

Nederlandse talkshows tijdens de verkiezingscampagne van 2025

Onderzoeksverantwoording

Amsterdam, november 2025

Nisanur Karadavut (MA)

Vaas de Wit (MA)

Dr. Nel Ruigrok

Inhoud

Inleiding	2
Dataverzameling	2
Methode	3

Inleiding

Het onderzoek naar de gasten van talkshows is uitgevoerd door studenten media en journalistiek van de Erasmus Universiteit Rotterdam onder leiding van Nel Ruigrok, Nisanur Karadavut en Vaas de Wit, in samenwerking met de Ombudsman voor de publieke omroepen. Het onderzoek is mede mogelijk gemaakt door het Commissariaat voor de Media en het Stimuleringsfonds voor de Journalistiek. De (ex)studenten die meewerkten: Gijs Berk, Rosa van der Vleuten, Frederique Zandvliet, Iza van Meggelen, Weia Tan, Niene Tempelman, Charlotte Soer, Christian Gravesteijn en Mare Hoekstra.

In deze onderzoeksverantwoording gaan we in op de dataverzameling en de verschillende methoden van onderzoek.

Dataverzameling

In totaal zijn tien talkshows geanalyseerd over de periode van 1 september tot en met 29 oktober 2025. De selectie bestond uit zowel programma's van publieke omroepen als van commerciële zenders. De geanalyseerde programma's van de publieke omroepen waren: *Goedemorgen Nederland*, *Pauw & De Wit*, *Buitenhof*, *WNL op Zondag*, *Eva*, *Café Kockelmann*, *Dit is de Week*, *Een buitengewoon gesprek* en *Ongehoord Nieuws*. Vanuit de commerciële zenders zijn *RTL Tonight*, *Nieuws van de Dag* en *Vandaag Inside* opgenomen in de analyse. Daarnaast zijn verschillende verkiezingsdebatten in aanloop naar de Tweede Kamerverkiezingen van 2025 meegenomen. Dit betrof zowel commerciële debatten (twee *RTL Verkiezingsdebatten*)¹ als publieke debatten (*NL Kiest* en het *EenVandaag Debat*).

In totaal ging het hierbij om 299 uitzendingen, 131 uitzendingen van de commerciële omroepen, 164 uitzendingen op de publieke omroepen en 4 debatten, 2 bij de commerciële omroepen en 2 bij de publieke omroepen.

Talkshows per broadcaster (2025)				
Episodes, time and speaking turns				
programma		n_episodes	hours	turns
Commercial				
	RTL Tonight	47	51.8	7,306
	Vandaag Inside	42	49.7	10,352
	Nieuws Van De Dag	42	41.6	5,953
Debates				
	Debates	4	6.5	1,058
NPO				
	Pauw En De Wit	36	29.4	5,532
	Eva	35	27.0	4,904
	Goedemorgen NL	43	14.2	10,616
	Ongehoord Nederland	15	10.2	2,179
	Wnl Op Zondag	8	7.9	1,411
	Cafe Kockelmann	8	7.4	1,862
	Buitenhof	8	7.4	1,203
	Dit Is De Week	8	5.8	1,265
	Buitengewoon gesprek	3	2.2	439
sum	—	299	261.1	54080

¹ *Het Debat van Nederland* is niet meegenomen in de uiteindelijke dataset, omdat de stemmen van presentator Wilfred Genee en Frans Timmermans onvoldoende betrouwbaar van elkaar te onderscheiden waren.

Methode

Voor de analyse is gebruikgemaakt van een combinatie van automatische en handmatige inhoudsanalyse. Om de spreektijd per gast vast te stellen, is diarization-technologie toegepast: een methode waarbij de computer automatisch verschillende sprekers in een uitzending identificeert en labelt. Het systeem detecteert wanneer er gesproken wordt en vertaalt stemkenmerken naar vector representaties. Op basis hiervan zijn de uitzendingen opgedeeld in spreekbeurten: segmenten waarin telkens één persoon aan het woord is. De vector representaties maken het mogelijk om verschillende stemmen automatisch van elkaar te onderscheiden en spreekbeurten die waarschijnlijk door dezelfde spreker zijn uitgesproken, te groeperen onder hetzelfde sprekers-nummer. In totaal gaat het in dit onderzoek om ruim 54 duizend spreekbeurten.

Vervolgens zijn alle spreekbeurten met behulp van Faster Whisper automatisch getranscribeerd. De transcripties, samen met de bijbehorende metadata (datum, start- en eindtijd van elke spreekbeurt, programma en sprekers-nummer), zijn opgeslagen in AmCAT – een database en computationele infrastructuur voor het verzamelen, beheren en analyseren van grootschalige tekstuele data (Van Atteveldt, 2017). In totaal gaat het om ruim 260 uur aan gesproken tekst.

Vervolgens zijn alle talkshows handmatig bekeken en werden de stemmen van alle gasten, inclusief presentatoren en vaste duiders, gekoppeld aan de sprekers nummers die in AmCAT zijn geregistreerd. Deze informatie is vervolgens gekoppeld aan de bijbehorende metagegevens van de personen die spraken, zoals gender, vakgebied etc.

Door deze combinatie van automatische spraakherkenning en handmatige verificatie is een betrouwbare dataset opgebouwd waarmee berekend kon worden welke gasten aan het woord waren en hoeveel spreektijd zij kregen in de verschillende programma's.