Rijksinstituut voor Volksgezondheid
en Milieu
*Ministerie van Volksgezondheid,
Welzijn en Sport*

Kennisnotitie

Naar effectievere signalering en prioritering van probleemstoffen in drinkwater: een verkenning

RIVM

A. van Leeuwenhoeklaan 9
3721 MA Bilthoven
Postbus 1
3720 BA Bilthoven
www.rivm.nl

T 088 689 89 89

Auteurs:
T.J.E. Janson
D.M. Narain-Ford
N.G.F.M. van der Aa
J. Hartmann
Centrum:
Duurzaamheid Milieu en Gezondheid (DMG)
Contact:
julia.hartmann@rivm.nl
Kenmerk:
KN-2026-0087
DOI:
10.21945/RIVM-KN-2026-0087
Datum:
1 oktober 2026

Aanleiding en context

Het Nederlandse drinkwater is gezond en van goede kwaliteit.[1] Tegelijkertijd staat de kwaliteit van de bronnen onder druk, waardoor de uitdaging toeneemt om ook in de toekomst een goede drinkwaterkwaliteit te waarborgen.[2–4] Dit vraagt een continue inspanning, waarbij het essentieel is om tijdig opkomende stoffen in drinkwater(bronnen) te signaleren en te beoordelen op mogelijke risico's. In deze kennisnotitie verwijst de term 'opkomende stoffen' naar alle stoffen waarvoor geen norm voor drinkwater(bronnen) bestaat.[a] Het betreft bijna altijd organische microverontreinigingen: synthetische, organische stoffen die voorkomen in drinkwater(bronnen) in concentraties van enkele nano- of microgrammen per liter.

Voor een effectieve aanpak van opkomende stoffen is het nodig om de periode te verkorten tussen het moment van signaleren en het uitrollen van effectieve maatregelen. Om maatregelen te kunnen nemen (bijvoorbeeld het verminderen van lozingen via industrieel afvalwater) zijn er drinkwaterrichtwaarden nodig. Drinkwaterrichtwaarden zijn gezondheidskundige grenswaarden. Ze worden doorgaans door het RIVM afgeleid en dienen als advies voor het Ministerie van Infrastructuur en Waterstaat (IenW). Het ministerie stelt uiteindelijk beleidsmatig de drinkwaterrichtwaarde vast. De drinkwaterrichtwaarde kan gebruikt worden voor het wegen van de risico's voor de drinkwaterkwaliteit bij het opstellen van lozingsvergunningen. Er is behoefte aan prioritering van stoffen waarvoor met voorrang drinkwaterrichtwaarden moeten worden afgeleid. Ook is er behoefte aan handvatten om de beschikbare informatie over deze stoffen optimaal te gebruiken in het beleidsproces, bijvoorbeeld voor het vaststellen van wettelijke kwaliteitseisen in drinkwater.

Drinkwaterbedrijven, waterschappen, provincies en Rijkswaterstaat spelen een belangrijke rol bij het identificeren van opkomende stoffen. Drinkwaterbedrijven voeren metingen uit in drinkwater(bronnen) aan de hand van het Richtsnoer risicogestuurd monitoren.[4] De door de provincies opgestelde gebiedsdossiers vormen hiervoor een belangrijke informatiebron.[5] Gebiedsdossiers bevatten alle relevante informatie over een drinkwaterwin-locatie en het bijbehorende grond- of oppervlaktewatersysteem. Ze geven inzicht in mogelijke risico's voor de waterkwaliteit, bijvoorbeeld door landbouw, industrie of verkeer. Het RIVM heeft in 2020 een overzicht gegeven van de staat van de Nederlandse drinkwaterbronnen op basis van de gebiedsdossiers met daarin ook een overzicht van de geïdentificeerde opkomende stoffen.[4] Rijkswaterstaat gebruikt de monitoringsgegevens van de drinkwaterbedrijven om opkomende stoffen te identificeren in oppervlaktewater dat wordt gebruikt voor de bereiding van drinkwater. Dit vormt onderdeel van het Protocol monitoring en toetsing drinkwaterbronnen[6].

[a] Deze groep stoffen wordt ook wel aangeduid als ongereguleerde stoffen, niet-wettelijk genormeerde stoffen, nieuwe verontreinigingen, niet-natuurlijke verontreinigingen, emerging contaminants, emerging substances, chemicals of emerging concern. Veel van deze termen worden ook buiten het kader van de drinkwaterregelgeving gebruikt.

Hoewel deze verschillende processen in de afgelopen jaren hebben geleid tot de identificatie van tientallen opkomende stoffen, zijn ze arbeidsintensief en missen ze mogelijk relevante stoffen.

Zowel de gebiedsdossiers als internationale onderzoeksnetwerken en -projecten rond opkomende stoffen, bieden aanvullende kansen om de signalering en aanpak van opkomende stoffen in drinkwaterbronnen verder te verbeteren. Het RIVM heeft in opdracht van het Ministerie van Infrastructuur en Waterstaat onderzocht hoe de gebiedsdossiers en internationale onderzoeksnetwerken en -projecten beter kunnen worden benut.

Kansen voor betere signalering en prioritering van probleemstoffen voor drinkwater

Voor een betere benutting van de gebiedsdossiers en internationale onderzoeksnetwerken en -projecten ziet het RIVM de volgende kansen:

1) Gebruiken kunstmatige intelligentie (AI) om de provinciale gebiedsdossiers effectiever te doorzoeken op probleemstoffen, zodat relevante informatie snel en effectief uit de meer dan 200 afzonderlijke dossiers kan worden gehaald.
2) Aansluiten bij toonaangevende internationale projecten en netwerken die zich richten op de signalering en prioritering van opkomende stoffen, namelijk:
   a) De werkgroep prioritering van het Norman Netwerk,
   b) Het Zweedse project 'New practice for assessing and managing risks of undesirable chemical substances in water',
   c) Het PARC Early Warning System (EWS).

De kansen pakken verschillende momenten in het signaleringsproces aan. Kans 1 richt zich op het signaleren van opkomende stoffen bij de bron. Kans 2a gaat over het prioriteren van stoffen die al gesignaleerd zijn. Kans 2b is het volgen van een Zweeds project dat zich richt op het signaleren en prioriteren van opkomende stoffen in drinkwater(bronnen). Kans 2c richt zich op het signaleren van nieuwe risico's door opkomende stoffen in de breedste zin van het woord. Dit is dus niet specifiek gericht op blootstelling via drinkwater, en is ook breder dan de definitie van opkomende stoffen die wordt gehanteerd in deze kennisnotitie.

Verkenning kansen

Kans 1: gebruiken AI om effectiever probleemstoffen te identificeren in gebiedsdossiers
De Europese drinkwaterrichtlijn verplicht een risicoanalyse en risicobeheer voor 'onttrekkingsgebieden voor onttrekkingspunten'.[7] Het Protocol Gebiedsdossiers voor Publieke Drinkwaterwinningen geeft hier invulling aan.[5] De gebiedsdossiers bevatten informatie over probleemstoffen die in winputten, gezamenlijk ruwwater of in waarnemingsputten worden gemeten. Probleemstoffen zijn opkomende stoffen (stoffen zonder bestaande wetgeving en/of norm) waarvan de aangetroffen concentratie een algemene signaleringswaarde van 0,1 µg/L overschrijdt. Stoffen met een concentratie hoger dan 75% van de signaleringswaarde worden als potentiële probleemstof geclassificeerd. Ook worden risico's, kenmerken, restopgaves en relevante ontwikkelingen van de drinkwaterwinning beschreven. Deze documenten zijn de bron geweest voor het rapport 'Staat Drinkwater-bronnen'[4].

Van Driezum et al.[4] gebruikten de gebiedsdossiers om de toestand van de drinkwaterbronnen in Nederland in beeld te brengen, waarbij ze onder meer lijsten maakten van huidige en potentiële probleemstoffen. Dit zijn stoffen die worden

aangetroffen in een concentratie die in de buurt van, of boven de geldende norm ligt. Deze lijsten bieden handvatten voor het signaleren van opkomende stoffen. De auteurs verwerkten hiervoor handmatig de informatie uit 216 gebiedsdossiers in een samenvattende tabel. Vanaf nu wordt naar deze tabel verwezen als 'de basistabel'.

In de basistabel zijn per gebiedsdossier (rijen) een aantal variabelen (kolommen) opgenomen, dit betreft een breed scala aan thema's uit de gebiedsdossiers. Het handmatig creëren van de basistabel is arbeidsintensief en foutgevoelig. Daarom is onderzocht of het mogelijk is om dit proces te automatiseren met behulp van 'Large Language Models' (LLM). LLM's zijn geavanceerde modellen die zijn getraind om menselijke taal te begrijpen en te genereren. Ze werken op basis van AI en worden getraind met enorme hoeveelheden tekst, waardoor ze in staat zijn om teksten te analyseren, samenvattingen te maken, vragen te beantwoorden of zelfs teksten te schrijven.

In deze verkenning is een workflow met 3 stappen opgezet, met behulp van GPT- 4.1 (taalmodel van OpenAI, beschikbaar binnen het RIVM [b]), waarbij de focus ligt op het identificeren en extraheren van informatie over probleemstoffen uit de gebiedsdossiers. Daarnaast worden ook de naam van de winning, het jaar van oprichting, de bron van het drinkwater en de classificatie van kwetsbaarheid uit de gebiedsdossiers gehaald. Deze gegevens zijn noodzakelijk om de resultaten vanuit deze workflow te kunnen vergelijken met de analyse van Van Driezum et al.[4]. Op verzoek van het ministerie van IenW wordt de workflow hieronder in detail toegelicht. De gehele code is terug te vinden op een Github pagina[8].

Workflow

1. In de eerste stap van de workflow worden de bestanden (pdf, docx) ingelezen. Uit elk document wordt eerst de inhoudsopgave uitgelezen. Vervolgens wordt het document per paragraaf opgesplitst. In deze stap worden tabellen door het taalmodel omgezet naar 'markdown', in deze vorm zijn tabelranden vervangen door streepjes (- & |), het taalmodel kan hierdoor de structuur van de tabel uitlezen. Deze markdown tabellen worden toegevoegd aan de opgesplitste paragrafen. Op deze manier komen ook tabellen als tekst in de paragrafen, wat leesbaar is voor het taalmodel. Iedere paragraaf, inclusief tekst en tabel, wordt daarna omgezet naar een numerieke vector die de semantische betekenis van de betreffende paragraaf vastlegt. Het resultaat is dat per gebiedsdossier een geïndexeerd bestand met numerieke vectoren ontstaat.
2. In de tweede stap van de workflow worden de bestanden gebruikt om specifieke vragen te beantwoorden. Allereerst wordt het bestand (uit stap 1) van een gebiedsdossier geopend. Vervolgens worden de vraag, de randvoorwaarden en de trefwoorden ingeladen en omgezet naar een numerieke vector die de semantische betekenis representeert. Met behulp van cosinusgelijkenis[9] en de numerieke vectoren kan er worden berekend welke paragraaf het meest overeenkomt met de vraag, randvoorwaarde en trefwoorden. De randvoorwaarden geven aan waar het antwoord mogelijk te vinden is en beschrijven de vraag in de vorm van een vaste regel, zonder vraagvorm. Trefwoorden zorgen ervoor dat de vergelijking met cosinusgelijkenis nauwkeuriger wordt omdat de trefwoorden de mogelijke locatie

[b] Het gebruik van deze modellen is binnen het RIVM toegestaan, maar er gelden wel een aantal voorwaarden en regels. De toepassing moet bijvoorbeeld voldoen aan wet- en regelgeving en een duidelijk doel hebben. Ook is bij sommige toepassingen een risicoanalyse nodig. Omdat in deze verkenning geen gevoelige data of persoonsgegevens in aanraking komen met taalmodellen is dit niet nodig. Alle gebiedsdossiers zijn namelijk openbaar beschikbaar.

van het antwoord in het document omschrijft De vijf paragrafen die het meest overeenkomen worden, samen met de vraag, randvoorwaarden en een set standaardregels, voorgelegd aan een taalmodel (GPT-4.1). Tot de standaardregels behoren o.a.: laat toevoegingen achterwege, geef prioriteit aan tabellen en noem geen verzameltermen. Welke variabelen, randvoorwaarden en trefwoorden in de workflow worden ingeladen staat beschreven in Bijlage 1. Het resultaat is een antwoord van het taalmodel dat gebaseerd is op de randvoorwaarden en de vooraf gedefinieerde regels. De pseudocode, oftewel de beschrijving van de python-code in begrijpelijke, leesbare taal, van de workflow wordt beschreven in Bijlage 2. De antwoorden worden op dezelfde wijze als de basistabel opgeslagen in een Excel bestand.

3. In de derde stap van de workflow wordt aandacht besteed aan mogelijke inconsistenties in de antwoorden van het taalmodel. Deze inconsistenties komen voort uit de variabiliteit van een taalmodel zoals GPT-4.1. Dit speelt met name bij de probleemstofvariabelen. Om de betrouwbaarheid van de resultaten te verhogen, wordt de vraag: "Welke stoffen worden boven de signaleringswaarde gemeten in [meetlocatie]?" vijf keer aan het taalmodel gesteld. Alleen de stoffen die in minimaal drie van de vijf antwoorden genoemd worden, zijn geselecteerd. Dit onderdeel wordt alleen uitgevoerd voor de probleemstoffen om variabiliteit in de antwoorden te verminderen.

Uitleg van de vraagstelling aan de hand van een voorbeeld
Als voorbeeld wordt in de volgende paragraaf de variabele 'huidige probleemstof' toegelicht. Hieronder staan voor deze variabele de vraagstelling, randvoorwaarde en trefwoorden beschreven die samen met 5 paragrafen worden voorgelegd aan het taalmodel.

Vraagstelling
Geef uitsluitend de individuele stoffen (geen stofgroepen/verzamelnamen) die expliciet gemeten zijn in de winning/winputten/winveld/pompput/in het gemengd ruwwater/innamepunt van de drinkwaterwinning en expliciet worden vermeld als overschrijding van de norm of signaleringswaarde (0,1 µg/L).

Randvoorwaarde
Geef alleen de stoffen die in het gemengd ruwwater, of in de winning, of in een pompput, of in het winveld, of in winputten, of in het innamepunt boven de norm of signaleringswaarde gemeten zijn. Geef dus geen stoffen die gemeten zijn in waarnemingsputten, waarnemingsfilters, het meetnet, peilbuizen of in infiltratiewater. Laat toevoegingen tussen haakjes weg. Zoek bij het beantwoorden van de vraag altijd eerst naar tabellen waarin stoffen die de norm of signaleringswaarde overschrijden bij het innamepunt, in de winning, in winputten, in het winveld, bij pompputten of in het gemengd ruwwater, expliciet zijn weergegeven. Neem elke stof die in zo'n Tabel als overschrijding wordt vermeld of een concentratie hoger dan 0,1 µg/L heeft op in je antwoord. Geef de stoffen die voldoen aan de eisen als opsomming gescheiden met een | (verticale streep)*. Noem alleen stoffen die met hun specifieke chemische naam in het document zijn benoemd. Noem <u>geen</u> groepen, klassen of verzameltermen zoals "overige verontreinigende stoffen" of "medicijnresten." Voeg geen meetresultaat toe. Als er geen stoffen voldoen aan deze eisen geef je een lege string als resultaat.*

Trefwoorden
Beschrijving van aangetroffen chemische stoffen in meetpunten van de drinkwaterwinning, zoals tabellen met meetconcentraties of opsommingen van chemische stoffen. Microverontreinigingen, overschrijdingen, signaleringswaarde, norm.

Validatie van de methode
De toetsing of met deze workflow de informatie voor deze gebiedsdossiers in de basistabel correct worden gereproduceerd, is handmatig uitgevoerd. Voor deze eerste toets zijn 11 dossiers geselecteerd van 6 verschillende organisaties/auteurs, met als doel een beeld te geven van de nauwkeurigheid. In Tabel 1 worden de probleemstoffen die het taalmodel gevonden heeft, vergeleken met de basistabel uit van Driezum et al. (2020) [4]. Op basis van deze dossiers en de resultaten is de methode verder aangescherpt. De uiteindelijke methode is dus geoptimaliseerd op deze 11 dossiers, omdat deze een goede afspiegeling zijn van alle gebiedsdossiers verwachten we dat de resultaten een goed beeld geven van de mogelijkheden die een taalmodel biedt.

Tabel 1 Vergelijking van resultaten uit het taalmodel en handmatig verwerkte resultaten direct uit de gebiedsdossiers. De werking van het model is getoetst door Precision / Recall (P / R) indicatoren te berekenen. Precision geeft aan welk deel van de door het model gevonden stoffen ook daadwerkelijk werden genoemd in het betreffende dossier. Recall geeft aan welk deel van de stoffen in het dossier door het model zijn geïdentificeerd. Precision [c] = True Positives / (True Positives + False Positives) en Recall [d] = True Positives / (True Positives + False Negatives)

Dossier	**Huidige probleemstof [P / R]**	**Potentiële probleemstof [P / R]**	**Huidige probleemstof in waarnemingsput [P / R]**
Haarlo Olden Eibergen (Royal HaskoningDHV)	1,00 / 1,00	0,80 / 0,92	1,00 / 0,28
Hanik (Royal HaskoningDHV)	1,00 / 1,00	1,00 / 1,00	1,00 / 1,00
Amsterdamse Waterleidings-Duinen (Grontmij)	1,00 / 1,00	1,00 /1,00	1,00 / 1,00
Eschede-Losser (Witteveen+Bos)	1,00 / 1,00	0,00 / -	1,00 / 1,00
Haamstede (Provincie Zeeland)	1,00 / 1,00	1,00 / 1,00	1,00 / 1,00
Sellingen (Royal HaskoningDHV)	1,00 / 1,00	1,00 / 1,00	1,00 / 1,00
Nietap (Royal HaskoningDHV)	1,00 / 1,00	1,00 / 1,00	1,00 / 1,00
Meijendel en Berkheide (Royal HaskoningDHV)	0,94 / 0,97	1,00 / 0,85	1,00 / 1,00
Vessem (Omgevingsdienst Zuidoost-Brabant)	0,00 / -	1,00 / 0,60	1,00 / 1,00

[c] Voorbeeld precision: stel het model vindt 100 stoffen en daarvan zijn er 60 genoemd in het dossier. Dan is de precision 60/100 = 0,6.
[d] Voorbeeld recall: stel er worden 100 stoffen genoemd in het dossier en het model vindt er 80. Dan is de recall: 80 / 100 = 0,8. Voor stoffen die niet in het gebiedsdossier worden genoemd, kan geen recall worden berekend. In Tabel 1 is dit aangegeven met een streepje (-).

Dossier	Huidige probleemstof [P / R]	Potentiële probleemstof [P / R]	Huidige probleemstof in waarnemingsput [P / R]
Biesbosch (Arcadis)	0,88 / 0,96	1,00 / 1,00	1,00 / 1,00
Noardburgum (Royal HaskoningDHV)	1,00 / 1,00	1,00 / 1,00	1,00 / 1,00

Uit Tabel 1 blijkt dat het taalmodel de stoffen uit de basistabel in de meeste gevallen goed kan reproduceren, er zijn echter wel een paar dossiers waar het niet goed gaat. Per dossier zijn de indicatoren precisie en recall berekend. Precision geeft aan hoeveel stoffen die het taalmodel noemt, ook voorkomen in de basistabel. Met recall wordt aangegeven hoeveel van de stoffen die het model noemt daadwerkelijk terug te vinden zijn in de dossiers. Een extra stof die niet in het dossier staat is hier ook fout.
Bij Meijendel en Berkheide mist het taalmodel een aantal stoffen en geeft het een aantal stoffen op die niet correct zijn, het model kan lastig onderscheid maken tussen potentiële en huidige probleemstoffen. In het gebiedsdossier van Vessem worden concentraties gerapporteerd van een paar stoffen, maar deze komen nooit boven de signaleringswaarde uit (0,1 µg/L). Het model kan dit slecht uit de tabel opmaken waardoor deze stoffen regelmatig worden geclassificeerd als 'huidige probleemstof', terwijl ze eigenlijk een 'potentiële probleemstof' zijn. In dit specifieke geval bevatte de tabel uit het gebiedsdossier synoniemen voor 'huidige probleemstof' & 'potentiële probleemstof'. Bij Haarlo Olden Eibergen worden veel stoffen in de tekst benoemd. Voor organische microverontreinigingen in waarnemingsputten wordt bijvoorbeeld een opsomming van stoffen gegeven die 'verhoogd' zijn aangetroffen. Een paar zinnen later wordt er beschreven dat bovengenoemde stoffen (de opsomming) worden aangetroffen boven de signaleringswaarde. Ondanks deze duidelijke zin en formulering heeft het taalmodel hier moeite mee en geeft vaak alleen het eerste deel van de opsomming.

Conclusies
Uit deze resultaten volgen een paar conclusies. Het model heeft moeite met onduidelijke tabellen. Als de tabel bijvoorbeeld als afbeelding met lage resolutie in het document is geplakt, kunnen lijnen wegvallen waardoor het model niet ziet in welke rij of kolom de waarde staat. Ook kan het model hierdoor stofnamen niet goed uitlezen waardoor spelfouten ontstaan. Verder heeft het model moeite met synoniemen van stoffen, het verschil tussen huidige en potentiële probleemstof en de wisselende structuur van gebiedsdossiers.

Om de workflow verder te toetsen is Tabel 5.1 en 5.2 uit Van Driezum et al.[4] gereproduceerd. Deze tabellen geven een top 10 van opkomende stoffen (potentiële en huidige probleemstoffen) voor zowel oppervlaktewater, grondwater als oevergrondwater en zijn gebaseerd op de informatie uit de basistabel. Voor deze toets zijn alle beschikbare dossiers (n=192) gebruikt. Het taalmodel deed er 3,5 uur over om alle variabelen (probleemstoffen, naam winning, type winning, jaar van oprichting en classificatie van kwetsbaarheid) te beantwoorden en op te slaan.

Het resultaat van deze toets staat in Tabel 2. Deze tabel laat zien dat het model vaak dezelfde stoffen in de gebiedsdossiers vindt als Van Driezum et al.[4]. De methode kan ook

de rangorde tot op zekere hoogte reproduceren. Deze methode kan een signalering en/of prioritering geven van opkomende stoffen op een landelijke schaal. Wel zijn er onzekerheden die het model minder nauwkeurig maken. Uit de analyse van de resultaten blijkt dat het model synoniemen van dezelfde stof als aparte stoffen ziet. Ook komen er spelfouten voor in de resultaten van het taalmodel, deze komen voornamelijk vanuit slecht leesbare tabellen in de gebiedsdossiers. Deze onnauwkeurigheid bleek ook al uit bovenstaande validatie met 11 dossiers en zorgt ervoor dat sommige stoffen nog minder zichtbaar worden als op landelijke schaal gekeken wordt zoals in de gereproduceerde tabellen 5.1 & 5.2.

Een andere reden waarom het model soms minder stoffen vindt is het wel of niet opsplitsen van dossiers. In Van Driezum et al.[4] zijn sommige gebiedsdossiers opgesplitst omdat het dossier meerdere winningen beschrijft (bijv. Bergen op Zoom – Diep & Bergen op Zoom – Middeldiep), hierdoor noemt Van Driezum et al.[4] 216 dossiers. In deze verkenning heeft het model deze splitsing niet gemaakt, omdat hiervoor de workflow aangepast moet worden. Hierdoor zijn er voor deze analyse 192 gebiedsdossiers bekeken. Een stof die wordt aangetroffen in Bergen op Zoom – Diep en in Bergen op Zoom- Middeldiep telt in Van Driezum et al dus twee keer, maar in deze analyse maar één keer, hierdoor kunnen resultaten lager uitvallen.

Tabel 2 Vergelijking van top 10 opkomende stoffen [a] per drinkwaterbron uit Van Driezum et al.[4] en het taalmodel.

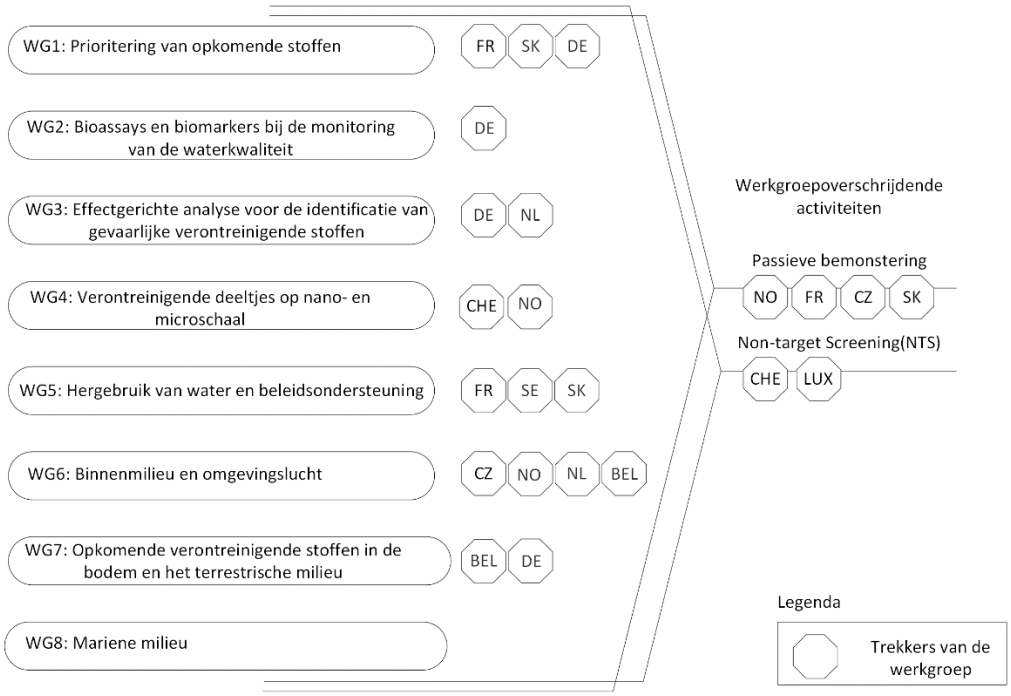
Opkomende stof	Aantal winningen waar de opkomende stof is aangetroffen					
	Oppervlakte-water [b]		Oever-grondwater		Grondwater	
	Tabel 5.1 [4]	Taal-model	Tabel 5.2 [4]	Taal-model	Tabel 5.8 [4]	Taal-model
2-Hy—4,6-bis(4sulfanilo)-1,3,5-triazine	-	-	14	8	-	-
4-Methyl-1H-Benzotriazool	6	4	-	-	-	-
Acesulfaam	8	4	7	7 of 6 [c]	12	7
Amidotrizoïnezuur	7	7	-	-	-	-
Bisfenol A	-	-	-	-	6	8
Carbamazepine	-	-	-	-	5	6
Chloorethaan	-	-	-	-	7	3
Chloormethaan	-	-	-	-	4	- [d]
Cyclamaat	6	6	-	-	6	6
Diethylftalaat	-	-	-	-	4	4
DIPB	-	-	-	-	5	4
DIPE	-	-	-	-	13	11
DTPA	-	-	8	8	-	-
Fenazon	-	-	-	-	4	5
Hydrochloorthiazide	8	7	-	-	-	-
Johexol	7	8	-	-	-	-
Jomeprol	9	10	-	-	-	-
Jopamidol	7	9	-	-	-	-
Jopromide	7	9	-	-	-	-
Metformine	9	9	-	-	-	-
Nitrobenzeensulfonaat	-	-	8	6	-	-
Pyrazol	-	-	8	6	-	-
Sucralose	8	11	7	8	-	-
Tert/butanol	-	-	-	-	5	6
Tetrahydrofuraan	-	-	-	-	36	26
Tri-glyme	-	-	13	8 [e]	-	-
Triphenylphospineoxide (TPPO)	-	-	7	4	-	-
Vigabatrine	-	-	7	7	-	-

[a] Huidige probleemstoffen en potentiële probleemstoffen zonder waarnemingsputten
[b] Voor oppervlaktewater is de top 11 weergegeven.
[c] Er is 7 keer acesulfaam en 6 keer acesulfaam-k gevonden.
[d] Er is wel 33 keer trichloormethaan gevonden.
[e] Er is ook 3 keer di- en triglyme gevonden.

Kans 2: Aansluiten bij internationale projecten en netwerken voor de signalering en prioritering van opkomende probleemstoffen.

Actieve deelname aan werkgroep prioritering van het Norman Netwerk
Het NORMAN-netwerk is een netwerk van referentielaboratoria, onderzoekscentra en aanverwante organisaties die zich bezighouden met het monitoren van opkomende milieustoffen. Het netwerk bevordert de uitwisseling van informatie over nieuwe stoffen in het milieu en stimuleert de validatie en harmonisatie van meetmethoden en monitoringinstrumenten en ondersteunt hiermee de risicobeoordeling en aanpak van opkomende stoffen. Het NORMAN-netwerk heeft acht werkgroepen (WG) en twee werkgroep-overschrijdende activiteiten die zich bezighouden met verschillende kwesties in verband met opkomende stoffen (Figuur 1).

Figuur 1 De acht werkgroepen en twee werkgroep overschrijdende activiteiten in het NORMAN-netwerk. In beide gevallen waar NL als trekker genoemd is, wordt de taak door de Vrije Universiteit Amsterdam ingevuld.



Het RIVM is via Bureau REACH lid van het NORMAN-netwerk en neemt deel aan verschillende werkgroepen. Uit een interne inventarisatie in november 2025 bleek dat het RIVM in ieder geval als agendalid betrokken is bij werkgroepen 1, 2, 4, 5 en 7 (zie Figuur 1 voor een beschrijving van de werkgroepen). Voor de werkgroep over nano- en microplastics vindt deelname plaats vanuit een lopend project. Voor de andere werkgroepen is echter geen budget gereserveerd. In het kader van deze opdracht is via het deelnemen aan de jaarlijkse algemene ledenvergadering[e] van 2025 en aan vergaderingen van WG1 verkend of actieve deelname aan het NORMAN netwerk specifieke kansen biedt voor het signaleren en prioriteren van opkomende stoffen.

[e] Tijdens de algemene ledenvergadering op 2 en 3 december 2025, is vastgesteld dat deelname aan WG 5 'Hergebruik van water en beleidsondersteuning' ook relevant kan zijn voor het RIVM drinkwaterteam, in het kader van het kennisbasis-onderwerp 'hergebruik van afvalwater'.

WG1 streeft er naar om vast te stellen welke stoffen de grootste bedreiging vormen voor ecosystemen en de menselijke gezondheid. Hiervoor heeft de werkgroep in 2013 een prioriteringsmethodiek opgeleverd in de vorm van een score die is gebaseerd op de volgende eigenschappen van een stof: P (persistentie), B (bioaccumulatie) en T (toxiciteit), samen met CMR (carcinogeen, mutageen, reprotoxisch) en ED ('Endocrine Disruption', hormoonverstoring).[10] De resultaten van deze prioriteringsmethodiek zullen beschikbaar worden gemaakt via een prioriteringstool op de NORMAN website.

Het RIVM heeft ook een prioriteringsmethodiek ontwikkeld voor PBT[11] en PMT[12] stoffen. De resultaten van de toepassing van deze methodieken zijn beschikbaar via de RIVM PBT & PMT screeningstool [f]. Er zijn plannen om de resultaten van de RIVM-tools op te nemen in de prioriteringstool van NORMAN, naast de resultaten van andere screeningsmethoden, zoals de PikMe tool van NILU, het Noors onderzoeksinstituut dat zich richt op onderzoek naar luchtkwaliteit, luchtvervuiling, klimaatverandering en aanverwante milieuproblemen.[13]

Dit is een belangrijke ontwikkeling, omdat het betekent dat de RIVM PBT & PMT screeningtool internationaal wordt geaccepteerd en gebruikt. Hiermee bevorderen we internationale afstemming en kunnen we voorkomen dat we in Nederland stoffen signaleren en aanpakken die elders worden doorgelaten of omgekeerd. Het RIVM pleit er daarom voor om bij de verdere ontwikkeling van de PBT & PMT screeningtool meer aansluiting te zoeken bij de internationale gemeenschap. Daarvoor is actieve deelname van het RIVM aan het NORMAN netwerk nodig. Dat biedt meer kansen dan enkel voor de PBT & PMT screeningtool, want de landelijke werkgroep Aanpak Opkomende Stoffen maakt voor het prioriteren van opkomende stoffen in oppervlaktewater ook al gebruik van de NORMAN systematiek. Het NORMAN netwerk is ook actief op andere relevante beleidsdossiers en is bovendien gesprekspartner van de Europese commissie, bijvoorbeeld rond de KRW. Actieve deelname geeft dus de mogelijkheid om de Nederlandse belangen internationaal in te brengen.

Zweeds project 'New praxis for assessing and managing risks of undesirable chemical substances in water'
Het RIVM heeft als een van de sprekers deelgenomen aan een workshop die is georganiseerd in het kader van het Zweedse project 'New praxis for assessing and managing risks of undesirable chemical substances in water'. Dit is onderdeel van het WaterWise Societies initiatief in Zweden. Het RIVM heeft een presentatie gegeven over onder andere de resultaten van het SPR Project PS-DRINK en de PBT & PMT screeningtool[14]. Helaas was de rest van de sessie in het Zweeds en konden de discussie en mogelijke signalen voor nieuwe stoffen niet worden opgepikt. Het definitieve rapport van het project is wel beschikbaar in het Engels.[15] Verder zal het RIVM ook deelnemen in de klankbordgroep van het vervolg op dit project. Zo kan het RIVM de ontwikkelingen in dat project volgen en mogelijke relevante opkomende stoffen zo vroeg mogelijk oppikken.

PARC Early Warning System Stakeholder Interaction Forum
PARC (Partnership for the Assessment of Risks from Chemicals) is een groot Europees samenwerkingsproject dat in 2022 is gestart en nog loopt tot 2029. Het doel van PARC is om kennis en onderzoek op het gebied van chemische risicobeoordeling te bundelen en te verbeteren. Het project wordt gefinancierd door de Europese Unie en betrokken landen en organisaties, waaronder het RIVM.

[f] https://rvszoeksysteem.rivm.nl/ScreeningTool

Werkpakket 8 binnen PARC richt zich op het ontwikkelen van toolboxen. In dit werkpakket wordt naast een toolbox voor Safe and Sustainable by Design (SSbD) ook een voorzet gedaan voor de wetenschappelijke en technische basis voor een Europees Early Warning System (EWS) voor chemische risico's [16,17]. Het werk dat in PARC gebeurt, is voorwerk voor het toekomstige 'Early warning and action system for emerging chemical risks', onderdeel van het grotere pakket 'One Substance, One Assessment'[18].

Het PARC EWS is bedoeld om op een systematische en gecoördineerde manier chemische dreigingen te detecteren en aan te pakken, nog voordat deze wijdverspreid worden. Het EWS is niet gericht op een bepaald milieucompartiment of specifieke blootstellingsroute, maar maakt gebruik van big data, gegevens uit verschillende bronnen, AI om automatisch en dynamisch zo breed mogelijk signalen van opkomende chemische risico´s te detecteren.

Het systeem werkt in drie stappen:
1) Signaleren en filteren van mogelijke risico's via diverse bronnen.
2) Controleren op bestaande wetgeving en maatregelen.
3) Vergaren en versterken van informatie rondom het signaal.

Om het systeem te testen worden praktijkproeven uitgevoerd, zoals chemische analyses van zwevend stof uit Duitse rivieren en vismonsters uit een Griekse rivier. Deze studies helpen bij het verfijnen van de methoden en de interpretatie van de data.

In het kader van deze opdracht heeft het RIVM (digitaal) deelgenomen aan de eerste Stakeholder Interaction Forum van het PARC Early Warning System. Het RIVM heeft ook het Stakeholder Interaction Forum op 13 maart bijgewoond. Net als voor de voorgaande activiteiten, is aanhaken bij dit project belangrijk om ervoor te zorgen dat het PARC EWS is afgestemd met/op onze nationale activiteiten. Ook zorgt dit ervoor dat drinkwaterbelang voldoende wordt meegenomen en aanwijzingen voor potentiële probleemstoffen zo vroeg mogelijk worden opgepikt.

Conclusie en aanbevelingen

Deelconclusie kans 1: gebruik AI om effectiever probleemstoffen te identificeren
In dit project hebben we een innovatieve aanpak getest waarbij met AI opkomende stoffen konden worden geïdentificeerd op basis van de gebiedsdossiers. De signalering en prioritering van opkomende stoffen in Van Driezum et al.[4] kan tot op zekere hoogte gereproduceerd worden. Voor een volgende analyse van een nieuwe generatie gebiedsdossiers is het op dit moment echter nog te foutgevoelig. Een verdere verfijning van de workflow kan veel van deze fouten verhelpen. Denk aan een betere methode om tabellen uit te lezen, het gebruik van stof-synoniemen lijsten en nieuwe taalmodellen. Daarnaast kan de methode ook in andere projecten worden ingezet door de workflow aan te passen om als een zoekmachine of AI-agent te dienen, eventueel voor een bredere selectie van documenten.

Uit deze verkenning zijn een aantal belangrijke aanbevelingen voor gebiedsdossiers naar voren gekomen. Zoals het toevoegen van een gestandaardiseerde tabel met probleemstoffen, meetlocaties, concentraties en CAS-nummers. Ook het schrijven van het CAS nummer in de tekst, achter een stofnaam, maakt het verwerken van de informatie efficiënter. Ook met AI-tools blijft het een grote inspanning om de relevante data uit de gebiedsdossiers te 'extraheren'. Het verzamelen van de onderliggende data in een voor het RIVM toegankelijke database zou de beste oplossing zijn om op een snelle manier overzicht te krijgen van (potentiële) probleemstoffen.

Deelconclusie kans 2: Aansluiten bij internationale projecten en netwerken voor de signalering en prioritering van opkomende probleemstoffen.
In deze verkenning is onderzocht wat de aansluiting bij toonaangevende internationale projecten en netwerken kan opleveren voor het signaleren en prioriteren van opkomende stoffen in drinkwater(bronnen). Daarbij is specifiek gekeken naar:

a. De werkgroep prioritering van het NORMAN-netwerk,
b. Het PARC Early Warning System Stakeholder Interaction Forum,
c. Het Zweedse project 'New practice for assessing and managing risks of undesirable chemical substances in water'.

Uit de verkenning blijkt dat deelname aan deze internationale netwerken bijdraagt aan het versterken van lopende nationale trajecten, het tijdig signaleren van opkomende stoffen, het borgen van het (Nederlandse) drinkwaterbelang en het op de hoogte blijven van relevante internationale ontwikkelingen. Binnen de reguliere (drink)wateropdrachten is echter weinig ruimte om actief deel te nemen aan de activiteiten van (met name WG1 van) het NORMAN-netwerk en andere internationale initiatieven op het gebied van signalering en prioritering. Daarom is het belangrijk om hiervoor structureel budget beschikbaar te stellen.

Referenties

1. Inspectie Leefomgeving en Transport (ILT). *Drinkwaterkwaliteit 2024*. https://www.ilent.nl/documenten/drinkwater/drinkwater/rapporten/drinkwaterkwaliteit-2024 (2025).
2. RIWA-Maas. *Jaarrapport 2024 De Maas*. https://www.riwa-maas.org/categories/jaarrapporten/ (2025).
3. RIWA-Rijn. *Jaarrapport 2024 De Rijn*. https://www.riwa-rijn.org/wp-content/uploads/2025/09/RIWA-2025-NL-Jaarrapport-2024-De-Rijn.pdf (2025).
4. van Driezum, I. *et al. Staat Drinkwaterbronnen*. https://rivm.openrepository.com/handle/10029/624657 (2020) doi:10.21945/RIVM-2020-0179.
5. IenW. *Protocol Gebiedsdossiers En Uitvoeringsprogramma's Drinkwaterwinningen*. https://tinyurl.com/25ms545c (2023).
6. Hartmann, J., Affourtit, F., Zijlstra, R. H., Janson, T. J. E. & Hofman-Caris, C. H. M. *Risicobeoordeling van Opkomende Stoffen in Oppervlaktewater Voor Drinkwater (2017-2020). Probleemstoffen Op Basis van KRW-Protocol*. (In voorbereiding).
7. EC. *Drinkwaterrichtlijn*. *Richtlijn 2020/2184* (2020).
8. Janson, T. J. E. RAG Gebiedsdossiers. (2026).
9. Krantz, T. & Jonker, A. What is cosine similarity.
10. Dulio, V. *et al.* NORMAN Prioritisation framework for emerging substances-Passive sampling interlaboratory study-data evaluation. (2013).
11. Rorije, E., Verbruggen, E., Hollander, A., Traas, T. & Janssen, M. *Identifying Potential POP and PBT Substances : Development of a New Persistence/Bioaccumulation-Score*. https://www.rivm.nl/bibliotheek/rapporten/601356001.html (2011).
12. Hartmann, J., Rorije, E., Wassenaar, P. N. & Verbruggen, E. Screening and prioritising persistent, mobile and toxic chemicals: development and application of a robust scoring system. *Environmental Sciences Europe* **35**, 40 (2023).
13. Wennberg, A. C., Rostkowski, P. & Reid, M. PikMe: a flexible prioritization tool for chemicals of emerging concern. *Environ Toxicol Chem* **44**, 3450–3459 (2025).
14. RIVM. PBT & PMT screening tool. https://rvszoeksysteem.rivm.nl/ScreeningTool (2023).

15. Sokolova, E. *et al. New Praxis for Assessment and Management of Risks with Unwanted Chemical Substances in Water*. https://waterwisesocieties.se/wp-content/uploads/sites/9/2025/09/Slutrapport_New-praxis-for-assessment-and-management-of-risks-with-unwanted-chemicals-in-water.pdf (2025).
16. PARC. PARC Early Warning System. https://tinyurl.com/3z3a38fw (2025).
17. PARC. The P-A-R-C SSbD Toolbox. https://www.parc-ssbd.eu/ (2025).
18. EC. Council and Parliament agree on simplified and more efficient handling of chemical assessment data. *Official website of the Council of the EU and the European Council* https://www.consilium.europa.eu/en/press/press-releases/2025/06/12/council-and-parliament-agree-on-simplified-and-more-efficient-handling-of-chemical-assessment-data/ (2025).

Bijlage 1: Variabelen, randvoorwaarden en trefwoorden die in de workflow worden ingeladen

https://www.rivm.nl/bibliotheek/rapporten/KN-2026-0087-bijlage.xlsx

Bijlage 2: Pseudocode

De workflow beschreven in Pseudocode. Deze Pseudocode beschrijft de gebruikte Python code in toegankelijke taal.

1. Initialisatie en Setup
Laad benodigde libraries
Stel API-keys, endpoints en clients in voor LLM en embeddings
Definieer bestandslocaties voor input, output en vector storage
Laad configuratiebestanden en prompts (Excel)

2. Document Parsing en Preprocessing
Voor elk document in de data directory:
Open PDF bestand
Probeer tabellen te extraheren:
- Zoek naar paginanummers met tabellen
- Maak afbeeldingen van deze pagina's
- Stuur afbeeldingen naar LLM voor reconstructie van tabellen in markdown
- Sla de tabellen op per pagina

Probeer inhoudsopgave (TOC) te reconstrueren:
- Zoek in de eerste pagina’s naar tekstpatronen voor TOC
- Als geen TOC gevonden, probeer outline uit PDF metadata te halen
- Bepaal start- en eindpagina’s van hoofdstukken

Voor elk hoofdstuk/paragraaf uit TOC:
- Extract tekstfragment (chunk) uit de relevante pagina’s
- Voeg indien aanwezig de gereconstrueerde tabellen toe aan de tekstfragmenten
- Sla het fragment op met metadata (titel, paginanummers, etc.)

3. Vectoriseren van Fragmenten
Voor elk tekstfragment:
- Genereer embeddings voor zowel de titel als de inhoud (met gekozen embedding model)
- Sla deze vectoren op samen met de metadata in een JSON-bestand

4. Vraag-Antwoord Workflow per Dossier
Voor elk gevectoriseerd dossier:
Laad de JSON met tekstfragmenten en embeddings

Voor elke prompt/vraag uit de Excel:
- Genereer een query embedding van de vraag + tags
- Bereken cosine similarity tussen query embedding en alle content embeddings
- Selecteer de top N meest relevante tekstfragmenten

Als 'multiple_runs' geactiveerd:
- Herhaal LLM-vraag N keer met de geselecteerde fragmenten
- Verzamel alle antwoorden
- Bepaal het meest voorkomende antwoord via majority voting
- Filter ongewenste/foute antwoorden

Anders:
- Stel de vraag 1 keer aan LLM met relevante fragmenten

Sla het resultaat op per dossier en per variabele

5. Resultaten Scoren en Opslaan
Voor elk dossier en variabele:
- Vergelijk het LLM-antwoord met een correct bestand
- Bereken scores (percentage juiste/foute antwoorden)
- Sla scores, resultaten en token usage op in DataFrames/Excel-bestanden

6. Rapportages en Tabellen Reconstrueren
Combineer de gescoorde variabelen om tabellen te reconstrueren:
- Maak lijsten van probleemstoffen per type winning
- Groepeer en tel frequenties voor samenvattende tabellen
- Sla deze tabellen op voor rapportage

7. (Optioneel) Interactieve Conversatie
Bied een functie aan om handmatig te converseren met het LLM via de gegenereerde fragmenten