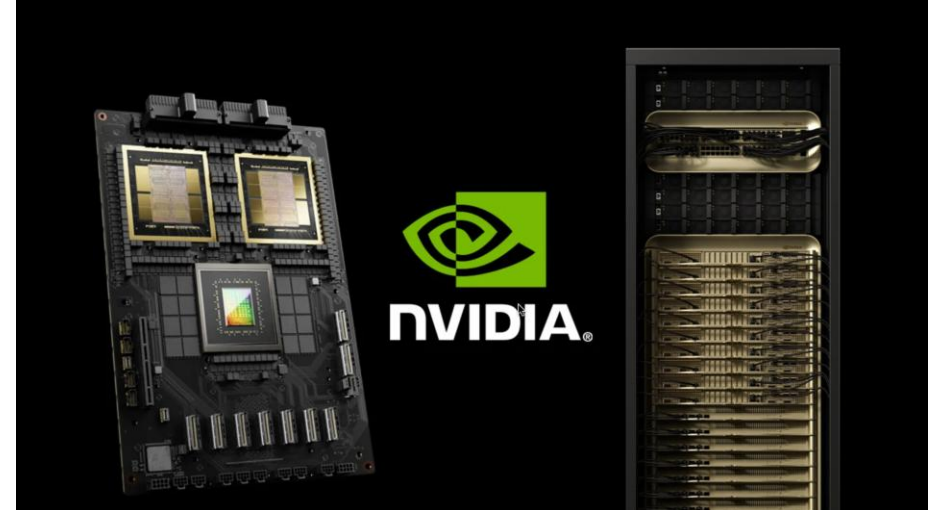
Artificial Analysis



Lokale AI

OPEN WEIGHT

- Open Weight $\neq$ Lokale AI
 - Grootste modellen eisen 1-3 TB aan RAM
 - 50-100 keer meer dan gemiddelde laptop
 - Lokale AI
 - Afhankelijk van investering in hardware
 - Typisch < 100-200 GB
 - Prestaties ook lager
 - Vergelijkbaar met topmodellen van ~1 jaar geleden



Min. €100k

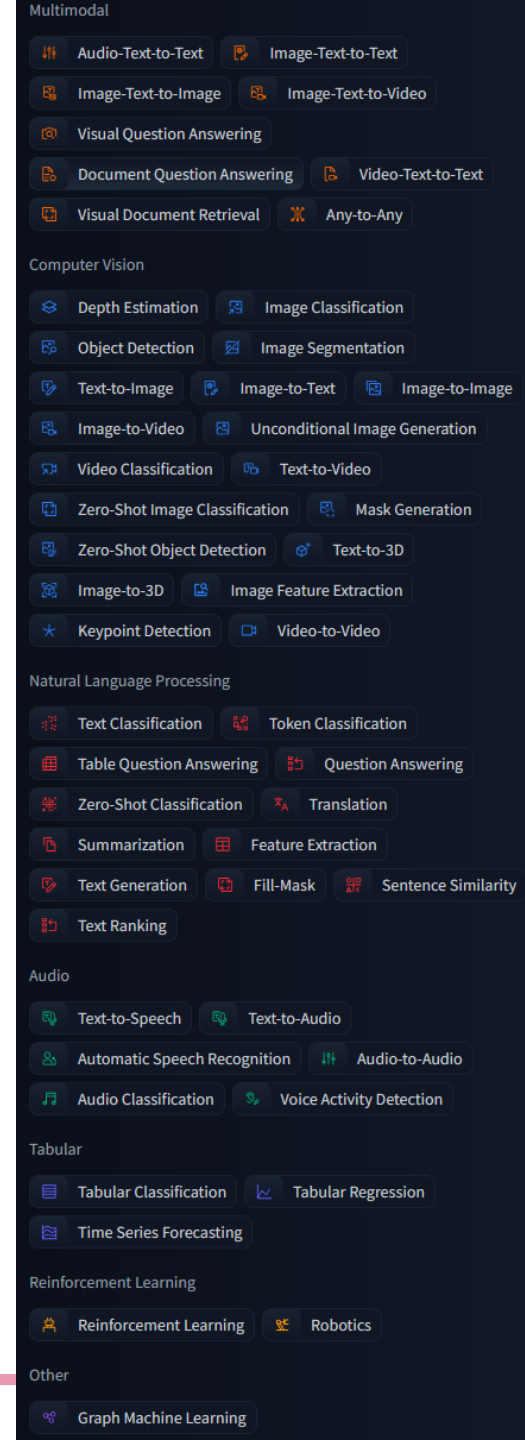


€4.5k

Lokale AI

OPEN WEIGHT

- Meer dan LLM's
 - Speech to text
 - Text to speech
 - Embedding
 - Document parsing
 - Image/video generation
 - "Klassieke" AI
 - Deep learning
 - Machine learning
 - ...
- Gespecialiseerde modellen zijn vaak beter dan grote AI-labs



Lokale AI

HARDWARE

- Voor elke token die gegenereerd wordt moet elke parameter van het model geladen worden.
- Limiterende factor: **RAM geheugen**
 - Hoeveelheid (GB)
 - Dicteert maximale grootte van het model
 - Snelheid (GB/s)
 - Dicteert maximale snelheid van het model



Lokale AI

HARDWARE – VOORBEELD 1

	Qwen 3.5 9B	Qwen 3.5 27B
# parameters	9 miljard	27 miljard
GB (@ 8-bit)	~10 GB	~30 GB
16 GB RAM	✓	✗
64 GB RAM	✓	✓
100 GB/s (typ. laptop)	10 tokens/s	3,3 tokens/s
450 GB/s (typ. "gamer" GPU)	45 tokens/s	15 tokens/s
1800 GB/s (prof. GPU)	180 tokens/s	60 tokens/s

GPT, Claude:
40-80 tokens/s

Lokale AI

HARDWARE – VOORBEELD 2

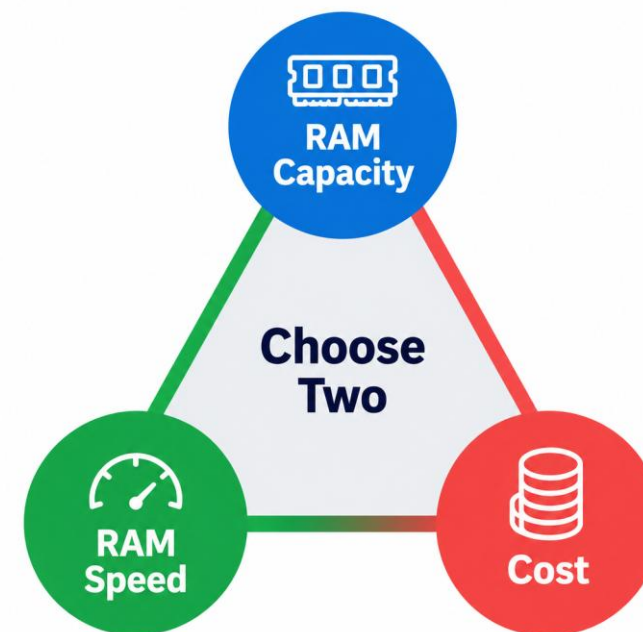
	Qwen 3.6 27B	Qwen 3.6 35B A3B
# totale parameters	27 miljard	35 miljard
# actieve parameters	27 miljard	3 miljard
GB (@ 8-bit)	~30 GB	~38 GB
16 GB RAM	✗	✗
64 GB RAM	✓	✓
100 GB/s (typ. laptop)	3,3 tokens/s	33 tokens/s
450 GB/s (typ. "gamer" GPU)	15 tokens/s	150 tokens/s 50-100 tokens/s
1800 GB/s (prof. GPU)	60 tokens/s	600 tokens/s 120-150 tokens/s

Aan deze snelheid is rekenkracht
opnieuw de limiterende factor ipv
RAM

Lokale AI

HARDWARE

- Altijd een compromis
 - CPU
 - Veel GB
 - Goedkoop
 - Lage snelheid
 - GPU
 - Beperkte GB
 - Duur
 - Zeer snel
 - Multi-GPU => capaciteit verhogen



Lokale AI

HARDWARE



- DGX spark & Mac mini / Mac studio
 - Goede balans tussen snelheid, capaciteit en prijs

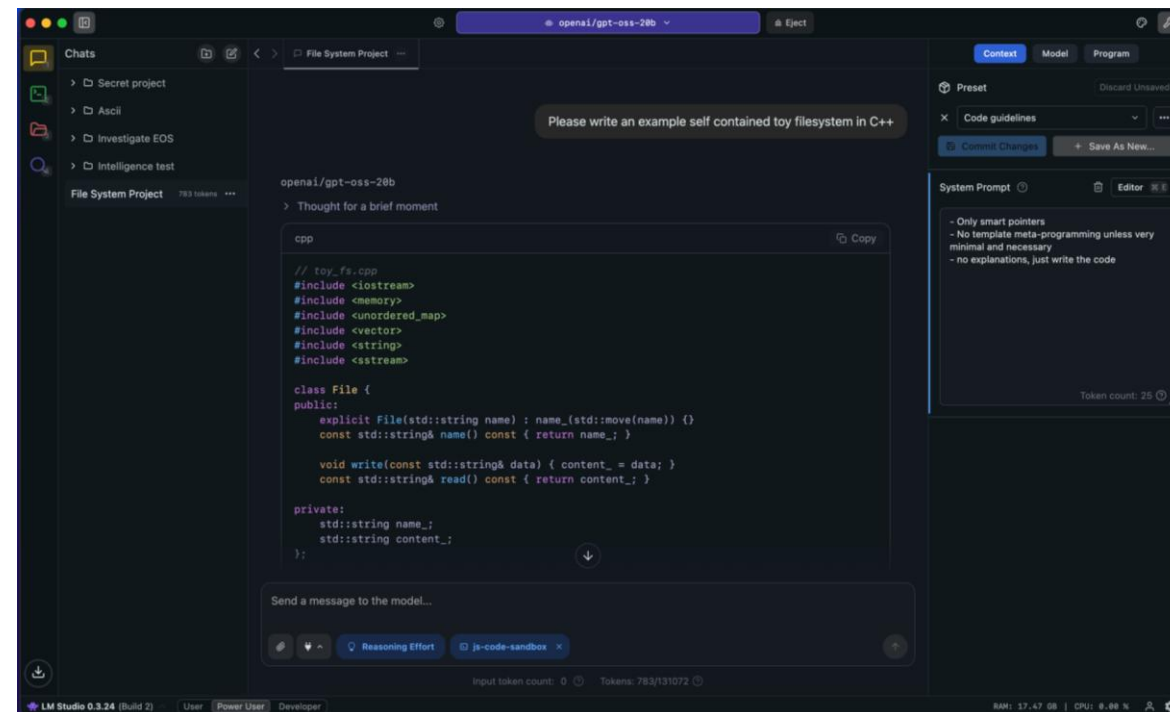
	DGX spark	Mac mini M5 Pro	Mac studio M5 Max	Mac studio M5 Ultra	Nvidia RTX 5090
GB	128	64	128	256	32
GB/s	273	307	614	1200	1800
GPU	Nvidia CUDA	Apple	Apple	Apple	Nvidia CUDA
Prijs	€4.500+	€3.200	€5.900	€11.000	~€5.500

CUDA ondersteuning heeft beste compatibiliteit

Lokale AI

HOE ZELF STARTEN

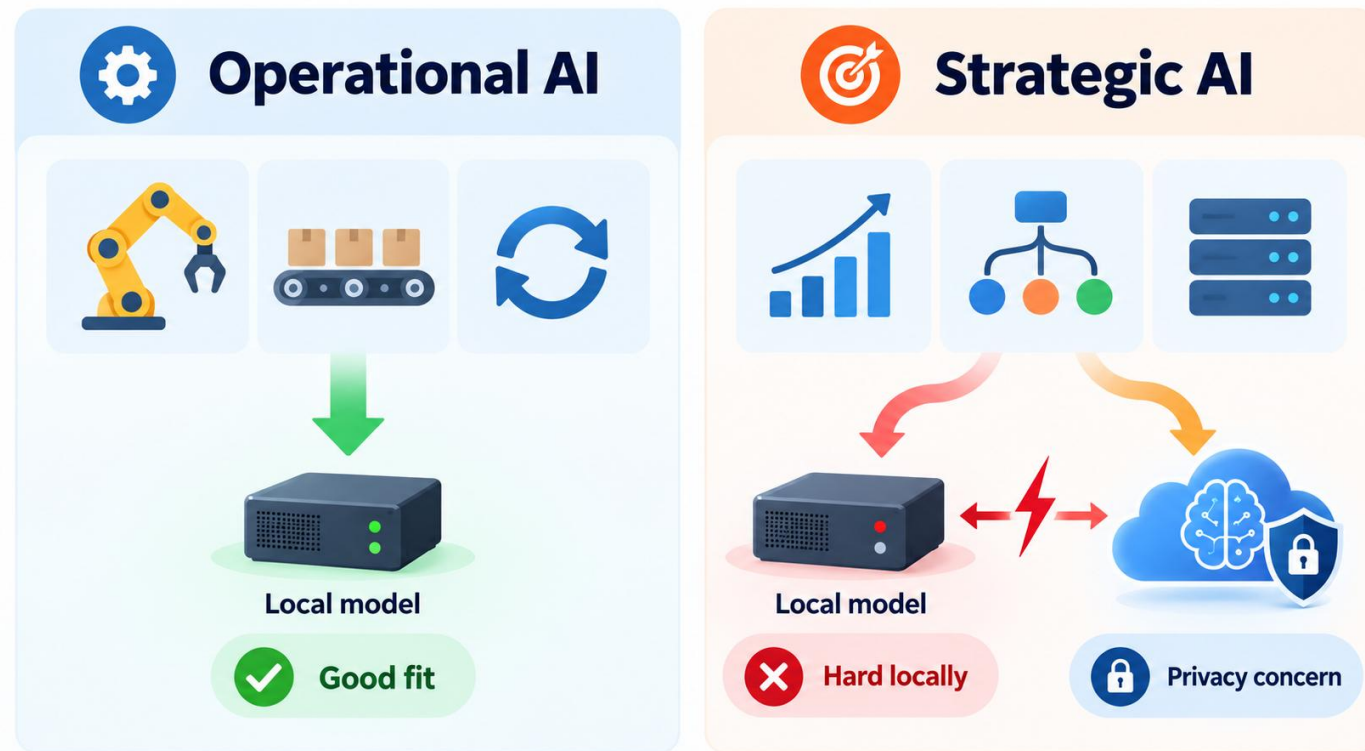
- LM Studio
 - Gebruiksvriendelijke app om te starten
 - Kies klein model
 - Qwen 3.5 4B (~4GB)
 - Gemma E2B (~5GB)
 - Draai op eigen laptop
 - Voeg MCP's toe en experimenteer
- vLLM & SGLang
 - Meer geavanceerde server installaties
 - Meer beschikbare modellen



Operationele & Strategische AI

MISMATCH TUSSEN LOKAAL EN CLOUD

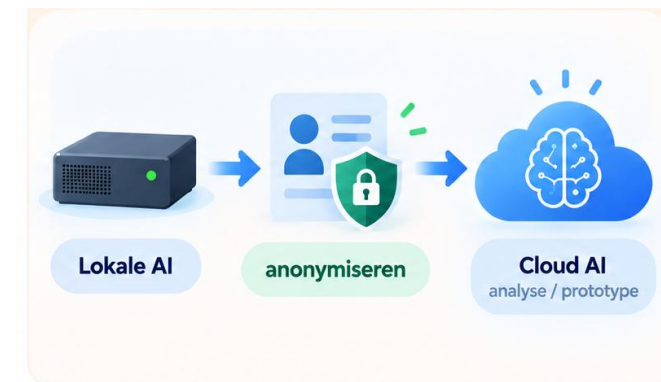
- Operationeel
 - Machinedata
 - Q&A
 - Planning
- Strategisch
 - Data analyse
 - Infrastructuur managen
 - => Slimste model nodig: cloud
 - => Meest gevoelige data: privacy?



Operationele & Strategische AI

MISMATCH TUSSEN LOKAAL EN CLOUD

- Oplossingen
 - Beter lokaal model (investeren, wachten)
 - Hybride:
 - Lokale AI om te anonimiseren
 - Cloud AI voor analyse / prototype
 - API provider
 - EU-modellen: Mistral
 - Openrouter, ...
 - "Private" cloud
 - Hardware huren per uur



Logging				
	OpenAI	Fast	50% off	\$8.00 \$4.00 \$40.00 \$20.00
	Amazon Bedrock			\$4.40 \$22.00
	Azure			\$5.50 \$33.00
	Azure			\$5.50 \$33.00
Regio No-logging				

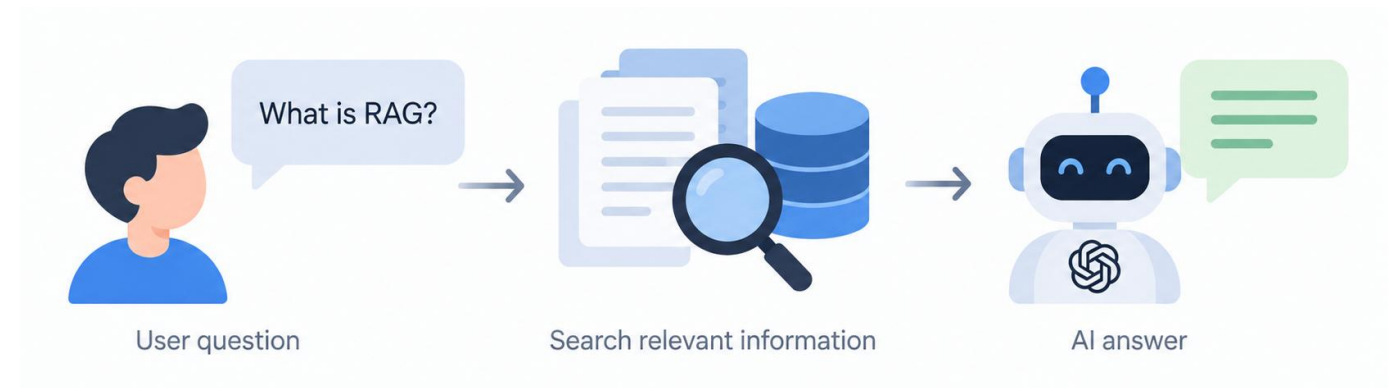
Lokale RAG

USE-CASE

Lokale RAG

INTRO

- **RAG: Retrieval Augmented Generation**
 - Eerst zoeken
 - Dan antwoorden
 - => Databronnen
 - Documenten
 - Machinedata: MCP, ...
 - Web
 - ...



Lokale RAG

INTRO

- Document uploaden in chat vs. RAG
 - Volledige tekst in context

Document uploaden	RAG
Volledige tekst in context	Enkel relevante stukken (chunks)
Weinig setup	Meer technische setup
Meest nauwkeruig*	Afhankelijk van implementatie
Duur / rekenkracht	Eenmalige verwerking
Beperkt schaalbaar	Goed schaalbaar voor veel documenten

Lokale RAG

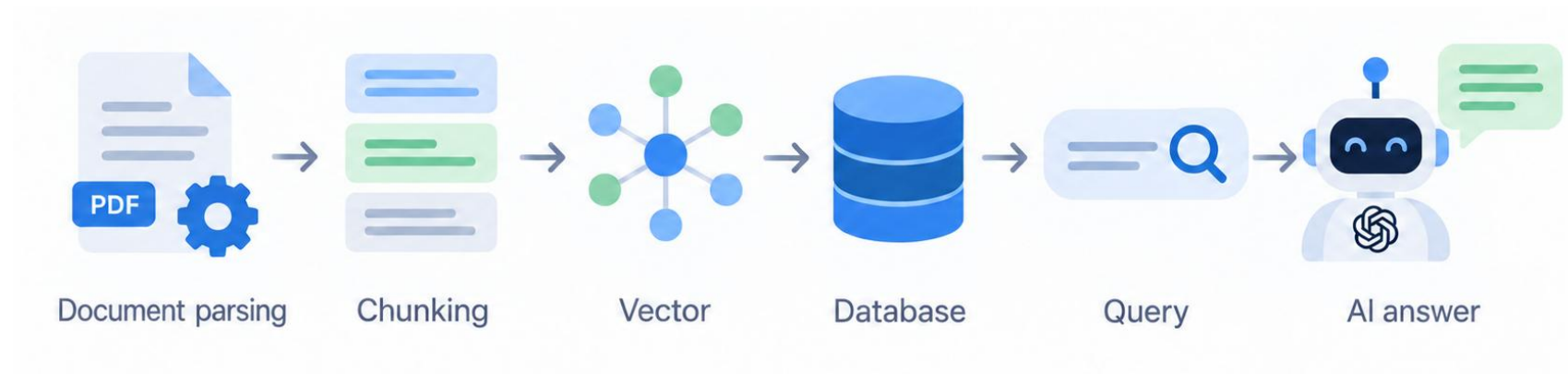
OPEN SOURCE OPLOSSINGEN

- Bestaande platformen
 - RAGFlow
 - Dify
 - Llamaindex
 - ...
- Eigen oplossing
 - Volledige controle proces
 - Custom UI
 - Extra functies



Lokale RAG

EIGEN OPLOSSING



Lokale RAG

STRATEGIEËN

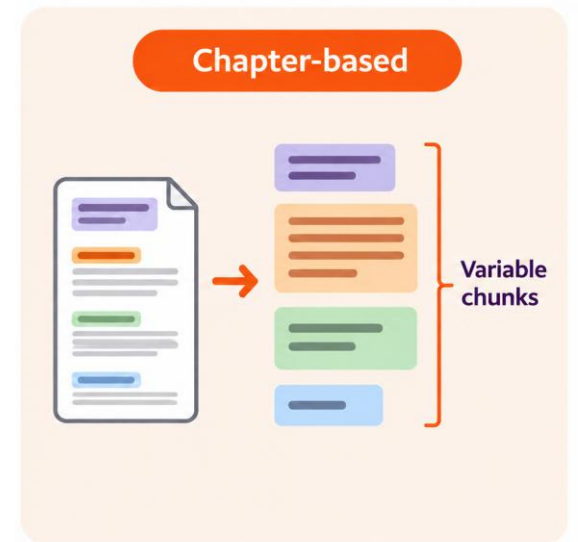
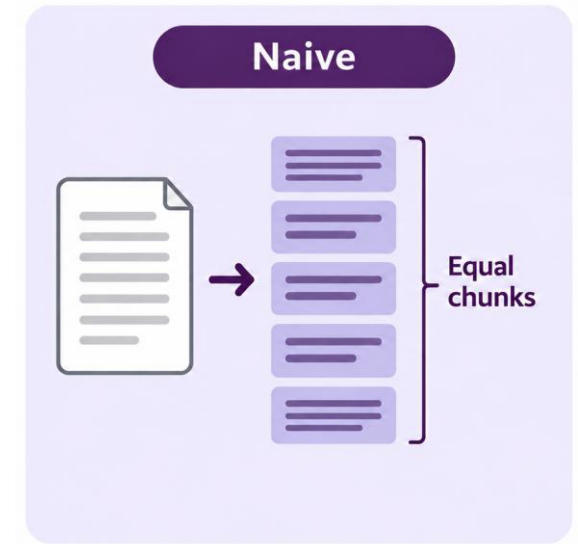
- Docling
 - Open source - IBM
 - Document parsing
 - Pdf
 - DOCX, PPT, XLSX
 - ...
 - Uniform formaat



Lokale RAG

STRATEGIEËN

- Chunking
 - Naïef
 - Verdelen in gelijke stukken
 - Hoofdstuk gebaseerd
 - Verdelen binnen tussentitels
 - => tussentitel herkenning docling!
- Chunkless / Agentic RAG
 - Mappenstructuur
 - AI kiest bestand en hoofdstuk

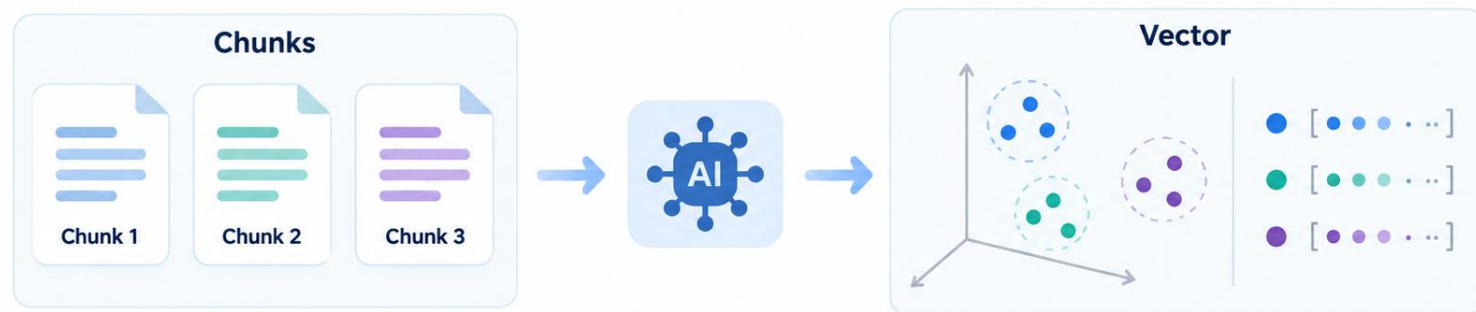


Lokale RAG

STRATEGIEËN

- Embedding & vector search
 - Gespecialiseerd AI model
 - Tekst naar “vector”
 - Behoud van betekenis

1 Chunks → Vector



2 Vraag → Vector → Zoek naar meest vergelijkbare vectoren



sir

 **Printelligence Library**
Document retrieval workspace

Documents

Refresh

princt

- › DIR prinect

PDF Speedmaster XL 106 - Onderhoud e...

PDF Speedmaster XL 106 - Wallscreen XL...

PDF Speedmaster XL 106-P - Specificatie...

princt

Use as chat scope

princt/Speedmaster XL 106 - Prinect Press Center - Bedieningshandleiding.pdf · page 180

[Open or download PDF](#)

Services ready

Ask the library

Answers are limited to retrieved sources.

Clear

All documents

You 02:26 PM

hoe kan ik een inktprofiel gebruiken voor een andere kleur?

Assistant 02:26 PM

✓ search_documents_by_content ✓ Done

▼ search_content ✓ Done

▼ 🔧 search_content ✓ Done

✓  get_chapter_content ✓ Done

Om een inktprofiel van de ene kleur naar een andere kleur te kopiëren, gaat u als volgt te werk (bijvoorbeeld in Prinect Press Center of Inpress Control):

1. **Selecteer de bronkleur:** Kies in het menu *Detaillaanzicht of Inktprofielen details* de kleurtoets waarvan u het inktprofiel wilt kopiëren.
2. **Activeer de kopiëerfunctie:** Druk op de toets **Inktprofiel kopiëren**. Het menu voor het kopiëren wordt nu weergegeven.
3. **Selecteer de doelkleur(n):** Kies in dit menu de kleur(en) waarvoor u het profiel wilt gebruiken. De kleur toets vanwaar u kopieert, wordt vaak grijs weergegeven (uitgegrijsd).
4. **Bevestiging:** Druk op de toets **OK** om de wijziging te bevestigen. Het inktprofiel wordt nu naar de geselecteerde kleur gekopieerd en overgenomen.

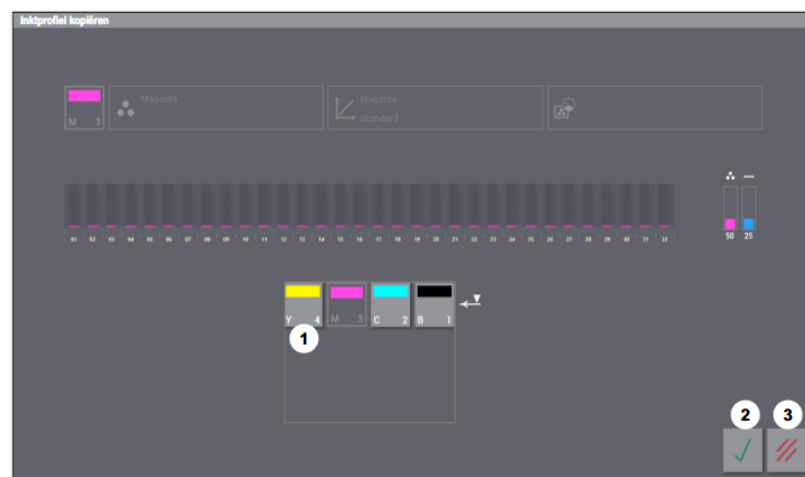
U kunt het kopiëren ook annuleren door op **Wissen** te drukken voordat u bevestigt.

Mocht u een profiel binnen dezelfde kant (schoondruk of weerdruk) willen kopiëren, of over de keerinrichting heen, let dan op dat bij machines met keerinrichting de keuze beperkt is tot bij een dezelfde afdruk (schoondruk of weerdruk). [\[S10\]](#)[\[S6\]](#)[\[S8\]](#)

Ask about maintenance, settings, procedures...

IAL • 25/09/2026

2.2 Inktprofiel kopiëren



Afb. 3

1. Selecteer de kleurtoets waarvan u de waarden wilt kopiëren (afb. 2)
2. Druk op de toets *Inktprofiel kopiëren* (afb. 2). Het menu **Inktprofiel kopiëren** opent (afb. 3) met het inktprofiel van de geselecteerde kleur.
3. Selecteer in het menu **Inktprofiel kopiëren** de kleurtoets (bijv. afb. 3/1), waarin u het inktprofiel wilt kopiëren.
4. Invoer bevestigen: druk op de toets *OK* (afb. 3/2). Het inktprofiel wordt gekopieerd en op het overeenkomstige drukwerk overgenomen
of
Invoer annuleren: druk op de toets *Wissen* (afb. 3/3).

E.1.4

FS.999.0010/14

sirris innovation
forward



[SIRRIS.BE](https://sirris.be)



[LINKEDIN.COM/COMPANY/SIRRIS](https://www.linkedin.com/company/sirris)



[FACEBOOK.COM/SIRRIS.BE](https://www.facebook.com/sirris.be)

