



2 SEPTEMBER 2026

Van AI-fabriek naar AI-productie

Nederland investeert terecht in een nationale AI-fabriek. Tegelijkertijd laat nieuw onderzoek van ABN AMRO zien dat traditionele datacenterontwikkeling steeds harder botst met grenzen aan stroom, netcapaciteit, water, ruimte en maatschappelijk draagvlak. Juist tussen die twee ontwikkelingen ontstaat ruimte voor een andere benadering: **centrale AI-ontwikkeling waar schaal nodig is, en modulaire production-inference waar data, energie en gebruiker samenkomen.**

Het Economisch Bureau van ABN AMRO legt in zijn analyse van 27 augustus 2026 een fundamenteel probleem bloot. De vraag naar datacentercapaciteit blijft door AI sterk groeien, maar de traditionele manier waarop we die capaciteit realiseren botst steeds harder met grenzen aan elektriciteit, netcapaciteit, ruimte, water, materialen, vergunningen en maatschappelijk draagvlak.

ABN AMRO verwacht dat de mondiale datacentercapaciteit tussen 2025 en 2030 bijna verdubbelt. Tegelijk stelt het rapport dat de strategische uitdaging voor Europa niet primair is om simpelweg méér datacenters te bouwen, maar om digitale groei te combineren met energiezekerheid, maatschappelijke acceptatie en concurrentiekracht.

Dat vind ik een belangrijk uitgangspunt. Want tegelijkertijd investeert Nederland terecht fors in eigen AI-capaciteit. Met de Nederlandse AI-fabriek in Groningen ontstaat een nationale infrastructuur voor soevereine AI, met een AI-supercomputer, een expertisecentrum en ondersteuning voor bedrijven, kennisinstellingen en publieke organisaties.

Deze twee ontwikkelingen lijken op het eerste gezicht misschien tegenstrijdig. Enerzijds hebben we dringend méér AI-rekenkracht nodig. Anderzijds wordt het steeds moeilijker om grote concentraties rekenkracht fysiek in te passen.

Volgens mij ligt precies daar een van de belangrijkste infrastructuurvragen voor de komende jaren: moeten we alle toekomstige AI-compute op dezelfde manier organiseren?

Mijn antwoord is nee. Voor het trainen van grote modellen zijn zeer grote, centrale en sterk gekoppelde AI-systemen logisch en noodzakelijk. Maar na training begint een andere markt: inference. Het dagelijks gebruiken van AI in ondernemingen, ziekenhuizen, overheden, industrie, logistiek, energievoorziening en andere operationele processen.

Voor die markt kunnen heel andere infrastructuurprincipes gelden. Dat is waar ik een mogelijke rol zie voor SOLOAI, in combinatie met Zirrow. Niet als concurrent van de Nederlandse AI-fabriek. Juist als mogelijke volgende schakel.

De Nederlandse AI-fabriek is meer dan alleen een grote computer

Om die positionering goed te begrijpen, moeten we eerst af van het beeld dat de Nederlandse AI-fabriek vooral een supercomputerproject is.

De Nederlandse AI-fabriek bestaat uit twee belangrijke onderdelen: een gespecialiseerde AI-supercomputer en een expertisecentrum. De supercomputer is bedoeld voor het ontwikkelen, trainen en fine-tunen van geavanceerde AI-modellen. Het expertisecentrum ondersteunt organisaties bij het ontwikkelen van bruikbare en verantwoorde toepassingen.

Dat is een wezenlijk andere rol dan een commerciële cloudomgeving. De toegang is primair gericht op ontwikkeling, training en R&D. Daarmee ontstaat vanzelf de vervolgvraag: wat gebeurt er wanneer een AI-model is ontwikkeld, getraind, getest en gevalideerd en vervolgens 24 uur per dag onderdeel moet worden van het bedrijfsproces van de gebruiker?

Daar begint een ander infrastructuurvraagstuk.

Van AI Factory naar AI Production

Stel dat een Nederlands maakbedrijf via de AI-fabriek een model ontwikkelt voor visuele kwaliteitscontrole. Of een ziekenhuis ontwikkelt een eigen medische taaltoepassing. Een logistiek bedrijf bouwt een AI-model voor planning. Een overheidsorganisatie ontwikkelt

een soevereine document-assistent. Of een industrieel bedrijf ontwikkelt een model voor predictive maintenance op basis van sensoren en historische productiedata.

In de ontwikkelfase is centrale toegang tot krachtige AI-compute zeer waardevol. Maar wanneer het model succesvol is, verandert de vraag.

Dan gaat het niet langer primair om: kunnen we dit model trainen? Maar om: hoe draaien we dit betrouwbaar, veilig, economisch en schaalbaar binnen onze dagelijkse operatie?

Dan worden latency, beschikbaarheid, cybersecurity, energieverbruik, SLA's, eigenaarschap, kosten, operationele data en levenscyclusmanagement belangrijk. En bij veel toepassingen wordt ook de vraag relevant of organisaties hun operationele data permanent naar een externe cloudomgeving willen sturen.

Voor gereguleerde sectoren, industrie, zorg, overheid, financiële dienstverlening en organisaties met hoogwaardige intellectuele eigendom is dat niet alleen een technologische vraag. Het kan een strategische, juridische en procurementvraag zijn.

Daarmee ontstaat een ontbrekende fysieke brug tussen AI-development en production inference.

NLAIF → SOLOAI → klant

Een mogelijke architectuur zou daarom veel verder kunnen gaan dan alleen: Nederlandse AI-fabriek → AI-model.

De volledige keten kan worden: Nederlandse AI-fabriek → ontwikkeling → training → fine-tuning → validatie → SOLOAI staging → deployment → lokale/private inference → bedrijfswaarde.

Binnen die keten blijft de Nederlandse AI-fabriek doen waar zij voor is opgezet: toegang bieden tot geavanceerde compute, kennis, expertise, experimentatie en ontwikkeling.

SOLOAI richt zich vervolgens op een ander deel van de levenscyclus: het fysiek operationeel maken van inference-capaciteit.

Dat onderscheid is belangrijk. SOLOAI moet daarom niet worden gepositioneerd als 'wij hebben ook een AI-datacenter' en al helemaal niet als 'wij kunnen de Nederlandse AI-supercomputer vervangen'.

De sterkere positionering is: SOLOAI kan de fysieke deploymentlaag worden tussen een succesvolle AI-ontwikkeling en productie-inference bij de organisatie die de AI daadwerkelijk gebruikt.

Dat is in essentie compute mobility from innovation to production.

Train centrally. Deploy where it makes sense.

Daarmee ontstaat een veel interessantere Europese AI-architectuur.

Niet iedere organisatie hoeft zelf een trainingscluster te bouwen. Grote modellen kunnen centraal worden ontwikkeld en getraind binnen NLAIF, andere Europese AI Factories of commerciële AI-infrastructuur.

Maar het eenmaal ontwikkelde model kan vervolgens draaien waar dat operationeel het meest logisch is: bij een onderneming, op een industrieel terrein, in een regionale AI-hub, bij een zorgcluster, naast een grote databron, in een sovereign cloudomgeving of op een locatie waar elektriciteit, glasvezel én een goede bestemming voor restwarmte aanwezig zijn.

De architectuur wordt daarmee: train centrally → validate → deploy selectively → infer locally or regionally.

Niet omdat lokaal altijd beter is, maar omdat de optimale locatie voor training niet automatisch de optimale locatie voor inference is.

Inference verandert de economische logica van AI-infrastructuur

Bij zeer grote training geldt een sterke schaalwet. Duizenden accelerators moeten tegelijkertijd samenwerken. Netwerkarchitectuur, interconnects, storage throughput en zeer hoge vermogensdichtheden zijn daarbij essentieel.

Inference is veel gevarieerder. Een inference-workload kan bestaan uit een knowledge assistant die documenten doorzoekt, een computer-visionmodel dat productielijnen bewaakt, agents die administratieve processen automatiseren of een enterprise-LLM dat honderden of duizenden medewerkers ondersteunt.

Niet iedere workload heeft dezelfde GPU, dezelfde latency of dezelfde vermogensdichtheid nodig. Daarom moeten we oppassen dat de infrastructuur voor de inferencemarkt niet wordt ontworpen volgens het principe: meer GPU's, meer MW's en meer kW per rack is per definitie beter.

Voor SOLOAI geldt juist het tegenovergestelde principe: less is more—zolang performance, beschikbaarheid en toekomstbestendigheid geborgd blijven.

De huidige ontwerpfilosofie gaat uit van modulaire configuraties in het gebied van ongeveer 0,3 tot maximaal 1,2 MW IT-load. Daarbij is 1,2 MW een maximale workload-enveloppe, niet het totale netaansluitvermogen. Ook zeer hoge rackdichtheden moeten als ontwerp- en kwalificatie-enveloppe worden gezien en niet als doel op zichzelf.

Een infrastructuur mag voorbereid zijn op zeer hoge dichtheden zonder dat iedere klant die capaciteit vanaf dag één moet installeren.

Waarom die modulaire benadering juist nu relevant wordt

De Nederlandse AI-fabriek laat ook zien waarom dit vraagstuk belangrijk wordt. De AI-hardwaremarkt verandert razendsnel. Hardware wordt krachtiger, duurder, schaarser en verandert bovendien sneller van generatie dan de fysieke infrastructuur eromheen.

Dat illustreert een structureel probleem van de hele AI-markt: compute-technologie verandert veel sneller dan fysieke infrastructuur.

Een GPU-generatie kan economisch binnen enkele jaren worden ingehaald. Een goed ontworpen elektrische installatie, constructie, koelinfrastructuur en warmte-interface moet veel langer meegaan.

Het is daarom onlogisch om een complete fysieke asset iedere keer mee te laten verouderen met de geïnstalleerde IT.

SOLOAI wordt juist vanuit dat principe ontwikkeld: de fysieke infrastructuurlaag moet meerdere generaties compute kunnen overleven. Racktechnologie, GPU's, CDUs, direct-to-chipkoeling, immersion-technologie, sensoren en elektrische componenten moeten binnen gecontroleerde interfaces kunnen worden vervangen of opnieuw gekwalificeerd zonder dat de volledige asset opnieuw moet worden ontworpen.

Dat maakt de infrastructuur niet hardware-onafhankelijk in absolute zin. Iedere hardwarewijziging moet technisch worden gevalideerd. Maar het ontwerp kan wel accelerator- en technologieflexibel worden gemaakt.

Van stranded capacity naar progressive provisioning

Ook financieel heeft dat gevolgen. Traditionele datacenterontwikkeling kent het risico dat grote hoeveelheden vermogen, koeling en elektrische apparatuur vooraf worden geïnstalleerd terwijl de daadwerkelijke IT-belasting pas later volgt.

Bij snel veranderende AI-hardware wordt dat risico groter. Daarom is progressive provisioning voor inference interessant.

Een SOLOAI kan bijvoorbeeld vanuit een eerste deelbelasting groeien wanneer daadwerkelijk workloads, klanten of contractuele capaciteit beschikbaar komen. De vaste infrastructuur wordt voorbereid op verdere groei, maar actieve componenten hoeven niet allemaal vooraf te worden geïnstalleerd.

De vraag wordt dan niet: hoeveel GPU's kunnen we vandaag maximaal kopen? Maar: hoe ontwerpen we een infrastructuur waarin we gedurende de komende tien of vijftien jaar steeds de economisch meest geschikte compute kunnen plaatsen?

Juist hier sluit het ABN AMRO-rapport aan

ABN AMRO laat zien dat klassieke datacenterontwikkeling steeds sterker wordt beperkt door de locatie.

Netcongestie is volgens het rapport inmiddels een van de belangrijkste selectiecriteria. Nederland heeft in de Europese vergelijking een zeer hoog netcongestierisico, terwijl het land tegelijkertijd uitzonderlijk sterk scoort op digitale infrastructuur en digitale concurrentiekracht.

Dat lijkt een paradox. Nederland heeft precies de ondernemingen, kennis, connectiviteit en digitale infrastructuur die nodig zijn om AI economisch te benutten. Maar juist op de plaatsen waar die economische activiteit geconcentreerd is, wordt grote nieuwe elektrische capaciteit moeilijker beschikbaar.

Een mogelijke oplossing is daarom niet uitsluitend nog grotere datacenters zoeken op locaties waar tientallen of honderden megawatts beschikbaar zijn.

Voor een deel van de inference-markt kunnen we ook andersom redeneren: waar bevinden zich beschikbare energie, glasvezel, gebruikers, data en potentiële warmteafnemers—en hoeveel inference-capaciteit past daar verantwoord bij?

Dat is een andere benadering van locatieontwikkeling.

SOLOAI creëert geen netcapaciteit

Daar hoort wel een belangrijke nuance bij. Een modulaire AI-infrastructuur lost netcongestie niet magisch op. Als op een locatie geen elektriciteit beschikbaar is, kan SOLOAI daar geen megawatt uit het niets creëren.

Dat is ook niet de claim.

De mogelijke waarde zit juist in het kunnen aanpassen aan verschillende locaties en in het voorkomen dat iedere AI-toepassing afhankelijk wordt gemaakt van één enorme centrale aansluiting.

SOLOAI kan worden ontworpen om rekening te houden met beschikbare netcapaciteit, lokale hernieuwbare productie, batterijopslag, glasvezel, koelcondities en lokale warmtevraag.

Daarmee verschuift het model van: zoek 100 MW en bouw daaromheen, naar: zoek de juiste combinatie van workload, energie, locatie en nuttige output.

Voor inference kan dat een belangrijk verschil worden.

Van NLAIF staging naar productie bij de klant

Een van de meest interessante vervolgstappen is het concept van een AI Deployment Zone. Niet als reeds afgesproken onderdeel van NLAIF, maar als een concept dat met het ecosysteem onderzocht zou kunnen worden.

Stel dat een organisatie via het NLAIF-expertisecentrum een succesvolle AI-toepassing heeft ontwikkeld. Het model is getest. De securityarchitectuur is bepaald. De inferencebehoefte is bekend.

Dan kan een SOLOAI-configuratie fysiek of operationeel worden gestaged. Daar worden hardware, softwareomgeving, security, workload, koeling, energieprofiel, monitoring en performance samen getest.

Daarna wordt dezelfde gestandaardiseerde infrastructuur naar de uiteindelijke productielocatie gebracht of volgens dezelfde gevalideerde configuratie elders opgebouwd.

De gedachte is dan: develop → train → validate → stage → deploy → operate.

Dat maakt van AI-compute als het ware een deployable asset. De innovatie zit dan niet alleen in de verplaatsbare box. De innovatie zit in het gestandaardiseerde proces

waarmee een AI-workload vanuit een ontwikkelomgeving naar operationele productie wordt gebracht.

De Nederlandse AI-fabriek kan daardoor méér economische impact creëren

Dat kan ook voor NLAIF interessant zijn.

Het succes van een AI Factory moet uiteindelijk niet alleen worden gemeten aan het aantal modellen dat is getraind, het aantal GPU-uren dat is gebruikt of het aantal organisaties dat toegang heeft gekregen.

De echte economische en maatschappelijke impact ontstaat pas wanneer toepassingen daadwerkelijk in productie komen.

Een model dat na een pilot op de plank blijft liggen, levert weinig productiviteitsgroei op. Een model dat veilig en economisch onderdeel wordt van een ziekenhuis, fabriek, overheidsdienst of mkb-onderneming wel.

Daarom zou naast de klassieke AI Factory ook het begrip AI Production Factory interessant kunnen worden. Niet als nieuw gebouw, maar als keten.

De Nederlandse AI-fabriek creëert succesvolle AI-toepassingen. Een deploymentlaag brengt ze naar productie. De klant gebruikt ze. En vervolgens meten we of zij daadwerkelijk waarde leveren.

Dat is uiteindelijk de stap van innovatie naar economisch verdienvermogen.

Er blijft bovendien een heldere publieke en commerciële scheiding

Juist de rolverdeling kan daarbij relatief zuiver blijven.

NLAIF heeft een publieke en strategische functie rondom AI-expertise, toegang tot compute, onderzoek, training, fine-tuning, validatie, responsible AI en het ondersteunen van innovatieve organisaties.

SOLOAI kan een commerciële fysieke infrastructuurlaag vormen voor productie, deployment, exploitatie, onderhoud, upgrades en SLA's.

De propositie hoeft dus nadrukkelijk niet te zijn: 'NLAIF moet SOLOAI's kopen.'

Een veel interessantere vraag is: kan het NLAIF-ecosysteem een gestandaardiseerde route ontwikkelen waarmee succesvolle Nederlandse AI-toepassingen na validatie naar soevereine productie-inference worden gebracht?

En kan SOLOAI daarin één van de fysieke deploymentopties worden?

Dat is naar mijn mening een veel sterker gesprek.

Zirrow kan de bewijslaag over de hele keten worden

Maar fysieke deployment alleen is niet genoeg.

We moeten voorkomen dat we een nieuwe generatie AI-datacenters bouwen en uiteindelijk opnieuw vooral meten hoeveel elektriciteit het gebouw binnengaat.

PUE blijft relevant, maar vertelt weinig over de vraag of de IT de energie ook daadwerkelijk productief gebruikt.

Bij inference wordt dat steeds belangrijker. Twee infrastructuren kunnen dezelfde PUE hebben, terwijl de ene aanzienlijk meer bruikbare AI-output per kWh levert dan de andere.

Daar komt Zirrow in beeld.

Zirrow kan worden gepositioneerd als een onafhankelijke Real-Data Evidence Layer die gegevens over energie, water, kosten, benutting, governance en workload-performance met elkaar verbindt. Het uitgangspunt is daadwerkelijke operationele data in plaats van alleen gemiddelden of theoretische nameplatewaarden.

Dat kan in combinatie met NLAIF bijzonder interessant worden.

Van modelbenchmark naar operationele benchmark

Tijdens ontwikkeling en training meten we allerlei AI-indicatoren: nauwkeurigheid, latency, modelkwaliteit, geheugengebruik, tokens per seconde en GPU-performance.

Maar na deployment ontstaat een andere set vragen. Wat kost deze workload per dag? Hoeveel energie wordt werkelijk gebruikt? Wat is het energieverbruik per transactie of per miljoen tokens? Hoeveel van de GPU-capaciteit wordt daadwerkelijk benut? Wanneer staat hardware grotendeels idle? Wat is de latency onder echte bedrijfsbelasting?

Wat gebeurt er met performance wanneer vermogen tijdelijk wordt begrensd? Hoeveel warmte wordt geproduceerd? Hoeveel daarvan wordt technisch opgevangen? En hoeveel daarvan wordt daadwerkelijk door een externe afnemer gebruikt?

Daarmee kan Zirrow zorgen voor een continuüm van bewijs: van technisch gevalideerd model naar aantoonbaar efficiënte productie.

Van PUE naar useful AI output

We moeten in digitale infrastructuur uiteindelijk meer gaan sturen op nuttige output per gebruikte resource.

Voor AI kan dat bijvoorbeeld betekenen dat—waar de interfaces dat toelaten—energie en kosten worden gerelateerd aan workload, transactie, inference, tokens, latency en SLA.

Dat is een belangrijke volgende stap. Want een AI-infrastructuur die weinig energie aan koeling besteedt maar waarvan de accelerators een groot deel van de tijd niets nuttigs doen, kan moeilijk optimaal duurzaam worden genoemd.

Een efficiënte machine die inefficiënt werk uitvoert, blijft verspilling.

Ook restwarmte moet bewijsbaar worden

ABN AMRO benadrukt daarnaast het toenemende belang van watergebruik, energie, ruimte en lokaal draagvlak.

Hetzelfde geldt voor restwarmte. In veel duurzaamheidsverhalen wordt gezegd dat warmte 'beschikbaar is voor hergebruik'. Maar beschikbaar zijn en daadwerkelijk worden gebruikt zijn twee verschillende dingen.

Een evidence-laag moet daarom onderscheid kunnen maken tussen warmte die wordt opgevangen, warmte die wordt overgedragen aan een externe interface en warmte die vervolgens aantoonbaar nuttig wordt gebruikt.

Dat lijkt een detail. Maar precies daar zit het verschil tussen duurzaamheidsmarketing en bewijs.

Locatiekeuze verandert daarmee fundamenteel

Wanneer we SOLOAI en Zirrow combineren met de lessen van ABN AMRO ontstaat uiteindelijk een andere methode voor de selectie van AI-locaties.

ABN AMRO beoordeelt locaties onder andere op elektriciteitsbeschikbaarheid, netcongestie, stroomprijs, hernieuwbare energie, digitale infrastructuur, arbeidsmarkt, grond, waterstress, klimaat, regelgeving, data-soevereiniteit en nabijheid van gebruikers. Voor distributed inference kun je daar een extra dimensie aan toevoegen: waar levert een MW AI-compute de meeste economische en maatschappelijke waarde?

Een locatie waar 1 MW beschikbaar is naast een industriële onderneming met grote lokale AI-workloads en een permanente warmtevraag kan economisch interessanter zijn dan een locatie waar 10 MW goedkope stroom beschikbaar is maar alle data, gebruikers en warmteafnemers honderden kilometers verderop zitten.

Niet voor iedere workload. Maar voor sommige wel. En dat onderscheid wordt in de komende jaren steeds belangrijker.

Nederland heeft daarvoor juist een interessante uitgangspositie

ABN AMRO plaatst Nederland ondanks de zeer hoge netcongestie hoog in zijn Europese vergelijking van vestigingscriteria voor datacenters. Nederland combineert zijn problemen op het elektriciteitsnet met een uitzonderlijk sterke digitale infrastructuur en nabijheid van gebruikers en economische activiteit.

We hoeven Scandinavië daarom niet te kopiëren. Voor hyperscale kan goedkope ruimte en overvloedige duurzame elektriciteit een enorm voordeel zijn. Maar voor een deel van de inference-markt heeft Nederland andere sterke kaarten.

We hebben een zeer hoge digitale dichtheid, sterke glasvezelconnectiviteit, universiteiten, SURF, TNO, een groot AI-ecosysteem, hightechindustrie, zorginstellingen, financiële instellingen, logistiek, energieclusters, glastuinbouw en veel economische activiteit op relatief korte afstand van elkaar.

En straks NLAIIF als belangrijke nationale AI-ontwikkelinfrastructuur.

Dat kan samen een heel andere infrastructuurstrategie opleveren dan uitsluitend proberen de grootste AI-campus te bouwen.

Een Nederlandse sovereign AI stack

De Nederlandse AI-fabriek moet daarom niet worden beoordeeld op de vraag of Nederland daarmee de Verenigde Staten of China kan verslaan in absolute GPU-aantallen. Dat is de verkeerde benchmark.

De kracht ligt juist in de combinatie. SURF brengt infrastructuur en supercomputing. TNO brengt toegepaste AI en het expertisecentrum. AIC4NL verbindt het ecosysteem. Samenwerking Noord verbindt regionale gebruikers en economische ontwikkeling. De Nederlandse AI-fabriek brengt dedicated AI-compute.

Voeg daar een private en regionale production-inference-laag aan toe en er ontstaat iets strategisch veel interessanter: een Nederlandse sovereign AI stack.

Niet één computer. Maar een keten.

Van nationale AI-infrastructuur naar Europees model

De Nederlandse AI-fabriek maakt deel uit van het Europese netwerk van AI Factories. Wanneer het concept van gestandaardiseerde production deployment werkt, kan hetzelfde model daarom theoretisch worden gerepliceerd.

European AI Factory → training en ontwikkeling. SOLOAI deployment infrastructure → productisation en staging. Regionale of private locatie → production inference. Zirrow → onafhankelijke operationele evidence.

NLAIF kan dan een Nederlandse pilotomgeving zijn voor een concept dat uiteindelijk breder binnen Europa toepasbaar is.

Dat is veel interessanter dan SOLOAI alleen als een transporteerbaar datacenter te presenteren. De box is het fysieke product. De echte propositie is de overgang: from national AI compute to local AI operations.

De concurrent van SOLOAI is daarom niet NLAIF

Dat brengt mij bij de belangrijkste positioneringskeuze.

SOLOAI moet niet proberen de Nederlandse AI-fabriek te verslaan. Ook niet de Europese AI Factories of Gigafactories. En ook niet proberen een kleinere hyperscaler te worden.

Die markten hebben andere schaalvoordelen en een andere functie.

SOLOAI richt zich juist op private, sectorale en regionale infrastructuur voor enterprise inference, RAG, agents, gereguleerde AI, AI-assisted engineering en geselecteerde fine-tuning, met een modulaire fysieke deployment.

De AI Factory bevindt zich voornamelijk upstream. SOLOAI kan zich downstream positioneren.

Dat maakt samenwerking veel logischer dan concurrentie.

De echte concurrent is de oude manier van denken

De fundamentele concurrent is misschien zelfs helemaal geen specifiek bedrijf.

Het is het idee dat iedere nieuwe digitale toepassing uiteindelijk vraagt om een groter gebouw, meer grond, meer gecontracteerd vermogen, meer koeling, meer hardware, en dat we daarna wel kijken wat er met de energie en warmte gebeurt.

Voor de volgende generatie AI-infrastructuur kunnen we beter precies andersom beginnen.

Welke workload willen we uitvoeren? Welke performance is daarvoor werkelijk nodig? Welke data mogen waar komen? Hoeveel capaciteit moet direct beschikbaar zijn? Welke capaciteit kan later worden toegevoegd? Welke energie is lokaal beschikbaar? Welke workloads zijn flexibel? Welke koeling past bij de hardware? Kan de warmte lokaal nuttig worden gemaakt? Kan de fysieke infrastructuur meerdere IT-generaties meegaan? En kunnen we al die prestaties vervolgens objectief meten?

Pas daarna bepalen we hoeveel infrastructuur we werkelijk nodig hebben.

Meer AI-waarde per beschikbare megawatt

Daarmee kom ik terug bij de conclusie van ABN AMRO. Europa's uitdaging is niet uitsluitend meer datacenters bouwen. We moeten digitale groei combineren met energiezekerheid, maatschappelijke acceptatie en concurrentiekracht.

Voor mij betekent dat dat we niet alleen moeten discussiëren over hoeveel MW AI Nederland nodig heeft, maar vooral over hoeveel economische, maatschappelijke en digitale waarde we uit iedere beschikbare MW kunnen halen.

De Nederlandse AI-fabriek kan daarbij een essentiële rol vervullen aan het begin van de keten. Daar kan AI worden ontwikkeld, getraind, getest en gevalideerd.

SOLOAI kan vervolgens worden onderzocht als mogelijke fysieke infrastructuurlaag waarmee succesvolle workloads vanuit die ontwikkelomgeving naar productie worden gebracht.

En Zirrow kan aantoonbaar maken wat er daarna werkelijk gebeurt: hoeveel energie wordt gebruikt, hoe goed hardware wordt benut, welke performance wordt geleverd, wat de workload werkelijk kost, hoeveel water nodig is, hoeveel warmte daadwerkelijk wordt hergebruikt en hoeveel nuttige AI-output uiteindelijk tegenover de gebruikte resources staat.

Dan ontstaat een compleet andere benadering van AI-infrastructuur.

Niet: train everything centrally and keep adding capacity.

Maar: develop centrally where schaal nodig is. Deploy selectively waar productie logisch is. Measure everything that matters.

Of nog korter: Train centrally. Deploy intelligently. Prove the performance.

Volgens mij kan juist daar een onderscheidende positie ontstaan voor Nederland — en mogelijk ook voor SOLOAI en Zirrow.

Niet door de grootste AI-fabriek van Europa te willen bouwen. Maar door een infrastructuurmodel te ontwikkelen waarmee de kennis en rekenkracht van de AI Factory daadwerkelijk, efficiënt en aantoonbaar worden omgezet in production AI bij Europese organisaties.