

Monitor pluriformiteit

Modellen om de diversiteit van
het nieuwsaanbod te monitoren



Voorwoord

Het belang van pluriform nieuwsaanbod

Beste lezer,

De pluriformiteit van het nieuwsaanbod is een belangrijk speerpunt voor het Commissariaat voor de Media. Niet voor niets luidt onze missie: 'Het Commissariaat bevordert betrouwbaar en pluriform media-aanbod voor iedereen'. Alleen bij een divers aanbod kunnen burgers zich een eigen mening vormen. Dat is essentieel voor onze democratie.

Het is goed om de pluriformiteit van het nieuwsaanbod te meten, zodat je de ontwikkeling ervan in de gaten kunt houden en bij kunt sturen daar waar nodig. Maar hoe meet je de pluriformiteit van het nieuwsaanbod? Dat is geen eenvoudige vraag om te beantwoorden. Daarom heeft het Commissariaat de afgelopen periode – in samenwerking met diverse betrokken partijen – drie pilots ontwikkeld en getest. De bevindingen presenteren we aan u in deze rapportage 'Monitor pluriformiteit'. Deze monitoring wordt onderdeel van de bekende, jaarlijkse Mediamonitor van het Commissariaat.

De wettelijke taak van de Mediamonitor schrijft dat het Commissariaat is belast met het onderzoek naar ontwikkelingen ten aanzien van concentraties en de gevolgen daarvan voor de pluriformiteit en onafhankelijkheid van de informatievoorziening.² Met de ontwikkeling van de Monitor pluriformiteit geven we naast de bestaande invulling van de Mediamonitor hier verder vorm aan.

In de meerjarenstrategie van het Commissariaat voor de komende vijf jaar (2025 – 2030) is één van de belangrijkste thema's: een robuuste nieuwsvoorziening. We willen de diversiteit aan geluiden in het medialandschap bevorderen, en waken voor opiniemonopolies. Zo zetten we ons in voor het zichtbaar en vindbaar maken van betrouwbaar nieuws op alle distributiekanaalen, ook digitaal. Dit doen wij zodat

Nederlanders kennis blijven nemen van pluriforme en onafhankelijke informatie en kwaliteitsjournalistiek.

Het Commissariaat ontwikkelde de pluriformiteitsmonitor op verzoek van het ministerie van OCW. Waarom deden we dat? Daarvoor gaan we terug naar de aanleiding. Eind 2021 werd de motie Kwint ingediend: Kamerlid Kwint verzocht de regering in kaart te brengen op welke manier de pluriformiteit van het media-aanbod beter beschermd kan worden. Die vraag kwam voort uit de steeds grotere mediaconcentratie: de zorg dat steeds meer nieuwsmedia in handen komen van een steeds beperkter aantal nieuwsbedrijven.

In december 2022 traptten we dit project af met een bijeenkomst in Amsterdam. Met in de zaal o.a. wetenschappers, onderzoekers, beleidsmakers van het ministerie van OCW, diverse nieuwsorganisaties en journalistieke organisaties. Toen hebben we gezamenlijk de contouren voor de ontwikkeling van een pluriformiteitsmonitor vastgesteld.

Eind januari van dit jaar zagen we dezelfde partijen weer terug in de OBA-bibliotheek in Amsterdam, bij de presentatie van de ontwikkelde Monitor pluriformiteit.

Graag bedank ik de wetenschappers, die voor ons de drie pilots hebben uitgevoerd: Mark Boukes, Nel Ruigrok, Elif Kılık en Knut De Swert. Ook wil ik de begeleidingscommissie heel hartelijk bedanken.

Ik wens u veel leesplezier.

Amma Asante,
voorzitter Commissariaat voor de Media

² Mediawet 2008, art 7.21

Inhoud

Voorwoord	2
Introductie	4
Samenvatting en advies	5
• Pilot 1: Ervaringen van pluriformiteit onder journalisten	5
• Pilot 2: Unieke en overeenkomstige berichtgeving	5
• Pilot 3: Diversiteit van thema's, bronnen en standpunten	6
• Inbedding van de Monitor pluriformiteit in de Mediamonitor	6
1 Aanleiding	8
1.1 Relevantie pluriform nieuwsecosysteem	9
1.2 Uitbreiding opzet Mediamonitor	10
1.3 Keuze voor drie pilots	11
2 Pilot 1: Ervaringen van pluriformiteit onder journalisten	12
2.1 Ontwikkeling van de vragenlijst	13
2.2 Bevindingen pilot 1	15
3 Pilot 2: Unieke en overeenkomstige berichtgeving	16
3.1 Selectie van nieuwsartikelen	16
3.2 Methode	17
3.3 Monitoring van diversiteit van nieuwsartikelen	19
4 Pilot 3: Diversiteit van thema's, bronnen en standpunten	20
4.1 Drie indicatoren voor diversiteit nieuwsinhoud	21
4.2 Methode	22
4.3 Monitoring van diversiteit van nieuwsinhoud	23
5 Reflectie op het proces	24
6 Reflectie op Monitor pluriformiteit – NDP Nieuwsmedia	26
7 Reflectie op Monitor pluriformiteit – WRR	28
8 Reflectie op Monitor pluriformiteit – NVJ	30
Dankwoord	33
Bijlage 1	34
Pilot 3, Diversiteit van thema's, bronnen en standpunten: coderingsproces en resultaten van de betrouwbaarheidsanalyse	34

Introductie

Wanneer we spreken over de pluriformiteit van het nieuwsaanbod wordt er vaak gekeken naar één aspect van het begrip. Marktconcentratie zou een negatieve werking hebben op de pluriformiteit van het aanbod. Of, zo redeneren anderen, concentratie zorgt er juist voor dat mediatitels behouden worden en draagt zo bij aan de pluriformiteit. Bij allebei de argumenten ligt de focus echter bij de diversiteit van het aanbod, terwijl pluriformiteit uit meer dimensies bestaat. In deze rapportage kijken we breder naar het begrip pluriformiteit.

Naar aanleiding van een workshop in december 2022 werkte het Commissariaat samen met wetenschappers aan drie pilotstudies om pluriformiteit van het nieuwsaanbod te onderzoeken.

De voortgang en inhoud van de studies bespraken wij met een begeleidingscommissie. In deze begeleidingscommissie waren de nieuwssector, beleidsmakers en wetenschappers vertegenwoordigd. In januari 2025 hield het Commissariaat

een bijeenkomst om de methodes van de pilots te presenteren en te bediscussiëren. Bij de bijeenkomst waren opnieuw de nieuwssector, beleidsmakers en wetenschappers aanwezig.

In dit rapport starten we met de samenvatting van de inhoud van de drie pilotstudies. We beschrijven welk deel van pluriformiteit we met deze studies in kaart brengen en geven we antwoord op de vraag: “welke onderzoeksmethodes zijn geschikt voor monitoring van de pluriformiteit van het nieuwsaanbod?” Hiervoor gebruikten we de input die wij kregen van de begeleidingscommissie en tijdens de bijeenkomst in januari 2025.

Daarna volgt in dit rapport een uitgebreide beschrijving van de pilotstudies en geven relevante partijen hun reflectie op de Monitor pluriformiteit. We vroegen vier partijen om een reflectie, drie partijen reageerden en deze reflecties zijn te vinden in deze rapportages. Met deze rapportage ronden we de pilotfase af.

Samenvatting en advies

In de jaarlijkse Mediamonitor wordt diversiteit van het aanbod en diversiteit van het media-gebruik in kaart gebracht. Deze vormen van diversiteit zijn belangrijke dimensies van pluriformiteit. Maar onder pluriformiteit verstaan we ook diversiteit van de inhoud en productiediversiteit. We onderzochten naar aanleiding van de motie Kwint met drie pilotstudies de mogelijkheden om deze vormen van diversiteit meetbaar te maken.

Met de afronding van de pilotstudies kwamen we tot methodes om diversiteit van de inhoud en de productiediversiteit meetbaar te maken. We onderzochten twee methodes om de diversiteit van de inhoud te meten. Eén methode die kijkt naar uniciteit en overlap in berichtgeving en één methode die kijkt naar verscheidenheid van onderwerpen, bronnen en standpunten binnen een onderwerp. Om productiediversiteit te meten werd een vragenlijst ontwikkeld. De vragenlijst is uitgezet onder een groot aantal journalisten en daarmee is al een start gemaakt in de uitvoering van deze methode.

De pilotstudies waren primair gericht op de ontwikkeling van de methode en daar gaan we in dit rapport op in. De volgende stap is om de gevonden methodes uit te gaan voeren en te implementeren in onze jaarlijkse Mediamonitor. In dit rapport schrijven we daarom een advies over welke methodes we willen voortzetten. Dit advies is gericht aan onze opdrachtgever, het ministerie van OCW.

Pilot 1: Ervaringen van pluriformiteit onder journalisten

Deze pilot bestaat uit het opzetten van een uitgebreide vragenlijst om de pluriformiteitsbeleving van journalisten te onderzoeken. De vragen gingen zowel in op werkomstandigheden als op maatschappelijke omstandigheden.

Met dit vragenlijstonderzoek onder journalisten verwachten wij inzicht te krijgen in de ervaren vrijheid van journalisten voor het kiezen van onderwerpen en invalshoeken waarover zij berichten. Dit onderzoek is uitermate geschikt voor monitoring. Het voorstel van de onderzoeker, de begeleidingscommissie en het Commissariaat is dit onderzoek drie- of vierjaarlijks uit te voeren. Als er gebeurtenissen plaatsvinden waarvan een grote impact op de werkomstandigheden van de journalisten wordt verwacht, kan het onderzoek eerder worden herhaald. Van groot belang voor een vergelijkbaarheid blijft een representatieve steekproef die alleen te realiseren is als de toegang tot journalisten voor het uitzetten van de vragenlijst

breed – door publieke en commerciële mediabedrijven en belangengroepen – wordt gesteund. Ook moet de inhoud van de vragenlijst identiek blijven.

Pilot 1 werd afgerond met het vaststellen van de vragenlijst en is begin 2024 als project voorgezet toen de onderzoeker de vragenlijst uitstuurde naar de leden van de NVJ. Begin 2025 komt daar de ondervraging bij van commerciële mediabedrijven en de regionale publieke omroepen. De eerste inhoudelijke rapportage over de resultaten volgt aan het einde van 2025. Dit is de nulmeting. Als we uitgaan van een vierjaarlijkse cyclus volgt de 1-meting begin 2029 en kan in datzelfde jaar worden gerapporteerd over de resultaten. Het advies van het Commissariaat is om deze vragenlijst een structureel karakter te geven en de uitvoering van de vragenlijst onder te brengen bij de wettelijke taak van de Mediamonitor.

Pilot 2: Unieke en overeenkomstige berichtgeving

In deze pilot keken we naar uniciteit en overlap in teksten van nieuwsberichten. Hoe minder overlap in nieuwsberichten hoe diverser het aanbod. Om overlap te bepalen gebruiken we taalkundige modellen. De mate van overlap is onderverdeeld in vier categorieën:

- Kopie (veel of volledige overlap)
- Overlap (aanzienlijke mate van overlap)
- Zelfde gebeurtenis (artikel gaat over dezelfde gebeurtenis maar met unieke redactionele content)
- Origineel (geen of nauwelijks overlap)

De taalkundige modellen zijn goed in staat om de overlap te berekenen. Het vaststellen van waar de grenzen liggen voor de verschillende categorieën is echter niet eenvoudig.

De begeleidingscommissie, het Commissariaat en de onderzoeker zijn het erover eens dat deze pilot een goede methode oplevert en een goed begin is om te komen tot monitoring van uniciteit en overlap in berichtgeving. De analyse kan jaarlijks worden herhaald. De onderzoeksperiode zal dan ten minste één maand zijn.

De modellen die werden gebruikt voor deze pilot zijn niet absoluut. Taalkundige modellen zijn op het moment enorm in ontwikkeling en de vooruitgang van deze modellen gaat heel snel. Toekomstige modellen kunnen later zorgen voor

efficiëntie omdat ze makkelijker te automatiseren zijn en minder computerkracht vragen.

De begeleidingscommissie en het Commissariaat hebben de nadrukkelijke wens om aanbod van regionale titels mee te nemen in de analyse en om de analyse niet alleen te beperken tot geschreven artikelen, maar ook audiovisueel nieuwsaanbod mee te nemen in de analyse. Dit is mogelijk door gebruik te maken van bijvoorbeeld ondertiteling. In de toekomst is het wenselijk om de analyse te splitsen naar katernen.

Wel is een belangrijke voorwaarde voor het uitvoeren van de inhoudsanalyse dat de toegang tot nieuwsartikelen gewaarborgd is.

Het advies is om pilot 2 voort te zetten als monitor met ingang van 2025. Indien mogelijk geven we met de eerste inhoudelijke rapportage ook inzicht in de overlap van regionale nieuwsmerken. In 2026 willen we de mogelijkheden onderzoeken voor het toevoegen van audiovisueel nieuwsaanbod.

Pilot 3: Diversiteit van thema's, bronnen en standpunten

De diversiteit van standpunten en invalshoeken valt uiteen in drie elementen:

- Onderwerpdiversiteit
- Diversiteit van bronnen
- Diversiteit van standpunten

In deze pilot zijn geautomatiseerde technieken voor tekstanalyse, waaronder taalmodellen, getest om vast te stellen of ze in staat zijn om de drie bovenstaande vormen van diversiteit te meten. Deze methode is in de pilotfase toegepast op het onderwerp corona.

Met een automatische inhoudsanalyse is de onderwerpdiversiteit goed vast te stellen. Ook de brondiversiteit wordt goed vastgesteld met automatische inhoudsanalyse. De automatische inhoudsanalyse bestaat uit de toepassing van technieken voor tekstanalyse. De diversiteit van standpunten is lastiger vast te stellen. Voor mensen is het moeilijk om standpunten uit media-inhoud te halen, dat betekent ook dat het moeilijk is voor automatische tekstanalyse.

Dat standpunten niet direct kunnen worden gemeten is een tekortkoming in deze methode volgens de begeleidingscommissie. Deze methode is – mede daardoor – minder geschikt voor monitoring, maar volgens het Commissariaat wel geschikt als ad hoc inzet om onderwerpdiversiteit en diversiteit van bronnen in kaart te brengen. Daarnaast kan standpuntdiversiteit benaderd

worden met de analyse van bronnen; de functies van bronnen, hun politieke affiliatie en de onderwerpen waarover ze aan bod komen zijn indicaties voor standpuntdiversiteit.

In principe kan de methode worden toegepast op elk onderwerp. Bij elke analyse van de drie vormen van diversiteit door middel van geautomatiseerde technieken dienen de modellen steeds gevalideerd te worden met behulp van door mensen gecodeerde nieuwsartikelen.

De geautomatiseerde technieken zijn momenteel niet in staat om standpunten rechtstreeks uit nieuwsartikelen te halen, wat ook voor menselijke codeurs een moeilijke taak blijkt te zijn. De ontwikkeling van taalmodellen (LLM's) gaat echter bijzonder snel en het is denkbaar dat deze modellen straks mogelijk wel standpunten kunnen meten. Maar dan is het nog niet vanzelfsprekend om LLM's te gebruiken voor deze analyse. Zelfs als de LLM's drastisch verbeteren en in staat zijn om standpunten uit nieuwsartikelen te halen, kan het gebruik van deze modellen voor de monitor gevoelig liggen omdat nieuwsartikelen niet zomaar door uitgevers beschikbaar mogen worden gesteld aan LLM's. De modellen zouden zichzelf ermee trainen en verrijken. De wetenschappers, begeleidingscommissie en het Commissariaat delen deze zorgen.

De toegang tot nieuwsartikelen is, net als bij pilot 2, ook hier een aandachtspunt. Voor deze methode geldt ook dat de wens is om niet alleen geschreven nieuwsartikelen maar ook audiovisueel nieuwsaanbod mee te nemen. Ook moet besproken worden met de sector en met beleidsmakers welk onderwerp geschikt is voor deze methode.

Om meer ervaring op te doen met deze methode is het advies van de onderzoeker en het Commissariaat om in 2026 één of meerdere onderwerpen te kiezen waar de mediabedrijven welwillend tegenover staan. De uitvoering van het project kan plaatsvinden in 2026 en de inhoudelijke rapportage in 2027. Wanneer er geen consensus wordt gevonden met de mediabedrijven over een geschikt onderwerp of wanneer we onvoldoende voorzorgsmaatregelen kunnen treffen om LLM's op een veilige manier te gebruiken is medewerking vanuit de mediabedrijven niet mogelijk en kan dit project niet worden uitgevoerd.

Inbedding van de Monitor pluriformiteit in de Mediamonitor

Daar waar we in deze rapportage ingaan op de methodes om pluriformiteit in kaart te brengen, gaan we in de projectfase samen met de begeleidingscommissie kijken naar de inhoudelijke resultaten van de meetmethodes. De resultaten daarvan krijgen een plek in de Mediamonitor van het Commissariaat.

De Mediamonitor is een rapport waarin uitgebreid aandacht is voor pluriformiteit. Tot op heden ging dat om pluriformiteit als diversiteit van het aanbod en de diversiteit van het gebruik. Met de Monitor pluriformiteit ontstaat ook inzicht in productiediversiteit en diversiteit van de inhoud.

Juist omdat pluriformiteit uit meerdere dimensies bestaat laat het begrip pluriformiteit zich niet kwantificeren en uitdrukken in één maat. Er bestaat niet zoiets als een ideale waarde die uitdrukt dat het goed of niet goed gesteld is met pluriformiteit. Juist door monitoring krijgen we inzicht in de verschillende diversiteitsdimensies; welke kant bewegen ze op en wat zijn verklaringen voor ontwikkelingen. Monitoring van de productiediversiteit en diversiteit van de inhoud kan in de toekomst naast de al bestaande monitoring van aanbiedersdiversiteit, diversiteit van (nieuws)merken en diversiteit van het gebruik, breed inzicht geven in welke bewegingen we vinden. Veranderingen in de diversiteitsdimensies moeten in de context van ontwikkelingen in de nieuwssector en in de samenleving worden gelezen.

Daarnaast moeten we aanmerken dat we niet compleet zijn met de opzet van onze meetmethodes. Het aanbod dat we meenemen in onze studies beperkt zich nu tot het aanbod van de voornaamste landelijke nieuwsmerken die gericht zijn op Nederland. Ook op sociale media is veel nieuwsaanbod te vinden, soms van journalisten afkomstig, soms van anderen die nieuws doorplaatsen. We moeten ons dus ook bewust blijven van de beperkingen en bezien hoe we daar in de toekomst

mogelijk een oplossing voor kunnen vinden. Dat sociale media naast traditionele media een belangrijke bron zijn voor de gebruiker is evident.

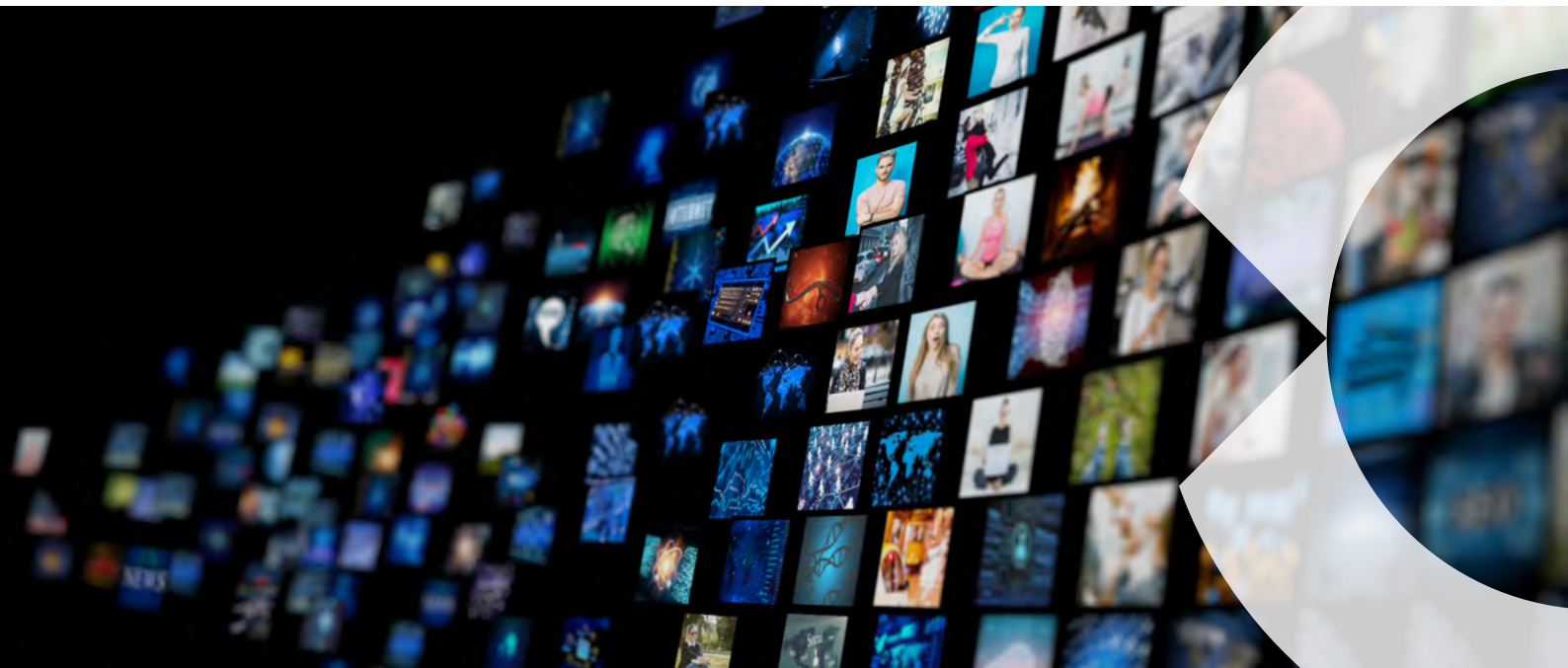
Met de toevoeging van de meetmethodes voortkomende uit pilot 1 en pilot 2 aan de Mediamonitor bouwen we kennis op en kunnen we ontwikkelingen van pluriformiteitsindicatoren volgen om doelgericht te adviseren hoe de productiediversiteit en inhoudelijke diversiteit van het nieuwsaanbod kan worden beschermd, wat het eigenlijke doel van de motie Kwint is.²

Op dit moment wordt ook het toetsingskader voor de zogenoemde pluriformiteitstoets, artikel 22 EMFA, ontwikkeld. Project 1 en 2 kunnen in aanvulling op verhoudingsgewijs harde cijfers over markt- en opiniemacht bijdragen aan het beoordelen van de gevolgen van mediaconcentraties op het media-aanbod en project 1 aan de beoordeling van de waarborgen voor redactionele onafhankelijkheid.

Niet alleen mediaconcentraties hebben effecten op de pluriformiteit van het nieuwsecosysteem. De digitale transformatie vraagt om andere aanpassingen van productie – denk aan de inzet van AI – , gebruik van nieuwe mediakanalen – denk aan sociale media – en beperkingen van toegang – denk aan algoritmes. Big tech, dus bedrijven die niet zelf content produceren, worden steeds belangrijkere gatekeepers en bepalen wie wat te zien krijgt. Daarom is ook, onafhankelijk van mediaconcentraties, het monitoren van de pluriformiteit van het nieuwsecosysteem van groot maatschappelijk belang.

² [Plan van aanpak persvrijheid en persveiligheid en motie Kwint](#)

1 Aanleiding



Het Ministerie verzocht in de zomer van 2022 het Commissariaat om samen met relevante partijen een model te ontwikkelen om de pluriformiteit van (het aanbod op) de nieuwsmarkt te monitoren.³ Zoals geschreven in het voorwoord was de aanleiding voor deze opdracht de motie Kwint die op 11 november 2021 werd aangenomen door de Tweede Kamer. De motie constateert “dat zowel in geschreven media als op televisie steeds meer aanbieders in handen komen van steeds minder partijen” en “verzoekt de regering in kaart te brengen op welke manier de pluriformiteit van het media-aanbod beter beschermd kan worden, en de Kamer hierover te informeren.”⁴ Pluriformiteit kreeg hiermee na jaren weer aandacht van de politiek.

Aandacht voor pluriformiteit was er tot 2011. Tot die tijd gold in Nederland de Tijdelijke wet mediaconcentraties (hierna: Twm). De Twm zorgde, net als in andere Europese landen, voor beperkingen voor fusies en overnames van

mediabedrijven. De Twm is in 2007 ingevoerd “uit zorg over de almaar toenemende concentraties van mediabedrijven. Die politieke zorg gold en geldt vooral over de gevolgen van dergelijke concentraties voor de pluriformiteit en de onafhankelijkheid van de nieuws- en informatievoorziening in ons land. Die politieke zorg is onverminderd gerechtvaardigd, met name op het terrein van de dagbladpers.”⁵ Deze politieke zorg is in 2010 naar de achtergrond verdwenen, de Twm werd voortijdig beëindigd.⁶ De ontstane ruimte wordt door verschillende fusies en overnames gevuld. De mediamarkten raken daardoor verder geconcentreerd en op de dagbladenmarkt is intussen in feite sprake van een duopolie.

Dat er een focus op de pluriformiteit van de nieuwsvoorziening ligt is logisch, gezien dat maatschappelijke risico's door een gebrek aan een pluriforme nieuwsvoorziening ontstaan en niet door gebrek aan pluriform vermaak of reclame.⁷

³ <https://open.overheid.nl/repository/ronl-e5dc6fd98ea2499f7a868cad7f270b6c7a411174/1/pdf/plan-van-aanpak-persvrijheid-en-persveiligheid.pdf>

⁴ <https://zoek.officielebekendmakingen.nl/kst-32827-236.html>

⁵ Kamerstuk 30921, nr. 9 | Overheid.nl > Officiële bekendmakingen

⁶ Staatsblad 2010, 835 | Overheid.nl > Officiële bekendmakingen

⁷ Focus op functies, p. 95. <https://www.wrr.nl/publicaties/rapporten/2004/12/14/focus-op-functies-uitdagingen-voor-een-toekomstbestendig-mediabeleid>

1.1 Relevantie pluriform nieuwsecosysteem

Pluriformiteit betekent op zichzelf niet meer dan verscheidenheid of diversiteit en is een abstracte term die moet worden geoperationaliseerd om betekenisvol te zijn. Volgens Jan van Cuilenburg, emeritus hoogleraar Communicatiewetenschap en oud-voorzitter van het Commissariaat, is pluriformiteit het onderling verschillen van analyse-eenheden naar een bepaald kenmerk of bepaalde kenmerken.⁸ Pluriformiteit moet dus worden gemeten en kunnen worden gerapporteerd in vorm van een statistisch maat. De pluriformiteit van de nieuwsmarkt laat zich niet met één kenmerk meten. De Wetenschappelijke Raad voor het Regeringsbeleid (WRR) onderscheidt vier

dimensies van pluriformiteit: aanbieder, aanbod, inhoud en gebruik.⁹ Binnen de opzet van de Mediamonitor is pluriformiteit een construct dat indirect wordt gemeten aan de hand van indicatoren, de verschillende diversiteitsdimensies. De diversiteitsdimensies hangen samen en interageren.

De pluriformiteit is hoog als er veel even grote aanbieders zijn, veel zelfstandige nieuwsmerken en nieuwsdiensten zijn, het aanbod divers is (onderwerpen, representatie van bevolkingsgroepen, etc.) en het gebruik divers is. Het zijn afzonderlijke dimensies die afzonderlijk en in samenhang met elkaar kunnen worden bekeken. Pluriformiteit is belangrijk zodat alle meningen gehoord worden en burgers zich een mening kunnen vormen en zich herkennen in het (nieuws)aanbod.

Hendrickx¹⁰ onderscheidt vijf dimensies van pluriformiteit die hij als tandwielen beschrijft en die met elkaar samenhangen. We grijpen voor de beschrijving van de diversiteitsdimensies zijn beeld op omdat een eenzijdige causale beïnvloeding tekort schiet. Leunend op Hendrickx onderscheiden we vijf dimensies:

- Diversiteit van de aanbieders: mediamerken zijn in handen van verschillende eigenaren en er is wet- en regelgeving om concentratie te beperken.
- Diversiteit van de nieuwsmerken: er zijn verschillende titels met een eigen identiteit binnen mediabedrijven en -landschappen.
- Diversiteit van de inhoud: in het totaalaanbod zijn veel verschillende invalshoeken, onderwerpen en meningen.
- Diversiteit van het gebruik: de gebruiker vertrouwt en volgt verschillende nieuwsmerken en journalisten.
- Productiediversiteit: de productiediversiteit staat centraal in het model omdat deze wordt beïnvloed door de andere vier dimensies en zelf invloed uitoefent op de vier dimensies. Veranderingen in bijvoorbeeld eigendom en gebruik beïnvloeden de productie waarbij de individuele werkende journalist centraal staat. Dit begrip wordt breed geïnterpreteerd, het kan gaan om de arbeidsomstandigheden, de diversiteit van de journalistieke medewerkers en om de journalistieke onafhankelijkheid.



⁸ Zie ook J.J. van Cuilenburg & D. McQuail (1982). Media en Pluriformiteit. Een beoordeling van de stand van zaken. Wetenschappelijke Raad voor het Regeringsbeleid (WRR), p. 24

⁹ Focus op functies, p. 31 <https://www.wrr.nl/publicaties/rapporten/2004/12/14/focus-op-functies-uitdagingen-voor-een-toekomstbestendig-mediabeleid>

¹⁰ https://www.researchgate.net/profile/Jonathan-Hendrickx-3/publication/359314426_Operationalising_news_diversity_A_comparison_of_Norway_and_Flanders/links/62347c36be72d414dacaf439/Operationalising-news-diversity-A-comparison-of-Norway-and-Flanders.pdf

Figuur 1. Diversiteitsdimensies



1.2 Uitbreiding opzet Mediamonitor

Het Commissariaat is in het kader van de opdracht van de Mediamonitor al “belast met het onderzoek naar ontwikkelingen ten aanzien van concentraties (...) en de gevolgen daarvan voor de pluriformiteit en onafhankelijkheid van de informatievoorziening.”¹¹ Het Commissariaat houdt met de Mediamonitor de ontwikkelingen in het medialandschap al sinds 2001 bij met een uitgebreid monitoringsmodel dat een beoordeling van diverse dimensies van de pluriformiteit van het nieuwsecosysteem mogelijk maakt.

Met de motie Kwint en de brief van de staatssecretaris aan de Tweede Kamer wordt duidelijk dat in Nederland, aanvullend op de huidige mediamonitoring, behoefte is aan een continue monitoring van de diversiteit van de productie en media-inhoud en vanwege het maatschappelijke belang met name van het nieuwsaanbod. Intussen is aanvullend ook Europese wetgeving, de EMFA, binnen enkele maanden van

kracht. In artikel 22 schrijft de EMFA voor dat lidstaten in hun nationale wetgeving materiële en procedurele regels moeten opnemen om concentraties op de mediamarkt te beoordelen die een wezenlijke invloed zouden kunnen hebben op de pluriformiteit van de media en de redactionele onafhankelijkheid. Deze regels worden ook wel aangeduid als de zogenaamde pluriformiteitstoets.¹² Ook in het regeerprogramma is aandacht voor de gevolgen van fusies voor de pluriformiteit van het media-aanbod.¹³

Dit alles vraagt om een structurele pluriformiteitsmonitor die indicatoren voor alle vijf diversiteitsdimensies in kaart brengt. Drie dimensies worden met de Mediamonitor structureel gevolgd. Dat zijn aanbiedersdiversiteit, diversiteit van de nieuwsmerken en de diversiteit van het gebruik. Twee dimensies die sterk met elkaar samenhangen zijn alleen incidenteel onderzocht: de productiediversiteit en de diversiteit van de inhoud.

¹¹ Mediawet 2008, art 7.21

¹² Verordening - EU - 2024/1083 - EN - EUR-Lex

¹³ Regeerprogramma kabinet-Schoof | Publicatie | Rijksoverheid.nl, p.87

Het langetermijndoel van het project is daarom indicatoren voor productiediversiteit en diversiteit van de inhoud van het nieuwsaanbod aan de structurele monitoring van de Mediamonitor toe te voegen. In de eerste fase hebben we met methodische pilots de mogelijkheden verkend om de productiediversiteit en de diversiteit van het nieuwsaanbod structureel in kaart te brengen en om een start te kunnen maken met monitoring.

1.3 Keuze voor drie pilots

Op 15 december 2022 hield het Commissariaat een workshop om het onderwerp met relevante organisaties te bespreken. De workshop was bedoeld als eerste stap naar een project. Het doel was om ideeën te krijgen over hoe invulling kan worden gegeven aan het monitoren van de diversiteit van het inhoudelijke nieuws- en opinieaanbod. Tijdens de workshop werden diverse opties genoemd om diversiteit van de inhoud en

productie van de verschillende type nieuwsaanbod te meten. Deze zijn in drie clusters onderzoek te groeperen:

1. Indicator productiediversiteit: onderzoek naar ervaringen van journalisten als indirecte indicator voor diversiteit van het aanbod, tegelijk indicator voor hoeveelheid en kwaliteit.
2. Indicator inhoudelijke diversiteit: onderzoek naar unieke en overeenkomstige berichtgeving.
3. Indicator inhoudelijke diversiteit: onderzoek naar sentimenten, opinie of invalshoeken ten opzichte van onderwerpen wordt heel vaak genoemd.

Onafhankelijke wetenschappers ontwikkelden pilots in opdracht van het Commissariaat. De voortgang van de pilots is besproken met een begeleidingscommissie die is samengesteld door het Commissariaat. De fase van methodeontwikkeling komt nu ten einde met dit voorgelegde rapport.

2 Pilot 1: Ervaringen van pluriformiteit onder journalisten



Geschreven door: Mark Boukes, Universiteit van Amsterdam

In de wetenschappelijke literatuur is al veel geschreven over hoe de pluriformiteit van de media over tijd verandert. Dit betreft vooral veel onderzoek dat is gedaan door middel van inhoudsanalyses, met vragen zoals: *Hoe is de diversiteit van nieuwsverslaggeving door de tijd heen veranderd?* Dit zal ook onderzocht worden in de andere twee pilots van dit project. Verder zijn er veel theoretische onderzoeken gedaan die door middel van conceptuele verkenningen theoriseerden hoe mediaconcentratie de pluriformiteit in een context zou kunnen beïnvloeden.¹⁴

Wat echter grotendeels ontbreekt in al het onderzoek naar mediapluriformiteit is het perspectief van de journalisten zelf hierop: hoe ervaren zij hun vrijheid om de onderwerpen en invalshoeken te kiezen waarover ze willen schrijven? Verandert deze vrijheid over de tijd heen en welke factoren bepalen hoe zeer zij zich (niet) belemmerd hierin voelen? Om een representatief beeld te schetsen van deze ervaringen is een grootschalig en herhaald enquête-onderzoek onder

journalisten nodig. De huidige pilot heeft de vragenlijst hiervoor ontwikkeld en poogt om een zo groot mogelijke groep journalisten deze in te laten vullen.

Monitoring op lange termijn is echter nodig om de trends in beeld te brengen. De eerste pilot binnen de ontwikkeling van de Monitor pluriformiteit kan daartoe het startschot zijn: de zogenaamde t0-meting. Door het onderzoek in latere jaren te herhalen, kan worden geanalyseerd of de pluriformiteitsbeleving van journalisten toeneemt of daalt, onder welke journalisten dit het geval is en hoe dit samenhangt met eventuele veranderingen van organisatiestructuren, zoals overnames in de mediasector of budgetwijzigingen bij de publieke omroep. Om dit onderzoek gedegen uit te voeren zijn twee randvoorwaarden nodig: (1) een veelomvattende vragenlijst en (2) de grootschalige medewerking van journalisten en/of journalistieke organisaties. De volgende secties bespreken beide elementen.

¹⁴ Bagdikian, B. H. (1984). The US Media: Supermarket or Assembly Line?. *Journal of Communication*, 35(3), 97-109.

2.1 Ontwikkeling van de vragenlijst

Om pluriformiteit effectief te kunnen monitoren is het nodig om een goede definitie te vinden voor dit concept. In eerdere workshops van het Commissariaat is hier ook een poging toe gedaan, echter blijft het altijd een abstract concept dat op meerdere manier te interpreteren valt. In deze pilot wordt *beleefde* pluriformiteit als afhankelijke variabele onderzocht door te focussen op de vrijheid die journalisten ervaren bij het selecteren van de nieuwsonderwerpen, de ervaren vrijheid bij het kiezen van de invalshoeken die ze daarbij benadrukken en de afwezigheid van zelfcensuur in hun werk.

Hoewel dit een indirecte benadering is van pluriformiteit – het onderzoekt niet direct de aanwezigheid van diverse standpunten en perspectieven binnen de verslaggeving in het Nederlandse medialandschap – staat het een meting toe in hoeverre Nederlandse journalisten zich al dan niet gehinderd voelen in het creëren van pluriforme nieuwsverslaggeving. Dit maakt een meerdimensionaal onderzoek naar pluriformiteit mogelijk in combinatie met de inhoudsanalysepilots (pilot 2 en pilot 3).

Om te onderzoeken of en hoe de belevenis van journalistieke (on)vrijheid in onderwerp-/invalshoekkeuze het gevolg is van de toenemende mediaconcentratie in Nederland (motie Kwint: “*zowel in geschreven media als op televisie steeds meer aanbieders in handen komen van steeds minder partijen*”) is een verkenning van de factoren die deze ervaren (on)vrijheid kunnen verklaren nodig. De keuze is daarbij gemaakt voor een breed kader aan onafhankelijke variabelen. De reden daarvoor is dat waarover en hoe journalisten verslag doen van

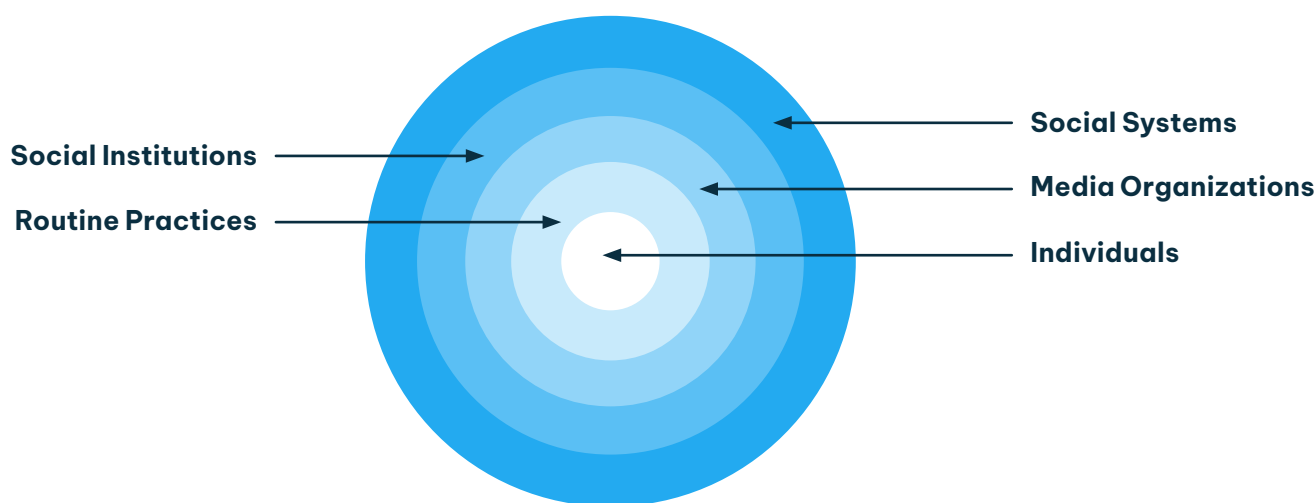
onderwerpen wordt bepaald door variabelen op verschillende niveaus. Deze worden gevisualiseerd in het Hierarchy of Influences Model van Shoemaker en Reese.¹⁵

Wanneer men in de vragenlijst alleen zou vragen naar de werkomstandigheden zoals die worden gecreëerd door de mediaorganisatie (en eventuele fusies), dan is het zeer waarschijnlijk dat andere daarmee samenhangende en/of meer bepalende factoren over het hoofd worden gezien. Daarom richt dit onderzoek zich op de verschillende niveaus – van de individuele journalist (microniveau) en de organisatie waarin zij werken (mesoniveau) tot de maatschappelijke omstandigheden (macroniveau) die de vrijheden in hun werk mede bepalen. Daarmee hoopt het onderzoek een algemeen beeld te kunnen schetsen van de staat van pluriformiteit op het moment dat het onderzoek wordt uitgevoerd; daarnaast zal het inzicht geven in de variabelen die bijdragen aan of afbreuk doen aan deze pluriformiteit. In de volgende secties wordt uitgelegd hoe de verschillende niveaus worden bevraagd in de vragenlijst en welke relevantie dit dient.

Bestaand onderzoek

Om een vragenlijst te creëren die de verschillende niveaus in het Hierarchy of Influences Model analyseert wordt in de huidige enquête een combinatie van vragen gebruikt die grotendeels afkomstig zijn uit drie bestaande onderzoeken die eerder de percepties van Nederlandse of Vlaamse journalisten in kaart brachten: de *World of Journalism Studies* (Nederlandse vertaling door Hermans et al., 2023), het *Portret van Belgische Journalisten in 2023* (door Vanhaelewyn et al., 2023), en het onderzoek *Nederlandse journalisten in 2010* (door Hermans et al., 2010). Deze vragenlijsten zijn reeds gevalideerd, bovendien

Figuur 2 Hierarchy of Influences Model, Shoemaker en Reese



¹⁵ Shoemaker en Reese, *Mediating the Message in the 21st Century, A Media Sociology Perspective*, p. 9 https://psipp.itb-ad.ac.id/wp-content/uploads/2020/10/Pamela-J.-Shoemaker-Stephen-D.-Reese-Mediating-the-Message-in-the-21st-Century_-A-Media-Sociology-Perspective-2013-Routledge.pdf

heeft dit twee voordelen: het maakt het mogelijk om onze bevindingen te vergelijken met bestaande kennis over Nederlandse journalisten in eerdere tijdsperiodes, bovendien staat het een internationale vergelijking met eerdere rapportages toe. Waar mogelijk zullen we ook proberen om de originele data van deze studies te verkrijgen, zodat ook statistische vergelijkingen gemaakt kunnen worden.

Naast de toepassing en het combineren van bestaande vragenlijsten is de vragenlijst aangevuld met vragen die specifiek van toepassing zijn op het onderwerp van dit onderzoek, namelijk pluriformiteit. Zo zijn er vragen toegevoegd over zelfcensuur, geografische herkomst, diversiteitskenmerken (microniveau), redactiestatuten en contentsharing, nevenwerkzaamheden en belangenverstrengeling (mesoniveau), externe beïnvloeding en bedreigingen voor de journalistiek (macroniveau).

Microniveau: Achtergrondkenmerken

“Who says what in which channel to whom with what effect?” (Lasswell, 1948) is één van de meest geciteerde quotes in journalistiek onderzoek.¹⁶ Het benadrukt dat “wie” de boodschap creëert ertoe doet voor de inhoud alsook voor de impact die het uiteindelijk kan hebben. In bestaand onderzoek naar pluriformiteit wordt echter vaak alleen onderzocht *hoe* (“what”) de journalistieke organisatie de inhoud beïnvloedt. Dit laat ongesproken dat diversiteit *onder* journalisten (en niet alleen onder de media-eigenaren) ook kan bijdragen aan de pluriformiteit van een medialandschap. Diverse achtergrondkenmerken van journalisten zullen in dit onderzoek worden uitgevraagd, zoals standaard is in vragenlijstonderzoek. Daarnaast worden diverse nieuwe elementen meegewogen, die nog niet eerder aan bod kwamen in onderzoek, zoals de woonplaats waarin iemand opgroeide en momenteel woont (naar aanleiding van het veelgehoorde verwijt dat veel journalistiek zich op de Randstad concentreert), sociaaleconomische klasse en het zijn van een eerstegeneratie-student¹⁷, als ook een open vraag die journalisten zelf laat beargumenteren waarom hij/zij/hen wel of niet als divers kan worden beschouwd.

Mesoniveau: De organisatie

Centraal in de vraagstelling van de motie Kwint staat dat *“steeds meer aanbieders in handen komen van steeds minder partijen.”* Het element van beïnvloeding door de mediaorganisatie tezamen met de eigenaarsstructuur wordt daarom uitgebreid bevraagd. Het argument dat veelvuldig door mediapartijen wordt gebruikt waarom meer geconcentreerd eigenaarschap de vrijheid van een redactie – en daarmee de pluriformiteit – niet direct zou bedreigen is het bestaan van de

redactiestatuten. Onbekend is echter of het theoretische mechanisme van deze redactiestatuten in de praktijk ook werkt: zijn journalisten bekend met het statuut van hun medium en is hun belevenis ook dat de aanwezigheid van een redactiestatuut de journalistieke onafhankelijkheid van hun organisatie waarborgt? Dit is een cruciale vraag die niet eerder vanuit het perspectief van journalisten beantwoord werd. Daarnaast wordt gevraagd naar een scala aan mogelijke factoren die van invloed kunnen zijn op de inhoudelijke werkzaamheden van journalisten (bijvoorbeeld onderwerpkeuze, toon, invalshoek) en of zij deze invloed als positief of negatief ervaren. Hier wordt expliciet gevraagd naar eigenaren van de nieuwsorganisatie, adverteerders en reclameoverwegingen en het redactiestatuut, naast veel andere factoren (zoals collega's, mediaregelgeving, het publiek, etc.). Dit maakt het mogelijk om te zien hoe factoren van eigenaarschap opwegen tegen diverse andere mogelijke invloeden.

Macroniveau: De omgeving

Omdat journalisten niet in een geïsoleerde omgeving werken, maar inherent onderdeel zijn van de maatschappij waar ze zelf verslag van doen, zal deze omgeving ook invloed hebben op de journalistiek. Wanneer het gaat om pluriformiteit en het ervaren van (on)vrijheid in het maken van journalistieke keuzes is het daarom belangrijk om dit ook mee te laten wegen in een monitoring op lange termijn. In de verschillende deelelementen van de enquête wordt daarom niet alleen gevraagd naar individuele kenmerken of invloeden vanuit de mediaorganisatie, maar waar mogelijk ook naar maatschappelijke invloeden, zoals die vanuit de politiek, ambtenarij, PR, kunstmatige intelligentie en economie (zoals werkdruk). Een speciale vraag is ingeruimd voor veiligheid en bedreigingen gericht aan journalisten.

Langetermijntrends

De resultaten zullen veel relevante inzichten bieden in welke factoren de (on)vrijheid van Nederlandse journalisten beïnvloeden. Dit levert cruciale inzichten op voor wetenschap, pers en beleidsmakers: welke elementen dragen bij aan pluriforme berichtgeving en welke vormen juist een bedreiging? Deze vragen zullen worden beantwoord door middel van verklarende analyses (veelal OLS regressies), waarin de invloed van verschillende factoren wordt vergeleken. Voor de beschrijvende analyses zoals welke factoren hoeveel invloed uitoefenen, zijn echter herhaalde metingen nodig. Een voorbeeld: als op de vraag “Dankzij de aanwezigheid van een redactiestatuut is de journalistieke onafhankelijkheid van mijn organisatie gewaarborgd” een gemiddelde score van 2,67 op een schaal van 0 tot 4 wordt gevonden is dat lastig te interpreteren. Maar als dit over

¹⁶ Lasswell, H. D. (1948), “The Structure and Function of Communication in Society,” in L. Bryson (ed.), *The Communication of Ideas*, pp. 32-51. New York: Harper.

¹⁷ Persoon die studeert maar waarvan de ouders geen studie genoten

een X-aantal jaar 3,00 is, dan weten we dat de werking van het redactiestatuut verbeterd is; en vice versa bij een daling. Het huidige vragenlijstonderzoek kan daarom het benodigde referentiekader zijn voor vervolgonderzoek om te monitoren hoe de beleefde pluriformiteit onder journalisten over tijd verandert. Het voorstel is om het onderzoek iedere 4 jaar te herhalen.

2.2 Bevindingen pilot 1

Maximaliseren van de respons

De huidige pilot heeft tot nu toe geleid tot een herbruikbare vragenlijst die de invloed van diverse factoren op de *beleefde* pluriformiteit onder journalisten meetbaar maakt. Het meet-instrument is daarmee gereed. Onmisbaar voor het slagen van het project is daarnaast de medewerking van een grote en representatieve groep journalisten. Waar in veel andere landen een openbare database bestaat van journalisten die voor een dergelijk onderzoek bereikt zouden kunnen worden, bestaat dit *niet* in Nederland. Daarom is het cruciaal dat de verschillende journalistieke stakeholders hun medewerking verlenen aan het onderzoek om de vragenlijst onder hun medewerkers en leden te verspreiden. Op dit moment is het onderzoek al verspreid onder de leden van de Nederlandse Vereniging van Journalisten (NVJ). Dit heeft geleid tot een respons van meer dan 500 journalisten. Omdat dit leden zijn van een beroepsorganisatie is het lastig hen te generaliseren naar de hele populatie van journalisten. Deze journalisten kunnen afwijken van de gehele groep journalisten.

Daarom is er veel tijd gestoken in gesprekken met NDP Nieuwsmedia, de brancheorganisatie voor nieuwsmedia in Nederland, om via de journalistieke werkgevers de vragenlijst ook te verspreiden onder hun vaste werknemers. Op het moment van schrijven (20 januari '25) wordt dit proces uitgevoerd. We hopen op een later moment de vragenlijst ook uit te kunnen zetten onder de journalistieke medewerkers van de NPO, maar daarover is nog geen besluit genomen. Omdat de NPO een grote speler is in het medialandschap en cruciaal is voor de pluriformiteit is het van groot belang dat hij zich ook achter het onderzoek schaaft. In een later stadium wordt ook gepoogd om de vragenlijst te verspreiden via de diverse lokale en regionale media in Nederland.

Frequentie

Zoals hierboven beschreven is het voorstel is om het onderzoek iedere 4 jaar te herhalen. Echter is dit afhankelijk van het budget en de medewerking van journalistieke partijen. Op deze

manier kan de variatie over tijd heen gemonitord worden van de specifieke vragen. Een periode van 4 jaar loopt ook in de pas met eventuele regeerperiodes/verkiezingen en het staat toe de impact van andere grote maatschappelijke gebeurtenissen te onderzoeken, zonder de journalisten over te bevragen met bijvoorbeeld een jaarlijkse meting.

In de vragenlijst hebben we de journalisten verzocht om hun emailadres achter te laten. Dit staat toe om hen over 4 jaar opnieuw te benaderen. Daarmee kunnen we niet alleen de gemiddelde veranderingen over de tijd heen meten, maar ook zien hoe dit voor individuele journalisten (nog steeds op anonieme basis) is verschoven of gelijk gebleven. Daarmee kunnen we preciezere inschattingen maken van hoe de pluriformiteits-beleefde veranderde onder journalisten en welke factoren daaraan bijdroegen.

Uitdagingen

De grootste uitdaging tot nu toe is het verkrijgen van medewerking van de journalistieke partijen. Om diverse redenen, en erg begrijpelijk, worden de thema's die worden bevraagd in de enquête beschouwd als een gevoelig onderwerp. Bovendien is het logisch dat men niet zomaar hun redacteuren beschikbaar stelt voor een project als dit. De ervaring is echter dat na verschillende overleggen (via mail, online of fysiek) zowel het maatschappelijke en wetenschappelijke belang van dit onderzoek begrepen wordt. We hopen ook de NPO binnenkort hiervan te kunnen overtuigen. Echter is het tot nu toe lastig gebleken om daar de verantwoordelijke persoon voor te spreken.

Dit toont ook aan dat er een reëel risico bestaat dat over vier jaar personeelwisselingen plaatsgevonden hebben, waardoor dit proces geheel opnieuw zou moeten starten, hetgeen wederom een grote tijdsinvestering zou vergen. Idealiter zouden we de afspraken op papier moeten vastleggen nadat pilot 1 succesvol is verlopen volgens alle betrokken partijen, zodat voor een tweede ronde dit proces gemakkelijker kan verlopen. Belangrijk om daarbij te benadrukken is dat in vervolgrondes dezelfde vragen gesteld worden, in dezelfde bewoordingen en op dezelfde wijze (bijvoorbeeld antwoordschalen). Alleen op deze manier kan een valide vergelijking gemaakt worden tussen de verschillende jaren. Wanneer andere vragen (of dezelfde vragen op een andere manier) gesteld worden, dan kan de vraagstelling een verklaring zijn voor gevonden verschillen in plaats van de veranderingen in de journalistieke beleving.

3 Pilot 2: Unieke en overeenkomstige berichtgeving



Geschreven door: Nel Ruigrok, Nieuwsmonitor B.V.

In dit hoofdstuk gaan we in op het onderzoek naar de mogelijkheden om de diversiteit van het nieuwsaanbod structureel te monitoren. Hiervoor is een maand nieuws, afkomstig van uitgeverconcerns DPG en Mediahuis, vergeleken met het nieuws van ANP en NOS.nl. De vraag die hierbij centraal stond was de volgende: “In hoeverre is er sprake van unieke en overeenkomstige berichtgeving in het nieuwsaanbod van verschillende media?”

We bespreken in dit hoofdstuk niet de uitkomsten van het onderzoek, maar gaan primair in op de opzet van het onderzoek en geven een methodologische toelichting. Daarnaast bespreken we de voor- en nadelen van de methode en geven we een overzicht van de mogelijkheden en uitdagingen om een dergelijke monitor structureel op te zetten.

Dataverzameling

In deze pilot hebben we gekeken naar de berichtgeving in de maand maart 2024, waarbij we zijn uitgegaan van data geleverd door de uitgevers zelf. Het gaat hierbij om het online nieuws van de volgende titels van DPG (AD, BN DeStem, Brabants Dagblad, De Gelderlander, De Stentor, Eindhovens Dagblad, Het Parool, NU.nl, PZC, Trouw,

Tubantia, de Volkskrant) en Mediahuis (NRC, De Telegraaf, Leidsch Dagblad, Dagblad voor Westfriesland, De Gooi- en Eemlander, Alkmaarsche Courant, Haarlems Dagblad, Dagblad Kennemerland, Helderse Courant, Schager Courant, IJmuider Courant, Dagblad Waterland, Dagblad Zaanstreek). Daarnaast hebben we ook gekeken naar het nieuws zoals gepubliceerd op NOS.nl en het nieuws afkomstig van het ANP.

3.1 Selectie van nieuwsartikelen

Hieronder beschrijven wij welke nieuwsartikelen buiten de analyse zijn gehouden en waarom.

Robotjournalistiek

In de data van DPG Media worden veel artikelen (deels) automatisch gegenereerd. Het gaat hierbij om zogenaamde 112-berichten waar nieuws over ongelukken, misdrijven en andere calamiteiten automatisch worden aangepast aan de specifieke plaats en omstandigheden. Ook nieuws op basis van bijvoorbeeld CBS-onderzoek over ontwikkelingen in het land worden op deze manier aangepast naar de plaats waar de online mediatitel zich op richt. Bij het AD wordt ongeveer 60% van de artikelen op deze manier geproduceerd. Deze

artikelen zijn niet identiek maar kennen wel een grote overlap en zijn in dit onderzoek achterwege gelaten. Het gaat vaak om heel regio-gebonden nieuws, zoals voetbaluitslagen van de lokale clubs.

In een analyse zouden dergelijke artikelen ervoor zorgen dat er minder diversiteit van het nieuws geboden wordt aan de nieuwsconsument. Echter, de specifieke nieuwe inhoud van een bericht is wel uniek voor de nieuwsconsument (bijvoorbeeld de uitslag van de plaatselijke voetbalclub). Bovendien zorgt een dergelijke service ervoor dat nieuwsconsumenten die eigenlijk vooral geïnteresseerd zijn in de plaatselijke voetbalclub door hun komst naar de website toch ook ander nieuws meekrijgen.

In dit onderzoek laten we automatisch gegenereerde nieuwsberichten achterwege en richten we ons op het nieuwsaanbod zoals door de verschillende journalisten geproduceerd.

Gedeelde artikelen binnen een concern

Tijdens het onderzoek stuitte we op een groot aantal artikelen dat intern door verschillende redacties wordt gedeeld. De afweging om artikelen in meerdere titels te gebruiken is een bedrijfsmatige afweging en geen journalistieke afweging. De verschillende afwegingen kunnen ook leiden tot verschillende uitkomsten wat betreft diversiteit van het nieuws. Zo kan iemand in Zeeland via de site AD.nl met keuzes voor verschillende gemeenten in Zeeland bijna alle artikelen lezen die ook via PZC.nl worden aangeboden. Deze afwegingen zijn niet per se slecht voor de diversiteit van het nieuws: het zorgt ervoor dat hetzelfde nieuws via verschillende kanalen te lezen is.

Deze artikelen worden niet meegenomen in het onderzoek omdat er geen journalistieke overwegingen worden gemaakt bij het produceren van het nieuws.

Artikelen met minder dan 20 woorden

We lieten ook artikelen achterwege die minder dan 20 woorden bevatten. Dit betreft veelal cartoons, colofons of hele korte aankondigingen.

Artikelen die overlap kunnen vertonen

In totaal hebben we 46.683 artikelen met elkaar vergeleken.

3.2 Methode

Om te kunnen bepalen in hoeverre er sprake is van overeenkomstige berichtgeving stellen we de volgende vraag: in hoeverre krijgt een nieuwsconsument eenzelfde beeld of verschillende beelden van het nieuws zoals op de verschillende mediatitels wordt aangeboden op iedere dag in maart? Hierbij gaan we expliciet uit van het aanbod dat aan de nieuwsconsument beschikbaar wordt gesteld.

Om de overeenkomsten te meten tussen de berichten van de diverse mediatitels kijken we naar de mate waarin alle nieuwsitems die gepubliceerd zijn op een dag op elkaar lijken. Vervolgens kunnen we het gemiddelde berekenen van alle dagen die zijn meegenomen in de dataverzameling, in dit geval alle dagen van maart 2024.

We bepalen per artikel in welke mate er sprake is van overlap tussen het artikel en alle andere berichten van die dag in de andere mediatitels. Hierbij wordt de overlap berekend op basis van directe overlap van n-grammen (zinsdelen van een of meer aansluitende woorden). We hebben hierbij gekeken naar verschillende maten van overlap:

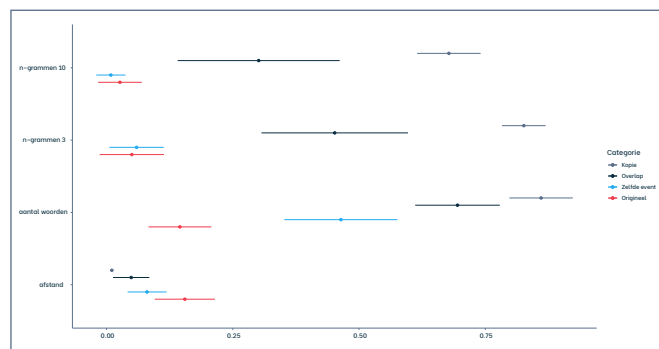
- Een vergelijking van alle woorden binnen een nieuwsbericht.
- Een vergelijking van het aantal n-grammen 3 (3 aaneengesloten overeenkomstige woorden) binnen een artikel.
- Een vergelijking van het aantal n-grammen 10 (10 aaneengesloten overeenkomstige woorden) binnen een artikel.
- Afstand: Een vierde methode die we hebben toegepast gaat uit van een vergelijking op basis van zogenaamde 'word embeddings'. Hierbij wordt een artikel omgezet in een representatie (vector) waardoor de afstand tussen vectoren, oftewel de artikelen, kan worden berekend. Hoe kleiner de afstand is, hoe meer ze op elkaar lijken.

Om de overeenkomsten in woorden en n-grammen te traceren is een aantal stappen nodig waarvoor verschillende programma's bestaan. In het bijzonder hebben we gebruikgemaakt van automatische taalkundige ontleding (Part-of-Speech tagging) van alle teksten. Bij deze ontleding wordt de woordsoort (werkwoord, eigennaam, etc.) van elk woord bepaald. Ook weten we van ieder woord waar het in een tekst staat, waardoor het ook mogelijk is om zogenaamde n-grammen-3 (3 aaneensluitende woorden) en n-grammen-10 (tien aaneensluitende woordcombinaties) te traceren.

Om overlap te bepalen in de berichtgeving hebben we vier categorieën onderscheiden:

- Kopie: Hierbij gaat het om een artikel dat een (bijna) letterlijke kopie is van het artikel waarmee het wordt vergeleken.
- Overlap: Een dergelijk artikel met overlap is niet helemaal een letterlijke kopie van het artikel waarmee het wordt vergeleken, maar heeft wel veel letterlijke overlap.
- Zelfde gebeurtenis: Hierbij gaat het om een artikel dat gaat over dezelfde gebeurtenis als het artikel waarmee het wordt vergeleken en er is wel enige overlap, maar geen hele letterlijke stukken tekst.
- Origineel: Een dergelijk artikel gaat echt over een ander onderwerp, maar heeft misschien wel wat dezelfde woorden of woordcombinaties in de tekst, omdat deze gebruikelijk zijn. Denk aan "het afgelopen jaar" of "volgens de woordvoerder van".

Figuur 3 Overzicht van de gemiddelde waarden van diverse metrics om overlap te bepalen per categorie



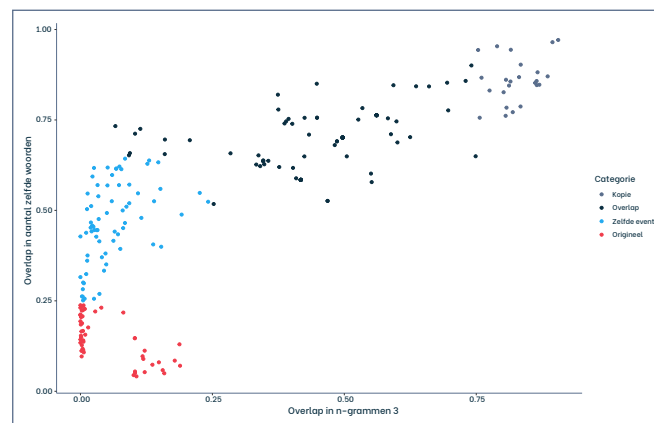
Hoewel we te maken hebben met vier verschillende categorieën kunnen we overlap eigenlijk zien als een schaal van 0 (geen overlap) tot 1 (volledige kopie) en het traceren waar precies de grenzen liggen tussen de verschillende categorieën is lastig. Om dit te doen, hebben we zo'n 500 artikelen vergeleken met andere artikelen en bepaald in welke categorie het artikel viel. Hierbij hebben we eerst 50 artikelen door twee verschillende codeurs laten beoordelen om ervoor te zorgen dat de interpretatie van de categorieën duidelijk was. De betrouwbaarheid van deze coderingen was zeer goed (alfa .93). In figuur 1 geven we per methode weer wat de gemiddelde score was van de gecodeerde artikelen per categorie. Dat is de middenstip van iedere lijn. De uiteinden van de lijnen geven de standaarddeviatie weer per categorie.

Een aantal conclusies kunnen we uit deze figuur trekken:

- De afstandsmaat ('afstand') is met de huidige modellen nog niet voldoende onderscheidend om te gebruiken.
- Dit geldt ook voor de maat waarbij we kijken naar 10 aaneensluitende woorden die gelijk zijn in de artikelen.
- De maat waarbij naar dezelfde woorden wordt gekeken geeft een goed onderscheid tussen artikelen die origineel zijn en die over dezelfde gebeurtenis gaan. Het onderscheid tussen een kopie en overlap is lastiger te zien.
- Wanneer we kijken naar de n-grammen-3 zien we dat het onderscheid tussen overlap en kopie ook lastig is, maar wel iets duidelijker dan bij alleen dezelfde woorden.

Al met al kunnen we hieruit concluderen dat een harde grens tussen de verschillende categorieën lastig te bepalen is. Dit is nog duidelijker te zien in figuur 2 waarbij we van de gecodeerde artikelen weergeven wat hun score was op de overlap in 3 aaneensluitende woorden (de x-as) en de score in de overlap in woordgebruik (y-as).

Figuur 4 Overzicht van de overlap in woordgebruik en de overlap van 3 dezelfde woorden per artikel



In de figuur is te zien dat er een duidelijk onderscheid is tussen de originele artikelen en de artikelen over dezelfde gebeurtenis in termen van 3 aaneensluitende woorden. Deze is altijd minder dan .25. Tegelijkertijd zien we dat artikelen die overlap vertonen of een kopie zijn, altijd meer dan .5 overlap aan woorden bevatten. Diffuser is het beeld wanneer we binnen deze twee groepen kijken. Zo is er een aantal artikelen dat over dezelfde gebeurtenis gaat (paarse punten in de figuur). Deze mengen zich met de artikelen die origineel zijn (groene punten). Dit is nog sterker wanneer we kijken naar de artikelen die overlap vertonen (blauwe punten) en de artikelen die een kopie zijn (rode punten).

Om toch in staat te zijn om onderscheid te maken tussen de groepen hebben we gekozen voor deze afbakening:

Soort artikel	Overlap woordgebruik	Overlap in 3 dezelfde woorden
Origineel artikel	kleiner dan .25	EN kleiner dan .25
Zelfde gebeurtenis	groter of gelijk aan .25 maar kleiner dan .75	EN kleiner dan .25
Overlap tussen artikelen	groter of gelijk aan .75 en kleiner dan .95	OF groter of gelijk aan .25 en kleiner dan .75
Kopie	groter of gelijk aan .95	OF groter dan .75

In overleg met de begeleidingscommissie hebben we deze afbakening gekozen. Deze afbakening geeft plausible resultaten waardoor het een valide methode is gebleken om overeenstemming tussen mediatitels te berekenen. Een iets andere afstemming tussen de verschillende categorieën gaf aan dat er weliswaar kleine verschuivingen zijn in de uitkomsten, maar dat deze wel hetzelfde beeld lieten zien, waardoor de methode als robuust kan worden gezien.

3.3 Monitoring van diversiteit van nieuwsartikelen

Om te komen tot een permanente monitor van de diversiteit van het nieuws is het belangrijk om een toekomstbestendige manier te vinden om dit te meten. Een aantal punten is hierbij van belang: de dataverzameling, het bepalen welke artikelen moeten worden vergeleken en de te gebruiken methode.

Dataverzameling

Tijdens de pilot was de dataverzameling een probleem omdat de data op verschillende manieren werd aangeleverd door de uitgevers. Een belangrijk probleem hierbij was dat bij de titels van DPG het niet te traceren was welke artikelen gebruikt waren voor de papieren versies van de verschillende titels, waardoor we genoodzaakt waren om de analyses te beperken tot de digitale versies van de titels. Ook bij een deel van de artikelen van Mediahuis was het onduidelijk of ze gebruikt waren in een online titel of niet.

Delen van artikelen

Zowel bij DPG als bij Mediahuis is sprake van het delen van artikelen. In verschillende mediatitels zijn dezelfde artikelen terug te vinden. Zo kennen veel regionale titels een katern met hetzelfde nieuws en een katern met specifiek nieuws over de eigen regio. Denk hierbij aan de diverse titels van het AD en het Noordhollands Dagblad. Het delen van deze artikelen is een bedrijfsmatige beslissing en geen journalistieke beslissing. Het is daarom aan te raden dergelijke artikelen niet mee te nemen in een permanente monitor naar diversiteit van het nieuws. Immers, als een lezer via AD.nl, door extra nieuws aan te klikken over diverse gemeenten, bijna allemaal dezelfde artikelen kan lezen als zijn gepubliceerd in PZC.nl, dan zegt dat meer over het bereik dan over de diversiteit van het nieuws.

Methode van vergelijking

We hebben tijdens het proces verschillende methodes bekeken waarvan de geschiktheid verschilt. Zo hebben we ook gekeken naar een vergelijking tussen artikelen op basis van het voorkomen van zelfstandige naamwoorden en eigennamen. Deze methode bleek echter minder geschikt dan de vergelijking tussen het aantal woorden dat overeenkomt en het aantal n-grammen-3 om de verschillende categorieën te bepalen. De combinatie van n-grammen-3 en alle woorden bleek het meest robuust.

De methode die voor deze pilot is gebruikt om op basis van 'embeddings' te komen tot een goede vergelijking tussen de artikelen, bleek nog niet geschikt genoeg om een duidelijk onderscheid te maken. Echter, op het gebied van zogenaamde Large Language Models (LLM's) zijn veel ontwikkelingen gaande en het is zeer voorstelbaar dat er in de nabije toekomst wel modellen zijn die dit onderscheid goed (genoeg) kunnen maken. Deze manier is te prefereren boven de methodes zoals wij hebben toegepast omdat het minder tijdsintensief is en makkelijker te automatiseren en daarmee ook toekomstbestendig is.

Conclusie

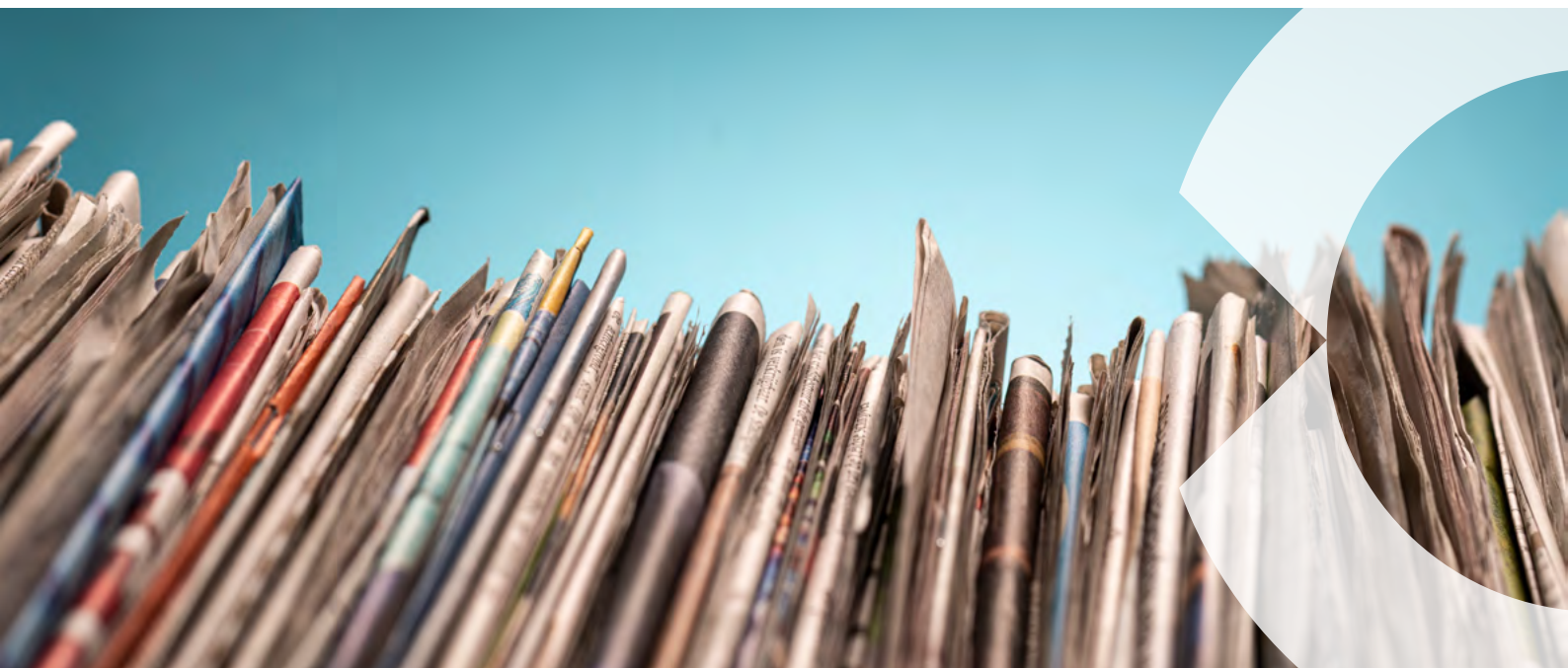
Om te komen tot een permanente monitor op het gebied van diversiteit is het aan te raden een dergelijke monitor op jaarbasis uit te voeren.

Vanuit een praktische overweging is het aan te raden om bij de opzet van de monitor in eerste instantie de focus te leggen op de landelijke titels en van deze media en van alle katernen de berichtgeving mee te nemen, waarbij afgesproken wordt dat deze wordt aangeleverd in een standaardformaat of dat deze gegevens worden verzameld door de onderzoekers zelf, zodat de vergelijkbaarheid gegarandeerd wordt.

Indien het mogelijk is om tot goede afspraken te komen met de uitgevers en het ook mogelijk is om de data op deze wijze te verzamelen van de regionale titels, dan is een dergelijke monitor te prefereren boven alleen de landelijke titels, omdat het een completer beeld geeft.

Tenslotte is het belangrijk om te komen tot een methode om de diversiteit te meten die robuust is, eenvoudig in gebruik en toekomstbestendig. Hoewel de methode om te werken met n-grammen-3 en overlap in woorden wel robuust is, is deze tijdsintensief. Daarbij komt dat de ontwikkelingen op het gebied van Large Language Models zo snel gaan dat het zinvol is om daarin te investeren en in te zetten. Door een grotere hoeveelheid artikelen te coderen waarbij het onderscheid tussen de verschillende categorieën duidelijk wordt gemaakt, is het mogelijk om een dergelijk model nauwkeuriger te trainen, zeker met de toekomstige modellen die nu al worden ontwikkeld in ander onderzoek en die open source beschikbaar zijn.

4 Pilot 3: Diversiteit van thema's, bronnen en standpunten



Geschreven door: Elif Kılık en Knut De Swert, Universiteit van Amsterdam

Voor een goed functionerende democratie is het essentieel dat burgers op een betrouwbare en juiste manier worden geïnformeerd over relevante kwesties en standpunten in de samenleving, zodat zij geïnformeerde keuzes kunnen maken. In het licht van deze normatieve verwachting vervullen nieuwsmedia een cruciale rol als een van de belangrijkste informatiebronnen in democratische systemen.¹⁸ Over de wenselijkheid van een pluriforme publieke sfeer is nagenoeg iedereen het eens, maar de manier waarop pluralisme moet worden ingevuld en gewaarborgd blijft nog altijd een omstreeden kwestie.¹⁹ Mediapluriformiteit blijft daarom een breed concept met verschillende benaderingen.

Recent onderzoek naar wetenschappelijke literatuur toont aan dat de begrippen mediapluriformiteit en mediadiversiteit vaak door elkaar worden gebruikt.²⁰ Maar er zijn ook onderzoekers die expliciet een onderscheid maken tussen de twee termen.^{19, 21}

Deze onderzoekers omschrijven diversiteit als een empirische benadering die zich richt op het analyseren van de mate waarin een breed scala aan onderwerpen en perspectieven daadwerkelijk in het media-aanbod aanwezig is, terwijl pluralisme verwijst naar het beoordelen van deze verschillen vanuit een meer normatieve, ideologische perspectief.^{19, 21} In onze empirische evaluatie van de automatische modellen om pluriformiteit van mediaberichtgeving te meten, gebruiken we in deze pilotstudie de term diversiteit.

In Nederland wordt mediapluriformiteit jaarlijks beoordeeld in het Mediamonitor-onderzoek. Sinds 2001 rapporteert het Commissariaat voor de Media over het media-aanbod, met speciale aandacht voor beschikbare middelen en eigendomsstructuren. Ons onderzoek bouwt voort op deze monitoring door zich te richten op de diversiteit van media-inhoud en de monitoring ervan. Het doel is om een geschikte aanpak te

¹⁸ McNair, Brian. "Journalism and democracy." In *The handbook of journalism studies*, pp. 257-269. Routledge, 2009.

¹⁹ Karppinen, Kari. *Rethinking media pluralism*. Fordham Univ Press, 2013.

²⁰ Loecherbach, Felicia, Judith Moeller, Damian Trilling, and Wouter van Atteveldt. "The unified framework of media diversity: A systematic literature review." *Digital Journalism* 8, no. 5 (2020): 605-642.

²¹ Raeijmaekers, Danielle, and Pieter Maesele. "Media, pluralism and democracy: what's in a name?." *Media, culture & society* 37, no. 7 (2015): 1042-1059.

ontwikkelen waarmee in de toekomst de diversiteit van media-inhoud nauwkeurig kan worden gemonitord.

Voor de vergelijking van verschillende automatische methodologieën voor de analyse van nieuwsdiversiteit hebben we als onderwerp de coronapandemie gekozen. Dit is om meerdere redenen een geschikt onderwerp. Ten eerste kreeg de coronacrisis gedurende een lange periode uitgebreide media-aandacht. Deze uitgebreide berichtgeving biedt ons een rijke dataset om diverse automatische methodologieën te testen die afhankelijk zijn van grote hoeveelheden tekstuele data. Ten tweede heeft de berichtgeving over de pandemie een duidelijk begin en einde en staat het niet langer hoog op de agenda, waardoor we een helder overzicht hebben van de kwesties en perspectieven rondom dit onderwerp. Hierdoor kunnen we ons volledig richten op de methodologische aspecten van dit onderzoek, zonder te worden beïnvloed door historische ontwikkelingen of lopende maatschappelijke discussies over het onderwerp. Ten slotte had de pandemie invloed op diverse aspecten van de samenleving, zoals sociale, economische, politieke en gezondheidszorg, waardoor we onze tests kunnen uitvoeren op een multidimensionaal thema met complexe aspecten. Van toekomstig onderzoek naar mediadiversiteit mag verwacht worden dat het ook andere belangrijke multidimensionale vraagstukken kan analyseren, dus moet een methodologie gezocht worden die dat aankan.

4.1 Drie indicatoren voor diversiteit nieuwsinhoud

Recente systematische literatuuroverzichten identificeerden drie belangrijke indicatoren van de diversiteit van media-inhoud: themadiversiteit, brondiversiteit en standpuntdiversiteit.^{22,23} Themadiversiteit, ook wel issue-diversiteit genoemd, verwijst naar de mate van variatie in de berichtgeving over verschillende onderwerpen en kwesties in het nieuws. Dit hangt af van het “onderwerp” en het niveau waarop de analyse plaatsvindt. Het niveau kan variëren van algemene genres zoals politiek, economie, sport, etc.²⁴ tot de verschillende aspecten van één enkel onderwerp.²⁵ De

coronapandemie is multidimensionaal van aard en kreeg langdurige media-aandacht, waardoor het essentieel is te onderzoeken hoe deze complexiteit in nieuwsberichtgeving werd weerspiegeld. Op basis hiervan maken we onderscheid tussen het hoofdthema, de coronapandemie dus, en de sub-onderwerpen die verschillende aspecten van de pandemie belichten, zoals gezondheidszorg, onderzoek, economie, politiek en sociale impact.

De tweede indicator richt zich op de diversiteit van journalistieke bronnen, namelijk personen en organisaties die in het nieuws informatie of opinies geven door een (direct of indirect) citaat. Een veelgebruikte definitie van bron- of actordiversiteit beschrijft het concept als een “spreiding van de representatie van affiliaties en statusposities van bronnen die worden gebruikt om een nieuwsproduct te creëren”.²⁶ Statusposities verwijzen naar de “elitestatus” van de bron en is gebaseerd op de normatieve verwachting dat verschillende groepen binnen de samenleving ruimte in de media moeten krijgen om informatie of een mening te geven. Een diverse vertegenwoordiging van bronnen omvat niet alleen elitebronnen, zoals overheids- en politieke bronnen, maar ook experts, het maatschappelijk middenveld en burgers. Met affiliatie wordt politieke oriëntatie bedoeld, dus welke politieke partijen vertegenwoordigd zijn in het nieuws. In onze grondige analyse van de methoden voor brondiversiteit in nieuwsberichten onderzoeken we zowel de modellen voor het extraheren van bronnen uit nieuwsteksten als voor het toekennen van statusposities aan deze bronnen.

De derde indicator van inhoudsdiversiteit is standpuntdiversiteit. In hoeverre laat het nieuws verschillende standpunten, meningen en invalshoeken over hetzelfde onderwerp zien? Deze dimensie wordt als het kernaspect van diversiteit gezien²⁷, maar is ook de meest complex om te meten²⁸. Bestaand onderzoek doet dit aan de hand van een frame-analyse^{28,29} of een analyse van de toon van berichtgeving, bijvoorbeeld van meningen voor en tegen immigratie^{27,30}. Omdat standpunten variëren afhankelijk van het onderwerp, beperkt het direct meten van standpunten onderzoekers tot de analyse van één specifiek thema, waardoor de resultaten

²² Loecherbach, Moeller, Trilling, and van Atteveldt. “The unified framework of media diversity: A systematic literature review.”

²³ Joris, Glen, Frederik De Grove, Kristin Van Damme, and Lieven De Marez. “News diversity reconsidered: A systematic literature review unraveling the diversity in conceptualizations.” *Journalism Studies* 21, no. 13 (2020): 1893-1912.

²⁴ Joris, Glen, Jonathan Hendrickx, Stefaan Vercoutere, Orphée De Clercq, and Lieven De Marez. “Paying for news diversity? A topic diversity analysis of free and paywalled online news.” *Journalism Practice* (2023): 1-20.

²⁵ Lörcher, Ines, and Monika Taddicken. “Discussing climate change online. Topics and perceptions in online climate change communication in different online public arenas.” *Journal of Science Communication* 16 (2017).

²⁶ Voakes, Paul S., Jack Kapfer, David Kurpius, and David Shano-yeon Chern. “Diversity in the news: A conceptual and methodological framework.” *Journalism & Mass Communication Quarterly* 73, no. 3 (1996): 582-593, 584.

²⁷ Masini, Andrea, and Peter Van Aelst. “Actor diversity and viewpoint diversity: Two of a kind?” *Communications* 42, no. 2 (2017): 107-126.

²⁸ Baden, Christian, and Nina Springer. “Conceptualizing viewpoint diversity in news discourse.” *Journalism* 18, no. 2 (2017): 176-194.

²⁹ Benson, Rodney. “What makes news more multiperspectival? A field analysis.” *Poetics* 37, no. 5-6 (2009): 402-418.

³⁰ Masini, Andrea, Peter Van Aelst, Thomas Zerback, Carsten Reinemann, Paolo Mancini, Marco Mazzoni, Marco Damiani, and Sharon Coen. “Measuring and explaining the diversity of voices and viewpoints in the news: A comparative study on the determinants of content diversity of immigration news.” *Journalism Studies* 19, no. 15 (2018): 2324-2343.

niet te veralgemenen zijn naar andere onderwerpen.³¹ Daarom meten onderzoekers standpunten soms via journalistieke bronnen, gebaseerd op de gedachte dat de aanwezigheid van bronnen uit verschillende delen van de samenleving voor meerdere perspectieven van een thema kan zorgen.^{27,32} Om standpunten centraal te houden in de meting, focust men dan graag op bronnen met een mening, in plaats van op alle actoren die in het nieuws voorkomen.³³ De politieke affiliaties van deze bronnen of de aanwezigheid van tegenstrijdige standpunten worden vervolgens gebruikt om de diversiteit in standpunten te beoordelen.

Er is geen wetenschappelijke eensgezindheid over de vraag of brondiversiteit als een indicator voor standpuntdiversiteit kan worden beschouwd. Voakes et al. (1996)³⁴ concludeerden dat deze concepten verschillend zijn, terwijl Masini en Van Aelst (2017)²⁷ besloten dat ze vaak samenhangen, waarbij artikelen met bronnen uit verschillende sociale groepen meestal ook een breder scala aan perspectieven bieden. In deze pilot onderzoeken we methoden voor het identificeren en categoriseren van bronnen én standpunten. Bij deze laatste focussen we om pragmatische redenen specifiek op standpunten over coronamaatregelen.

4.2 Methode

De meeste onderzoeken naar diversiteit maken gebruik van kwantitatieve inhoudsanalyse als de primaire methode om de drie bovengenoemde indicatoren van diversiteit te meten. Door de recente en snelle digitalisering openen zich steeds meer mogelijkheden om computationele technieken in te zetten voor het analyseren van nieuwsteksten op grotere schaal en in kortere tijd. Doordat de hoeveelheid beschikbare inhoud exponentieel toeneemt, liggen in deze evolutie wellicht ook nieuwe kansen voor het onderzoek naar pluriformiteit. Gezien het doel van deze studie om een geschikte methode te vinden voor de periodieke analyse van nieuwsinhoud in Nederland nu en in de nabije toekomst, richten wij ons in de eerste plaats op deze computationele technieken. Daarom voorzien we een systematische analyse van verschillende automatische technieken om de drie indicatoren van diversiteit in nieuwsinhoud te meten. We valideren daarbij de uitkomst van de computationele methoden met menselijke coderingen.

Data

Voor deze pilotstudie hebben we artikelen verzameld van NOS.nl, de website van de Nederlandse publieke omroep voor nieuws, sport en evenementen. Volgens de jaarlijkse Mediamonitor^{35,36,37} had NOS.nl tijdens de analyseperiode het grootste bereik in Nederland. Daarnaast biedt de NOS, als publieke omroep, gratis toegang tot nieuwsartikelen zonder betaalmuren, wat de dataverzameling aanzienlijk faciliteert. Om relevante artikelen te verzamelen, gebruikten we de zoekfunctie van de website met de trefwoorden *corona**, *covid**, *sars-cov-2*. We verzamelden de titel, tekst, datum, categorie en, waar van toepassing, trefwoorden. We beperken de analyse tot artikelen tot en met eind mei 2022, toen de laatste coronamaatregelen werden opgeheven ($N=10.662$).

Manuele inhoudsanalyse

Automatische analyse van tekst maakt gebruik van menselijke gecodeerde artikelen, voor het trainen van de modellen en voor de validatie van de resultaten. Voor deze studie hebben we een gouden standaard dataset samengesteld op basis van menselijke coderingen. Een gouden standaard vereist betrouwbare en valide metingen, beoordeeld door overeenstemming tussen codeurs. Om de betrouwbaarheid te waarborgen, hebben we het codeboek iteratief ontwikkeld en verfijnd met tussentijdse testen en feedback. De primaire codeur, een research-masterstudent van de UvA, heeft uitgebreid training gekregen over het codeboek en heeft 796 artikelen gecodeerd. Vervolgens kreeg de secundaire codeur, een onderzoeksassistent van de Universiteit Antwerpen, een identieke training en codeerde die vervolgens 120 artikelen. Om ervoor te zorgen dat de metingen op dezelfde manier worden geïnterpreteerd en dus reproduceerbaar zijn, hebben we Krippendorff's alfa berekend voor alle variabelen. Informatie over het coderingsproces en de resultaten van de betrouwbaarheidsanalyses zijn terug te vinden in de bijlage. Daarnaast beoordeelde één van de onderzoekers de discrepanties tussen de twee codeurs en codeerde deze gevallen handmatig. Deze correcties werden gebruikt voor het evalueren van geautomatiseerde technieken.

Automatische Inhoudsanalyse

Voor de geautomatiseerde analyse naar onderwerpen,

³¹ Steiner, Miriam, Melanie Magin, Birgit Stark, and Stefan Geiß. "Seek and You Shall Find? A Content Analysis on the Diversity of Five Search Engines' Results on Political Queries." *Information, Communication & Society* 25, no. 2 (2022): 217–41.

³² Gans, Herbert J. "Multiperspectival news revisited: Journalism and representative democracy." *Journalism* 12, no. 1 (2011): 3–13.

³³ Magin, Melanie, Birgit Stark, Olaf Jandura, Linards Udris, Andreas Riedl, Miriam Klein, Mark Eisenegger, Raphael Kösters, and Brigitte Hofstetter Furrer. "Seeing the whole picture. Towards a multi-perspective approach to news content diversity based on liberal and deliberative models of democracy." *Journalism Studies* 24, no. 5 (2023): 669–696.

³⁴ Voakes, Kapfer, Kurpius, and Chern, "Diversity in the News" 582–593.

*Om de politieke partijen van de bronnen te identificeren testen we geen modellen, aangezien deze eenvoudig kunnen worden gekoppeld aan bestaande datasets van Nederlandse politieke partijen en politici, zoals die van Wikidata of de Tweede Kamer.

³⁵ Commissariaat voor de Media, *Mediamonitor 2020* (Hilversum: Commissariaat voor de Media, 2020).

³⁶ Commissariaat voor de Media, *Mediamonitor 2021* (Hilversum: Commissariaat voor de Media, 2021).

³⁷ Commissariaat voor de Media, *Mediamonitor 2022* (Hilversum: Commissariaat voor de Media, 2022).

bronnen en standpunten, maken we gebruik van verschillende benaderingen in *Natural Language Processing (NLP)*: *supervised classification*, *topic modelling*, *named entity recognition (NER)* en *text annotation with Large Language Models (LLM's)*. *Supervised classification*, waarbij de computer leert van menselijke annotaties (coderingen), wordt toegepast met twee benaderingen: een statistische methode met een *SVM-classifier* en een *deep learning*-methode met een vooraf getraind taalmodel, *RobBERT-2023*. Deze modellen worden getraind op onze specifieke data voor tekstclassificatie, waarbij *RobBERT* de context van woorden beter begrijpt, wat nauwkeurigere classificaties mogelijk maakt. Daarnaast gebruiken we *topic modeling* om automatisch onderwerpen in teksten te identificeren zonder vooraf gedefinieerde labels. Dit gebeurt via de *BERTopic*-methode, die tekst op paragraafniveau clustert om meerdere onderwerpen per artikel te identificeren. Verder passen we *named entity recognition (NER)* toe om entiteiten zoals mensen, organisaties en locaties uit de teksten te extraheren, waarbij het *Stanza*-model het meest geschikt bleek voor het extraheren van de meeste en beste entiteiten. Ten slotte hebben we generatieve *Large Language Models (LLM's)* getest voor het automatisch annoteren van teksten, met gebruik van open-source modellen zoals *Starling-LM-7B-alfa* en *Mistral-7B-Instruct-v0.2*. Gedetailleerde informatie over deze technieken en over de verschillende stappen in onze analyse is beschikbaar in de bijlage. De uitkomsten van alle analyses zijn opgenomen in de bijlage.

4.3 Monitoring van diversiteit van nieuwsinhoud

In deze pilotstudie hebben we verschillende state-of-the-art automatische tekstanalysemethoden vergeleken om onderwerpen, bronnen en standpunten in nieuwsartikelen te identificeren, met als doel een methodologie te ontwikkelen voor de periodieke evaluatie van de diversiteit in het Nederlandse nieuws. Onze bevindingen tonen aan dat we automatische technieken kunnen gebruiken om de diversiteit van onderwerpen en journalistieke bronnen te meten. Voor het meten van standpunten is deze methode (nog) niet geschikt. Dit komt door de complexiteit en veelzijdigheid van standpunten. Daarom adviseren we om bronnen die informatie en opinie geven via een citaat, de functies en politieke affiliaties van deze bronnen te gebruiken als een breed toepasbaar en valide alternatief voor het evalueren van de diversiteit aan standpunten.

Op basis van onze tests blijkt het voorgetrainde Nederlandstalige *RobBERT*-model het meest geschikt voor het monitoren van inhoudsdiversiteit. Hoewel generatieve *LLM's* zoals *ChatGPT* mogelijk beter presteren, raden we het

gebruik van propriëtaire modellen af vanwege zorgen over datadeling, privacy en kosten. In onze ervaring zien we dat de mediabedrijven vaak terughoudend zijn in het gebruik van deze modellen en wij delen hun bezorgdheid. *Open-source* modellen zijn veilige alternatieven, maar op basis van onze tests presteren deze modellen minder goed dan *RobBERT* in het uithalen van informatie uit Nederlandstalige nieuwsberichten. Bovendien vereisen ze meer rekencapaciteit dan het finetunen van het *RobBERT*-model. Het finetunen van *RobBERT* vereist menselijke gecodeerde data, wat ook kosten met zich meebrengt. Aangezien de technologie snel evolueert en nieuwe, betere open-source *LLM's* en betaalbare rekencapaciteit regelmatig beschikbaar komen, verwachten we dat analyses binnenkort efficiënter kunnen worden uitgevoerd met bijvoorbeeld een Nederlandstalige *LLM*. Onderzoek naar verschillende technieken moet daarom geïntegreerd worden in het proces van deze monitor.

Dit onderzoek kan periodiek uitgevoerd worden als onderdeel van de bestaande Mediamonitor. Het toevoegen van de analyse van de diversiteit in nieuwsinhoud zou de monitor verrijken door waardevolle inzichten te bieden in de media-pluriformiteit in Nederland. Dankzij de geautomatiseerde technieken kan de berichtgeving van complexe vraagstukken, die hoog op de sociale en politieke agenda's staan, periodiek op een efficiënte manier geanalyseerd worden. Om dit onderzoek uit te voeren, is samenwerking met mediaorganisaties van essentieel belang, aangezien dit onderzoek niet kan worden uitgevoerd zonder toegang tot de nieuwsartikelen die relevant zijn voor de analyse. Aan de andere kant dient het delen van gegevens plaats te vinden in veilige omgevingen en moeten de gegevens worden verwerkt volgens de afspraken tussen alle betrokken partijen.

Voor vervolgonderzoek kan een analyse van de nieuwsberichtgeving in Nederland in 2023 uitgevoerd worden, gezien de cruciale rol van mediadiversiteit voor, tijdens en na nationale verkiezingen. Vier onderwerpen, die als de meest prominente op de publieke agenda in 2023 werden gedefinieerd volgens de European Barometer en het Nationaal Verkiezingsonderzoek, kunnen worden geanalyseerd: de kosten van levensonderhoud/stijgende prijzen, het milieu en klimaatverandering, immigratie en huisvesting. Voor elk van deze onderwerpen kunnen de subthema's worden geïdentificeerd om te begrijpen in hoeverre verschillende aspecten aanwezig of prominent waren in de berichtgeving. Vervolgens kunnen de geciteerde bronnen worden onderzocht om een overzicht te geven van wie informatie en opinie deelt in het nieuws en of de statusposities en politieke affiliaties van deze bronnen divers zijn. Dit onderzoek kan fungeren als een pilotstudie voor de periodieke monitoring van de diversiteit in nieuwsinhoud in Nederland.

5 Reflectie op het proces



Het uitvoeren van de pilots bracht moeilijkheden met zich mee en daarom reflecteren wij in deze paragraaf op het proces tot dusver. De volgende onderwerpen komen achtereenvolgens aan bod: de rol van de begeleidingscommissie, de samenwerking met de sector, planning en een vooruitblik naar de toekomst.

Rol begeleidingscommissie

Met de begeleidingscommissie wilden wij de journalistieke praktijk en beleidsmedewerkers betrekken bij de ontwikkeling van de pluriformiteitsmonitor. Ook wilden wij via de begeleidingscommissie de mediabedrijven verzoeken toegang te krijgen tot nieuwsaanbod om de pilots uit te voeren.

Samenwerking met de sector

Met branchevereniging NDP Nieuwsmedia vonden gedurende de uitvoering van de pilots meerdere gesprekken plaats. De gesprekken waren soms inhoudelijk van aard en soms procesmatig. Zowel NDP Nieuwsmedia als het Commissariaat toonden zich bereidwillig maar dit leidde niet altijd tot een bevredigend resultaat voor beide partijen.

Na ruim een jaar is besloten om de begeleidingscommissie uit te breiden met een plek voor DPG Media. De nieuwssector werd hierdoor, naast een plek voor branchevereniging NDP Nieuwsmedia, directer betrokken bij de ontwikkeling van de monitor.

ANP vond dat voor de ontwikkeling van de monitor werd gevraagd om bedrijfsgevoelige informatie van ANP en besloot daarop in eerste instantie niet mee te werken. Dit was voor het Commissariaat geen reden om de inhoud van de pilots aan te passen of te stoppen met de ontwikkeling van de monitor. ANP stelde zich in een latere fase in de ontwikkeling van de monitor wel bereid mee te werken aan pilot 1 en pilot 2.

Het Commissariaat en de betrokken wetenschappers vroegen in een vroeg stadium van de pilots om toegang tot nieuwsartikelen aan de nieuwssector en om toegang tot journalistieke medewerkers. Deze toegang was nodig om de pilots te kunnen uitvoeren. Dit bracht uitdagingen met zich mee die vooraf niet waren voorzien.

Pilot 1 ronden wij af op het moment dat de vragenlijst was ontwikkeld en ging over in de projectfase toen de vragenlijst werd uitgestuurd naar journalistieke medewerkers. Het doel was om de vragenlijst onder een brede groep van journalistieke medewerkers uit te zetten. Echter bleken er in die fase opnieuw vragen te ontstaan over de inhoud van de vragenlijst vanuit een deel van de nieuwssector. Daardoor konden we alleen starten met het uitsturen van de vragenlijst naar journalistieke medewerkers die zijn aangesloten bij de NVJ. Het uitsturen van de vragenlijst naar journalistieke medewerkers van de resterende nieuwssector heeft vervolgens lange tijd stilgelegen. Pas na nieuwe gesprekken tussen de

nieuwssector en de betrokken wetenschapper Mark Boukes werd de vragenlijst uitgezet.

Met de nieuwssector is voor de uitvoering van pilot 2 een overeenkomst gesloten voor de uitwisseling van nieuwsartikelen. Toegang tot nieuwsartikelen voor pilot 3 is slechts gedeeltelijk geslaagd omdat een deel van de nieuwssector deze pilot niet ondersteunde vanwege het onderwerp.

Planning

Het uitvoeren van de pilotfase heeft meer tijd in beslag genomen dan aanvankelijk gedacht. De belangrijkste reden

daarvoor is dat het onderwerp gevoeliger was dan vooraf gedacht en er meer tijd nodig was voor de uitwisseling van nieuwsartikelen en samenwerking met de nieuwssector voor de pilots.

Toekomst

Vanwege de gevoeligheid van het onderwerp en de gewenste duiding vanuit de nieuwssector op de inhoudelijke resultaten, vroeg het Commissariaat aan de begeleidingscommissie om aangesloten te blijven bij het vervolgtraject van deze Monitor pluriformiteit. De begeleidingscommissie stemde hiermee in en blijft betrokken bij de monitor.



6 Reflectie op Monitor pluriformiteit – NDP Nieuwsmedia



Het Commissariaat vroeg alle sectorpartijen uit de begeleidingscommissie om een bijdrage aan deze rapportage. In de reflectie op de Monitor pluriformiteit zetten de leden namens de organisatie die zij vertegenwoordigen uiteen welk belang zij hechten aan de Monitor pluriformiteit.

Geschreven door: Herman Wolswinkel, NDP

Nieuwsmedia

Een pluriform nieuwsaanbod is cruciaal voor onze democratische samenleving. Journalisten die onafhankelijk en op verschillende wijze verslag doen van wat er om ons heen gebeurt, verrijken de blik van de burger. Te midden van alle uitdagingen waarvoor de journalistiek zich gesteld ziet, is het hooghouden van een pluriform nieuwsaanbod voor nieuwsuitgevers een belangrijk doel. NDP Nieuwsmedia, waarin de Nederlandse private nieuwsuitgevers zijn verenigd, maakt zich daar sterk voor en participeert daarom graag in de begeleidingscommissie van de ontwikkeling van een pluriformiteitsmonitor.

Het verwarrende aan het begrip pluriformiteit

Pluriformiteit is een complex begrip, dat door beleidsmakers, politici, toezichthouders, wetenschappers en journalisten verschillend wordt geïnterpreteerd en uitgelegd. De verwantschap met begrippen als pluralisme en pluraliteit leidt al snel tot een spraakverwarring, om nog maar te zwijgen over het internationale debat over pluralism. Dat het Commissariaat voor de Media bewust kiest voor 'diversiteit van de inhoud van het nieuwsaanbod' zet een standaard neer die duidelijk maakt dat

het hier niet over eigenaarschap van nieuwsmedia gaat. Een goed afgebakende pluriformiteitsmonitor kan helpen het debat daadwerkelijk te voeren over de kernvraag: is er voldoende pluriform nieuwsaanbod beschikbaar voor de burger? Dat debat heeft baat bij meer feiten en minder meningen. Daarin kan een goede pluriformiteitsmonitor straks voorzien.

Dat het vergrootglas ligt op de nieuwsinhoud, brengt ook risico's met zich mee. Onderzoek naar journalistieke inhoud mag nimmer aanleiding geven tot (overheids)bemoeienis met die inhoud. De overheid moet altijd op afstand blijven van het inhoudelijke werk van journalisten en juist de vrije pers waarborgen. Als de pluriformiteitsmonitor in de toekomst zou laten zien dat de pluriformiteit onder druk staat, moet dat leiden tot een systeemdiscussie hoe een vitaal journalistiek landschap mogelijk blijft – en beslist níét tot een discussie over de journalistieke inhoud zelf.

Een tweede risico is dat diversiteit van nieuwsaanbod als een te absolute indicator wordt gehanteerd. We leven in een aandachtseconomie waar content op sociale netwerken op steeds meer aandacht kan rekenen en de platforms er alles aan doen om die aandacht van gebruikers zo lang mogelijk vast te houden, om reclame-inkomsten te maximaliseren. De beschikbaarheid van een divers nieuwsaanbod zegt daarom nog niets over de wijze waarop de volledige breedte van het publiek zich nu en straks daadwerkelijk laat informeren. Evenmin zegt het iets over de algoritmen waarmee dat nieuwsaanbod zichtbaar wordt gemaakt voor het bredere publiek. Aanbod, zichtbaarheid en consumptie van het nieuws moeten daarom altijd in samenhang worden geanalyseerd.

Wie de pluriformiteit vormgeven

In het debat over pluriformiteit wordt al snel gekeken naar de grote nieuwsuitgevers, waar in het afgelopen decennium met name de regionale dagbladen een nieuw journalistiek thuis hebben gevonden. Van de totale informatievoorziening waarvan de burger gebruikmaakt, is dat echter maar een select deel. Pluriformiteit krijgt vorm door het geheel aan nieuwsaanbod dat voor het publiek beschikbaar is. Het publiek leest, luistert of kijkt niet alleen bij private nieuwsmedia en publieke omroepen, maar ook steeds vaker via sociale netwerken, hoe discutabel de kwaliteit en onafhankelijkheid van dit laatste type content vaak ook zijn. In de ontwikkeling van de pluriformiteitsmonitor zijn private en publieke nieuwsmedia wél betrokken, maar de content op sociale media niét. Dat als gevolg alleen een selectie van de informatievoorziening betrokken is, is een gemis. Dat geldt zeker waar het jongere publieksgroepen betreft, die nieuwsmedia minder goed kunnen bereiken, zoals ook het Commissariaat voor de Media recent concludeerde.³⁸

Meewegen van de ontwikkelingen in het nieuwslandschap

Consolidatie in de dagbladsector lijkt een belangrijke aanjager te zijn geweest voor het debat over pluriformiteit. Als je het vluchtig bekijkt, is dat navolgbaar. Maar wie de echte context betreft, kan niet anders dan concluderen dat consolidatie voor in het bijzonder regionale nieuwstitels een belangrijk middel is geweest om die nieuwstitels toekomstbestendiger te maken. Daarmee heeft consolidatie juist bijgedragen aan het behoud van pluriformiteit. De sterke Nederlandse traditie van redactiestatuten, die de onafhankelijkheid van redacties waarborgen, is daarbij altijd beschermd en verankerd in de nieuwsbedrijven (dat was, is en blijft overigens zo). Consolidatie is daarbij geen doel op zich geweest, maar een middel om journalistieke titels op de langere termijn te behouden en daarmee de grondstof voor meningsvorming te blijven leveren. De noodzaak daarvoor is ook economisch glashelder: afgezet tegen de situatie in 2006 verliezen nieuwsuitgevers jaarlijks 500 miljoen euro aan advertentie-inkomsten aan de mondiale techbedrijven, wat ruim meer is dan het opgetelde salaris van alle journalisten die op hun redacties werkzaam zijn. En wat staat ons te wachten met de in potentie disruptieve ontwikkelingen rond generatieve AI of het afbouwen van contentmoderatie door de platforms? Het is daarom onontbeerlijk dat resultaten uit de te ontwikkelen pluriformiteitsmonitor straks ook met die bril worden bekeken. In welke context vinden de onderzochte pluriformiteitsontwikkelingen plaats? Cruciaal is dat uitgeefstrategieën ook in dit licht worden gezien en als context dienen bij het meten van pluriformiteit.

Neem bijvoorbeeld de beperkte praktijk waarbij nieuwsartikelen tussen titels worden uitgewisseld. Redacties doen

dat om hun schaarse capaciteit in te kunnen zetten op journalistieke producties waarmee ze zich daadwerkelijk kunnen onderscheiden – en waarvoor lezers bereid zijn een abonnement af te sluiten. Voor het behoud van pluriformiteit is dat een goed te rechtvaardigen werkwijze, maar in de beeldvorming zijn nieuwsuitgevers met deze werkwijze juist de gebeten hond. Context van de uitgeefpraktijk en dilemma's van nieuwsmedia verdienen daarom een plaats in een pluriformiteitsmonitor. Bovendien is een voortdurend gesprek met nieuwsbedrijven nodig over de uitgeefpraktijk.

Geeft het een getrouw beeld van de journalistieke praktijk?

Een goede weergave van de journalistieke praktijk is van het grootste belang bij de ontwikkeling van de pluriformiteitsmonitor. Gaandeweg de ontwikkeling ontstaat gemakkelijk afstand tussen de oorspronkelijke vraag vanuit de Tweede Kamer, de vertaalslag die wetenschappers maken voor hun onderzoek en de daadwerkelijke journalistieke praktijk. Het is belangrijk dat te voorkomen. NDP Nieuwsmedia ziet het mede als haar taak om in de begeleidingscommissie de afstand met de journalistieke praktijk zo klein mogelijk te houden.

Tijdens de ontwikkeling van de monitor nemen onderzoekers begrijpelijkerwijs afslagen. Steeds moet de vraag op tafel komen of dan nog steeds gemeten wordt wat beoogd werd gemeten te worden. Dat klinkt vanzelfsprekend, maar is het niet. Bijvoorbeeld bij Pilot 1 is kennis van de journalistieke mores die op de redactievloer geldt belangrijk, om journalisten de juiste vragen te kunnen stellen over pluriformiteit. Zonder een gedegen beeld van hoe journalistieke besluitvorming op onafhankelijke redacties tot stand komt, kan al snel een verkeerde conclusie over individuele redactionele vrijheid worden getrokken. Pilot 2, het onderzoek naar unieke en overlappende berichtgeving, is een ander voorbeeld: als uit de onderzoeksresultaten enige mate van overlap blijkt, hoeft dat helemaal niet te wijzen op een gebrek aan pluriformiteit. Artikelen van verschillende nieuwsmedia die dezelfde nieuwsgebeurtenis beschrijven vertonen logischerwijs overeenkomsten, maar zijn juist een indicatie van pluriformiteit. Kortom, wordt steeds gemeten wat beoogd wordt gemeten te worden?

Het is daarom een goede zaak dat het Commissariaat voor de Media de ontwikkeling van de monitor met zorgvuldigheid ter hand neemt en daarbij altijd de context van de journalistieke praktijk in ogenschouw wil nemen. Alleen dan kunnen de onderzoeksresultaten een getrouw beeld geven van de journalistieke praktijk.

³⁸ [Jongeren, nieuws en sociale media – Commissariaat voor de Media](#)

7 Reflectie op Monitor pluriformiteit – WRR



Het Commissariaat vroeg alle sectorpartijen uit de begeleidingscommissie om een bijdrage aan deze rapportage. In de reflectie op de Monitor pluriformiteit zetten de leden namens de organisatie die zij vertegenwoordigen uiteen welk belang zij hechten aan de Monitor pluriformiteit.

Geschreven door: Alyt Damstra, WRR

Media zijn onmisbaar in een democratische samenleving. Dat is de openingszin van het rapport *Aandacht voor media. Naar nieuwe waarborgen voor hun democratische functies*, dat de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) in 2024 publiceerde.³⁹ Die onmisbare rol komt tot uitdrukking in drie fundamentele functies. In de eerste plaats informeren media de samenleving. Ze brengen nieuws over relevante vraagstukken en zorgen voor duiding. Daarnaast monitoren ze de macht. Ten slotte bieden media een podium voor publiek debat. Ze geven ruimte aan de vele ideeën en opvattingen die de samenleving rijk is. Langs deze functies dragen media bij aan de democratische meningsvorming. Burgers kunnen op basis van kennis – over wat er in het land gebeurt en hoe daarover wordt gedacht – hun eigen mening vormen en deze omzetten in politieke betrokkenheid, voorkeuren en stemgedrag.

Maar hoe media aan deze functies invulling geven is geen uitgemaakte zaak. Media kunnen het democratisch proces in positieve zin beïnvloeden, maar ze kunnen het evenzogoed ondermijnen of ontwrichten, bijvoorbeeld door onware informatie te verspreiden. In het WRR-rapport wordt verwezen naar vier criteria die bijdragen aan een adequate invulling van deze functies. Media dienen onafhankelijk, betrouwbaar en toegankelijk te zijn en ze moeten ter verantwoording kunnen worden geroepen. Media zijn immers zelf óók een factor van macht. Als aan deze voorwaarden wordt voldaan, is de kans groot dat het media-aanbod als geheel ook voldoet aan misschien wel de belangrijkste normatieve vereiste: pluriformiteit.

Externe en interne pluriformiteit

Pluriformiteit is een wezenskenmerk van een democratisch functionerend mediasysteem. Het belang dat aan pluriformiteit wordt gehecht komt onder meer tot uitdrukking in de Mediawet, waarin veelvuldig wordt verwezen naar de pluriformiteit van het aanbod, de pluriformiteit van de informatievoorziening en de pluriformiteit van de pers.⁴⁰ Pluriformiteit is geen eenduidig begrip en kent verschillende dimensies. Zo verwijst interne pluriformiteit naar de diversiteit aan inhoud en standpunten die door een enkel medium wordt aangeboden. Externe pluriformiteit gaat over de structuur van mediamarkten als geheel. Hoe groter het aantal mediapartijen

³⁹ <https://www.wrr.nl/publicaties/rapporten/2024/10/03/aandacht-voor-media>

⁴⁰ <https://wetten.overheid.nl/BWBR0025028/2024-01-01>; ook internationaal is het pluriformiteitsbegrip wettelijk verankerd, bijvoorbeeld in het Europees verdrag tot bescherming van de rechten van de mens en de fundamentele vrijheden (EVRM)

dat op de markt opereert, hoe groter de kans dat zij een pluriform aanbod van diverse inhoud en standpunten voortbrengen.⁴¹ Hieruit volgt dat een sterk geconcentreerde mediamarkt met slechts een handjevol grote spelers in potentie gevolgen heeft voor de pluriformiteit van het aanbod. In ieder geval waar het de externe pluriformiteit betreft. Of en in hoeverre dit ook gevolgen heeft voor de interne pluriformiteit valt vervolgens te bezien, dat hangt er immers maar vanaf hoe het aanbod van individuele titels binnen deze markt tot stand komt. Bovendien kan concentratie er ook toe leiden dat individuele titels juist voort kunnen bestaan, wat dan weer ten goede komt aan de diversiteit van het aanbod.

In Nederland is er al jaren sprake van een steeds verdergaande concentratie van de mediamarkt. Op het moment van schrijven zijn er vijf grote mediaorganisaties die op één of meerdere productmarkten domineren: DPG Media en Mediahuis (dagbladen en radio), NPO, RTL en Talpa (radio en televisie).⁴² De hoge mate van concentratie vormde in 2021 aanleiding voor de motie Kwant, waarin de regering werd verzocht om in kaart te brengen op welke manier de pluriformiteit van het Nederlandse media-aanbod beter beschermd kan worden.

Belang van de Monitor pluriformiteit

Het Commissariaat voor de Media is vervolgens met deze taak belast, wat concreet resulteerde in de start van de pilot Monitoring Pluriformiteit.

De WRR onderschrijft het belang van deze monitor. De democratische functies van media bestaan bij de gratie van pluriformiteit. Of het nu gaat over de informatievoorziening of het publieke debat, burgers dienen toegang te hebben tot een diversiteit aan aanbieders en aanbod. De relevantie van deze monitor is echter niet enkel gelegen in de toenemende concentratie van de markt. In het eerdergenoemde WRR-rapport wijst de raad op de platformisering van de online publieke ruimte en de grote gevolgen die dit heeft voor de positie

van Nederlandse mediaorganisaties. Digitale advertentie-inkomsten die wegvloeien naar het buitenland, afnemend mediagebruik door jongere generaties, afhankelijkheid van de infrastructuur die platforms bieden. Het zijn juist deze ontwikkelingen, waar de toenemende concentratie deels een reactie op is, die een systematische monitoring van de kwaliteit van het Nederlandse media-aanbod zeer relevant maken.

Invulling geven aan het meten van pluriformiteit

De keuze voor een langer lopend traject met drie wetenschappelijke pilots en meerdere tussenevaluaties doet recht aan de uitdagingen die het operationaliseren van pluriformiteit met zich meebrengt. Pluriformiteit is immers niet in een enkele meting te vangen. De eerste pilot richt zich op diegenen die het Nederlandse media-aanbod gezamenlijk tot stand brengen: journalisten. Door middel van vragenlijsten wordt zicht verkregen op de samenstelling en diversiteit van deze beroepsgroep en de context waarbinnen journalisten hun werk doen. Door de vragenlijst na verloop van tijd opnieuw uit te sturen, kunnen ook veranderingen over tijd in beeld worden gebracht. De tweede pilot kijkt naar (het ontbreken van) pluriformiteit door de mate van overlap in het nieuwsaanbod in kaart te brengen. Het gaat hier dus niet zozeer om diversiteit in termen van zienswijzen, maar om de originaliteit van content. Hiermee adresseert deze opzet een veelgehoorde zorg rond concentratie: het toenemen van kopij-uitwisseling tussen titels die tot hetzelfde concern behoren. De derde pilot richt zich op pluriformiteit in het media-aanbod zelf, door te kijken naar diversiteit binnen berichtgeving, onder meer van bronnen en standpunten.

De WRR schetste de voorwaarden waaronder media invulling kunnen blijven geven aan hun democratische functies. Met de Monitor pluriformiteit kan het Commissariaat voor de Media de diversiteit van het aanbod monitoren. Hoe snel de informatie-omgeving ook mag veranderen, het democratisch belang van de journalistiek blijft onveranderd groot.

⁴¹ Kooter (2023)

⁴² <https://www.cvdm.nl/nieuws/mediamonitor-2024-verschuiving-in-mediagebruik-naar-internationale-platformen-vormt-risico-voor-geinformeerdheid-van-burgers/>

8 Reflectie op Monitor pluriformiteit – NVJ



Het Commissariaat vroeg alle sectorpartijen uit de begeleidingscommissie om een bijdrage aan deze rapportage. In de reflectie op de Monitor pluriformiteit zetten de leden namens de organisatie die zij vertegenwoordigen uiteen welk belang zij hechten aan de Monitor pluriformiteit.

Geschreven door: Thomas Bruning, NVJ

Het belang van pluriformiteit en monitoring

Het mag duidelijk zijn dat de NVJ als beroepsvereniging van journalisten een groot belang hecht aan een pluriform journalistiek landschap.

Uniforme journalistiek, het andere uiterste van de weegschaal, zou het wezen en de meerwaarde van het vak onderuithalen.

De essentie van ons vak is het op onafhankelijke wijze informeren van het publiek, waarbij ernaar wordt gestreefd alle relevante feiten en gezichtspunten boven tafel te krijgen. Een streven naar waarheidsvinding, noemen we dat. Maar alhoewel de feiten eenduidig kunnen zijn, is de weging daarvan, het perspectief erop en de perceptie daarvan zeker niet eenduidig.

Het is dus nogal een risico als we die taak in handen leggen van een eenvormige groep professionals, die daarbij dezelfde instructies en kaders krijgt.

Er bestaat immers niet zoiets als één juiste invalshoek of één alwetende bron, waar hoor en wederhoor gehaald kan worden.

Precies daar ligt het gevaar van vergaande vormen van mediaconcentratie die de journalistiek kwetsbaar en manipulatief kunnen maken.

Juist de veelkleurigheid aan stemmen, invalshoeken en onderwerpen, dat een pluriform medialandschap garandeert, zorgt ervoor dat de burger goed geïnformeerd wordt, maar ook dat de journalistiek zelf op de meest gezonde wijze gecontroleerd wordt, namelijk door het publiek en door media onderling.

Maar hoe stel je nu op de meest objectieve wijze vast dat een land, regio of stad over een pluriform aanbod beschikt?

De pilot voor een eerste “Monitor pluriformiteit” van het Commissariaat van de Media doet daar een eerste poging toe.

Pluriformiteit gaat zeker niet alleen om het aantal media-bedrijven dat in een land of regio actief is. Binnen een mediabedrijf kunnen immers heel wel een reeks aan onafhankelijke titels worden ondergebracht, die op geheel eigen en onafhankelijke wijze het nieuws brengen. Maar dan moet de onafhankelijkheid van die titel zowel in theorie als in praktijk wel worden nageleefd.

In pilot 1 van de monitor wordt dan ook terecht gekeken naar de praktijk achter de redactionele titels. Wat is er vastgelegd over de onafhankelijkheid en in hoeverre wordt die theoretische werkelijkheid nageleefd.

Met name in de regio zou dit waardevolle data kunnen opleveren. Daar is immers het aantal concurrerende titels zeer beperkt. Naast een regionale krant is er meestal alleen de regionale of streekomroep, die voor een andere invalshoek of onderlinge controle kan zorgen.

Daarbij is de vraag relevant of de betrokken redacties werkelijk vrij hun keuzes kunnen maken. Beschikken ze over een eigen budget, redactiestatuut en kunnen ze vrij van overheids- of bedrijfsbemoeienis hun verhalen publiceren? En hoe ontwikkelt het budget en de omvang van de redactie zich over de jaren?

Ook op landelijk niveau is dit onderdeel van de pilot relevant. Vooral bij de publieke omroep vermoeden we een pijnpunt: de vaak zo geroemde interne pluriformiteit binnen de publieke omroep is weinig waard als redacties instructies van hun verenigingsvoorzitter moeten volgen of redactionele zeggenschap feitelijk ontbreekt.

Er is dus een verschil tussen theorie en praktijk. Het bestaan van een redactiestatuut sluit niet uit dat er informele druk bestaat om bepaalde verhalen juist wel of niet te publiceren. Daardoor kunnen invalshoeken en onderwerpkeuzes eenvoudig worden.

De onafhankelijkheid van een redactie is relatief als bepaalde artikelen altijd voorrang krijgen omwille van de clicks of een beproefd format.

En uiteindelijk bepaalt het redactionele budget van een titel de werkelijke onafhankelijkheid. Een statuut beschermt immers wel tegen inhoudelijke bemoeienis, maar niet tegen sanering of opheffing van een titel.

Ook is relevant wat er wordt betaald aan de medewerkers en freelancers, die aan deze titels meewerken. Stelt de geboden vergoeding hen in staat werkelijk onafhankelijk te werken? Of beschouwen ze de publicaties vooral als een marketing-instrument voor het binnenhalen van commerciële klussen in hetzelfde aandachtsgebied?

En wat als blijkt dat sommige titels nog slechts façades zijn en de inhoud vooral bestaat uit veredelde persberichten of centraal aangeleverde kopij?

Daarom is het goed om niet alleen het aantal titels, maar ook de redactieomvang en freelancebudgetten per titel in beeld te brengen, en de mores en afspraken binnen de redactie van deze titels te onderzoeken.

Maar daarvoor is het dan wel noodzakelijk dat de (grote!) mediabedrijven hun medewerking verlenen aan dit soort onderzoek.

Het is ronduit teleurstellend dat de uitgevers verenigd binnen NDP en de belangrijkste private speler in de regio, DPG, aanvankelijk niet volledig wilden meewerken aan deze eerste pilot en de in dit verband zeer relevante NVJ-arbeidsmarktmonitor, want juist de data van deze bedrijven kunnen een beter beeld geven op de praktijk.

Een volgende pluriformiteitsmonitor zou kunnen aantonen dat regionale journalistiek bij een groot bedrijf als DPG goed gewaarborgd is, maar door aanvankelijk zo defensief op de onderzoeksvragen in de pilot en in de NVJ-arbeidsmarktmonitor te reageren, lijkt nu het omgekeerde het geval.

In pilot 2 wordt getracht de mate van uniciteit van berichtgeving in kaart te brengen. Dat wordt in deze fase op landelijk niveau onderzocht, waar de keuze en verscheidenheid nog best groot is. Ik verwacht daar dan ook weinig revolutionaire resultaten.

Een dergelijk onderzoek zou vooral op regionaal of lokaal niveau tot verhelderende inzichten kunnen leiden. Daar is de kans veel groter dat een artikel met een bepaald nieuwsfeit daar min of meer hergebruikt wordt door andere titels en dus niet voor werkelijke verscheidenheid richting het publiek zorgt. Er is regionaal immers veel minder journalistieke menskracht om zelf naar eigen bronnen en invalshoeken op zoek te gaan.

Eerder heeft het Stimuleringsfonds voor de Journalistiek op regionaal niveau het aantal uniek gepubliceerde artikelen per titel in kaart gebracht. Daar zou in pilot 2 op kunnen worden voortgebouwd, waarbij aangetekend dat dit arbeidsintensief is en daarom slechts steekproefsgewijs kan worden uitgevoerd. En daarbij hoeft de mate van originaliteit weinig te zeggen over de journalistieke verscheidenheid en diepgang.

In de derde pilot van het Commissariaat wordt getracht de diversiteit in standpunten en invalshoeken in kaart te brengen binnen één groter onderwerp.

Dit zou iets kunnen zeggen over de mate waarin de journalistieke titels, ondanks een goed gewaarborgde onafhankelijkheid, toch de neiging hebben om dezelfde bronnen te gebruiken en eenzelfde soort vragen te stellen en zo tot eenzelfde soort verhalen komen. Grote vraag is welk onderwerp zich goed leent voor een dergelijk onderzoek. Bij een ramp of groot incident is het aantal relevante bronnen immers beperkt en de verslaglegging rondom corona, die voor de pilot als onderwerp is gekozen, is dan weer zo atypisch, dat er ook moeilijk bredere conclusies uit te trekken vallen.

Uiteindelijk staat of valt het mijns inziens met een andere vraag, de tot op heden minst gemeten en misschien juist wel

meest relevante parameter voor gezonde pluriformiteit; het aantal actieve journalisten per aandachtsgebied of regio.

Journalistiek is immers mensenwerk, waarbij handen aan het bed het verschil maken. Natuurlijk kan je achter je laptop en met je telefoon bergen verzetten en veel kopij produceren. De hoeveelheid artikelen van zelfs de kleinste lokale titel kan indrukwekkend zijn. Maar de publieke meerwaarde van deze brei aan berichten is veelal beperkt, als de tijd ontbreekt om daadwerkelijk met bronnen en het publiek in contact te treden.

Uiteindelijk liggen de meest relevante verhalen op straat, ontstaan ze door gedegen onderzoekwerk en gesprekken met mensen, die niet zo mondig of makkelijk bereikbaar zijn. Dat vraagt om menskracht, die nu te veel wordt ingezet om de hele dag door te produceren vanuit de kantoortuin, terwijl het journalistiek in contact staan met een wijk, stad of regio meer publieke waarde en verbinding oplevert en ook een veelkleuriger beeld verschaft.

En met als waardevolle bijvangst meer begrip voor de rol die de journalistiek kan spelen binnen een democratische samenleving.

Bij een beperkt aantal journalisten in een stad of regio weten we zeker dat bepaalde nieuwsfeiten, misstanden of delen van die stad of regio nooit aandacht krijgen, laat staan dat daarover langs meerdere invalshoeken en bronnen verhalen worden verteld.

Zoals hierboven betoogd is deze parameter, het aantal actieve journalisten, tot op heden maar matig onderzocht en ontbrak zij ook in de pilot van de eerste pluriformiteitsmonitor. Met de recent gepubliceerde NVJ Journalistieke arbeidsmarktmonitor heeft de NVJ een eerste ruwe poging gedaan om de aantallen in kaart te brengen, maar bij het meer nauwkeurig rubriceren van de aantallen vaste en freelancemedewerkers per titel, stuiten we bij nogal wat uitgevers op de nodige onwil.

Freelance/redactiebudgetten en redactie-omvang op titelniveau zouden volgens uitgevers concurrentiegevoelige informatie betreffen.

Dat is jammer, want juist deze gegevens zouden het beeld van een kwakkelende regionale journalistiek wel eens kunnen nuanceren. Zo zijn er regio's, waar de journalistiek best goed lijkt te zijn vertegenwoordigd, zeker met de aanvullende gelden, die de overheid beschikbaar heeft gesteld voor streekomroepen en onderzoeksjournalistiek.

Maar er zijn ook regio's, dorpen en stadsdelen waar de controlerende en verslagleggende rol echt is uitgehold.

Er is ons het begin van inzicht beloofd, althans door de grootste speler in de markt, DPG, die dit jaar voor het eerst een journalistiek jaarverslag wil publiceren, waarin dit soort zaken in de regio's waar zij actief is, wel wordt vrijgegeven.

Dat lijkt ons een goede zaak, want alleen dan kunnen we werkelijk zien waar de journalistiek en pluriformiteit een impuls verdient en waar het publiek echt onderbedeeld wordt.

Deze gegevens kunnen wat ons betreft als vertrekpunt worden gebruikt in een volgende pluriformiteitsmonitor, waar de aantallen journalisten en de onderliggende diversiteit meer centraal komen te staan.

Maar daarbij dient de sector in onze optiek minder terughoudend te zijn in het verschaffen van data, net zo goed als zij dat ook verwacht van de publieke partijen, die zij dagelijks de maat nemen.

Het kan ertoe leiden dat we een beter beeld krijgen van het journalistieke landschap in Nederland; de kracht die we moeten koesteren en faciliteren en de zwaktes, die we onder ogen moeten zien en moeten aanpakken.

Dankwoord

Velen hebben bijgedragen aan de uitvoering van de pilotonderzoeken en de totstandkoming van dit rapport. Wij vinden het gepast om iedereen daarvoor op deze plek te bedanken.

Wetenschappers

- Mark Boukes, Universitair hoofddocent
Communicatiewetenschap, Universiteit van Amsterdam
en verantwoordelijk voor pilotonderzoek 1 'Ervaringen van
pluriformiteit onder journalisten'
- Nel Ruigrok, Mediawetenschapper, verbonden aan de
Erasmus Universiteit, eigenaar van Nieuwsmonitor B.V. en
verantwoordelijk voor pilotonderzoek 2 'Unieke en over-
eenkomstige berichtgeving'
- Elif Kilik, promotie onderzoeker, junior docent Politieke
Communicatie en Journalistiek, Universiteit Antwerpen,
Universiteit van Amsterdam en medeverantwoordelijk
voor pilotonderzoek 3 'Diversiteit van thema's, bronnen en
standpunten'
- Knut De Swert, Universitair docent Politieke Communicatie
en Journalistiek, Universiteit van Amsterdam en mede-
verantwoordelijk voor pilotonderzoek 3 'Diversiteit van
thema's, bronnen en standpunten'

Begeleidingscommissie

- Alyt Damstra, WRR
- Herman Wolswinkel, NDP Nieuwsmedia
- Leonie Koot-Warris, NPO
- Maarten Glorie, Ministerie van Onderwijs, Cultuur en
Wetenschap
- Paul Hek, NPO
- Philippe Remarque, DPG Media
- Thomas Bruning, NVJ

Overige betrokkenen

- Bastiaan Toering, NPO
- Charlotte Wolff, Ministerie van Onderwijs, Cultuur
en Wetenschap
- Corine de Vries, Mediahuis
- Erik Schrijvers, WRR
- Ernest Slot, Ministerie van Onderwijs, Cultuur en
Wetenschap
- Esmée Dijk, ACM
- Freek Staps, ANP
- Gijs Hesselink, Ministerie van Onderwijs, Cultuur en
Wetenschap
- Giselle van Cann, NOS
- Jasja Pen, RPO
- Joost Ramaer, o.a. Follow the Money
- Karlijn Goossen, NPO
- Leoni van de Schraaf, ANP
- Melanie Groenendijk, ACM
- Nienke Venema, Stichting Democratie en Media
- Peter Smet, SVDJ
- Quint Kik, SVDJ
- Rens Vliegenthart, Universiteit van Wageningen
- Rien van Beemen, Mediahuis
- Stefan Heijndael, NDP Nieuwsmedia
- Theresa Seipp, IViR
- Thomas Koole, NPO
- Wilma Haan, NOS
- Alle betrokken collega's van het
Commissariaat voor de Media

Bijlage 1

Pilot 3, Diversiteit van thema's, bronnen en standpunten: coderingsproces en resultaten van de betrouwbaarheidsanalyse

Om de betrouwbaarheid van onze metingen te evalueren, berekenen we Krippendorff's alfa, een veelgebruikte maat voor het meten van de overeenstemming tussen codeurs. Een alfa-score boven 0,80 is vereist voor consistente interpretatie, terwijl scores boven 0,67 acceptabel worden beschouwd. Bij coderingen van complexe concepten kan een α van 0,60 nog acceptabel worden gezien.⁴³

Onderwerpen

De codering begon met het bepalen of de coronapandemie het hoofdonderwerp van het artikel was. Dit werd gedefinieerd als het onderwerp het grootste deel van de tekst wordt besproken. Alleen artikelen met coronapandemie als hoofdonderwerp werden opgenomen in de volgende analyses. Voor deze variabele was Krippendorff's alfa $\alpha=0,89$, wat wijst op een hoge betrouwbaarheid tussen de primaire en secundaire codeur. Van de 120 artikelen in de betrouwbaarheidsdata werden er 76 gecodeerd als voornamelijk over de coronapandemie, en deze artikelen zijn meegenomen in de verdere analyses.

Codeurs beoordeelden veertien vooraf gedefinieerde subcategorieën, waarbij een subcategorie als substantieel besproken werd gemarkeerd als deze in minimaal twee zinnen werd besproken. Deze veertien subcategorieën zijn geselecteerd op basis van eerder onderzoek naar berichtgeving over pandemieën en verder verfijnd via een pilotstudie van het codeboek. Per artikel konden meerdere subcategorieën worden gecodeerd.

De betrouwbaarheidsscores voor subcategorieën varieerden van $\alpha=0,48$ tot $\alpha=0,93$, wat aangeeft dat sommige subcategorieën moeilijker te coderen waren dan andere. Subcategorieën met een betrouwbaarheid lager dan $\alpha=0,67$ zijn nader onderzocht. Vergelijkingen met door onderzoekers gecorrigeerde coderingen brachten inconsistenties aan het licht: beide codeurs maakten verschillende fouten, voornamelijk door onzorgvuldigheid. Twee subcategorieën – de maatschappelijke gevolgen van de pandemie en de impact op mentale gezondheid – zijn samengevoegd vanwege de lage coderingsfrequentie van de tweede categorie en hun

inhoudelijke overlap. De samengevoegde subcategorie behaalde een acceptabele betrouwbaarheid ($\alpha=0,67$ tussen codeurs; $\alpha=0,72$ tussen de primaire codeur en de onderzoeker). De overige 12 subcategorieën zijn gebruikt voor verdere analyse. Ondanks de lage betrouwbaarheid tussen codeurs werden geen subcategorieën uitgesloten van de analyses. De betrouwbaarheid tussen de primaire codeur en de onderzoeker was hoog, en we gebruikten de coderingen van de primaire codeur voor training en de coderingen van de onderzoeker om te testen. De subcategorie over misinformatie, desinformatie en complottheorieën werd echter wel uitgesloten, omdat deze slechts zelden voorkwam in de coderingen van de onderzoeker ($N=2$).

Bronnen

Informatiebronnen zijn geanalyseerd via actoren, gedefinieerd als personen en organisaties die geciteerd worden in het nieuwsartikel. De primaire codeur identificeerde 329 actoren, de secundaire codeur 301. Om de betrouwbaarheid te beoordelen, vergeleken we het aantal geciteerde actoren per artikel na afstemming van de coderingen. Krippendorff's alfa toonde een goede betrouwbaarheid voor organisaties ($\alpha=0,73$) en een hogere voor individuen ($\alpha=0,86$). Voor de betrouwbaarheid van actorniveau analyseerden we gegevens van 265 actoren, gecodeerd door beide codeurs, ingedeeld in 17 functiecategorieën op basis van beroep, titel of rol. Deze categorieën, gebaseerd op eerder onderzoek, omvatten onder omvatten Nederlandse politieke/staatspersonen, internationale politieke organisaties, justitiële actoren, wetenschappelijke/medische entiteiten, publieke/semi-publieke organisaties, NGO's, religieuze instellingen, bedrijven, journalisten, mediapersoon en burgers. Voor niet-geclassificeerde actoren was een "overige"-categorie beschikbaar. De functiecategorieën vertoonden een goede betrouwbaarheid ($\alpha=0,77$).

Daarnaast hebben de codeurs ook aangeven bij welke politieke partij een actor hoorde, als deze viel binnen de functiecategorie politieke actoren. Deze variabele werd gecodeerd met een zeer hoge betrouwbaarheid ($\alpha=0,94$). Ten slotte gaven de codeurs per actor aan of deze iets vermeldde

⁴³ De Swert, Knut. "Calculating inter-coder reliability in media content analysis using Krippendorff's Alfa." *Center for Politics and Communication* 15, no. 1-15 (2012): 3.

of besprak over coronamaatregelen. Deze variabele fungeerde als een filter om te bepalen welke actoren moesten worden meegenomen in de analyse van diversiteit naar standpunten en vertoonde een hoge betrouwbaarheid ($\alpha=0,84$).

Standpunten

In dit onderzoek werden standpunten gedefinieerd als steunend of kritisch ten opzichte van coronamaatregelen. De codering vond plaats per maatregel en per actor, zodat meningen over meerdere maatregelen per bron konden worden vastgelegd. Codeurs bepaalden voor elke maatregel of de bron een steunend, kritisch of neutraal standpunt innam. Voor de analyse werden deze coderingen per bron samengevoegd tot binaire variabelen: een bron werd als steunend (1) gemarkeerd als deze ten minste één maatregel steunde, en als kritisch (1) als deze ten minste één maatregel bekritiseerde.

In vergelijking met de codering van onderwerpen en actoren bleek het coderen van standpunten een uitdagende taak. Codeurs kwamen niet goed overeen over de vraag of een bron steun of een positieve mening gaf over de maatregelen ($\alpha=0,52$) of kritiek of een negatieve mening gaf ($\alpha=0,60$). Bij de onderzoekscorrecties toonde verdere analyse van de meningsverschillen aan dat de lage alfa-scores te wijten waren aan onoplettendheid of verkeerde interpretaties van de coderingsregels. De onderzoekscorrecties kwamen grotendeels overeen met de standpuntcoderingen van de primaire codeur, met hoge alfa-scores voor steunende standpunten ($\alpha=0,89$) en kritische standpunten ($\alpha=0,81$). Hierdoor werden de coderingen van de primaire codeur geschikt geacht voor het trainen en valideren van geautomatiseerde technieken. De coderingen van de onderzoeker werden als testdata gebruikt om de modellen te vergelijken.

Uitleg van de verschillende toegepaste methodologieën en de verschillende taken in de analysepijplijn

Supervised classification is een *machine learning*-techniek waarbij de computer wordt getraind met voorbeelden die al door mensen zijn voorzien van de juiste categorieën (*labels*). De computer leert vervolgens patronen in de tekst te herkennen die horen bij een bepaalde categorie. We implementeren twee technieken om teksten te classificeren:

een statistische benadering met een *SVM* (*Support Vector Machine*) -*classifier* en een *deep learning*-benadering met een vooraf getraind taalmodel. Voor de *SVM* verwerken we de teksten door ze op te schonen en vectorrepresentaties te genereren met *Term Frequency-Inverse Document Frequency* (*Tf-Idf*), waarbij minder voorkomende, maar meer onderscheidende woorden zwaarder worden gewogen. Vervolgens wordt de *SVM* getraind om deze patronen te herkennen en toe te passen op nieuwe teksten voor verschillende analyses in onze studie. Ten tweede maken we gebruik van een vooraf getraind *transformer*-model voor onze *deep learning*-benadering. *Transformer*-modellen, zoals *BERT*⁴⁴ en *RoBERTa*⁴⁵, creëren betere tekstrepresentatie door rekening te houden met de context van woorden, wat zorgt voor nauwkeuriger begrip van de betekenis. Deze modellen worden vooraf getraind op grote tekstverzamelingen en vervolgens gefinetuned voor specifieke taken. Wij gebruiken *RobBERT-2023*, een *RoBERTa*-gebaseerd Nederlands taalmodel, oorspronkelijk getraind in 2019⁴⁶ en geüpdatet in 2022 en 2023 voor betere prestaties, inclusief bij tekstclassificatie gerelateerd aan de coronapandemie.^{47,48} Voor elk classificatieprobleem fine-tunen we *RobBERT-2023* met onze trainingdata.

Topic modeling is een techniek van *unsupervised machine learning* voor het identificeren van onderwerpen in grote verzamelingen teksten. *Unsupervised learning* vereist geen training met menselijke labels zoals *supervised* modellen, maar identificeert patronen door de numerieke representaties van woorden en zinnen in de nieuwsteksten te analyseren. Deze data-gedreven aanpak heeft geen vooraf gedefinieerde onderwerpen nodig, wat de kans op onderzoeker bias vermindert, maar beperkt het model tot het extraheren van alleen de onderwerpen die aanwezig zijn in de dataset, zonder inzicht te geven in onbesproken onderwerpen. Voor *topic modeling* maken we gebruik van *BERTopic*⁴⁹ met gebruik van *RobBERT Dutch sentence transformer*⁵⁰, die creëert numerieke representaties van tekst op zinsniveau. *BERTopic* gebruikt deze representaties voor hiërarchische clustering, waarbij elk cluster een onderwerp vertegenwoordigt. Aangezien *BERTopic* het meest waarschijnlijke onderwerp aan elk document toewijst, hebben we het model op paragraafniveau uitgevoerd in plaats van op artikelniveau, waardoor meerdere onderwerpen voor elk artikel toegewezen kunnen worden. De output van

⁴⁴ Devlin, Jacob. "BERT: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

⁴⁵ Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. "RoBERTa: A robustly optimized BERT pretraining approach." *arXiv [Preprint]* (2019). *arXiv preprint arXiv:1907.11692* (1907).

⁴⁶ Delobelle, Pieter, Thomas Winters, and Bettina Berendt. "Robbert: a dutch roberta-based language model." *arXiv preprint arXiv:2001.06286* (2020).

⁴⁷ Delobelle, Pieter, Thomas Winters, and Bettina Berendt. "Robbert-2022: Updating a dutch language model to account for evolving language use." *arXiv preprint arXiv:2211.08192* (2022).

⁴⁸ Delobelle, Pieter, and François Remy. "Robbert-2023: Keeping dutch language models up-to-date at a lower cost thanks to model conversion." *Computational Linguistics in the Netherlands Journal* 13 (2024): 193–203.

⁴⁹ Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." *arXiv preprint arXiv:2203.05794* (2022).

⁵⁰ Netherlands Forensic Institute. *robbert-2022-dutch-sentence-transformers* (Revision cdf42f6). 2024. Hugging Face. <https://huggingface.co/NetherlandsForensicInstitute/robbert-2022-dutch-sentence-transformers>. <https://doi.org/10.57967/hf/3457>.

deze modellen geeft onderwerpen en de hiërarchische relaties tussen de onderwerpen waardoor we bepaalde topics kunnen combineren en enigszins vergelijkbare resultaten laten zien ten opzichte van vooraf gedefinieerde subthema's.

Om de bronnen te onderzoeken, maken we gebruik van *named entity recognition (NER)*, een veelgebruikte *NLP*-techniek voor het extraheren van de namen van entiteiten, waarbij entiteiten verwijzen naar mensen, organisaties, locaties en geopolitieke entiteiten.⁵¹ In de sociale wetenschappen wordt *NER* gebruikt met verschillende benaderingen, variërend van het vinden van vooraf gedefinieerde benoemde entiteiten in teksten, ook wel *dictionary-based*-benadering genoemd, tot *machine-learning*-technieken.⁵² We hebben drie verschillende modellen getest voor het uithalen van *named entities* die direct beschikbaar zijn voor gebruik voor de Nederlandse teksten: *Flair*⁵³, *spaCy*⁵⁴ en *Stanza*⁵⁵ en vonden *Stanza* het meest geschikt model in termen van aantal geëxtraheerde entiteiten en de kwaliteit van deze entiteiten, wat in lijn is met de bevindingen van een eerdere studie die deze modellen vergelijkt.³⁴ Deze modellen zijn voorgetraind en klaar om *named entities* uit tekst te halen.

Generatieve *Large Language Models (LLM's)*, geïntroduceerd door de *Generative Pre-Trained Transformer (GPT)*, hebben een revolutie teweeggebracht in geautomatiseerde tekstanalyse.⁵⁶ Deze modellen, die eerst functioneerden als een "geavanceerde autocomplete"⁵⁷, zijn getraind om tekst te voorspellen en aan te vullen. Grotere versies zijn verder verfijnd om instructies te volgen, waardoor ze bijzonder effectief zijn voor het annoteren van tekst. In tegenstelling tot *supervised*-modellen, waarvoor uitgebreide en dure, door mensen

geannoteerde trainingsgegevens nodig zijn, zijn generatieve *LLM's* uitstekend in *zero-shot learning*: ze voeren diverse *NLP*-taken uit zonder training door simpelweg instructies op te volgen.³⁶ Sommige onderzoekers waarschuwen echter voor problemen op het gebied van ethiek, reproduceerbaarheid, privacy en auteursrecht bij het gebruik van *proprietary*-modellen die alleen kunnen worden gebruikt via tools van de bedrijven, zoals de *GPT*-modellen van *OpenAI*. Ze pleiten voor *open-source*-modellen die kunnen worden gebruikt op lokale of zelfgehoste omgevingen.^{58,59,60} Om deze redenen hebben we twee *open-source LLM's* getest, *Starling-LM-7B-alfa* ontwikkeld door *UC Berkeley* en *Mistral-7B-Instruct-v0.2* ontwikkeld door *MistralAI*. Beide modellen werden gedownload van *HuggingFace* en draaiden in een werkruimte met *GPU*-beschikbaarheid, gehost door *SURF*, de *ICT*-coöperatie van Nederlandse onderwijs- en onderzoeksinstituten.

Beoordeling en vergelijking van modellen

We evalueerden de prestaties van de classificatiemodellen met twee statistieken: *accuracy* en *f1-score*. *Accuracy* geeft het percentage correct geclassificeerde artikelen aan, maar kan misleidend zijn bij ongebalanceerde data. Daarom rapporteren we ook de *f1-score*, die zowel *precision* als *recall* combineert. *Precision* toont de verhouding van correct geclassificeerde positieve gevallen ten opzichte van alle voorspelde positieve gevallen, terwijl *recall* de verhouding weergeeft tussen correct geclassificeerde positieve gevallen en alle werkelijke positieve gevallen in de data. Een *f1-score* boven de 0,7 wordt doorgaans als goed beschouwd, met scores boven de 0,8 die wijzen op een sterke modelprestatie. Deze waarden kunnen echter variëren afhankelijk van de complexiteit van de taak.

⁵¹ Jurafsky, Daniel, and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 3rd ed. 2025. Online manuscript, released January 12, 2025. <https://web.stanford.edu/~jurafsky/slp3/>. Chapter 17.3.

⁵² Balluff, Paul, Hajo G. Boomgaarden, and Annie Waldherr. "Automatically Finding Actors in Texts: A Performance Review of Multilingual Named Entity Recognition Tools." *Communication Methods and Measures* (2024): 1-19.

⁵³ Akbik, Alan, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. "FLAIR: An easy-to-use framework for state-of-the-art NLP." In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations)*, pp. 54-59. 2019.

⁵⁴ Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adrian Boyd. "spaCy: Industrial-Strength Natural Language Processing in Python." 2020. <https://doi.org/10.5281/zenodo.1212303>.

⁵⁵ Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. "Stanza: A Python natural language processing toolkit for many human languages." *arXiv preprint arXiv:2003.07082* (2020).

⁵⁶ Alamm, Jay, and Maarten Grootendorst. *Hands-On Large Language Models: Language Understanding and Generation*. "O'Reilly Media, Inc.", 2024.

⁵⁷ Törnberg, Petter. "Large language models outperform expert coders and supervised classifiers at annotating political social media messages." *Social Science Computer Review* (2024): 08944393241286471, page 3.

⁵⁸ Ollion, Étienne, Rubing Shen, Ana Macanovic, and Arnault Chatelain. "The dangers of using proprietary LLMs for research." *Nature Machine Intelligence* 6, no. 1 (2024): 4-5.

⁵⁹ Törnberg, "Large Language Models"

⁶⁰ Weber, Maximilian, and Merle Reichardt. "Evaluation is all you need. prompting generative large language models for annotation tasks in the social sciences. a primer using open models." *arXiv preprint arXiv:2401.00284* (2023).

Tabel 1 Taakomschrijvingen, geteste methodologieën en modellen

Variabelen	Beschrijving van de taak	Geteste modellen
Hoofdonderwerp	<i>Binary classification</i> (0: hoofdonderwerp niet corona, 1: hoofdonderwerp wel corona)	• <i>Tf-Idf vectors + SVM</i> • <i>Fine-tuned RobBERT</i> • <i>Starling-LM-7B-alfa</i> • <i>Mistral-7B-Instruct-v0.2</i>
Subthema's	<i>Binary classification</i> voor elke vooraf gedefinieerde subthema (1: subthema voldoende besproken, 0: subthema onvoldoende/niet besproken)	• <i>Tf-Idf vectors + SVM</i> • <i>Fine-tuned RobBERT</i> • <i>Starling-LM-7B-alfa</i> • <i>Mistral-7B-Instruct-v0.2</i>
	<i>Unsupervised topic modelling</i>	• <i>BERTopic & robert sentence transformer</i>
Bronnen: geciteerde mensen en organisaties	<i>Named entity recognition + binary classification</i> van elke bron (1: geciteerd, 0: niet geciteerd)	• <i>Stanza + Tf-Idf vectors + SVM</i> • <i>Stanza + Fine-tuned RobBERT</i>
	Alleen geciteerde bronnen direct uit de tekst halen via LLM's	• <i>Starling-LM-7B-alfa</i> • <i>Mistral-7B-Instruct-v0.2</i>
Classificatie van bronfuncties	<i>Multiclass classification</i> van bronnen in een van de vooraf gedefinieerde bronfunctiecategorieën	• <i>Tf-Idf vectors + SVM</i> • <i>Fine-tuned RobBERT</i> • <i>Starling-LM-7B-alfa</i> • <i>Mistral-7B-Instruct-v0.2</i>
Bron noemt coronamaatregelen	Filtervraag, <i>binary classification</i> voor elke bron (1: bron noemt maatregelen, 0: bron noemt geen maatregelen)	• <i>Tf-Idf vectors + SVM</i> • <i>Fine-tuned RobBERT</i> • <i>Starling-LM-7B-alfa</i> • <i>Mistral-7B-Instruct-v0.2</i>
Standpunt t.o.v. coronamaatregelen	<i>Binary classification</i> voor ondersteunende standpunten (1: ondersteunend, 0: niet ondersteunend/neutraal) <i>Binary classification</i> voor kritische standpunten (1: kritisch, 0: niet kritisch/neutraal)	• <i>Tf-Idf vectors + SVM</i> • <i>Fine-tuned RobBERT</i> • <i>Starling-LM-7B-alfa</i> • <i>Mistral-7B-Instruct-v0.2</i>

Overzicht van de analyse en modelprestaties

Thema-analyse

Als eerste stap hebben we nieuwsartikelen geclassificeerd om te bepalen of het hoofdthema de coronapandemie was, zodat niet-relevante artikelen konden worden uitgesloten. De testresultaten zijn opgenomen in Tabel 2. Uit de resultaten blijkt dat beide LLM's vergelijkbaar presteren en het beste resultaat

behalen, gevolgd door het *fine-tuned RobBERT*-model. Hoewel het *SVM*-model iets minder nauwkeurig presteert, is het nog steeds in staat deze classificatie uit te voeren en vormt het een kosteneffectieve optie, omdat het geen krachtige hardware vereist en aanzienlijk sneller te trainen is. Voor het filteren van artikelen die niet primair over de pandemie gingen, hebben we het *Starling LLM* ingezet. Dit resulteerde in een selectie van 6.491 artikelen over de coronapandemie tussen januari 2020 tot en met mei 2022.

Tabel 2 Prestaties van verschillende modellen om hoofdonderwerp te classificeren

Modelnaam	Aantal artikelen (Aantal corona-artikelen)			Modelprestaties	
	Training	Validation	Test	Accuracy	f1-macro
Tf-IDF + SVM	540 (331)	136 (79)	120 (78)	0,84	0,83
RobBERT	540 (331)	136 (79)	120 (78)	0,90	0,89
Mistral	-	-	120 (78)	0,97	0,96
Starling	-	-	120 (78)	0,95	0,95

Na het filteren van artikelen met slechts korte vermeldingen van het coronavirus/pandemie hebben we binaire *SVM*- en *RobBERT-classifiers* gebruikt om vooraf gedefinieerde subthema's te detecteren, waarbij meerdere subthema's per artikel mogelijk waren. Daarnaast hebben we *LLM*'s ingezet, waarbij beide modellen de prompt kregen om subthema's te identificeren die ten minste twee zinnen beslaan, in lijn met de instructies aan menselijke codeurs.

De resultaten tonen uiteenlopende prestaties per subthema. De *SVM* presteerde sterk bij 8 van de 12 subthema's en overtrof

RobBERT bij onderwerpen zoals politieke discussies, internationale reacties en samenwerkingen. *Fine-tuned RobBERT* was beter in het classificeren van subthema's zoals pandemiestatistiek, medische respons en onderzoek maar had, net als de *SVM*, moeite met vier uitdagende subthema's. *LLM*'s, toegepast in een zero-shot setting zonder training, presteerden opvallend goed bij complexe subthema's zoals rechten en vrijheden en buitenlandse reacties. *Starling* bleek onder de *LLM*'s het meest effectief en liet sterke prestaties zien bij complexe thema's zoals sociale en politieke gevolgen. Een overzicht van de prestaties van alle modellen is te vinden in Tabel 3.

Tabel 3 Prestaties van verschillende modellen om subthema's te classificeren

Subthema's	Aantal artikelen (per subthema)		Tf-Idf + SVM		RobBERT		Starling		Mistral	
	Training*	Test	Acc.	f1	Acc.	f1	Acc.	f1	Acc.	f1
Statistieken en status van de pandemie	410 (117)	78 (21)	0,90	0,80	0,94	0,90	0,92	0,90	0,87	0,85
Coronamaatregelen	410 (119)	78 (30)	0,84	0,76	0,82	0,77	0,74	0,74	0,76	0,75
Coronatesten en testprocedures	410 (58)	78 (12)	0,93	0,73	0,94	0,78	0,91	0,77	0,97	0,95
Vaccinaties en vaccinatiecampagnes (incl. 2G & 3G)	410 (98)	78 (16)	0,93	0,81	0,88	0,79	0,9	0,82	0,88	0,80
Long covid en langdurige gevolgen op gezondheid	410 (26)	78 (8)	0,98	0,92	0,98	0,92	0,96	0,87	0,99	0,96
Uitdagingen in de zorg en medische respons	410 (26)	78 (14)	0,92	0,77	0,96	0,88	0,94	0,89	0,94	0,90
Onderzoek en kennis over virus & ziekte (incl. varianten)	410 (29)	78 (14)	0,89	0,54	0,91	0,69	0,94	0,88	0,88	0,83
Impact op economie, bedrijven en herstel	410 (40)	78 (5)**	0,96	0,80	0,95	0,77	0,99	0,94	0,96	0,85
Sociale impact en mentale gezondheid	410 (40)	78 (13)	0,86	0,51	0,77	0,50	0,88	0,77	0,81	0,59
Gevolgen voor rechten & vrijheden	410 (11)*	78 (13)	0,90	0,54	0,88	0,53	0,94	0,86	0,90	0,83
Politieke impact en discussies	410 (29)	78 (6)**	0,97	0,74	0,92	0,62	0,94	0,82	0,81	0,59
Internationale respons op de pandemie en samenwerking	410 (8)*	78 (8)**	0,95	0,69	0,93	0,48	0,95	0,82	0,97	0,92

* Trainingdata werd gebruikt alleen voor SVM- en RobBERT-modellen. Testdata werden gelijk gehouden voor alle modellen.

** Laag aantal artikelen in training of testdata.

Uit onze bevindingen blijkt dat de classificatie-aanpak met een vooraf gedefinieerde lijst van subthema's zowel voordelen als nadelen heeft. Een belangrijk voordeel is dat deze aanpak het mogelijk maakt om specifiek te bepalen wanneer een subthema daadwerkelijk besproken wordt. Artikelen worden alleen geclassificeerd onder een subthema als er voldoende inhoudelijke bespreking plaatsvindt (minimaal twee zinnen). Daarnaast zorgt deze methode ervoor dat over het hoofd geziene of niet-gerapporteerde subthema's in de analyse worden geïdentificeerd. Een nadeel is dat genuanceerde subthema's buiten de lijst worden gelaten en thema's met weinig aandacht zelden voorkomen in door mensen

geannoteerde data. Hierdoor is er onvoldoende trainings- en testmateriaal. In onze steekproef was de berichtgeving over desinformatie, rechten en vrijheden, en internationale samenwerking schaars. Selectie op kernwoorden loste dit niet op; bijvoorbeeld, desinformatie kon niet worden getest vanwege het lage aantal artikelen. Daarom hebben we een *unsupervised*-techniek onderzocht. *BERTopic*, toegepast op 6.491 artikelen over de pandemie, identificeerde 63 onderwerpen. De representatieve kernwoorden van deze 63 onderwerpen staan in Tabel 4. Aan de hand van de hiërarchische relaties en de representatieve kernwoorden en teksten gecreëerd door het model, hebben we 16 onderwerpcategorieën vastgesteld.



Dit zijn:

1. Coronamaatregelen
2. Onderzoek & kennis over virus & ziekte (incl. varianten)
3. Vaccinaties en vaccinatiecampagnes (incl. 2G & 3G)
4. Coronatesten en testprocedures
5. Impact op sociale leven (sport, evenementen, cultuur) en op kwetsbare groepen
6. Statistieken en status van de pandemie
7. Uitdagingen in de zorg en medische respons
8. Impact op economie, bedrijven en herstel
9. Internationale reacties op de pandemie & samenwerkingen
10. Corona app
11. Gevolgen voor rechten & vrijheden
12. Uitzending corona
13. Gevolgen voor de dierenindustrie
14. Misinformatie over coronavirus/pandemie & complottheorieën
15. Koninklijke familie
16. Long covid en langdurige gevolgen op gezondheid

Verschillende benaderingen voor de onderwerpanalyse leveren verschillende resultaten op. *BERTopic* genereerde meer themaclusters dan de classificatie op basis van de vooraf gedefinieerde lijst, waaronder clusters die niet door menselijke codeurs waren gecodeerd. Daarentegen creëerde deze methode geen specifieke cluster voor politieke discussies en de impact van de pandemie op politieke processen, en voor de impact van de pandemie op mentale gezondheid. Politieke discussies, zoals parlementaire debatten, werden in plaats

daarvan voornamelijk gegroepeerd met andere subthema's, zoals coronamaatregelen. De *BERTopic*-aanpak resulteerde in meer subthema's per artikel; 79% van de artikelen kreeg meer subthema's toegewezen dan bij *LLM*-classificatie. Dit is te verklaren omdat we *LLM*'s (en menselijke codeurs) expliciet hebben geïnstrueerd om alleen een subthema toe te wijzen bij voldoende bespreking, wat bij de *BERTopic*-methode niet mogelijk is. Daarnaast hebben we *BERTopic* op paragraafniveau toegepast, wat het mogelijk maakt dat aan elk artikel meer dan een thema wordt toegekend. Bovendien liet de binaire classificatie met *LLM*'s zien dat 1.231 artikelen geen enkel onderwerp kregen toegewezen, terwijl alle artikelen met *BERTopic* ten minste één onderwerp kregen.

Over het algemeen blijkt *BERTopic* een efficiënte aanpak voor het analyseren van thema-diversiteit op grote schaal, omdat het geen menselijke codering of vooraf gedefinieerde lijsten vereist. Hoewel de *LLM*'s zeer goed presteren in het detecteren van subthema's, vereist de analyse meer rekenkracht en duurt de binaire classificatie van elk subthema langer in vergelijking met de *BERTopic*-benadering. We kunnen met dezelfde computer (met twee A10 GPU's), de subthema's van 6.491 artikelen in slechts enkele minuten berekenen met *BERTopic*, terwijl de *LLM*'s voor elke subthema 4-5 uur nodig hebben, omdat deze modellen elke artikelen individueel "lezen" en classificeren. Wel vereist *BERTopic* kritische analyse door onderzoekers, aangezien de methode afhankelijk is van de data die hiervoor worden gebruikt en de neiging heeft om subthema's te overschatten. Vergelijking met werkelijke politieke en maatschappelijke discussies blijft essentieel, aangezien het model geen subthema's kan detecteren die niet al in het nieuws aan bod komen.

Tabel 4 De representatieve kernwoorden van thema's geïdentificeerd door BERTopic

Thema	Top 3 Woorden	Thema	Top 3 Woorden
0	vaccinaties, pfizer vaccin, vaccin	34	aantal incidenten, overlast, geweld politie
1	betaald voetbal, wedstrijden, toernooi	35	zogenoemde besmettingsgetal, 100 besmettelijke, besmettingsgetal
2	weer persconferentie, persconferentie, overleg	36	aantal doden, aantal overledenen, 69 doden
3	werkloosheid, steunmaatregelen, uitkering	37	vaccinatiecampagne, vaccineren, vaccinatie
4	patiënten spreiding, ziekenhuisopnames, intensivereafdelingen	38	aantal sterfgevallen, aantal doden, besmettingen china
5	code oranje, code geel, geel oranje	39	viering prinsjesdag, bezoek koning, anders prinsjesdag
6	scholen sluiten, scholen open, scholen weer	40	meerdaagse festivals, festivalseizoen, eendaagse festivals
7	testcapaciteit, testbeleid, testlocatie	41	nederland aantal, aantal besmettingen, aantal coronabesmettingen
8	ooit corona, corona gehad, corona hebt	42	vion slachterij, vleesbedrijf vion, slachterij vion
9	zeggen, blijven, komt	43	gezondheid moeder, vertelt dochter, ziek
10	uur open, openingstijden, sluitingstijd	44	djokovic gevaccineerd, djokovic mag, gaf djokovic
11	verspreiding virus, virus we, ontwikkeling virus	45	19 infectie, 19 symptomen, patiënten covid
12	dragen mondmasker, verplicht mondkapje, dragen mondkapje	46	nepnieuws coronavirus, nepnieuws verspreiden, nepnieuws verspreid
13	variant besmettelijker, afrikaanse variant, indische variant	47	gevangenen tijdelijk, mochten gedetineerden, gevangenen quarantaine
14	zegt china, shanghai, chinees nieuwjaar	48	sneltesten antigeentesten, test pcr, test sneltest
15	corona apps, corona app, coronamelder app	49	golf waarschijnlijk, golf waarschuwd, golf komt
16	tests positief, percentage positieve, tests per	50	bezoekregeling verpleeghuizen, bezoekverbod verpleeghuizen, verruimd bezoekregeling
17	enkele demonstranten, demonstraties, protesten	51	2g 3g, 3g, 3g systeem
18	sterftecijfers, aantal coronadoden, sterftecijfer	52	behandeling covid, ontstekingsremmer dexamethason, geneesmiddelen
19	trump weer, president trump, trump	53	testverplichting reizigers, test quarantaine, testuitslag tonen
20	geldt avondklok, avondklok ingevoerd, avondklok ingesteld	54	productie beademingsapparaten, 100 beademingsapparaten, beademingsapparaten afgeleverd
21	minder trainen, openbaar vervoer, minder reizigers	55	bolsonaro president, president bolsonaro, verkiezingen bolsonaro
22	patiënten corona, corona ziekenhuis, corona afdeling	56	virus nederland, verspreiding virus, nederland afneemt
23	brexit onderhandelingen, brexit, hongarije	57	komt long, krijgen long, patiënten long
24	nertsenfokkers, nertsenhouderij, nertsenhouderijen	58	inkomsten musea, deel musea, musea komt
25	premier johnson, johnson minister, johnson bood	59	compensatie zorgmedewerkers, regeling zorgmedewerkers, zorgmedewerkers long
26	live volgen, tv uitzending, nieuws via	60	opgenomen verpleegafdelingen, patiënten opgenomen, verpleegafdelingen nieuwe
27	moskeebezoek, moskeeën, kerken moskeeën	61	limburg carnaval, bond carnavalsverenigingen, start carnaval
28	scanner app, scanner, codes	62	aantal coronabesmettingen, aantal besmettingen, nieuwe coronabesmettingen
29	goede ventilatie, lucht blijven, luchtcirculatie		
30	kerstvakantie, sinterklaas, houden kerst		
31	nieuwe coronabesmettingen, bevestigde coronabesmettingen, werden gemeld		
32	omzet supermarkten, online omzet, omzet		
33	besmettingen italië, virus italië, noord italië		

Bronnen identificeren en categoriseren

Om de journalistieke bronnen, gedefinieerd als personen en organisaties die geciteerd worden (direct of indirect), te extraheren, hebben we twee benaderingen getest. Eerst vroegen we de *LLM*'s om direct en indirecte citaten uit de nieuwsartikelen te extraheren en om de namen van de personen en organisaties te identificeren, samen met de aanduiding of het om een persoon of een organisatie ging. Vervolgens gebruikten we *Named Entity Recognition (NER)* om namen van alle personen en organisaties te extraheren, waarna we een *SVM*- en een *RobBERT*-classificatiemodel trainden om te bepalen welke van deze bronnen geciteerd waren. De resultaten werden vergeleken met de *gold standard*.

De *Starling LLM* extraheerde 61% van de bronnen, terwijl de *Mistral LLM* 73% van de bronnen juist extraheerde. Beide modellen presteerden beter bij het identificeren van personen (*Starling*: 63%, *Mistral*: 77%) dan organisaties (*Starling*: 55%, *Mistral*: 67%). Het *NER*-model extraheerde 501 entiteiten. Met de *RobBERT*-classificatie voor citaten konden we 79% van de bronnen juist identificeren, terwijl de *SVM* 74% van de bronnen in de *gold standard* data identificeerde. Beide modellen identificeerden personen en organisaties met een vergelijkbare nauwkeurigheid: *RobBERT* herkende 78% van de personen en 81% van de organisaties, terwijl *SVM* 74% van de personen en 75% van de organisaties correct vaststelde.

Daarnaast analyseerden we de niet-geïdentificeerde bronnen. Ongeveer twee derde van niet-geïdentificeerde bronnen door *LLM*'s waren personen, vaak aangeduid met algemene termen zoals 'experts', 'bewoners' of 'dokters', terwijl sommige bronnen met een eigenaam niet als bron geïdentificeerd werden, maar wel herkend hadden moeten worden. Aan de andere kant waren de bronnen die niet werden geïdentificeerd door de *NER*-classificatiemethode voornamelijk anonieme bronnen, terwijl de personen en organisaties met eigenaam met enkele uitzonderingen correct werden geïdentificeerd.

Bovendien onderzochten we de "extra" bronnen die de modellen identificeerden, maar die niet in de *gold standard* dataset voorkwamen. Alle geteste modellen hebben extra bronnen geïdentificeerd; de *NER+RobBERT*-methode identificeerde het minste aantal extra bronnen, terwijl de *NER+SVM*-methode het meeste aantal extra bronnen classificeerde. Als we deze bronnen verder controleren zien we dat de modellen soms bronnen breder interpreteren dan strikt de definitie van citaat. Als een persoon of organisatie informatie geeft zonder expliciete citaten, wordt deze ook als bron geïdentificeerd. Bijvoorbeeld, alle modellen classificeerden "het kabinet" als een bron in de zin: "Het kabinet wil dat alle

jongeren van 12 tot en met 17 jaar de mogelijkheid krijgen om zich te laten vaccineren."

Samenvattend blijkt uit onze resultaten dat het *NER*-model in combinatie met de *RobBERT*-model het beste presteert voor het identificeren van journalistieke bronnen. Hoewel de *LLM*'s ook in staat zijn om bronnen te identificeren zonder training, zijn ze minder goed in het herkennen van organisaties als bronnen. Dit kan te maken hebben met hun training, die voornamelijk is gebaseerd op Engelstalige teksten. De *NER+RobBERT*-methode identificeert niet alleen correct de namen, maar kan ook goed onderscheid maken tussen entiteiten die genoemd worden en entiteiten die informatie en meningen verstrekken via citaten. De anonieme/collectieve bronnen blijven nog moeilijk voor al geteste automatische modellen om uit tekst te halen.

Functies van de bronnen

Na het extraheren van de bronnen hebben we aan de *LLM*'s instructies gegeven om de bronnen in te delen in vier functie-categorieën: overheid en politiek, professionals en experts, maatschappelijke organisaties en burgers. Hiervoor gebruikten we de naam van de bron en de zinnen uit het nieuwsartikel waarin de bron werd genoemd als context. Met dezelfde input voerden we een *multiclass*-classificatie uit van functies met het *SVM*-model en een *fine-tuned RobBERT*-model. Het aantal bronnen per categorie in de trainings-, validatie- en testsets (voor de classificatiemodellen), samen met de *f1-scores* per functie en de algemene nauwkeurigheid en *f1-scores* van de modellen, zijn te vinden in de tabel 5.

Het *RobBERT*-model presteerde over het algemeen het beste. Hoewel alle modellen succesvol waren in het classificeren van politieke bronnen en experts, had *RobBERT* een duidelijke voorsprong bij het nauwkeurig classificeren van maatschappelijke organisaties, waar de *LLM*'s en het *SVM*-model moeite mee hadden. Het classificeren van burgers als bronnen bleek voor alle modellen een uitdaging, en de resultaten voor deze categorie moeten met voorzichtigheid worden geïnterpreteerd, gezien het lage aantal burgerbronnen in de *gold standard* dataset. Omdat het identificeren van burgerbronnen over het algemeen moeilijk is, blijft de automatische analyse van deze groep problematisch. In de toekomst zouden alternatieve (en mogelijk niet-automatische) methoden kunnen worden overwogen om burgerbronnen in nieuwsartikelen te identificeren, mocht deze categorie worden opgenomen in de monitoring van brondiversiteit. Over het algemeen toont *RobBERT* zich een betrouwbare methode voor het identificeren van bronnen en het toewijzen van hun functiestatus, vooral voor politieke bronnen, experts en vertegenwoordigers van maatschappelijke organisaties.

Tabel 5 Prestaties van modellen om bronnen in functiecategorieën in te delen

Functie	Aantal bronnen per functie*			Starling	Mistral	SVM	RobBERT
	Training	Validatie	Test	f1	f1	f1	f1
Overheid en politiek	967 (332)	415 (142)	294 (95)	0,80	0,81	0,75	0,92
Professionals en deskundigen	967 (495)	415 (213)	294 (136)	0,80	0,78	0,72	0,84
Maatschappelijke organisaties	967 (78)	415 (33)	294 (36)	0,58	0,49	0,41	0,66
Burgers	967 (62)	415 (27)	294 (27)	0,50	0,59	0,28	0,46
Modelnauwkeurigheid	-	-	-	0,76	0,71	0,67	0,82
Macro-gem. f1-score	-	-	-	0,67	0,67	0,54	0,72

* Het aantal bronnen per categorie in de trainings-, validatie- en testsets is gerapporteerd op basis van de NER+SVM- en NER+RobBERT-methodologieën.

Analyse naar de standpunten over coronamaatregelen

In deze pilot definieerden we standpunten als ondersteunend en/of kritisch ten opzichte van de coronamaatregelen. Voordat we de standpunten analyseerden, lieten we *LLM's* bepalen of de bron enige coronamaatregelen, adviezen of regels noemt in de zinnen waarin de bron wordt genoemd. Vervolgens vroegen we de *LLM's* om aan te geven of de bron ondersteunende en/of kritische standpunten inneemt. Voor de *SVM* en *fine-tuned RobBERT* trainden we aparte classificatiemodellen voor het herkennen van genoemde

maatregelen en het identificeren van standpunten, omdat een bron in hetzelfde artikel zowel positieve als negatieve standpunten kan geven. Alle modellen ontvingen bronnaam en zinnen waar de bron aan bod komt als input. De prestaties van de modellen staan in tabel 6. Over het algemeen waren deze taken uitdagend voor zowel de *LLM's* als *SVM* en *RobBERT*. *RobBERT* presteerde het beste bij het bepalen of de bron maatregelen noemt, met de hoogste nauwkeurigheid in vergelijking met de andere modellen, die lagere scores behaalden. Tabel 6 toont de resultaten.

Tabel 6 Prestaties van modellen voor het classificeren of de bron coronamaatregelen noemt

Modelnaam	Aantal artikelen (Aantal corona-artikelen)			Modelprestaties	
	Training	Validatie	Test	Nauwkeurigheid	f1-macro
Tf-IDF + SVM	1.105 (560)	277 (140)	294 (150)	0,68	0,68
RobBERT	1.105 (560)	277 (140)	294 (150)	0,75	0,75
Mistral	-	-	294 (150)	0,58	0,56
Starling	-	-	294 (150)	0,61	0,61

De prestaties van de modellen variëren bij het bepalen of een bron een ondersteunend of kritisch standpunt inneemt ten opzichte van coronamaatregelen, wat de complexiteit van de taak weerspiegelt. Voor het classificeren van ondersteunende standpunten presteerde het SVM-model het beste, gevolgd door RobBERT en Starling, en Mistral. Bij het detecteren van kritische standpunten presteerden alle modellen minder goed, met de laagste score voor SVM en de hoogste voor RobBERT, gevolgd door Starling en Mistral. Alle modellen hadden moeite met het herkennen van negatieve standpunten. De prestaties van de modellen staan in Tabel 7 en 8.

Deze taak was ook moeilijk voor menselijke codeurs. De grootste discrepantie tussen de primaire en secundaire codeurs

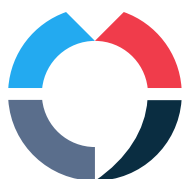
ontstond bij deze variabelen. Hoewel we kozen voor een eenvoudige definitie van standpunt (ondersteunend versus kritisch), blijft de taak vatbaar voor verschillende interpretaties. Als een taak complex is voor mensen, blijft deze ook uitdagend voor geautomatiseerde modellen. Daarom vereist het meten van standpunten een genuanceerde benadering, waarbij alternatieve methoden, zoals kwalitatieve inhoudsanalyse, mogelijk beter de complexiteit kunnen vangen. Aangezien we de diversiteit van standpunten periodiek willen monitoren, is een flexibele methodologie nodig die breed toepasbaar is. Het gebruik van bronnen, hun status en politieke affiliaties, en de combinatie daarvan met subthema's biedt daarom een schaalbare en generaliseerbare manier om de diversiteit naar standpunten te beoordelen.

Tabel 7 Prestaties van modellen bij het classificeren of een bron steunende standpunten over maatregelen inneemt

Modelnaam	Aantal artikelen (Aantal corona-artikelen)			Modelprestaties	
	Training	Validatie	Test	Nauwkeurigheid	f1-macro
Tf-IDF + SVM	560 (215)	140 (54)	150 (52)	0,73	0,67
RobBERT	560 (215)	140 (54)	150 (52)	0,64	0,63
Mistral	-	-	150 (52)	0,64	0,53
Starling	-	-	150 (52)	0,81	0,63

Tabel 8 Prestaties van modellen bij het classificeren of een bron kritische standpunten over maatregelen inneemt

Modelnaam	Aantal artikelen (Aantal corona-artikelen)			Modelprestaties	
	Training	Validatie	Test	Nauwkeurigheid	f1-macro
Tf-IDF + SVM	560 (125)	140 (31)	150 (38)	0,72	0,46
RobBERT	560 (125)	140 (31)	150 (38)	0,76	0,60
Mistral	-	-	150 (38)	0,81	0,53
Starling	-	-	150 (38)	0,78	0,57



Commissariaat voor de Media

Dit is een uitgave van het
Commissariaat voor de Media
Maart 2025

Vormgeving
de Toekomst

Commissariaat voor de Media

Hoge Naarderweg 78
1217 AH Hilversum

Postbus 1426
1200 BK Hilversum

035 773 77 00
cvdm@cvdm.nl
www.cvdm.nl