



12 november NFIR
Cybersecurity in het AI-tijdperk

SOEVEREINE AI

IMG

Ludo Baauw



INTERMAX GROUP

- 7 bedrijven, ~275 medewerkers
- Managed Cloud & Security voor overheden, zorg en vitale sectoren
- 100% Nederlands, al 32 jaar

DAARNAAST

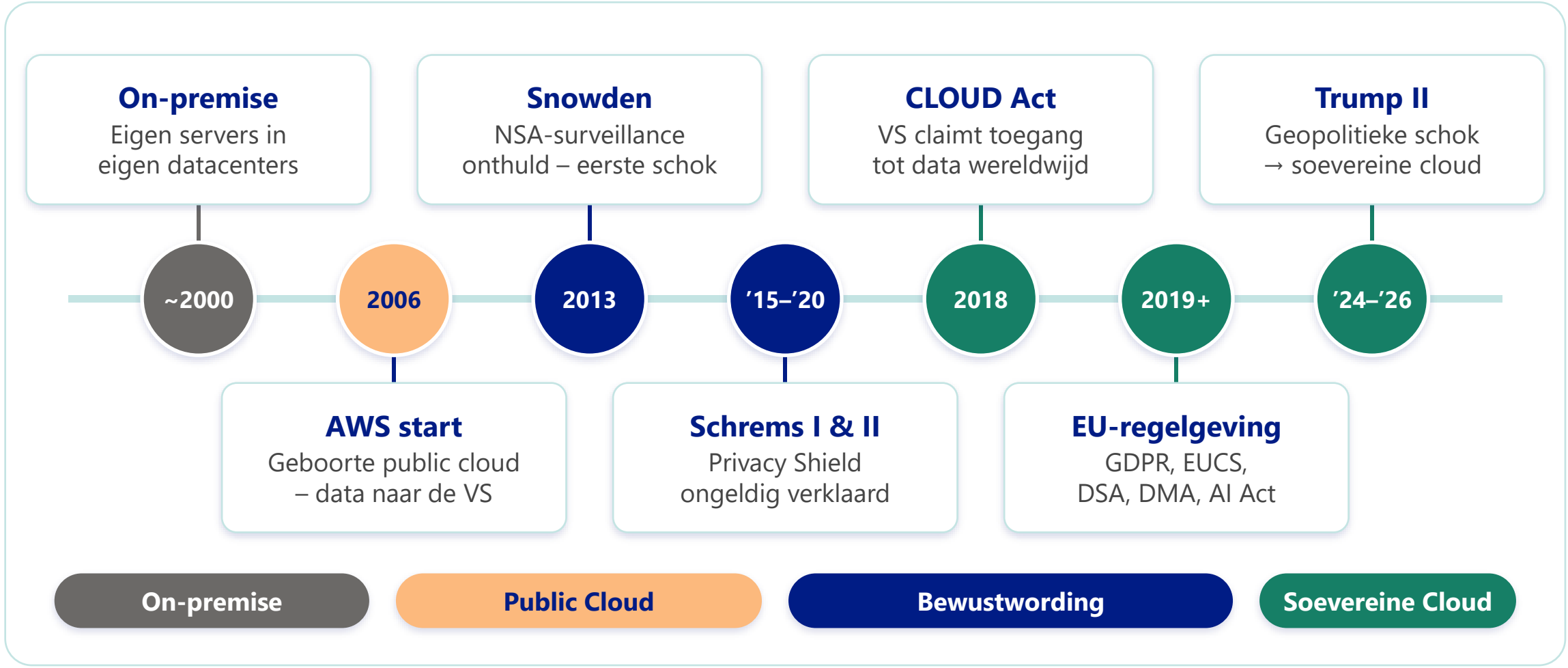
- Medeoprichter **Open Cloud Alliantie**
- Voorzitter **NBIP**, bedenker NaWas.nl
- Founding partner **DEC-Alliance**

IMG

intermax GUIDA gridly

GUARDIAN 360° bizway NFIR CYBER SECURITY EXPERTS

Hoe zijn we hier gekomen?



Het leeft nogal

IN GESPREK MET DE POLITIEK







RAPPORTEN, NOTA'S EN MANIFEST









’de Nederlandse cloud’ in de media

TV & RADIO

Nieuwsuur
13 maart 2025



Lubach
22 mei 2025



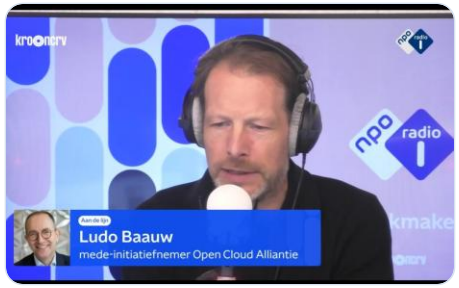
EenVandaag
16 april 2026



Buitenhof
3 mei 2026



NPO Radio 1
22 oktober 2026



KRANTEN

fd.

Bedrijven pauzeren al plannen om nog meer te doen in de Amerikaanse cloud, vernam Ludo Baauw, oprichter en ceo van het Rotterdamse cloudbedrijf Intermax. ‘Overal vraagt men zich af: waar draaien onze IT-diensten? De pleuris is uitgebroken.’ Nog weinig ondernemingen besloten alles naar een Europese cloud te brengen, aldus Baauw. ‘Bedrijven zijn vooral bezig met het opstellen van een plan B, voor als er wat gebeurt rond die Amerikaanse clouddiensten.’

deVolkskrant

Enkele overheidsdiensten hebben volgens Baauw ‘net op tijd’ de overstap naar de Amerikaanse cloud stop kunnen zetten, waaronder een van de meest kritieke Rijksorganisaties die hij niet bij naam kan noemen. Baauw: ‘Maar partijen die al wel waren overgegaan, hebben nu een probleem. Het is zoals The Eagles zingen in Hotel California: *you can check-out anytime you like, but you can never leave.*’

The New York Times

European tech companies see an opportunity to win customers from their American rivals. Cloud service providers like Intermax Group, based in the Netherlands, and Exoscale, based in Switzerland, said they had seen a jump in new business.

“A few years ago, everyone was saying, ‘They’re our trusted partners,’” Ludo Baauw, Intermax’s chief executive, said of U.S. tech companies. “There’s been a radical change.”

IMG

Soevereiniteit is niet hetzelfde als autonomie



Soevereiniteit

Wiens recht geldt?

Risico: jurisdictie



Autonomie

Kun je weg?

Risico: lock-in

Zie ook: ACM, AFM, AP, DNB en RDI, De route naar digitale autonomie (10 jul. 2026)

Security is eng. Weerbaarheid is hip.

92%

van het ICC werk hangt af van Amerikaanse bedrijven

Weerbaarheid is kunnen uitwijken

2025

Mailbox aanklager op zwart

Mei 2026

Namen ACM-ambtenaren naar VS

9 okt. 2026

Hele Strafhof op de lijst

7 apr. 2027

Vrijstelling IT en cloud verloopt

Bronnen: OFAC, GL 13 en 14 (9 okt. 2026); Justice Info (okt. 2026); NL Times (26 okt. 2026); NOS en Vrij Nederland (mei 2026)

IMG

Elke grootmacht heeft wetten die de staat toegang geven

Verenigde Staten	China	Europese Unie
CLOUD Act, FISA §702, EO 12333, RISAA	Inlichtingenwet, Databeveiligingswet, Cyberwet	AVG, nationale bewaarplicht, NIS2, Data Act, EUCS, CSF, CADA

Bronnen o.a. China Law Translate, Wikipedia (DSL/NIL), wirescreen.ai. Handhaving en reikwijdte zijn omstreden.

Vandaag een iets ander verhaal

CLOUD-SOEVEREINITEIT

Waar je data staat

Hyperscalers, opslag, jurisdictie. CLOUD Act, FISA 702, EU Data Boundary, DigiD, Rijkscloud…

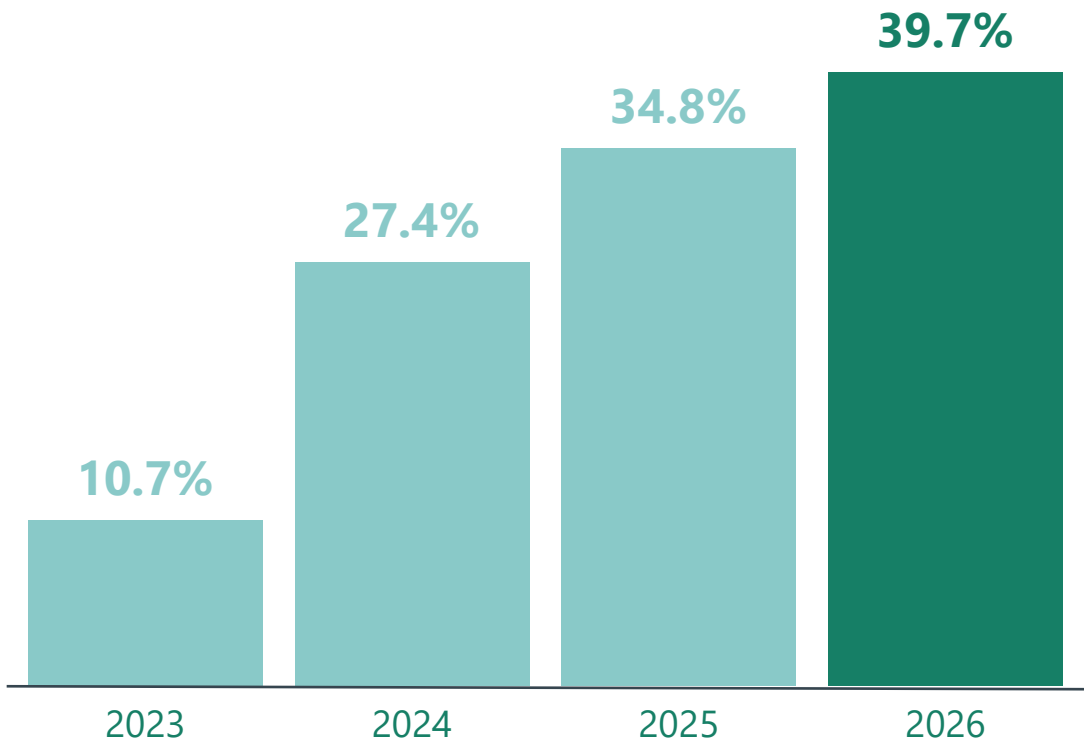
AI-SOEVEREINITEIT

Waar je over nadenkt

Elke prompt = je intentie, context, strategie. Dezelfde leveranciers, maar een veel ‘intiemer verhaal’.

"*Terwijl wij de cloud-discussie nog voeren, is AI ons voorbij gerend.*"

Bijna 4 op de 10 AI-interacties bevatten gevoelige data



39.7%

bevat van alle AI-interacties gevoelige data

- Elke 3 dagen per medewerker
- Bijna 4× zoveel als in 2023

Bron: Cyberhaven AI Adoption & Risk Report 2025 en 2026

Precies de dingen die je niet kwijt wil



Samen 90% van alle AI-beleidsovertredingen: code, klantdata en bedrijfsgeheimen.

Bron: Netskope Cloud and Threat Report 2026 % van datalekovertredingen bij GenAI-apps

Shadow-AI loopt grotendeels via privé-accounts.

ChatGPT

32.3%

via persoonlijke accounts

Claude

58.2%

via persoonlijke accounts

Perplexity

60.9%

via persoonlijke accounts

Wel een kentering: het aandeel AI-gebruikers met privé-apps daalde van 78% (2024) naar 47% (2025).[1]

Geen SSO. Geen audit-log. Geen retentiebeleid. Geen DPIA mogelijk.

Bron: Cyberhaven AI Adoption & Risk Report 2026 % van gebruik via persoonlijke accounts [1] Netskope Cloud and Threat Report 2026

Hoezo is dat belangrijk dan?

...zei de leerling-parachutist, en sprong zonder rugzak het vliegtuig uit...

Shadow-AI-lekken: van Samsung tot Eindhoven

1

Eindhoven, 2025

Ruim 1.000 documenten met persoonsgegevens in één maand naar publieke AI.

2

CISA (VS), 2025

Directeur uploadt geheime contractstukken naar publieke ChatGPT.

3

Samsung, 2023

Engineers plakken broncode en notulen in ChatGPT. Samsung verbiedt AI.

Een verbod helpt niet. Eindhoven en Samsung kozen wel een eigen AI-omgeving.

Bron: Gemeente Eindhoven / NOS; Politico; AI Incident Database #768

IMG

Brussel en Den Haag stellen nu eisen aan AI

CADA

Europese Commissie

Wetsvoorstel van juni 2026. Deelt clouddiensten in vier soevereiniteitsniveaus in. Overheden gaan daarop inkopen.

Verbod

AI-standpunt overheid

Ambtenaren mogen sinds 1 januari 2026 geen gratis AI-tools meer gebruiken voor hun werk, ook niet op een eigen account.

Groen licht

SLM-Rijk Copilot DPIA

Na de update van mei 2026 blijven 2 middelhoge risico's over. Wel nodig: eigen DPIA, AI-beleid en Purview op orde.

Bronnen: Europese Commissie COM(2026) 502, Rijksoverheid, SLM Rijk DPIA 2e update (mei 2026)

Zes AI-diensten, zes risicoprofielen

	ChatGPT gratis / Plus	ChatGPT Business	ChatGPT Enterprise	Microsoft 365 Copilot	Claude in Copilot	Claude direct
Training op je data	Ja, tenzij opt-out	Nee	Nee	Nee	Nee	Nee (zakelijk)
Retentie	Standaard	Door beheerder	Instelbaar, eigen sleutels (EKM)	M365 audit-log	Standaard	Beperkt; ZDR via API
Jurisdictie	VS	VS, EU-opslag deels	VS-bedrijf; EU-opslag en verwerking mogelijk	EU*, met uitzonderingen voor beveiliging en piekbelasting	VS, EU gepland eind 2026	VS; EU via Bedrock of Vertex

**De EU Data Boundary beschermt niet tegen de CLOUD Act.*

Bron: Microsoft EU Data Boundary-documentatie (nov 2026), OpenAI data residency, Anthropic privacybeleid

IMG

EU Data Boundary ≠ jurisdictie-immuniteit

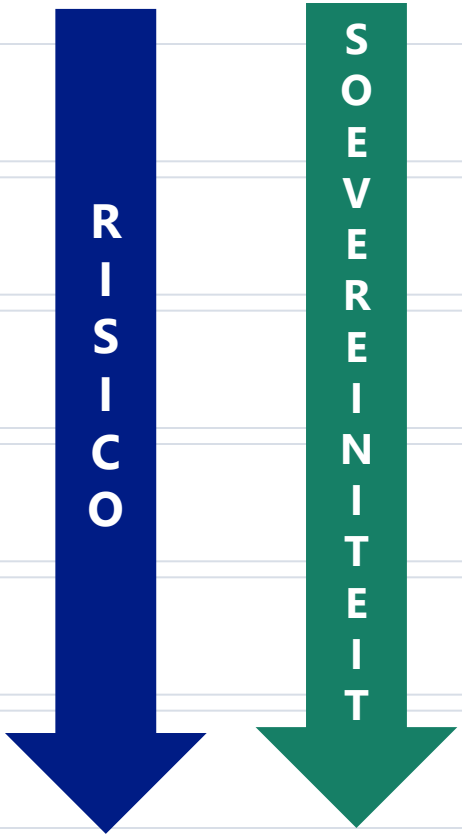
> ***"Voor global security operations kan data buiten de EU worden verwerkt. Klanten kunnen daar niet uit opt-outen."***
>
> **— Microsoft EU Data Boundary FAQ, februari 2025**

- Boundary af sinds feb 2025: klant-, pseudonieme en supportdata in EU
- **Carve-out voor security: data flows naar VS, geen opt-out mogelijk**
- Copilot bij piekbelasting: verwerking in VS, CA, AU
- FISA 702 verlopen jun 2026, surveillance loopt door tot mrt 2027

Maar Ludo, wat moeten we dan?!

Van slecht naar soeverein

1	**Gratis consumer-tools**	ChatGPT free, Gemini free`; niet voor werk
2	**Enterprise-tier + DPA**	ChatGPT Enterprise, Claude Team/Enterprise, Copilot M365
3	**+ PII-stripping / redaction**	Presidio, OpenAI Privacy Filter, LLM Guard, Grepture (EU)
4	**EU-gehoste closed APIs**	Mistral Studio, STACKIT, IONOS, OVHcloud, Scaleway
5	**Open-weight op EU-soevereine GPU**	Mistral, Qwen, DeepSeek op Nebul / Intermax.ai
6	**On-prem / air-gapped**	Eigen GPU, BBN3, als je Ministerie van Defensie bent ofzo



IMG

PII-stripping helpt een beetje, maar is geen oplossing

Presidio

OPEN SOURCE

Herkent PII op patronen en namen. Gratis, maar mist ruim 1 op 5 in rommelige tekst.

OpenAI Privacy Filter

OPEN SOURCE

Draait lokaal, Apache 2.0. Leest context beter, maar werkt best in het Engels.

Grepture

EU-SAAS

Haalt PII en sleutels uit de prompt voordat die de LLM bereikt. Open source, gehost in Frankfurt.

Skyflow

TOKENISATIE

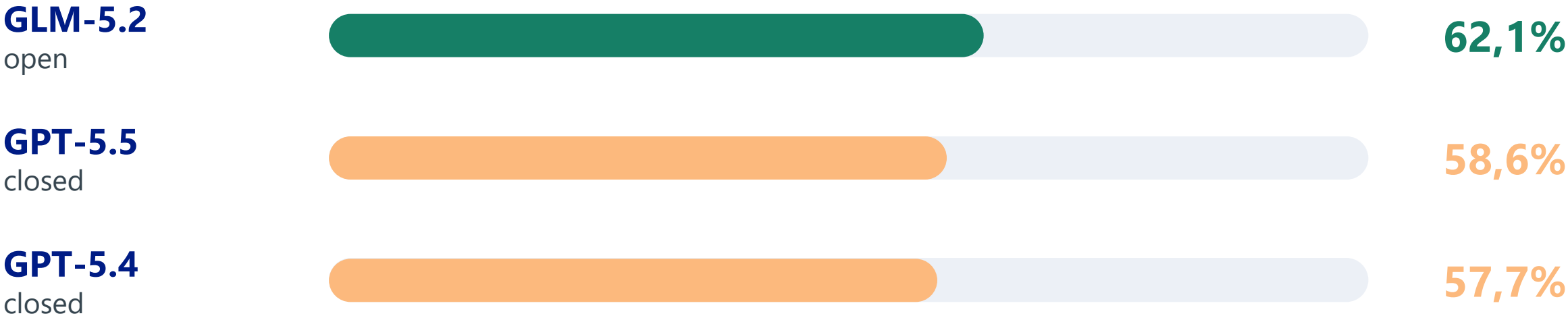
Vervangt PII door tokens en zet ze daarna terug. Opslag in VS-cloud.

Combineer detectie (Presidio/Privacy Filter) met tokenisatie (Skyflow) voor betere bescherming.

IMG

Open modellen zijn goed genoeg voor zakelijke taken

SWE-bench Pro, % softwaretaken opgelost, jun 2026



Daarnaast: *op de brede AA-index loopt open ~12 punten achter, GLM-5.2 kost ~1/7 per token, Mistral Large 4 is het sterkste Europese open model*

Bronnen: Z.ai / VentureBeat (jun 2026), Digital Applied Q2 2026, Artificial Analysis Intelligence Index v5 (okt 2026)

#	Model	Land	Index
1	Kimi K3.1 Pro		46
2	DeepSeek V4.1 Flash		33
3	GLM-5.3		32
4	MiMo-V2.6-Pro		30
5	Ling 3.1 Flash		30
6	MiniMax-M3.1-Flash		29
7	Mistral Large 4		26
8	Muse Spark Open		25
9	Qwen3.8 27B		20
10	K2 Horizon 375B		18

Zoek de overeenkomsten!

Bron: Artificial Analysis Intelligence Index v5, open-weights modellen, 9 nov 2026

IMG

Het werkt al. In Nederland. Vandaag.

IND

Immigratie- en Naturalisatiedienst

Productie-AI-applicaties op UbiOps, draaiend in eigen Nederlandse datacenter. Geen US-cloud-afhankelijkheid.

Hoge Raad

Raad voor de Rechtspraak

Interne juridische chatbot voor rechters, gehost op UbiOps via secure Dutch cloud. Eigen data, eigen model.

KNB

Koninklijke Notariële Beroepsorganisatie

Centraal platform voor 2 miljoen notariële aktes per jaar draait sinds 2024 bij Intermax.

Geen byte naar de VS. Volledige stack in NL. Herkenbare toepassingen.

IMG

De data bepaalt welke AI-tool je mag gebruiken

	TIER 1 **Publiek**	TIER 2 **Vertrouwelijk**	TIER 3 **Gevoelig**
Data	Wat je ook op LinkedIn zou zetten	Interne stukken, zonder BSN of bijzondere persoonsgegevens	Alles met BSN, broncode, juridische of medische info
Tool-keuze	Vrije keuze, gratis tools mogen ook	Zakelijke versie, met verwerkersovereenkomst en DPIA	Alleen soevereine AI, op Europese of Nederlandse servers
Voorwaarden	Een AI-beleid en uitleg aan medewerkers	Logging, vaste bewaartermijn en eigen sleutels	Eigen sleutels, geen Amerikaanse verwerker, alles controleerbaar

IMG

Nog ruim een jaar tot de high-risk-regels van de AI Act



Uitgesteld: high-risk-regels gelden pas vanaf 2 december 2027

Digital Omnibus, in werking sinds 27 juli 2026. Biometrie, kritieke infra, onderwijs, werk en publieke diensten. AI in gereguleerde producten (bijlage I): 2 augustus 2028.

Bron: Verordening (EU) 2026/1744 (Digital Omnibus on AI), Publicatieblad 24 juli 2026

DUS:

Verbieden duwt medewerkers naar eigen tools (shadow AI)

Niets doen laat bedrijfsdata weglekken naar gratis tools

Wat werkt: per use case bepalen welke tool mag

Welke use case deel jij morgen als eerste in?

FULL DISCLOSURE VOORDAT HET VERPLICHT WORDT:

AI checkte de feiten en poetste de slides op.

De rest is ambachtelijk gebakken door de spreker.

APPENDIX

Voor de techs in de zaal

Open-weight stack · RAG · PII-stripping · jurisdictie-matrix · hardware-economie





IMG

API's worden goedkoper, GPU's duurder: self-hosting loont pas bij groot volume

$9,50

/uur

B300, mediaan on-demand

$3,36

/uur

H100 ter vergelijking (mediaan)

288 GB

per GPU

B300: tot 125B in FP16

$0,10

/1M tokens

Claude Haiku 5.5 API (input)

Let op: Een B300 draait 70B zonder opsplitsen, maar kost bijna 3× een H100.

Bronnen: getdeploying.com en gpus.io (B300- en H100-prijzen, nov 2026), Anthropic API-prijzen (nov 2026)

Een model draaien is nog geen productie-AI

Application & UI	**Klantapp, chatbot, agent**
Security & governance	**PII-filter, rate-limits, audit-log, prompt-injectie**
RAG-laag	**Vector store, retrieval, reranking, evaluatie**
MLOps platform	**Versiebeheer, monitoring, eval-pipelines (bijv. UbiOps)**
Model serving	**vLLM of SGLang (continuous batching)**
GPU-infrastructuur	**Intermax, Nebul (H200, B200, B300), onder EU-recht**

UbiOps neemt MLOps en model serving uit handen. Daar zit het meeste beheerwerk.

IMG

APPENDIX A: SERVING STACK

Inferentie-engines en modelgrootte vs hardware

INFERENTIE-ENGINES

vLLM
Standaard voor productie. Continuous batching, paged attention.

SGLang
Sterk bij agents en complexe pipelines.

TensorRT-LLM
Engine van NVIDIA zelf. Haalt het meeste uit H100 en B300.

Ollama / llama.cpp
Voor lokaal, dev en edge. Niet voor productievolume.

MODELGROOTTE → HARDWARE

Tot ~30B (Qwen3.8 27B) → 1× RTX 5090 of DGX Spark / GX10

30-70B Q4 → 1× H100

70-125B FP16 → 1× B300 (288 GB)

Tot ~460B Q4 → 1× B300

Grote MoE (DeepSeek V4.1) → Meerdere B300's

Quantisatie: GGUF, AWQ, GPTQ. Q4 kost weinig kwaliteit; Q8 of FP16 voor productie als het geheugen het toelaat.

APPENDIX B: RAG-ARCHITECTUUR

De kwaliteit van RAG hangt vooral af van retrieval

1	**Ingestion**	Chunks van 512-1024 tokens met overlap, plus een embedding-model
2	**Vector store**	Qdrant, Weaviate, pgvector of Milvus; kies op latency en schaal
3	**Retrieval**	Hybride zoeken (dense + BM25) met een reranker
4	**Generation**	Open-weight model met system prompt en de opgehaalde context
5	**Evaluation**	RAGAS: faithfulness, answer relevance, context precision en recall

Veelgemaakte fout: Te grote chunks vervuilen de context, te kleine versnipperen hem. Begin bij 512-1024 tokens.

Frameworks: LangChain, LlamaIndex of Haystack. Evaluatie met RAGAS.

APPENDIX C: JURISDICTIE-MATRIX

Wie is verwerker, wie heeft toegang, vanwaar?

Provider	Hosting	EU controller	CLOUD Act	Notitie
OpenAI ChatGPT / API	Azure / eigen; EU-regio	OpenAI Ireland	Ja	EU-residency na goedkeuring
Microsoft 365 Copilot	Azure, EU DB	Microsoft Ireland	Ja*	*Beveiligingsuitzondering
Anthropic Claude direct	VS; EU via Bedrock	n.v.t.	Ja	Geen training; ZDR mogelijk
Anthropic Claude via Copilot	Nu buiten EU DB	Via Microsoft	Ja	EU DB gepland eind 2026
Mistral La Plateforme	EU (FR)	Mistral AI SAS	Nee	Deels open-weight
Open-weight bij Intermax	NL	Klant zelf	Nee	Volledige controle

De vraag is niet "staat data in EU?", maar "onder welke jurisdictie valt de verwerker?"

EU DB = EU Data Boundary; ZDR = Zero Data Retention. Alle Amerikaanse aanbieders vallen onder de CLOUD Act, waar de data ook staat.

APPENDIX D: HARDWARE & BRONNEN

Referentietabel en verder lezen

HARDWARE-ECONOMIE (nov 2026)

B300 cloud on-demand	**$5,80 - 9,50 / uur**
H100 ter vergelijking	**$3,36 / uur (mediaan)**
B300 geheugen	**288 GB; 125B FP16 op één GPU**
70B Q4 baseline	**50-80 tok/s op H100**
70B vLLM batch=32	**1.800-2.000 tok/s aggregaat**
Claude Haiku 5.5 API	**$0,10 in / $0,50 uit per 1M**
Breakeven self-host	**Per use case narekenen**

VERDER LEZEN

- Cyberhaven 2025 AI Adoption Report
- Lewis Silkin: Italiaanse Garante, EUR 15M boete OpenAI
- Microsoft Copilot Blog: Claude in EU Data Boundary, 5 nov 2026
- SLM-Rijk DPIA Microsoft 365 Copilot v2
- Artificial Analysis: open-weights ranglijst, nov 2026
- gpus.io en getdeploying.com: B300- en H100-prijzen, nov 2026
- Anthropic: API-prijzen, nov 2026
- Intermax / UbiOps: NL soevereine infra
- Verordening (EU) 2026/1744 (Digital Omnibus on AI)





Ludo Baauw

Ludo@img.nl

Linkedin.com/in/ludobaauw/