



De energie-, water- en grondstoffenvoetafdruk van generatieve AI

Auteur: dr. Peter van Thienen

Samenvatting

Kunstmatige intelligentie en met name de generatieve variant hiervan duikt op steeds meer plaatsen in ons dagelijks bestaan op. De techniek belooft oplossingen voor allerlei problemen en we staan nog maar aan het begin van een revolutie. Maar hier hangt een prijskaartje aan: een voetafdruk met betrekking tot energie, water en grondstoffen. Deze trendalert signaleert de groei van het besef van dit prijskaartje en onderzoekt in welke mate dit zodanig groot is, in de context van technische en maatschappelijke ontwikkelingen, dat deze voetafdruk in de afweging of AI überhaupt moet worden toegepast een rol zou moeten spelen.

Consequenties voor u

	Laag	Middel	Hoog	Beknorte uitleg
Impact				Opname in besluitvorming is onvermijdelijk
Zekerheid				Mate is afhankelijk van maatschappelijke en technologische ontwikkelingen



Afbeeldingen gegenereerd door Dall-E.

Trendbeschrijving en achtergrond

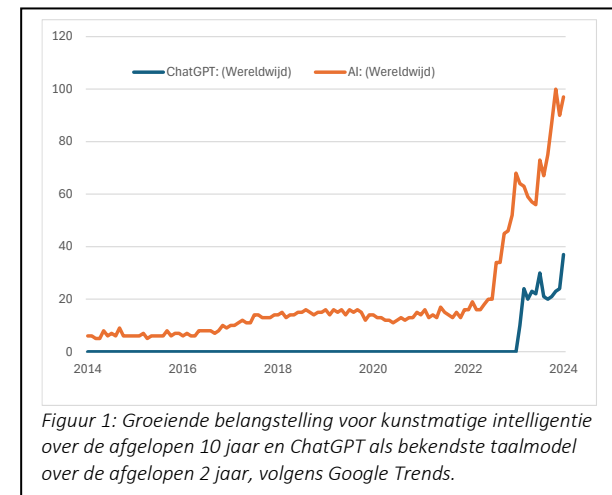
AI in opkomst

Kunstmatige intelligentie (AI, *artificial intelligence*) en met name de generatieve variant hiervan (die in staat is om tekst of afbeeldingen op verzoek te genereren, bijvoorbeeld ChatGPT) duikt op steeds meer plaatsen in ons dagelijks bestaan op. Waar AI tot 2010 vooral in het lab vertoefde, laat zij zich vanaf die tijd steeds vaker in het vrije veld zien (WRR, 2021).

Met name de laatste jaren neemt de belangstelling hiervoor een hoge vlucht (zie Figuur 1). Ook voor de watersector worden veel kansen geïdentificeerd voor toepassing (Makropoulos & Savić, 2019; Doorn, 2021). Het is een sleuteltechniek in de bredere ontwikkeling van de digitalisering van het waterbedrijf, die door sommigen wordt gezien als onontbeerlijk voor het omgaan met toekomstige uitdagingen gerelateerd aan stijgende watervraag en klimaatverandering (Daniels et al., 2023). De Wetenschappelijke Raad voor het Regeringsbeleid beschouwt AI als een systeem-technologie, vergelijkbaar met de stoommachine, elektriciteit, verbrandingsmotor en de computer (WRR, 2021). De afgelopen jaren zijn met name de belangstelling voor en de kracht van generatieve AI, waaronder taalmodellen als GPT, explosief toegenomen. Zowel in de wetenschap als in de maatschappij, en daarmee is het gebruik ervan genormaliseerd. De

techniek belooft oplossingen voor allerlei problemen en we staan nog maar aan het begin van een revolutie. Maar hier hangt een prijskaartje aan. Bij monde van OpenAI-voorman Sam Altman wordt vanuit de AI-industrie zelfs onderkend dat de energievraag van toekomstige kunstmatige intelligentie zo groot zal zijn, dat een doorbraak in duurzame energieopwekking noodzakelijk is (Reuters, 2024). Maar het gaat niet alleen om energie; de datacenters waarop de AI (naast andere energieslurpers als cryptovaluta en streamingdiensten) draait gebruiken ook grote hoeveelheden zoet water voor koeling (Crawford, 2024). Verder zijn voor hun productie grote hoeveelheden grondstoffen nodig, waaronder veel metalen die uit mijnbouw worden gewonnen. Bovendien worden apparaten waar AI op draait na een relatief korte levensduur afgedankt (Hildebrandt en Jung, 2023). Deze negatieve effecten dreigen, wederom, geëxporteerd te worden naar opkomende economieën in het mondiale zuiden (Mongabay, 2024). De ontwikkelingen gaan hard, en vooralsnog lijkt de energiehonger van modellen steeds groter te worden, o.a. omdat deze modellen steeds groter worden (Ludvigsen, 2023).

“Als we deze niet in de hand houden, zou de milieu-impact van AI rampzalig kunnen zijn. Een exponentiële groei in de energiebehoefte zal de bronnen onder druk zetten, klimaatverandering



Figuur 1: Groeiende belangstelling voor kunstmatige intelligentie over de afgelopen 10 jaar en ChatGPT als bekendste taalmodel over de afgelopen 2 jaar, volgens Google Trends.

verergeren, en een nieuwe klasse van datavervuilers voortbrengen. We bevinden ons op een tweesprong: we kunnen het potentieel van AI op onverantwoordelijke wijze omarmen en daarmee onherstelbare schade riskeren, of we kunnen haar kracht in de hand houden voor een groenere toekomst.” (Mustafa, 2023)

Trendalert

Deze trendalert signaleert niet zozeer de opkomst van generatieve AI – deze is overduidelijk en vergt geen nadere toelichting – maar veeleer de groei van het besef dat er een prijskaartje aan deze ontwikkeling hangt. Het onderzoekt in welke mate de voetafdruk van generatieve AI met betrekking tot energie, water en grondstoffen



zodanig groot is, dat deze in de afweging of AI moet worden toegepast een rol zou moeten spelen en of dat deze te verantwoorden is. Ook worden ontwikkelingen in de techniek en maatschappij besproken die zullen bepalen hoe groot generatieve AI en haar prijskaartje daadwerkelijk zullen zijn.

Relevantie

De watersector onderkent de mogelijkheden die AI haar biedt en zet eerste stappen om deze in de bedrijfsvoering op te nemen. Tegelijkertijd hebben de waterbedrijven en waterschappen vaak relatief ambitieuze doelstellingen m.b.t. broeikasgasemissies geformuleerd (in ieder geval in vergelijking met veel andere sectoren). De vraag hoe deze twee ontwikkelingen zich tot elkaar verhouden, en of de eerstgenoemde de tweede in de weg kan zitten, is daarom uiterst relevant.

In het onderstaande beschouwen we eerst hoe we naar de milieu-impact van het gebruik van AI kunnen kijken. We richten ons daarbij specifiek op generatieve AI in de vorm van grote taalmodellen (die in antwoord op vragen of bevelen in een natuurlijke tekst genereren, zoals het eerder genoemde ChatGPT). Vervolgens proberen we deze te kwantificeren in algemene zin en werken we

enkele getalsmatige voorbeelden voor de watersector uit.

Aspecten en fasen

Aan het gebruik van een AI-model gaan noodzakelijkerwijs een aantal stappen vooraf: het model moet worden ontwikkeld en getraind, hetgeen in datacenters gebeurt en veel rekentijd, en daarmee energie en water, vergt. Deze datacenters zelf zijn opgebouwd uit vele servers, waarvan de productie zelf ook energie en water vergt, en bovendien de nodige grondstoffen. Die laatste moeten aan het eind van de levensduur van een server ook verwerkt worden.

Voor de volledige milieuvoetafdruk van de toepassing van AI moeten we daarom naar de volgende aspecten kijken (Wu et al., 2022, Ludvigsen, 2023b):

1. extractie van grondstoffen;
2. productie van materialen;
3. productie van rekeninfrastructuur (computers, etc.);
4. experimenteren met en trainen van modellen;
5. gebruik van modellen;
6. einde levensduur rekeninfrastructuur.

In dit trendalert voeg ik alle aspecten m.b.t. de hardware (1,2,3,6) samen in één categorie en beschouwen we deze in verhouding tot de andere twee (4 en 5). Het

experimenteren met modellen voorafgaand aan de trainingsfase wordt door veel auteurs niet apart genoemd. Dit aspect wordt in deze gevallen verondersteld onderdeel te zijn van de trainingsfase.

Energie- en CO₂-voetafdruk

Het is dus zinvol om onderscheid te maken tussen het trainen van een model (stap 4), dat zeer energie-intensief is, maar in de basis eenmalig, en het gebruik van het model (stap 5), dat veel minder energie kost, maar heel vaak gebeurt (Verdecchia et al., 2023). De emissies die met beide stappen gepaard gaan, kunnen worden geclassificeerd volgens de gangbare methode van 3 zogenaamde scopes volgens het Greenhouse Gas (GHG)-protocol:

- scope 1: directe emissies van eigen activiteiten;
- scope 2: indirecte emissies van aangekochte energie; in dit geval vooral elektriciteit.
- scope 3: indirecte emissies in de waardeketen.

Voor een datacenter vallen de emissies van training en gebruik in scope 2; voor een aanbieder van AI-modellen of een gebruiker hiervan vallen zij in scope 3. Ook de emissies gepaard gaande met de winning van grondstoffen, productie van materialen en infrastructuur, en het afdanken hiervan, vallen voor beide onder scope 3.



Watervoetafdruk

Voor de watervoetafdruk van AI moeten we kijken naar de datacenters waar deze op draaien. Li et al. (2023) geven een overzicht hiervan. Het waterverbruik van een datacenter kan, net als de uitstoot van broeikasgassen, worden opgedeeld in 3 scopes:

- Scope 1: waterconsumptie in/door het datacenter zelf, t.b.v. koeling van servers. Hierbij gaat water deels verloren naar de atmosfeer door verdamping in open (recirculerende) koeltorens en wordt water deels geloosd naar het oppervlaktewater als het door indamping niet meer aan de kwaliteitseisen voor het koelsysteem voldoet (spui); en voor bevochtiging van lucht bij luchtkoeling in koelere klimaten.
- Scope 2: watergebruik bij de productie van de elektriciteit waarop het datacenter draait – ook

Kader 1: CO₂-equivalenten

Hoewel CO₂ het bekendste broeikasgas is, zijn er ook andere gassen (zoals lachgas, methaan, etc.) en activiteiten (zoals ontbossing) die een opwarmend effect op de atmosfeer hebben. Om deze met elkaar te kunnen vergelijken, worden de effecten vaak omgerekend naar een hoeveelheid CO₂ die een vergelijkbaar klimaateffect heeft (op een gekozen tijdschaal): CO₂-equivalenten of CO₂eq.

dit water wordt met name voor koeldoeleinden gebruikt.

- Scope 3: gebruik van water in de productie van de componenten van het datacenter – met name de productie van halfgeleiders heeft een grote watervoetafdruk.

Grondstoffen en afval

Omdat de focus van beschouwingen van de milieuvoetafdruk van AI op dit moment ligt bij energie in de eerste plaats en bij water in de tweede plaats, is er minder cijfermatig materiaal beschikbaar in de literatuur over de impact (bijvoorbeeld op landgebruik en biodiversiteit, Sonter et al., 2020) van de winning van grondstoffen en de verwerking (of niet) van elektronisch afval aan het einde van de levenscyclus van servers (bijv. vervuiling van ecosystemen, Purchase et al., 2020). Hildebrandt en Jung (2020) geven een meer anekdotisch overzicht van deze effecten. In deze trendalert beperken we ons daarom tot het benoemen van het bestaan van deze effecten en de noodzaak naar meer onderzoek hiernaar en een betere kwantificering hiervan.

Getalsmatige inschattingen energieverbruik en broeikasgasemissies

Er zijn de laatste jaren enkele publicaties uitgekomen in de formele en informele literatuur die een inschatting proberen te maken van het energieverbruik en de

bijbehorende broeikasgasuitstoot die gepaard gaan met het trainen en het gebruiken van AI-modellen. Voor de trainingsfase van bekende grote taalmodellen BLOOM, GPT-3, Gopher en OPT worden energiehoeveelheden in het bereik van 300-1.300 MWh gerapporteerd (De Vries, 2023). De bijbehorende uitstoot van broeikasgassen ligt in het bereik van 30-550 ton CO₂eq (zie Kader 1) (Luccioni et al., 2022). Voor de gebruiksfase liggen de schattingen voor ChatGPT3 en BLOOM op ca. 3-4 Wh per interactie ('request'); dat is ca. 10 maal zoveel als voor een traditionele Google-zoekopdracht (De Vries, 2023; Luccioni et al., 2023). Voor GPT-4 ligt dit nog een factor 3 hoger (Schreiner, 2023). Met het toenemen van het aantal parameters in nieuwere grote taalmodellen, neemt ook de energiebehoefte van het trainingsproces toe. Voor GPT-4 is ingeschat dat er 52.000-62.000 MWh aan stroom is gebruikt (Ludvigson, 2023), een ordegrrootte hoger dan voor de vorige generatie, en hetzelfde geldt voor de klimaatvoetafdruk (tot 15.000 ton CO₂eq). Waarbij de auteur nog opmerkt: *"Hoewel deze aantallen klein lijken, verbleken ze in vergelijking met de impact op het milieu van andere stadia van de levenscyclus van een AI-model, zoals het uitrolstadium."*

Tabel 1: Overzicht van inschattingen van energie- en waterverbruik en CO₂-voetafdrukken van generatieve AI-modellen in de trainings- en gebruiksfase. Inschattingen ontleend aan de literatuur zijn normaal weergegeven en voorzien van een bronverwijzing (letter). Door de auteur van dit trendalert afgeleide getallen zijn cursief weergegeven. a: Luccioni et al. (2023); b: Li et al. (2023); c: Ludvigsen (2023); d: De Vries (2023); e: Ludvigson (2024); f: Schreiner (2023); g: Luccioni et al. (2024)

model	aantal parameters	trainingsfase (+finetuning)				watergebruik <i>m3 (scope 2)</i>		gebruiksfase		CO2eq-voetafdruk <i>g CO2eq per interactie (inference)</i>	watergebruik <i>ml per interactie (inference)</i>	
		energiegebruik <i>MWh</i>		CO2eq-voetafdruk <i>ton CO2eq</i>				energiegebruik <i>kWh per interactie (inference)</i>				
(Chat)GPT-3	175 miljard	1287	a	552	a	700	b	0.00239-0.0029	c,d	0.07-1.69	10-61	b
BLOOM	176 miljard	433	a	30	a	238-6699		0.00397	a	0.11-2.31	17-83	
GPT-4	1,8 biljoen	52000-62000	e	1200-15000	e	28000-964000		0.00717-0.0087	f	0.2-5.06	30-182	
BLOOMz-560M	560 miljoen	11.048	g	0.3-6.4		6-171		5.40E-05	g	0.002-0.031	0.2-1.1	
BLOOMz-1B	1 miljard	18.133	g	0.5-10.6		10-281		6.20E-05	g	0.002-0.036	0.3-1.3	
BLOOMz-3B	3 miljard	28.876	g	0.8-16.8		16-447		7.30E-05	g	0.002-0.042	0.3-1.5	
BLOOMz-7B	7 miljard	59.257	g	1.7-34.5		33-917		1.00E-04	g	0.003-0.058	0.4-2.1	

Berekening van de afgeleide parameters:

- CO2eq-voetafdruk: vermenigvuldiging van het energieverbruik met lage en hoge koolstofintensiteit voor de opwekking hiervan, namelijk 28 g/kWh (IJsland) en 582 g/kWh (China) (bron: Our World In Data, 2024)
- Waterverbruik scope 1 (datacenter): geschaald met energieverbruik per query op basis van waterverbruik volgens Li et al. (2023).
- Waterverbruik scope 2 (elektriciteitsproductie): vermenigvuldiging van energieverbruik met waterintensiteit van de opwekking hiervan, tussen de 550 en 15470 ml/kWh (Lohrmann et al., 2021)
- Energieverbruik GPT-4: drie maal dat van GPT-3 (Schreiner, 2023)



Om een gevoel bij deze getallen te geven:

- het zetten van een kopje thee of het opladen van een smartphone kost ca. 10-15 Wh, hetgeen dus vergelijkbaar is met 1-3 interacties met een taalmodel;
- Waterbedrijf Groningen rapporteert voor 2022 een totale CO₂-voetafdruk van iets minder dan 10.000 ton CO₂eq; het trainen van GPT-4 heeft dus een grotere klimaatvoetafdruk dan het voorzien van een hele provincie van water voor een heel jaar.

De productie van een server resulteert naar schatting in de uitstoot van 3.700 kg CO₂-eq (Ludvigsen, 2023, op basis van documentatie van de HP Enterprise ProLiant DL345 server). Er zouden bijna 3.600 van deze servers nodig zijn om 195 miljoen dagelijkse interacties met ChapGPT te verwerken (Patel en Ahmad, 2023).

Getalsmatige inschattingen waterverbruik

Over het waterverbruik van datacenters is veel minder bekend dan over hun energieverbruik. Li et al. (2023) hebben als eerste een inschatting hiervan gemaakt. Zij concluderen dat er bij het uitvoeren van 10-50 interacties (inferences) met ChatGPT3, 500 ml water wordt verbruikt. Het gaat hierbij om verbruikt water in de datacenters (scope 1) en om water dat is verbruikt bij de elektriciteitsproductie (scope 2). Beide kunnen sterk

variëren met de locatie van het datacentrum en de wijze van elektriciteitsopwekking.

Overzicht

De hierboven beschreven getallen geven nog geen compleet overzicht, maar wel voldoende handvatten om een inschatting van het waterverbruik te maken. In Tabel 1 zijn de resultaten van de inschatting van het waterverbruik van generatieve AI modellen in de trainings- en gebruiksfase weergegeven. Hierbij zijn sommige getallen uit de literatuur overgenomen en sommige ingeschat op basis van andere getallen.

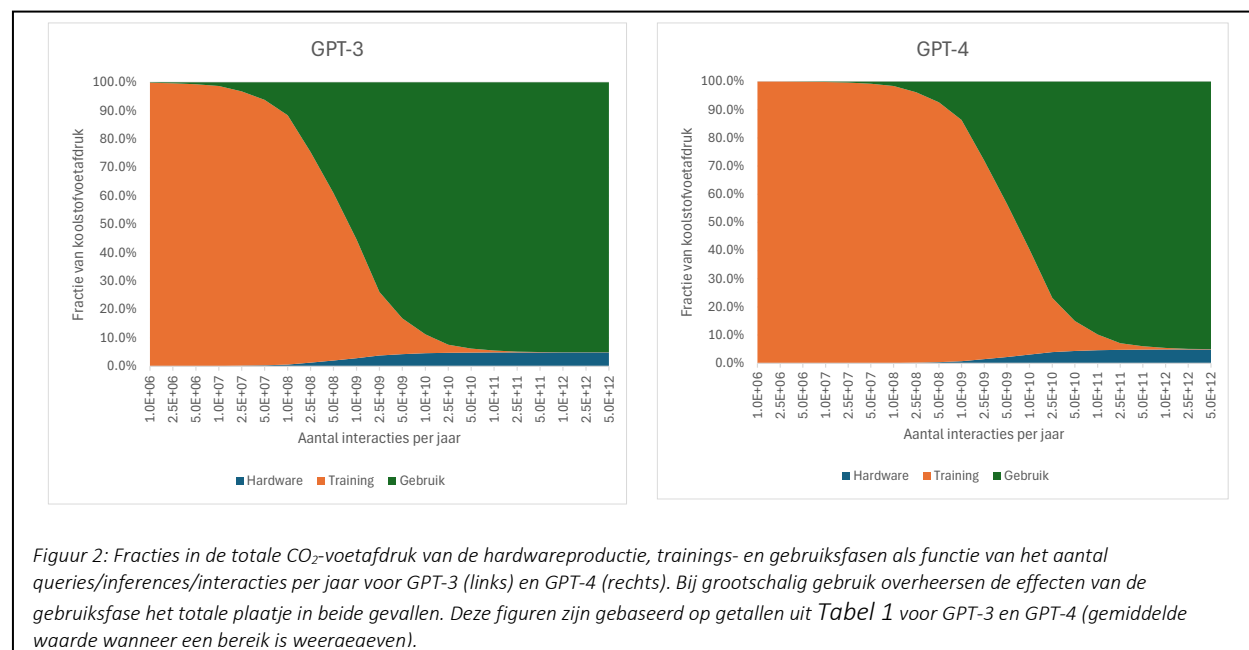
Verhoudingen tussen hardware, training en gebruiksvoetafdrukken

De omvang van de uitstoot van broeikasgassen die aan de gebruiksfase kan worden toegekend, hangt uiteraard af van de mate van gebruik, ofwel hoeveel vragen er aan het model worden gesteld. Hoe de verhouding tussen de uitstoot gerelateerd aan hardware, training en gebruik varieert als functie van het aantal interacties per jaar, is weergegeven in Figuur 2, voor GPT-3 en GPT-4. Voor GPT-3 wordt de CO₂-voetafdruk van het gebruik dominant bij meer dan ca. 1 miljard interacties per jaar (5 dagelijkse interacties voor ca. een half miljoen gebruikers); voor GPT-4 ligt deze grens een factor 10 hoger, op ca. 10 miljard interacties per jaar (5 dagelijkse interacties voor ca. vijf miljoen gebruikers). Hier blijft de

verhouding tussen de CO₂-voetafdrukken van hardware en energieverbruik constant, omdat extra gebruik ook extra servers vergt om deze modellen op te draaien. In de meest extreme gevallen, aan de rechterkant van deze diagrammen, stellen 2,75 miljard aardbewoners dagelijks 5 vragen aan de modellen. Hiervoor is dan 0,07% (GPT-3) respectievelijk 0,20% (GPT-4) van de mondiale elektriciteitsproductie van 2023 nodig (29.479 TWh, Our World in Data, 2024).

Groei van AI

AI zou een van de grootste en snelst groeiende gebruiken van datacenters zijn (Li et al., 2023). Deze auteurs schatten in dat de wereldwijde energiebehoefte van AI in 2027 93,5-147,4 TWh zal zijn. Ter vergelijking: Nederland produceerde in 2023 122 TWh aan stroom (Our World in Data, 2024). De bijbehorende waterbehoefte zal volgens Li et al. (2023) 4,2-6,0 miljard m³ (inname) respectievelijk 0,38-0,60 miljard m³ (consumptie) zijn. Opnieuw ter vergelijking: de Nederlandse drinkwaterbedrijven zetten in 2023 1,1 miljard m³ water af (Vewin, 2024). De Vries (2023) rekent voor dat op basis van de productiegetallen voor rekeneenheden voor servers verwacht kan worden dat jaarlijks servers met een energiebehoefte van 85-134 TWh worden geïnstalleerd – hoewel het de vraag is of deze alle tot hun volle capaciteit zullen worden benut en daarmee zoveel stroom zullen vragen. Ook het



Internationaal energieagentschap (IEA, 2024) komt op een vergelijkbare schatting voor de energievraag uit. Met deze getallen wordt duidelijk dat de wereldwijde grootschalige inzet van AI op korte termijn al tot de elektriciteits- en waterconsumptie kan leiden die vergelijkbaar is met die van een land als Nederland.

Groene AI

Sinds er aandacht voor de milieu-impact van kunstmatige intelligentie is gekomen, is er eveneens een onderzoeksveld ontstaan dat zich hierop richt: groene

kunstmatige intelligentie (“Green AI” of “Groene AI”). Dit veld kan als volgt worden gedefinieerd (Verdecchia et al., 2023): “Groene AI beschouwt praktijken die zich richten op het gebruik van AI om de impact die mensen hebben op het milieu in de zin van het gebruik van natuurlijke hulpbronnen te mitigeren, en/of om de impact van de AI zelf op het milieu te mitigeren.”

Verbetering van efficiëntie AI

Een veelgenoemde manier om de milieu-impact te verminderen, is het vergroten van de efficiëntie van AI.

Een voorbeeld hiervan aan de softwarekant is de toepassing van kleinere, taakspecifieke modellen in plaats van grote, generieke modellen als GPT-4, die een veel kleinere energiebehoefte hebben (Luccioni et al., 2024). Er zijn daarnaast diverse ontwikkelingen gaande die resulteren in gespecialiseerde chips, zoals TPU’s (Tensor Processing Units) die meer bewerkingen voor AI-toepassingen kunnen uitvoeren bij een lager energieverbruik (Davies, 2024). Hierdoor neemt de energie-efficiëntie van de hardwarekant toe. Een meer gedetailleerd overzicht van mogelijke technische maatregelen om de efficiëntie van AI te vergroten wordt geboden door Wu et al. (2022).

Het waterverbruik van AI hangt direct samen met het energieverbruik, dus een verbetering van de energie-efficiëntie zou ook moeten resulteren in een verbetering van de waterefficiëntie. In tegenstelling tot de (fossiele en nucleaire) energiesector *verbruikt* de halfgeleiderindustrie maar weinig water van zijn totale *gebruik* – het meeste is daarna opnieuw beschikbaar. De sector zet in op een grotere efficiëntie van het waterverbruik en op de bouw van afvalwaterzuiveringen die als bron van proceswater kunnen dienen (Mone, 2024).

Wright et al. (2023) laten echter zien dat alleen het efficiënter maken van AI niet voldoende is om deze duurzaam te maken. Zij benoemen drie discrepanties, namelijk:



- 1) rekefficiëntie, energie-efficiëntie en koolstofefficiëntie zijn om uiteenlopende redenen niet hetzelfde;
- 2) efficiëntiemaatregelen kunnen op verschillende wijzen doorwerken in de verschillende fasen van de levenscyclus van een AI-model;
- 3) de milieu-impact van de hardware waar de AI op draait kent meer aspecten dan alleen dat van energie en broeikasgasuitstoot.

Deze auteurs roepen daarom op om weliswaar werk te maken van efficiëntie, maar zich bewust te blijven van de beperkingen van dit perspectief, en daarbij een breder systeem perspectief te beschouwen.

Paradox van Jevons

In de 19^e eeuw stelde econoom William Stanley Jevons dat een toename in efficiëntie van het gebruik van een delfstof (in zijn geval steenkool) kan leiden tot een *toename* van het gebruik hiervan in plaats van een *afname*, omdat de efficiëntie de prijs van het gebruik drukt en de toegankelijkheid daarmee vergroot – de paradox van Jevons. Deze paradox is mogelijk eveneens van toepassing op generatieve AI. Omdat AI repetitieve taken van de mens kan overnemen en deze daardoor efficiënter werkt, zou kunnen worden verondersteld dat het aantal banen zal dalen. Echter, doordat AI nauwere integratie tussen verschillende sectoren mogelijk maakt,

en vanwege de mensen die zowel technische als sociale en beleidsmatige zaken rondom AI regelen, zou zij juist nieuwe banen creëren (Agostini, 2023).

Een variant van deze redenering is dat de perceptie dat AI tot een vergroting van efficiëntie van bestaande processen leidt, zal leiden tot toepassing van AI op een brede waaier van processen die mogelijk niet allemaal veel ruimte tot optimalisatie bieden (omdat ze al redelijk dicht bij optimaal zijn), en daarmee netto juist tot een afname van efficiëntie zou kunnen leiden.

En ook voor de efficiëntievergroting van de AI-modellen zelf (waarover later meer) kan worden gesteld dat deze aan de paradox van Jevons onderhevig zijn (De Vries, 2023). In ieder geval dragen de integratie van AI in diverse digitale producten (Copilot, etc.) eenvoudige gebruikersinterfaces en de mogelijkheid om eenvoudige of repeterende taken te automatiseren bij aan een brede acceptatie en inzet van AI.

Er zijn tekenen dat deze paradox ook weer optreedt bij de digitalisering van onze maatschappij. Santarius et al. (2020) laten zien dat de digitalisering in ieder geval niet per se tot een ontkoppeling van economische groei en stroomverbruik leidt door een sterke groei van de vraag.

Milieu-economische Kuznetscurve

De milieu-economische Kuznetscurve is een hypothese die stelt dat bij toenemende economische productiviteit de milieubelasting eerst toeneemt, maar bij een verdere

groei weer afneemt volgens een omgekeerde U-vorm doordat er bij voldoende welvaart een groter belang aan milieukwaliteit wordt gehecht (Stern, 2017). Hoewel dit waar lijkt te zijn voor een aantal soorten van schadelijke uitstoot (zoals bijvoorbeeld SO₂), lijkt deze hypothese niet universeel op te gaan, en blijkt er in veel gevallen sprake te zijn van een N-vormige curve, oftewel een curve die uiteindelijk weer verder omhoog gaat (Wang et al., 2024). Deze auteurs waren niet in staat om eenduidige conclusies te trekken over de vorm van de curve voor de milieubelasting door AI. Of anders gezegd, we weten nog niet of de grootschalige introductie van AI uiteindelijk tot een afname of een toename van haar milieuvoetafdruk zal leiden.

Samenvattend

De voetafdruk van AI m.b.t. de uitstoot van broeikasgassen is groot en de verwachting is dat deze sterk zal groeien. Het is op dit moment niet duidelijk of diverse technische en andere ontwikkelingen het mogelijk zullen maken om deze impact binnen acceptabele grenzen te houden.



Omgang en handelingsperspectief

Voor CIO's en IT beslissers in en buiten de watersector is het niet alleen belangrijk om de technische aspecten te beschouwen, maar ook maar ook hoe organisaties strategieën kunnen ontwikkelen om AI duurzaam in te zetten. Dit omvat onder andere de monitoring en rapportage van het water- en energieverbruik, de CO₂-uitstoot, het kiezen van groene cloud-datacenters, het uitvoeren van milieu-impactanalyses en het afleggen van verantwoording van de duurzame inzet van AI. Dit alles binnen het bredere kader van een organisatie (en de samenleving als geheel).

Kaack et al. (2022) hebben een raamwerk opgesteld om de effecten van machinelearning (ML) op broeikasgasemissies in beeld te brengen. Dit raamwerk kijkt niet alleen naar de uitstoot die direct voortkomt uit het uitvoeren van ML-berekeningen, maar ook naar de bredere effecten van ML-toepassingen. Dit kan zowel positieve effecten hebben, zoals bij onderzoek naar groene technologie, als negatieve effecten, zoals bij het ondersteunen van olie- en gaswinning. Daarnaast wordt er ook rekening gehouden met systeemniveau-effecten, zoals de Jevons-paradox, waarbij efficiëntieverbeteringen kunnen leiden tot een toename in energieverbruik. Zij stellen een aantal mogelijke maatregelen voor, variërend van economische instrumenten en regulering tot monitoring en *corporate climate action*.

Kader 2: Scenario's voor de CO₂-voetafdruk van AI in de watersector

Om de in dit trendalert besproken getallen te plaatsen in de context van de watersector, wordt in dit kader ter illustratie een aantal scenario's voor het gebruik van AI in waterbedrijven geschetst en wordt ingeschat hoe de totale CO₂-voetafdruk hiervan zich verhoudt tot die van de productie en distributie van water. Hierbij wordt uitgegaan van dezelfde koolstofintensiteit voor elektriciteit t.b.v. productie, afzet, en datacenters. Drie scenario's gaan uit van een 90% reductie van de koolstofintensiteit (OER: operationele emissiereductie). We gaan uit van de energie-intensiviteit van GPT-3, GPT-4, en een hypothetische nieuwe generatie AI (nextgen) die nog eens een factor 10 meer energie gebruikt.

Scenario	Uitgangspunten voor energieverbruik voor productie en afzet ^a	Uitgangspunt voor AI-gebruik
basis	jaar 2023	10 queries/medewerker /dag
AI spaarzaam	jaar 2023	2 queries/medewerker /dag
AI overal	jaar 2023	1000 queries/medewerker /dag
OER	jaar 2023 met 90% reductie koolstofintensiteit	10 queries/medewerker /dag
OER, AI spaarzaam	jaar 2023 met 90% reductie koolstofintensiteit	2 queries/medewerker /dag
OER, AI overal	jaar 2023 met 90% reductie koolstofintensiteit	1000 queries/medewerker /dag

^aop basis van Vewin Kerngegevens Drinkwater 2023

Bij deze scenario's zien we dat alleen in de twee scenario's met alomtegenwoordige AI het energieverbruik hiervan boven de 1% van het gesommeerde energieverbruik (productie, afzet en AI) uitkomt. De voorzichtige conclusie die hieruit getrokken kan worden is dat de watersector weliswaar vanuit algemene principes, elders in dit trendalert besproken, naar weloverwogen toepassing van AI zou moeten streven, maar dat toepassingen van AI die bijdragen aan het verminderen van emissies of het oplossen van andere milieuproblemen waarschijnlijk in de meeste gevallen meer zullen opbrengen dan dat zij milieuschade opleveren. Vanzelfsprekend is dit iets dat van geval tot geval zou moeten worden beoordeeld.





In ieder geval zouden klimaatoverwegingen een centraal uitgangspunt moeten zijn in beleid m.b.t. de ontwikkeling en toepassing van AI. Hierbinnen valt 1) het bevorderen van onderzoek naar en ontwikkeling van klimaatpositieve toepassingen van ML, 2) het vereisen van transparantie en verantwoordelijkheid, en 3) het toepassen van technologiebeoordelingen die klimaateffecten meewegen op gebruiksscenario's die normaal gesproken niet binnen klimaatbeleid vallen, maar waarvan wel gevolgen voor toekomstige klimaatimpact kunnen worden verwacht. De AI Act van de Europese Unie bevat wat dit betreft aanzienlijke lacunes (Alder et al., 2024).

Volgens Wu et al. (2022) is het eerste dat organisaties die AI willen toepassen zouden moeten ontwikkelen een duurzaamheidsmentaliteit. Op basis hiervan kunnen allerlei technische maatregelen genomen worden om de efficiëntie van de toepassing van AI te maximaliseren en daarmee de voetafdruk *per gebruik* te minimaliseren. Zij betogen ook dat het nodig is juist ook dat deel van de CO₂-voetafdruk van AI-systemen aan te pakken dat niet met efficiëntiemaatregelen opgelost kan worden.

Hildebrandt en Jung (2023) betogen dat het bepaald geen gegeven is dat de introductie van AI leidt tot een groene toekomst. Zij menen daarom dat deze ontwikkeling sterker begeleid en vormgegeven moet

worden op het niveau van de volledige maatschappij, op basis van democratische processen en een machtsbalans tussen de burgermaatschappij en grote technologiebedrijven, met globale solidariteit langs de volledige toeleveringsketens, en binnen planetaire grenzen. Khanal et al. (2024) beschrijven hoe deze machtsbalans (verder) aan het verschuiven is in het voordeel van de technologiebedrijven.

Naast technologische ontwikkelingen die AI energiezuiniger maken, is het ook noodzakelijk dat AI-bedrijven transparanter worden over de milieuvoetafdruk van hun diensten, en dat overheden hier kaders voor opstellen (Crawford, 2024).

Handelingsperspectief voor de watersector

Het hierboven genoemde kader van Kaack et al. (2022) biedt ook de watersector een aantal beleidsmatige knoppen om aan te draaien om de milieu-impact van haar eigen AI-gebruik te minimaliseren:

1. voer een interne koolstofprijs in die ook emissies veroorzaakt door berekeningen en applicaties meeneemt (scope 1, 2, 3, inclusief (cloud)diensten);
2. minimaliseer het overbodige hertrainen en gebruik van modellen;
3. maak energie-efficiëntie een sleutelcriterium bij de beoordeling van modellen;

4. draai modellen op locaties en tijdstippen met een lage koolstofintensiteit van de gebruikte elektriciteit;
5. vermijd toepassingen van AI die een toename van broeikasgasemissies veroorzaken en bevorder toepassingen die emissies helpen verminderen;
6. bepaal en rapporteer het aandeel van AI in de totale voetafdruk van het bedrijf.

Het lijkt op dit moment echter niet waarschijnlijk dat gerichte of zelfs bredere toepassing van AI door de watersector ertoe zal leiden dat dit een belangrijke factor wordt in haar uitstoot van broeikasgassen (zie Kader 2).

Het echte probleem van grootschalige implementatie en gebruik van AI zit hem echter in de schaal. Daarom is het in algemene zin verstandig om AI daar in te zetten waar het positieve impact heeft, daar waar het duidelijke toegevoegde waarde heeft in het oplossen van een probleem, het vergroten van veerkracht of het verkleinen van uitstoot. En daarom is het verstandig om het gebruik van AI voor triviale taken of taken zonder duidelijke positieve impact te vermijden. Dit geldt in de eerste plaats voor de maatschappij als geheel (hoewel deze over de middelen noch de zelfbeheersing lijkt te beschikken om een dergelijke ambitie te realiseren; hiervoor lijkt een vooruitziend en gedurfd politiek



ingrijpen noodzakelijk), en afgeleid daarvan ook voor de waterbedrijven.

In lijn met deze gedachtegang benadrukt De Vries (2023) dat het verstandig is om zich niet alleen te richten op de efficiëntie van AI, maar om zich in de eerste plaats af te vragen of voor de beoogde toepassing überhaupt AI nodig is.

Meer informatie

- Agostini, M. (2023) AI and the Future of Work: Harnessing Jevons Paradox and Moore's Law for Employment Growth. <https://medium.com/@tarifabeach/ai-and-the-future-of-work-harnessing-jevons-paradox-and-moores-law-for-employment-growth-df1778c7dd6e>, bezocht op 22 mei 2024.
- Alder, N., Ebert, K., Herbrich, R., & Hacker, P. (2024). AI, Climate, and Transparency: Operationalizing and Improving the AI Act. *arXiv preprint arXiv:2409.07471*.
- Crawford, K. (2024) Generative AI is guzzling water and energy. *Nature*, vol. 626, doi: <https://doi.org/10.1038/d41586-024-00478-x>
- Daniel, I., Ajami, N. K., Castelletti, A., Savic, D., Stewart, R. A., & Cominola, A. (2023). A survey of water utilities' digital transformation: drivers, impacts, and enabling technologies. *npj Clean Water*, 6(1), 51. <https://doi.org/10.1038/s41545-023-00265-7>
- Doorn, N. (2021). Artificial intelligence in the water domain: Opportunities for responsible use. *Science of the Total Environment*, 755, 142561. <https://doi.org/10.1016/j.scitotenv.2020.142561>
- Hacker, P. (2024). Sustainable AI regulation. *Common Market Law Review*, 61(2). <https://arxiv.org/pdf/2306.00292>
- Hildebrandt, F., & Jung, M. (2023). Digital, grün, global gerecht: Mit Digitalisierung das Klima retten?. *Forschungsjournal Soziale Bewegungen*, 36(3), 392-403.
- International Energy Agency (2024) Electricity 2024. Analysis and forecast to 2026. <https://iea.blob.core.windows.net/assets/18f3ed24-4b26-4c83-a3d2-8a1be51c8cc8/Electricity2024-Analysisandforecastto2026.pdf>, bezocht op 8 juli 2024.
- Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 12(6), 518-527. <https://doi.org/10.1038/s41558-022-01377-7>
- Khanal, S., Zhang, H., & Taihagh, A. (2024). Why and how is the power of Big Tech increasing in the policy process? The case of generative AI. *Policy and Society*, puae012.
- Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making ai less "thirsty": Uncovering and addressing the secret water footprint of ai models. *arXiv preprint arXiv:2304.03271*. <https://arxiv.org/pdf/2304.03271>
- Lohrmann, A., Child, M., & Breyer, C. (2021). Assessment of the water footprint for the European power sector during the transition towards a 100% renewable energy system. *Energy*, 233, 121098.
- Luccioni, A. S., Viguier, S., & Ligozat, A. L. (2023). Estimating the carbon footprint of bloom, a 176b parameter language model. *Journal of Machine Learning Research*, 24(253), 1-15. <https://www.jmlr.org/papers/v24/23-0069.html>
- Luccioni, S., Jernite, Y., & Strubell, E. (2024, June). Power hungry processing: Watts driving the cost of AI deployment?. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 85-99). <https://dl.acm.org/doi/pdf/10.1145/3630106.3658542>
- Ludvigson, K. (2023) The carbon footprint of GPT-4. <https://towardsdatascience.com/the-carbon-footprint-of-gpt-4-d6c676eb21ae>, bezocht op 27 mei 2024.
- Ludvigson, K. (2023b) Environmental Impact of Ubiquitous Generative AI.



- <https://medium.com/p/9e061bac6800> , bezocht op 27 mei 2024.
- Makropoulos, C., & Savić, D. A. (2019). Urban hydroinformatics: Past, present and future. *Water*, 11(10), 1959. <https://doi.org/10.3390/w11101959>
 - Mone, G. Thirsty Fabs: How the surprising water footprint of semiconductor manufacturing, datacenters, and AI solutions is sparking sustainable development. <https://dl.acm.org/doi/full/10.1145/3640352>
 - Mogabay (2024) Critics fear catastrophic energy crisis as AI is outsourced to Latin America. <https://news.mogabay.com/2024/03/critics-fear-catastrophic-energy-crisis-as-ai-is-outsourced-to-latin-america/>
 - Mustafa, N. (2023) Beyond Hype, Beyond Footprint: Unlocking AI's Green Potential. *AI Mind*. <https://pub.aimind.so/beyond-hype-beyond-footprint-unlocking-ais-green-potential-ba7ca43eecd2> , bezocht op 31 mei 2024.
 - Our World in Data (2024) Electricity production by source. <https://ourworldindata.org/grapher/electricity-production-source-stacked>, bezocht op 8 juli 2024.
 - Patel, D., en A. Ahmad (2023) The inference cost of search disruption – large language model cost analysis. <https://www.semianalysis.com/p/the-inference-cost-of-search-disruption>, bezocht op 1 juli 2024.
 - Purchase, D., Abbasi, G., Bisschop, L., Chatterjee, D., Ekberg, C., Ermolin, M., ... & Wong, M. H. (2020). Global occurrence, chemical properties, and ecological impacts of e-wastes (IUPAC Technical Report). *Pure and Applied Chemistry*, 92(11), 1733-1767.
 - Reuters (2024) OpenAI CEO Altman says at Davos future AI depends on energy breakthrough. <https://www.reuters.com/technology/openai-ceo-altman-says-davos-future-ai-depends-energy-breakthrough-2024-01-16/>
 - Santarius, T., Pohl, J., & Lange, S. (2020). Digitalization and the decoupling debate: can ICT help to reduce environmental impacts while the economy keeps growing?. *Sustainability*, 12(18), 7496.
 - Sonter, L. J., Dade, M. C., Watson, J. E., & Valenta, R. K. (2020). Renewable energy production will exacerbate mining threats to biodiversity. *Nature communications*, 11(1), 4174.
 - Stern, D. I. (2017). The environmental Kuznets curve. In *Oxford research encyclopedia of environmental science* (p. 401). Oxford, UK: Oxford University Press.
 - Tomlinson, B., Black, R. W., Patterson, D. J., & Torrance, A. W. (2024). The carbon emissions of writing and illustrating are lower for AI than for humans. *Scientific Reports*, 14(1), 3732. <https://doi.org/10.1038/s41598-024-54271-x>
 - Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(4), e1507. <https://wires.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/widm.1507>
 - Vewin (2024) Kernegevens drinkwater 2023. https://www.vewin.nl/sitecollectiondocuments/publicaties/5283-vewin-kernegevens-2023-a6-staand_nl-web.pdf
 - de Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, 7(10), 2191-2194. <https://doi.org/10.1016/j.joule.2023.09.004>
 - Wright, D., Igel, C., Samuel, G., & Selvan, R. (2023). Efficiency is Not Enough: A Critical Perspective of Environmentally Sustainable AI. *arXiv preprint arXiv:2309.02065*. <https://arxiv.org/pdf/2309.02065>
 - Wu, C. J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., ... & Hazelwood, K. (2022). Sustainable ai: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems*, 4, 795-813. https://proceedings.mlsys.org/paper_files/paper/2022/file/462211f67c7d858f663355eff93b745e-Paper.pdf



- WRR (2021) Opgave AI. De nieuwe systeemtechnologie. WRR-rapport 105.
https://www.wrr.nl/binaries/wrr/documenten/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie/WRRRapport_%20Opgave%20AI_De%20nieuwe%20systeemtechnologie_NR105WRR.pdf

Sleutelwoorden

AI; kunstmatige intelligentie; energievoetafdruk; watervoetafdruk

Auteur: dr. Peter van Thienen

Thematisch Trendalert: Hydroinformatica

Projectnummer: 404300/084/003