



Universiteit
Leiden
The Netherlands

Opinion diversity through hybrid intelligence

Meer, M.T. van der

Citation

Meer, M. T. van der. (2025, March 26). *Opinion diversity through hybrid intelligence*. *SIKS Dissertation Series*. Retrieved from <https://hdl.handle.net/1887/4209024>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4209024>

Note: To cite this publication please use the final published version (if applicable).

Samenvatting

Probleem In een representatieve democratie kiezen burgers vertegenwoordigers die in hun belang handelen bij het vormgeven van overheidsbeleid. Moderne democratieën kampen met een kritiek probleem van afnemende burgerparticipatie, wat leidt tot een kloof tussen burgers en hun gekozen vertegenwoordigers. Een mogelijke aanpak voor dit probleem is via een deliberatieve democratie, waarin open dialogen en wijdere participatie wordt aangemoedigd. Traditionele fysieke deliberatiemethoden kennen echter aanzienlijke uitdagingen, zoals het betrekken van deelnemers met verschillende perspectieven en achtergronden, en het waarborgen van gelijkwaardige inspraak in de discussies. Als alternatief bieden online social-mediaplatformen een mogelijkheid voor grootschalige deliberatie, omdat het discussies met een groter publiek en met snelle toegang tot informatie mogelijk maakt. Er bestaan echter zorgen over de vraag of deze platformen inclusieve en diverse discussies kunnen bevorderen. Het kan bijvoorbeeld moeilijk zijn om bijdragen van relevante perspectieven te vinden vanwege de grote hoeveelheden versnipperde informatie. Bovendien kunnen interacties op platformen tot echokamers leiden die polarisatie opdrijven, wat vervolgens de egalitaire basis van de discussies bedreigt. Het gebruik van Kunstmatige Intelligentie (KI), en met name Natural Language Processing (NLP), voor het faciliteren van online tekstgebaseerde discussies is aantrekkelijk geworden als oplossing hiervoor. De impact van deze technologie op de diversiteit van perspectieven bij online deliberaties is tot dusver nog niet onderzocht.

Methoden Dit proefschrift (1) identificeert de uitdagingen die komen kijken bij het faciliteren van online discussies op grote schaal met behulp van NLP, (2) suggereert oplossingen voor deze uitdagingen door het bewerkstelligen van hybride mens-KI-technologie, en (3) onderzoekt wat deze technologieën kunnen ophelderen over individuele perspectieven in online discussies. Om perspectieven te representeren stellen wij een drielaagse hiërarchie voor die kan worden verkregen door een combinatie van menselijke intelligentie en Large Language Models (LLM's). Deze combinatie noemen wordt ook wel Hybride Intelligentie (HI) genoemd. We illustreren hoe deze representaties ons inzicht kunnen bieden in de diversiteit van perspectieven, en hoe ze ons toestaan interacties in online discussies te karakteriseren.

- In Deel I van dit proefschrift laten we zien hoe bestaande methoden voor het analyseren van meningen, specifiek wanneer ze gebruikmaken van LLM's, beperkt zijn in het begrijpen van de perspectieven die geuit worden door minderheden. Desalniettemin maakt de kracht van deze modellen om tekstgebaseerde data op grote schaal te verwerken ze aantrekkelijk voor het analyseren van online discussies. We deze modellen effectief toepassen wanneer we weinig brondata tot onze beschikking hebben ondanks de complexiteit van het interpreteren van ongestructureerde meningen. LLM's zijn, in het bijzonder, in staat om abstracte taken effectief te voltooien en missende informatie zelfstandig aan te vullen. Hun gevoeligheid benadrukt daarentegen wel dat LLM's menselijk toezicht nodig hebben om goed te presteren in verschillende contexten. LLM's gedragen zich anders dan mensen:

ze kunnen menselijke meningsverschillen vaak niet begrijpen, en maken andere soorten fouten dan mensen. Dit maakt dat ze zorgvuldige menselijke begeleiding nodig hebben.

- In Deel II gebruiken we de potentie van HI-systemen voor het analyseren van meningen. Door middel van het strategisch combineren van menselijke input met dat van LLM's, en door het bevorderen van een heen-en-weer proces tussen mensen en KI, kunnen HI-systemen ontworpen worden die diverse meningen op een precieze en efficiënte manier kunnen vatten. HI vereist een zorgvuldige taakverdeling. Onze methoden maken gebruik van menselijke annotateerders die een genuanceerd begrip van bijvoorbeeld argumenten kunnen leveren. Tegelijkertijd kunnen we NLP-technieken gebruiken voor het selecteren van interessante meningen uit een grote dataset. Beide kanten kunnen door herhaalde interactie continu aanpassen en van elkaar leren. Onze hybride opzet baant de weg voor een toekomst waarin mensen en NLP-technologie de handen ineenslaan om een dieper begrip over de veelzijdige aard van online discussies te cultiveren.
- In Deel III laten we zien dat HI-systemen de *Perspectief Hiërarchie* van een individu kunnen extraheren, gebaseerd op de taken uit Deel I. Voor het extraheren van deze hiërarchieën gebruiken we de complementaire kracht van mensen en NLP-modellen. We tonen aan dat hoe de houding, argumenten, en persoonlijke waarden van een individu kerncomponenten zijn van de hiërarchie door een directe relatie tussen waarden en houdingen te onderzoeken.

Bevindingen We constateren dat er fundamentele problemen zijn voor het bevorderen van diversiteit in de analyse van social media platformen. Deze problemen omvatten (1) het waarborgen dat gemarginaliseerde standpunten deel nemen op de platformen, (2) een nadruk op vaak herhaalde meningen die de politieke kloof verergeren, en (3) een verergering van dit probleem door de gemakkelijke toepassing van LLM's voor het analyseren van meningen. HI kan helpen om deze problemen te verlichten. In onze aanpak voor HI moedigen wij explicietere communicatie tussen mensen aan door middel van herhaalde interactie met LLM's. Dit alles voedt verbeteringen aan twee kanten: mensen profiteren van expliciete communicatie en kunnen elkaars uitgangspunten nauwgezet overwegen terwijl hun redeneringen gebruikt kunnen worden om betere KI-modellen te trainen.

Dit proefschrift biedt een van de eerste demonstraties van de integratie van mensen en KI door het samenwerkingsvermogen van mensen te combineren met LLM's. Toch blijft het aantonen dat HI leidt tot een verbeterde en diversere discussie moeilijk. Bestaande evaluatieparadigma's voor dergelijke systemen zijn ontoereikend om te meten hoe HI leidt tot verbeteringen ten opzichte van alleen KI of handmatige aanpakken, wat de noodzaak voor meer dynamische en context-specifieke evaluaties onderstreept.