



AI met hulp van AI

mei 2025

Onderdeel van het rapport
Bouwen aan de gemeente organisatie van morgen

Inhoud

Onderzoeksfase 1 (juli-september 2024)	2
Werk sessie “hoe verder?” (oktober 2024)	3
Verdere analyse- en validatiefase (oktober-november 2024)	4
Rapportagefase (november-december 2024)	5
Communicatiefase (december 2024 - maart 2025)	5

AI met hulp van AI

Het waarderend onderzoek (Appreciative Inquiry) naar de waarde en waardering van de gemeentebambtenaar was volstrekt onmogelijk geweest zonder intelligente automatisering (die andere AI). Dat we het internet, e-mail, Word, Powerpoint en Teams meetings gebruiken om een onderzoeksvorstel te maken en te bespreken, vinden we natuurlijk volstrekt normaal. Het bestoken van Google en Google Scholar met diverse ‘search strings’ - op zoek naar relevante bronnen - is wetenschappelijk gemeengoed. We hadden echter niet kunnen bedenken dat AI-tools zo belangrijk bleken te zijn in het uitvoeren van het onderzoek. Onmisbaar, eerlijk gezegd. We hebben een journaal gemaakt van onze stappen, met name waar het de inzet van AI (KI) applicaties betrof. Je vindt het hieronder in een ruwe vorm.

Onderzoeksfase 1 (juli-september 2024)

1. Opdrachtgever: “Zoek een prominent die een visie schrijft”
2. Onderzoeker: “Wat is er eigenlijk al over geschreven?”
3. Op zoek naar “voorlopige bronnen” (Google, Google Scholar)
4. Ontstaan van voorlopige “stellingen” bij het lezen van de (50) artikelen (Grounded Theory)
5. Start met handmatig matchen van artikelen en stellingen aan de hand van een daartoe gecreëerde XLS.
6. Mini-workshop “spel met stellingen” op de K80 conferentie

Werksessie “hoe verder?” (oktober 2024)

7. Onderzoek naar de mogelijkheid om AI-tools in te zetten. ChatGPT brengt ons op het pad van Elicit.com als beste optie voor het ‘matchen’.

8. Wij doen allereerst een test met Elicit (o.a. met onze eerste handmatige analyse)

9. Om de bronnen aan de stellingen te analyseren zijn in Elicit kolommen gecreëerd die de stelling verwoorden. Elicit geeft de mogelijkheid om per kolom instructies voor de analyse te geven, een analyse ‘prompt’, als het ware. De volgende prompt is gebruikt om de bronnen aan de stellingen te analyseren: “We doen onderzoek naar de waardering van de ambtenaar. Daarbij doen we een brononderzoek, waarbij we onze stellingen willen toetsen aan de hand van bronnen. Zoek fragmenten uit de bronnen die de volgende stelling ondersteunen of weerleggen: [Stelling]. Geef daarbij specifieke citaten uit de bron die deze stelling weerleggen of ondersteunen.”

10. Elicit produceert per artikel, per stelling de mate van ondersteuning, en vervolgens in meer detail relevante quotes, met reasoning en voorzien van confidence factor. De volledige analyse van Elicit is in een XLS-document te downloaden.

11. De bronanalyse van Elicit is in een Excel document gezet, geordend per cluster. De kolommen vertegenwoordigen de stellingen. Dit is, per stelling, inclusief een kolom Reasoning en een kolom Quotes. De rijen vertegenwoordigen de artikelen, en de cellen bevatten de bronanalyse per artikel, per stelling.

12. We hebben ChatGPT gevraagd om die (52) samenvattingen per stelling samen te vatten in ruim 200 woorden (prompts door Nina, ChatGPT gaf eerst nog zelf verzonden bronnen, dus die moest in toom worden gehouden ...). Daarbij is ChatGPT gevoed met het Excel document. Met de volgende prompt is ChatGPT bevraagd om van de grote hoeveelheid data van Elicit een samenvatting per stelling te maken: “De tekst in dit document bevat een extract van tekstdata van diverse bronnen die de stelling [STELLING] ondersteunen óf weerleggen, met specifieke citaten. Kan je uit de bronnen de belangrijkste inzichten samenvatten die deze stelling ondersteunen én weerleggen?”.

13. ChatGPT oppert vervolgens een antwoord met een aantal van de belangrijkste inzichten die de stelling ondersteunen en weerleggen. Om hier een goed verwoorde samenvatting van te krijgen is een vervolgvraag nodig: “Kan je in 200-300 woorden samenvatten wat de bronnen zeggen?”

14. Als laatste gebruiken we, ter versterking van de samenvatting door middel van een aantal citaten, ChatGPT om drie citaten uit de bronnen te halen die van toepassing zijn: “Geef 3 citaten die deze samenvatting ondersteunen.”

15. Wanneer de antwoorden van ChatGPT inhoudelijk niet accuraat of concreet genoeg zijn, kan worden doorgevraagd, bijvoorbeeld om de samenvattende analyse uit te breiden, inhoudelijk meer te richten op een bepaald onderwerp, of om nog meer citaten te geven. Hier komt ook weer het mensenwerk bij: het overgrote deel van de bronnen hadden wijzelf al gelezen, dus hadden wij een duidelijk beeld van de inhoud van de bronnen. Dit heeft ervoor gezorgd dat wij goed konden bepalen of de antwoorden van ChatGPT inhoudelijk nauwkeurig genoeg waren.

16. De quotes per stelling zijn kwalitatief geanalyseerd (tellen van citaten per artikel) met meeweging van de Confidence levels. Zo kom je op een (relatief) aantal quotes per stelling plus een gewogen gemiddeld confidence level.

17. Volgende stap: de 200 woorden per stelling bundelen tot een bloemlezing per categorie, liefst aangevuld met kwantitatieve info per stelling. En dan op zoek naar aansprekende quotes om daadwerkelijk te citeren, met bronvermelding. En dan een referentielijst van alle bronnen. De samenvattingen van elke stelling per cluster is aan ChatGPT gevoed en vervolgens in ChatGPT gevraagd om daar een verhaal van te maken in maximaal 600 woorden.

18. En dan weer even stilstaan. Hoe willen we met ons rapport de lezer niet alleen informeren maar vooral ook activeren. First things first: eerst nog onze eigen resultaten goed checken.

Verdere analyse- en validatiefase (oktober-november 2024)

19. Kwantitatieve analyse kan volgens Cees op 2 niveaus: (a) op artikelniveau wordt gezegd of het de stelling volledig of gemixt of niet ondersteunt; dat is te kwantificeren; en (b) de gewogen factoren doen dan dienst als confidence factor bij die uitspraken.

20. De nieuwe “W stelling van Alex Peters” levert geen ‘hits’ op. Prima uitlegbaar overigens. Mede omdat de stelling pas later is toegevoegd en we onze bronnen niet speciaal hebben uitgebreid. Dit inspireerde wel tot het ‘testen’ van Elicit met een ‘bewust vreemde stelling’: “De kleur van de verlichting op de parkeerplaats van het gemeentehuis zou geel moeten zijn.” Elicit levert hier een nul-resultaat en dat biedt vertrouwen.

22. We hebben ook nog eens ChatGPT aan de tand gevoeld omtrent het ongevraagd aanhalen van andere bronnen en dito citaten. Doorvragen (inclusief excuses van ChatGPT) levert op dat die bronnen niet eens echt bestaan en dat ChatGPT de citaten gewoon creëert. (Zou het zo erg kunnen worden dat ChatGPT er achteraf nog het bewuste bronartikel bij zou kunnen produceren? Dat is echt freaky.)

23. Verder doen we nog een extra check door (drie items uit) onze aanvankelijke handmatige selectie te vergelijken met de Elicit uitkomsten (actie Nina). Hieruit is gebleken dat Elicit eenzelfde soort conclusie trekt over de ondersteuning (of weerlegging) van de bronnen per stelling, én dezelfde citaten eruit haalt. Echter, de analyse van Elicit is een stuk uitgebreider dan die van onszelf; Elicit vindt veel meer citaten en relateert aan meer stellingen. Er is dus geen tegenstrijdigheid gevonden, maar wel vastgesteld dat Elicit ruim analyseert.

24. Tenslotte ontstond het idee om de dataset en vraagstelling nog eens in Co-Pilot over te doen. Daar zijn we uiteindelijk niet meer aan toegekomen, mede door beperkte toegang tot Co-Pilot.

25. Omwille van pragmatische overwegingen ervoor gekozen het gemiddelde Confidence Level van de verzamelde comments bij een stelling niet nog eens te laten wegen door het tellen van de citaten per artikel. De afwijking blijkt marginaal.

26. Tweede overweging: kan AI nog helpen bij het “tellen” van de mate van support, zoals weergegeven in de eerste kolommen van elk werkblad. Of moet het toch handmatig gevinkt worden?

27. ChatGPT heeft geholpen de “Support statements” te ‘tellen’ met de prompt: “Kijk naar column [columnnummers]. Elke column is een stelling. Analyseer voor elke column hoeveel bronnen de stelling ondersteunen. Geef daarnaast ook de ondersteuningspercentages.”

28. Door per stelling “ #supporting articles x average confidence level” te vermelden hebben we voldoende gedaan om onze data te valideren.

29. En niet te vergeten de cross-check met onze handmatige analyses in het begin van het proces (acties voor alle drie)

Rapportagefase (november-december 2024)

30. We laten de AI-tools nu achter ons en gaan de summary teksten per categorie tot een verhaal breien. Karin Leiderschap, Cees Processen en Nina Gedrag. Beoogd product beslaat 500-600 woorden en bevat 1 a 2 'streamers'.

31. Nou ja ... niet dus. ChatGPT bleek prima in staat een uitgebreide bloemlezing uit de Elicit sheet te destilleren, en vervolgens het geheel nog eens naar onze kaderstelling in te dikken. De samenvattingen van de stellingen zijn per cluster in een document gezet. Vervolgens is het document aan ChatGPT gevoed en gevraagd: "Dit document bevat de kerninzichten van de stellingen [bijvoorbeeld G1 t/m G5], die zich richten op de waarde en waardering van de ambtenaar. Kun je deze inzichten samenvoegen tot een vloeiend, samenhangend verhaal van maximaal 600 woorden? Het verhaal moet een verbindende lijn hebben tussen de stellingen, met de belangrijkste kerninzichten op een manier die inspirerend en waardierend is voor de rol van de ambtenaar. Zorg ervoor dat het als een geheel leest." Dat leverde een goede basis op die hooguit nog wat ontduubeling en eindredactie vergde, en aldus is ingepast in het voortgangsrapport dat overigens wel geheel door mensenhersens en -vingers is geproduceerd.

32. Gek genoeg bleken al onze AI-tools, wat we ook probeerden, geen antwoord te kunnen geven op de vraag hoeveel pagina's onze 50 bronnen beslaan, of hoeveel woorden ze samen bevatten.

33. Voor de referentielijst dachten we ook dankbaar gebruik te kunnen maken van een AI-tool, maar dat ging niet op rolletjes. Het is uiteindelijk handmatig gedaan. Zo gaf ChatGPT bijvoorbeeld niet verder dan een lijst van 30 bronnen, en daarop 'aangesproken' kwam het programma opeens met 60 bronnen... Het zegt maar weer dat de AI-instrumenten uiterst zorgvuldig bediend dienen te worden, en altijd een menselijke check behoeven

Communicatiefase (december 2024 - maart 2025)

34. Voor de volledigheid vermelden we de overweging dat dit (generatief actie-)onderzoek niet eindigt maar welbeschouwd pas begint bij het uitbrengen van de bevindingen en aanbevelingen.

35. Op 7 februari mogen we een workshop geven op de tweedaagse conferentie van de Vereniging Directeuren Publieksdiensten (VDP). Daar zien we het generatieve effect ontstaan van de 23 bouwstenen van de toekomstige gemeenteorganisatie. Het publiek denkt en droomt enthousiast mee. We gebruiken Padlet.com om hun ideeën te verzamelen. Die leggen we vervolgens in een 'live demo' voor aan ChatGPT. Dat levert natuurlijk een krachtig resumé op van de meest haalbare actievoorstellen. En ook een 'profielschets' van het publiek dat de respons aanleverde. Leuk en leerzaam.

36. Hoe bondig en beknopt we het onderzoeksrapport (genaamd 'voortgangsverslag') ook maakten, het was nog steeds te veel tekst naar onze smaak. Met Adobe als familie van allerlei geautomatiseerde grafische hulpmiddelen gaat vormgever Vincent van der Geest aan de slag om een interactieve PDF te produceren. De vorm dwingt ons nogmaals tot het volledig herindelen en indikken van de tekst. Schrijven is schrappen.

Dit verhaal is hopelijk nooit af, en vervolgt zich idealiter in de dagelijkse praktijk van de organisatie als werkomgeving van onze gemeenteamttenaren. We schrijven op dit moment 15 april 2025 en over een maand voorzien we de officiële lancering van het 'begin' van dit onderzoek.



**Vereniging van
Nederlandse Gemeenten**

Nassaulaan 12
2514 JS Den Haag
+31 70 373 83 93

info@vng.nl

mei 2025