De kansen en risico's van Large Language Models bij onderzoek

Arvid van Dam, Xin Tian (KWR)

Large Language Models (LLM's) zoals GPT en Copilot bieden ook in de watersector nieuwe mogelijkheden voor onderzoek. Door hun vermogen om grote hoeveelheden data te verwerken en nieuwe informatie te genereren zijn ze op verschillende manieren toepasbaar in de verschillende fasen van het onderzoeksproces. In elke onderzoeksfase brengen LLM's ook specifieke risico's met zich mee, zoals mogelijke misinformatie, bias, gebrek aan moreel inzicht en milieubelasting. Hoewel sommige beperkingen steeds beter beheerst worden, blijft voorzichtigheid geboden om de kwaliteit, creativiteit en milieuvriendelijkheid van het onderzoek te behouden.

De watersector is een belangrijke opdrachtgever, uitvoerder, en gebruiker van onderzoek. De opkomst van generatieve AI, en in het bijzonder Large Language Models (LLM's) als GPT van OpenAI, kan grote invloed hebben op de manier waarop onderzoek wordt opgezet, uitgevoerd en gerapporteerd. LLM's bieden mogelijkheden voor nieuwe manieren van werken, dataverwerking en informatievoorziening. Ook voor het onderzoek in de watersector biedt dit kansen.
Tegelijkertijd zijn er belangrijke ethische en maatschappelijke kanttekeningen te plaatsen bij het gebruik van LLM's. Zo kunnen ze misinformatie genereren en bevooroordeeld zijn, kunnen ze soms privacygevoelige informatie ontsluiten en hebben ze een enorme milieu-impact door het hoge energie- en waterverbruik.
Dus hoe zijn LLMs op een verantwoordelijke manier in te zetten in onderzoek in de watersector? Hoe kunnen onderzoekers en kennisgebruikers deze techniek gebruiken en waar moeten ze op letten? In dit artikel wordt dieper ingegaan op hoe LLMs werken en hoe zij zich verhouden tot kennisontwikkeling. Daarna worden de kansen en risico's van LLMs in de verschillende stappen van het onderzoeksproces besproken. Tot slot worden enkele aanbevelingen gedaan voor onderzoekers en opdrachtgevers in de sector.

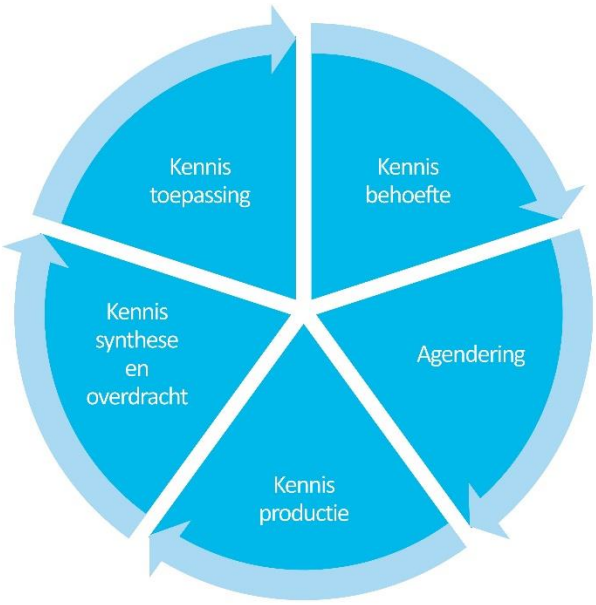
Wat 'begrijpt' een LLM?

Taal is het belangrijkste communicatiemiddel tussen mensen. Het is een complex systeem dat voortdurend evolueert, zich aanpast en verandert. Lange tijd waren computers niet in staat om menselijke talen op dezelfde manier te begrijpen. Met de ontwikkeling van LLMs zijn computers nu in staat om teksten te begrijpen (*Natural Language Understanding*) en te genereren (*Natural Language Generation*). Maar hoe 'intelligent' zijn LLM's nou eigenlijk?
Om te begrijpen wat de impact van LLM's zal zijn op de productie, de overdracht en het gebruik van kennis, moeten we eerst (kort) nagaan wat we onder 'kennis' verstaan en hoe LLM's zich hiertoe verhouden. Een interessant perspectief hierop komt van theoretisch taalkundige Noam Chomsky. Voor Chomsky en zijn collega's [1] is menselijk redeneren wezenlijk anders dan de taalverwerking van LLM's. LLM's analyseren enorme hoeveelheden gegevens om statistische patronen te identificeren en op basis daarvan nauwkeurige voorspellingen doen over het volgende woord in een reeks. Deze voorspellingen worden gedaan op basis van statistische waarschijnlijkheid, afgeleid uit bestaande teksten en aangevuld met de feedback van mensen waarmee het model getraind wordt.

Dit staat in contrast met het diepere begrip dat ten grondslag ligt aan de menselijke cognitie. De menselijke geest kan namelijk intelligent werken met zeer kleine hoeveelheden informatie. Ons brein, zo stelt Chomsky, "probeert geen brute correlaties tussen datapunten af te leiden, maar verklaringen te creëren." En juist dit vermogen om te verklaren is volgens hem de sleutel tot ware intelligentie.
LLM's zijn dus in staat om naar waarschijnlijkheid voorspellingen te doen, maar missen het vermogen om uit te leggen wat onwaarschijnlijk of onmogelijk is. Om een hypothetisch voorbeeld te geven: op basis van de beschikbare schriftelijke informatie in het begin van de zestiende eeuw zouden LLM's, puur gebaseerd op statistische taalvoorspelling, waarschijnlijk het idee hebben gereproduceerd dat de zon om de aarde draait. De heliocentrische theorie van Copernicus was hoogst uitzonderlijk—en controversieel—maar zou de manier waarop we het universum nu begrijpen fundamenteel veranderen.
Dit heeft te maken met het feit dat de menselijke intelligentie juist het vermogen is om ideeën te verwoorden. Met andere woorden: voor mensen liggen ideeën ten grondslag aan de formulering in taal. Deze ideeën kunnen statistisch onwaarschijnlijk, maar toch inzichtelijk zijn. Dit fundamentele verschil tussen mensen en LLM's betekent dat de verwerking van teksten door LLM's niet moeten worden verward met begrip. LLM's kennen alleen de formulering in taal, niet het idee. Ze missen het contextuele, 'real-world'-begrip dat mensen hebben.
Een ander fundamenteel verschil is dat mensen moreel denken, terwijl LLM's dat niet kunnen. Ze zijn niet immoreel (als in: fout), maar amoreel: simpelweg niet in staat om vanuit morele principes te redeneren en fundamenteel onverschillig voor de uitkomsten van hun voorspellingen. Dit betekent dat het voor LLM's moeilijk is om statistische waarschijnlijkheid te onderscheiden van voor mensen acceptabele kennis. Daarom worden de belangrijkste LLM's die momenteel in gebruik zijn, zoals ChatGPT, door hun ontwikkelaars beperkt om geen controversiële output te leveren. Maar dit roept nieuwe vragen op, bijvoorbeeld op wiens morele kader deze beperkingen gebaseerd zouden moeten zijn, en hoever ontwikkelaars hierin moeten gaan.
Kortom, theorie en kennis zijn subjectief en creatief, wat de beperkingen van LLM's laat zien als het gaat om het produceren van nieuwe kennis. Terwijl LLM's informatie verwerken, wordt kennis geproduceerd tussen mensen en wordt deze altijd in context geplaatst. Anderzijds biedt de enorme capaciteit van LLM's om grote hoeveelheden informatie te verwerken en hier een hoogstwaarschijnlijke interpretatie van te geven zeker kansen in het onderzoeksproces.

Kansen en risico's van LLM's in het onderzoeksproces

Om beter in beeld te krijgen hoe LLM's ingezet kunnen worden bij onderzoek, is het handig het onderzoeksproces zelf iets te ontleden. Toegepast onderzoek, zoals veel in de watersector wordt ingezet, verloopt in een aantal fases, van kennisbehoefte en agendering, via kennisproductie, synthese en overdracht naar de uiteindelijke toepassing waar vaak weer nieuwe kennisvragen uit voortkomen. Deze fases zijn gedefinieerd in de valorisatiecyclus (zie afbeelding 1) [2]. In elk van deze fases wordt op verschillende manieren informatie verwerkt. Dat betekent dat LLM's ook op verschillende manieren kunnen worden toegepast gedurende het onderzoeksproces en dat elke stap met specifieke kansen en risico's gepaard gaat.



Afbeelding 1. Onderzoeksfases zoals gedefinieerd in de valorisatiecyclus

Kennisbehoefte

Uit de praktijk komen continu kennisvragen. In deze fase definiëren opdrachtgevers en onderzoekers de doelen, vragen en gewenste uitkomsten van het onderzoek, op basis van kennisbehoeften en maatschappelijke ontwikkelingen.

LLM's kunnen in deze fase worden gebruikt om bestaande informatie te scannen. Dit kan bijvoorbeeld de bibliotheek van een onderzoeksinstituut zijn, of de interne database van een drinkwaterbedrijf of waterschap. Deze kennis kunnen ze samenvatten in thema's en onderzoekslijnen, en visualiseren in bijvoorbeeld een tijdlijn of kennisgrafieken. Zo kunnen ze helpen om de bestaande kennis in de organisatie of in de sector beter in beeld te brengen en ontbrekende kennis te identificeren, wat weer kan leiden tot nieuwe onderzoeksvragen of inzicht kan geven in problemen. Ze kunnen ook scenario's en hypothetische situaties opstellen om nieuwe vragen aan het licht te brengen. Praktisch gezien kunnen LLM's helpen bij het schrijven van projectvoorstellen en het formuleren van (deel)vragen.

Daar staat tegenover dat door LLM's gegenereerde onderzoeksvragen disproportioneel gericht zijn op al bestaande kennis, omdat de modellen afhankelijk zijn van de bestaande data waarop ze getraind zijn. Het kan dus lastig zijn om met echt vernieuwende ideeën te komen. Ook de kans op vooroordelen en het gebrek aan contextueel begrip van LLM's kunnen leiden tot betekenisloze of ongepaste formuleringen van kennisbehoeften.

Agendering

Als de kennisbehoefte duidelijk is, wordt deze geagendeerd en geprioriteerd in onderzoeksprogramma's. LLM's kunnen nuttig zijn bij het samenvatten van verschillende onderzoeksideeën en om passende project- of programmastructuren te ontwikkelen. Opdrachtgevers kunnen LLM's in deze fase gebruiken om te helpen bij het beoordelen van projectvoorstellen.

De prioritering van kennisbehoeften vereist echter een aanzienlijk contextueel begrip, wat LLM's missen. Daardoor kunnen ze bijvoorbeeld de relevantie van een voorstel dat niet aan alle criteria voldoet, maar wél belangrijk is, over het hoofd zien.

In deze fase hangt veel af van de machtsstructuren, waarden en belangen van de betrokken personen en organisaties. Hierdoor kan de op waarschijnlijkheid gebaseerde benadering van LLM's een beperkende factor zijn. Dit beperkt hun bruikbaarheid in deze fase en kan, bij onzorgvuldig gebruik, leiden tot voorbarige in- en uitsluiting van ideeën. Bovendien kunnen bevooroordeelde trainingsdata de resultaten van de prioritering verstoren.

Kennisproductie

Als een onderwerp eenmaal geagendeerd is, de rollen zijn verdeeld en tijd en middelen zijn toegekend, kan het onderzoeksproject beginnen. In deze fase wordt kennis gegenereerd, wat uiteindelijk leidt tot de onderzoeksresultaten. Soms is dit het werk van onderzoekers, maar in veel gevallen zijn ook opdrachtgevers en eindgebruikers betrokken als procesbegeleiders, in klankbordgroepen, of leveren zij een actieve bijdrage in co-creatieprocessen.

Een belangrijk voordeel van LLM's is hun vermogen om relevante informatie te verwerken, iets dat cruciaal is in de kennisproductie. Ze kunnen de benodigde codering voor modellen genereren en grote datasets verwerken. In sociaalwetenschappelijk onderzoek kunnen LLM's helpen bij het ontwikkelen van vragenlijsten en kwalitatieve data sneller verwerken door interviews te transcriberen, onderwerpen te coderen en interviewverslagen op te stellen. Het zal interessant zijn om te zien in hoeverre LLM’s in de nabije toekomst kunnen worden gebruikt in co-creatieprocessen (zoals bij ontwerponderzoek of actieonderzoek). LLM's zouden nuttige hulpmiddelen kunnen worden in brainstormsessies en ideeëngeneratie.

Het gevaar hierbij is dat LLM's misinformatie kunnen produceren (‘hallucinatie’ genoemd), wat kan leiden tot fouten in de dataverwerking en interpretatie van resultaten. Door het gebrek aan transparantie van LLM's is het niet goed mogelijk om de stappen tussen data en resultaten te traceren. In de sociale wetenschappen kan de beperkte toegang van LLM's tot menselijke ervaringen een beperkende factor zijn, waarbij vooroordelen een grote rol spelen. Een andere belangrijke zorg is de gevoelige aard van data. Privacy en veiligheid kunnen in het geding komen als data worden verwerkt door een publiek toegankelijk model.

Kennissynthese en -overdracht

In deze fase vatten onderzoekers de resultaten samen en vertalen deze naar output die is afgestemd op de ontvanger. Deze output (vaak in de vorm van een rapport of artikel) wordt vervolgens gedeeld via direct of indirect contact tussen onderzoekers en de potentiële gebruikers van die kennis.

LLM's kunnen zeer praktisch zijn bij het ontwikkelen van output die afgestemd is op de behoeften van verschillende doelgroepen, bijvoorbeeld door op basis van een onderzoeksrapport blogs, socialmediaposts en artikelen te schrijven. Bovendien kunnen LLM's worden ingezet om narratieven en scenario's te genereren die onderzoeksresultaten tastbaarder te maken voor de doelgroepen. De mogelijkheid om visualisaties te genereren (grafieken, infographics) kan zeer krachtig zijn voor het verspreiden van onderzoeksbevindingen.

Ook in deze fase geldt dat de toepasbaarheid van LLM's wordt beperkt door de bias en misinformatie die LLM's kunnen genereren. Op een meer fundamenteel niveau hangt kennisoverdracht grotendeels af van menselijke interacties, iets wat LLM's niet kunnen vervangen.

Kennistoepassing

In deze fase wordt de nieuwe kennis verspreid, opgenomen in trainingen, tools of beleid, vertaald naar besluitvorming en investeringen, en in de praktijk gebracht. Hierbij kunnen LLM's bijvoorbeeld nuttig zijn om verbanden te identificeren tussen het onderzoek en de bredere kennisbasis van een organisatie. Ze kunnen ook worden gebruikt om bestaande databases bij te werken met de nieuwe kennis.

Auteursrechten en hallucinaties zijn belangrijke zorgen in deze fase, omdat kennisgebruikers (zowel onderzoekers als opdrachtgevers en andere belanghebbenden) moeten kunnen vertrouwen op de validiteit van de kennis.

LLM's gebruiken in onderzoek?

Concluderend wordt uit een analyse van de manieren waarop informatie wordt verwerkt in de verschillende fasen van onderzoek, duidelijk dat LLM's een zeer krachtig hulpmiddel kunnen zijn. Onderzoeksinstellingen wereldwijd verkennen momenteel de mogelijkheden en ontwikkelen toepassingen. Ook het onderzoek in de watersector kan hiervan profiteren.

De ontwikkeling van LLM's gaat snel en naar verwachting zullen er steeds betere modellen beschikbaar komen. De toepasbaarheid van LLM's voor onderzoek zal dan ook toenemen. Recente ontwikkelingen pakken ook al een aantal van de in dit artikel genoemde inherente beperkingen aan.

Retrieval-Augmented Generation (RAG), bijvoorbeeld, combineert LLM's met relevante data van buiten de trainingsgegevens. Dit vermindert hallucinaties en zorgt ervoor dat het model nauwkeurigere, actuelere antwoorden kan bieden. GraphRAG bouwt hierop voort door gestructureerde informatie zoals kennisgrafieken op te nemen, waardoor het model de context en logische consistentie binnen complexe structuren kan behouden. Een ander voorbeeld is LangChain, een raamwerk voor ontwikkelaars om LLM's te integreren met verschillende externe gegevensbronnen, wat ook resulteert in meer genuanceerde en contextueel passende interacties.

Door LLM's te koppelen aan betrouwbare externe gegevensbronnen en gestructureerde informatie, zorgen deze technieken ervoor dat antwoorden meer gefundeerd en contextbewust zijn. Hierdoor worden ze beter geschikt voor onderzoekstoepassingen waarbij nauwkeurigheid en traceerbaarheid cruciaal zijn.

Toch blijft het belangrijk om in gedachten te houden dat mensen informatie anders verwerken dan LLM's. Voor sommige taken zijn LLM's niet geschikt en zullen mensen nodig blijven— met name als het gaat om taken die contextueel begrip vereisen, of die vragen om normatieve keuzes. De toepassing van LLM's in onderzoek (of de verwachting hiervan door opdrachtgevers) kan ook druk leggen op projectbudgetten, omdat menselijke arbeid in het onderzoeksproces mogelijk wordt verminderd. De vraag is dan of de focus komt te liggen op efficiëntie of kwaliteit en creativiteit. Ook de milieu-impact van grootschalig gebruik vraagt om serieuze afwegingen—zeker voor de watersector is de watervraag van LLM-servers een gewetensvraag. Hoewel LLM's sommige stappen in het onderzoeksproces kunnen verbeteren en versnellen, moet het gebruik ervan bewust gebeuren en in evenwicht worden gebracht met menselijk toezicht om de nauwkeurigheid en integriteit van het onderzoek te waarborgen.

Referenties

1. Chomsky, N., Roberts, I. and Watumull, J. (2023). 'The False Promise of ChatGPT'. *The New York Times*, March 8. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html, geraadpleegd op 1 november 2024.

2. Munaretto, S., Mooren, C.E. and Hessels, L.K. (2022). 'Valorization of transdisciplinary research: An evaluation approach and empirical illustration'. *Research Evaluation* 31(3): 355–371. https://doi.org/10.1093/reseval/rvac019