

AI als schrijfassistent voor heldere gemeentelijke brieven, gebaseerd op B1-niveau Nederlands

KAAN GÖGCAY, AMBER VAN HASSEL, AMIR JACOBS, TERENCE TERZOL, and BIBIËNE WÜST*,
Hogeschool van Amsterdam, Nederland

ABSTRACT (SAMENVATTING)

Additional Key Words and Phrases: KEYWORDS

*Alle auteurs hebben in gelijke mate bijgedragen aan dit onderzoek.

Authors' Contact Information: Kaan Gögcay, kaan.gogcay@hva.nl; Amber van Hassel, amber.van.hassel@hva.nl; Amir Jacobs, amir.jacobs@hva.nl;
Terence Terzol, terence.terzol@hva.nl; Bibiëne Wüst, bibiene.wust.wust@hva.nl,
Hogeschool van Amsterdam, Amsterdam, Nederland.

1 Introductie

1.1 Context

In Nederland hebben overheden de taak om inwoners duidelijk te informeren over wetgeving, voorzieningen en procedures [37]. Schriftelijke communicatiemiddelen, zoals brieven, spelen hierbij een belangrijke rol, omdat deze vaak (juridische) informatie bevatten over rechten en plichten (Bijlage A). Op nationaal niveau is er toenemende aandacht voor begrijpelijke taal binnen publieke organisaties. Dit is belangrijk, omdat ongeveer 2,5 miljoen mensen in Nederland moeite hebben met lezen en schrijven [14]. Het is essentieel dat iedere inwoner overheidscommunicatie goed kan begrijpen. Daarom gebruiken veel gemeenten een schrijfwijzer die bepaalt hoe zij teksten op B1-niveau opstellen. Zo heeft onder andere de Gemeente Amsterdam een eigen schrijfwijzer ontwikkeld [29]. Ook de OVER-gemeenten (Gemeenten Oostzaan en Wormerveer) maken gebruik van een schrijfwijzer om medewerkers te ondersteunen bij hun communicatie met inwoners [30]. Daarnaast verkent de OVER-gemeenten de inzet van nieuwe technologieën, zoals kunstmatige intelligentie (AI), ter ondersteuning van haar communicatieve processen.

1.2 Probleem

Ondanks bestaande richtlijnen voor begrijpelijke communicatie [29, 30], blijken gemeentelijke brieven in Nederland in de praktijk vaak moeilijk te begrijpen voor een groot deel van de bevolking. Dat dit geen incidenteel probleem is, blijkt uit onderzoek naar Nederlandse publieke organisaties: een analyse van 240 teksten van 70 organisaties toont aan dat veel officiële communicatie te complex is, onder andere door lange zinnen en ingewikkeld woordgebruik [52]. Dit kan leiden tot misverstanden en problemen, zoals het niet lezen van essentiële informatie of het missen van kansen doordat mensen niet begrijpen wat er van hen wordt verwacht [28, 37].

Ook binnen de OVER-gemeenten speelt dit vraagstuk. Medewerkers beoordelen elkaar teksten op of deze voldoet aan de regels van de OVER-schrijfwijzer, maar doordat de schrijfwijzer omvangrijk is en niet altijd consequent wordt toegepast, kunnen deze beoordelingen onbetrouwbaar zijn. Hierdoor ontvangen inwoners soms brieven die zij niet goed begrijpen, wat de kans op miscommunicatie vergroot. Daarnaast blijkt uit een interview (Bijlage A) dat het evalueren en herschrijven van teksten volgens de schrijfwijzer extra tijd kost.

Deze combinatie van een algemeen landelijk probleem en knelpunten binnen de OVER-gemeenten creëert de behoefte aan een ondersteunend systeem dat medewerkers helpt bij het vereenvoudigen of beoordelen van teksten, zodat brieven begrijpelijker worden en de communicatie met inwoners verbeterd.

1.3 Bestaand werk over het probleem

Om de begrijpelijkheid van overheidscommunicatie te verbeteren, hebben verschillende gemeenten de afgelopen jaren maatregelen genomen. Voorbeelden hiervan zijn lokale schrijfwijzers [29, 30], algemene richtlijnen voor duidelijke taal [27] en landelijke initiatieven zoals Direct Duidelijk [28]. Direct Duidelijk is een programma van de Rijksoverheid dat organisaties helpt hun communicatie toegankelijker te maken voor iedereen. Deze hulpmiddelen bieden praktische adviezen over woordkeuze, zinslengte en toon, maar de toepassing in de praktijk blijkt wisselend.

Binnen het onderzoeksgebied van Automatic Text Simplification (ATS) zijn al technologische ontwikkelingen gedaan. Taalmodellen zoals T5 en GPT worden internationaal gebruikt voor het vereenvoudigen van vooral Engelstalige teksten [45, 46]. Hoewel deze modellen laten zien dat AI in staat is teksten te analyseren en te versimpelen, worden zij binnen de OVER-gemeenten nog niet ingezet (Bijlage A). Overheidsorganisaties kunnen hun documenten vanwege privacy- en veiligheidsredenen niet zomaar aanbieden aan externe platforms [19]. Dit kan ertoe leiden dat privacygevoelige informatie uitlekt. Daarnaast zijn bestaande modellen niet specifiek getraind op de eisen en schrijfrichtlijnen die gelden voor Nederlandse gemeentelijke communicatie. Daarom is er behoefte aan een veilige AI-oplossing die kan helpen bij het herschrijven of beoordelen van brieven, waarbij bestaande gemeentelijke richtlijnen worden gevolgd en de menselijke controle behouden blijft.

1.4 Wetenschappelijk gat in bestaand werk

Het wetenschappelijke gat ligt vooral bij de beperkingen van huidige methoden voor automatische tekstvereenvoudiging wanneer deze worden ingezet voor Nederlandse gemeentelijke communicatie met een juridische inhoud. Bestaand onderzoek naar Automatic Text Simplification (ATS) en Lexical Simplification (LS) richt zich vooral op Engelstalige domeinen. Daarbij wordt weinig rekening gehouden met teksten die juridische gevolgen kunnen hebben [33, 58].

Vanuit recent Nederlands onderzoek is op verschillende manieren gekeken naar tekstcomplexiteit en tekstvereenvoudiging. Zo is in de taalmodellen "Geen makkie" en "LSBertje" onderzocht welke taalkenmerken bijdragen aan de moeilijkheid van Nederlandse zinnen en teksten [36], en beschrijft een recent benchmarkonderzoek hoe complex gemeentelijke teksten gemiddeld geschreven zijn [53]. Monitoronderzoek laat daarnaast zien hoe burgers overheidsteksten ervaren en waar zij vastlopen [23, 24], terwijl schrijfwijzers en richtlijnen uitwerken wat precies onder B1-niveau wordt verstaan en hoe overheden begrijpelijk kunnen schrijven [29] [30] [27] [28]. Deze studies richten zich echter vooral op algemene leesbaarheid en doelgroepgericht schrijven, en veel minder op juridisch beladen gemeentelijke brieven waarin formuleringen directe rechtsgevolgen kunnen hebben. Daardoor sluiten ze nog niet volledig aan bij de eisen van gemeentelijke communicatie, waarin juridische nauwkeurigheid essentieel is.

In dit onderzoek wordt daarom ingezoomd op drie tekortkomingen in de bestaande wetenschap:

(1) Behoud van betekenis: hoe combineer je eenvoud met juridische nauwkeurigheid?

Bestaand onderzoek naar ATS en LS richt zich voornamelijk op leesbaarheid, bijvoorbeeld door kortere zinnen en eenvoudigere woorden [36, 53], maar gemeentelijke brieven bevatten vaak juridische bepalingen, uitzonderingen en verplichtingen. Een kleine herformulering kan leiden tot een andere interpretatie, met mogelijke gevolgen voor besluiten, bezwaarprocedures of rechten van inwoners. Het huidige onderzoek biedt geen aanpak om automatisch te controleren of een vereenvoudiging de semantische inhoud volledig behoudt, en voldoet aan de juridische formuleringen die gemeenten verplicht moeten communiceren.

(2) Gebruikersgerichte AI-assistentie: hoe blijft de medewerker eindverantwoordelijk?

Veel ATS-onderzoek gaat uit van volledig automatische output [58]. Voor de OVER-gemeenten is dat onwenselijk: medewerkers moeten altijd kunnen beoordelen, corrigeren en verantwoorden wat er in een brief staat. Human-centered AI benadrukt daarom transparantie, uitlegbaarheid en samenwerking tussen de mens en model [39, 54].

Huidig onderzoek beschrijft nog geen concreet ontwerp voor een Nederlandse human-in-the-loop-schrijfassistent die:

- (a) suggesties geeft in lijn met een lokale schrijfwijzer,
- (b) uitlegt waarom een zin complex is of tegen een regel ingaat,
- (c) en de medewerker actief ondersteunt in het verbeteren van een tekst zonder autonomie over te nemen.

(3) Domein-specifieke evaluatie: hoe bepaal je of een tekst echt voldoet aan de regels van de OVER-schrijfwijzer en juridisch correct is?

Gangbare evaluatiemethoden, zoals SARI of n-gram-metrics [53, 58], meten voornamelijk taalkundige eenvoud. Deze evaluaties zeggen echter weinig over de mate waarin een tekst daadwerkelijk voldoet aan B1-criteria uit Nederlandse schrijfwijzers [27, 29, 30], noch over het behoud van feiten, juridische voorwaarden en uitzonderingen. Evenmin geven ze inzicht in de vraag of een voorgestelde vereenvoudiging uitvoerbaar en verantwoord is binnen gemeentelijke praktijk.

Daarom ontbreekt op dit moment een domeinspecifiek evaluatiekader dat niet alleen de afstand tot de regels van de OVER-schrijfwijzer betrouwbaar kan bepalen volgens gemeentelijke richtlijnen, maar tegelijkertijd kan toetsen of de oorspronkelijke betekenis en juridische juistheid behouden blijven.

1.5 Voorstel

Het voorstel is om een AI-gestuurde applicatie (AI-tool) te ontwikkelen die medewerkers van de OVER-gemeenten ondersteunt tijdens het schrijven van brieven aan de bewoners. Hierbij wordt een taalmodel gebruikt om bijvoorbeeld automatisch moeilijke woorden en lange zinnen te signaleren. Hierbij worden automatisch verbeteringssuggesties gegeven die direct aansluiten op de schrijfwijzer van de OVER-gemeenten. Bij elke suggestie krijgt de medewerker een toelichting en verwijzing naar de regel in de schrijfwijzer waar de suggestie betrekking tot heeft. Hierdoor wordt de tekst verbeterd en leren medewerkers ook beter begrijpen waarom een bepaalde formulering past binnen de gemeentelijke richtlijnen.

De toepassing wordt ontwikkeld voor gebruik in een lokale omgeving, zodat teksten en persoonsgegevens binnen de gemeentelijke infrastructuur blijven en niet worden gedeeld met externe partijen. Er wordt onderzocht of de AI-tool kan worden geïntegreerd als extensie binnen Microsoft Word, zodat medewerkers het hulpmiddel direct in hun vertrouwde werkomgeving kunnen gebruiken. Er is voor Word gekozen omdat dit programma dagelijks door medewerkers van de OVER-gemeenten wordt gebruikt voor het opstellen van brieven (Bijlage A). Om het hulpmiddel gebruiksvriendelijk en flexibel inzetbaar te maken, wordt de AI-tool zo ontworpen dat deze aan- en uitgezet kan worden. Medewerkers stellen eerst de volledige brief op, waarna de AI-tool wordt ingezet om de tekst te beoordelen.

Met behulp van deze AI-tool kunnen medewerkers praktisch en betrouwbaar geholpen worden bij het schrijven van begrijpelijke brieven aan inwoners. Hiermee wordt de kwaliteit van gemeentelijke brieven verhoogd en wordt de dienstverlening aan inwoners toegankelijker en duidelijker.

2 Achtergrond

2.1 Basiskennis

2.1.1 Maatschappelijke context.

In Nederland moeten gemeenten inwoners duidelijk informeren over hun rechten en plichten. Communicatie moet begrijpelijk zijn, zodat inwoners daarop adequaat kunnen handelen [28]. De OVER-gemeenten streven daarom naar B1-taalgang niveau, zoals blijkt uit een gesprek met medewerkers van de OVER-gemeenten (Bijlage A). Uit deze vragensessie bleek ook dat het consistent toepassen van B1-richtlijnen tijd kost en vraagt om taalvaardigheid die niet iedere medewerker bezit. Hierdoor ontstaat een maatschappelijke behoefte aan technologische ondersteuning om brieven voor inwoners beter begrijpelijk te maken.

2.1.2 Bestaande oplossingen bij de OVER-gemeenten.

Binnen de OVER-gemeenten worden verschillende middelen ingezet om de begrijpelijkheid van brieven te verbeteren (Bijlage A). Medewerkers lezen elkaar brieven na om te controleren of deze voldoet aan de regels van de OVER-schrijfwijzer en juridisch juist zijn. Daarnaast maken de medewerkers gebruik van de gemeentelijke schrijfwijzer, waarin richtlijnen staan voor het formuleren van teksten op B1-niveau. Ook worden online hulpmiddelen, zoals [51], gebruikt om woordkeuze te beoordelen op begrijpelijkheid (Bijlage A). Hoewel deze middelen ondersteuning bieden, blijkt het lastig om ze toe te passen. Voornamelijk wanneer de medewerkers een hoge tijdsdruk ervaren of niet precies weten welke onderdelen van de tekst wel of niet voldoen aan de B1-taalgang niveau.

2.1.3 Herschrijven van tekst met AI.

Het automatisch versimpelen van teksten is een taalverwerkingsprobleem waarvoor taalmodellen in de praktijk worden ingezet [32, 59]. Verschillende studies laten zien dat taalmodellen alternatieve woorden of zinnen kunnen aanbieden voor het herschrijven van complexe formuleringen [32, 59]. In [59] wordt bijvoorbeeld een model voorgesteld dat zinnen vereenvoudigt door een Transformer-architectuur te combineren met een externe parafrase-database (Simple PPDB). In plaats van simpelweg ‘moeilijke’ woorden te vervangen door ‘makkelijkere’ synoniemen, leert het model welke vereenvoudigingsregels het meest passend zijn en past deze gecontroleerd toe tijdens het genereren van de nieuwe zin. Taalmodellen kunnen dus eventueel gebruikt worden als ondersteuning voor het (her)schrijven van (begrijpelijke) gemeentelijke brieven.

2.1.4 State-of-the-art.

Er is een aanzienlijke hoeveelheid onderzoek verricht naar tekst-versimpeling [22]. Traditioneel werden hiervoor klassieke tekstverwerkingsmodellen ingezet, zoals bag-of-words-modellen. Deze modellen negeren echter de volgorde van woorden in een zin, waardoor belangrijke contextuele informatie verloren gaat. Modellen, zoals N-grammodellen, vormen een gedeeltelijke verbetering doordat zij korte reeksen opeenvolgende woorden meenemen. Toch kunnen N-grammodellen nog steeds geen langere verbanden in een tekst goed vastleggen.

De interesse in sequentiemodellen nam in 2015 toe, met de heropleving van Recurrent Neural Networks (RNN's) [22]. RNN's kunnen informatie over langere reeksen woorden verwerken en daarmee context behouden die verder reikt dan de lokale structuur van N-grammodellen. Ondanks deze vooruitgang kampen RNN's met beperkingen, zoals het moeilijk vasthouden van informatie over zeer lange tekstreeksen en hoge rekenkosten bij het trainen.

Daarna kwam de introductie van de Transformer-modellen. Dit volgende een doorbraak binnen Natural Language Processing (NLP) [22]. Deze modellen maken gebruik van (self-)attention, waardoor zij relaties tussen woorden over de gehele tekst kunnen herkennen. Ze zijn efficiënter te trainen en beter schaalbaar dan de eerder genoemde modellen. In deze context is het van belang dat de AI-tool relaties tussen woorden in de hele brief kan herkennen, zodat de suggesties voor de medewerker logisch aansluiten bij de rest van de context. Hierdoor zijn Transformer-modellen, binnen de context van de OVER-gemeenten, het meest geschikt voor toepassing in de AI-tool die gemeentelijke brieven herschrijft. Deze Transformer-modellen kunnen gebruikt worden op meerdere manieren, zoals in een naïve-Retrieval-Augmented Generation (naïve-RAG). implementatie, waarbij externe informatie wordt toegevoegd aan de input dat het model krijgt [43].

2.2 Stakeholderanalyse

In Tabel 1 staat een overzicht van de stakeholders, hun belangen, ervaren problemen en modelvereisten binnen het huidige proces van het herschrijven en controleren van brieven op B1-niveau.

Table 1. Stakeholderanalyse - belangen, ervaren problemen en modelvereisten

Stakeholder	Belangen	Ervaren problemen	Modelvereisten
OVER-gemeenten	Een snelle tool die ervoor zorgt dat de werkdruk van de medewerkers minder belast wordt, waardoor het werk sneller afkomt.	Op dit moment kost het controleren en schrijven van deze brieven veel tijd. Ook hebben niet alle medewerkers dezelfde (specialistische) kennis over het B1-taalniveau.	Een snelle tool die de medewerkers van de gemeente helpen om de werkdruk te verlagen en werk sneller af te ronden.
Inwoners van de OVER-gemeenten	Begrijpelijke teksten waarin duidelijk is wat de mededeling is. Daarnaast moeten de teksten ook inclusief geschreven zijn; niemand zou zich buitengesloten of gediscrimineerd moeten voelen.	Verwarring over de beschreven situatie en vervolgstappen, waardoor problemen ervaren worden bij het krijgen van hulp op bijvoorbeeld sociaal vlak en bij betalingen.	Dat privacygevoelige informatie niet wordt gelekt en de teksten inclusief geschreven zijn. Daarnaast moet de brief een persoonlijke toon bevatten, als de context dat vraagt.
Medewerkers van OVER-gemeenten die de brieven schrijven/controleren	Een tool die de medewerkers helpt om de brieven begrijpelijker te verwoorden met gebruik van de schrijfwijzer die gebaseerd is op B1-niveau. Hierbij moet rekening gehouden worden met juridische aspecten en het beschermen van privacygevoelige informatie.	Op dit moment kost het controleren en schrijven van deze brieven veel tijd. Ook hebben niet alle medewerkers dezelfde (specialistische) kennis over het B1-taalniveau.	De verbeterde teksten moeten gebaseerd zijn op de schrijfwijzer die de OVER-gemeenten aanbiedt. De verbeterde teksten moeten daarbij ook refereren naar de schrijfwijzer.
ICT-afdeling	De tool moet lokaal gerund worden, of op de eigen server, en iedereen die brieven schrijft binnen de OVER-gemeenten kan deze gebruiken. Hierbij moet voorkomen worden dat privacygevoelige gegevens gelekt worden.	Als data extern verwerkt wordt of als er onvoldoende controle is over wie de tool gebruikt, kunnen beveiligingsrisico's ontstaan.	Het model moet lichtgewicht zijn, zodat het lokaal kan draaien op de bestaande infrastructuur van de gemeente.
Overheid	De tool moet voldoen aan de wetten en regelgeving die door de overheid zijn opgesteld m.b.t. technische, ethische en juridische eisen.	Inwoners van gemeenten begrijpen niet altijd waar ze recht op hebben, doordat zij brieven niet begrijpen.	De brieven die verstuurd worden moeten op B1-niveau geschreven zijn, volgens de richtlijnen van de overheid [2].

Figuur 1 toont daarnaast de positie van de stakeholders in de macht-belangdiagram. Medewerkers die brieven schrijven en controleren, evenals de OVER-gemeenten, hebben zowel veel belang als veel macht en spelen daarom een centrale rol in de ontwikkeling van de AI-tool. De ICT-afdeling beschikt over veel macht vanwege de technische randvoorwaarden, maar heeft minder direct inhoudelijk belang. Inwoners van de OVER-gemeenten en de overheid hebben juist veel belang bij duidelijke communicatie, maar weinig macht over het ontwerp van de tool.



Fig. 1. Macht-belangdiagram

2.3 Huidige situatie

Bij het opstellen van een brief doorloopt de OVER-gemeenten een twee-staps controle. Een medewerker schrijft de brief eerst inhoudelijk correct en herschrijft deze vervolgens naar begrijpelijke taal met hulpmiddelen zoals de OVER-schrijfwijzer. Daarna controleren ten minste twee medewerkers de begrijpelijkheid van de brief en plaatsen opmerkingen in een Word-document. Deze wordt vervolgens teruggestuurd naar de medewerker die de brief heeft opgesteld, zodat diegene de feedback kan verwerken. Na de feedback-verwerking ontvangt de inwoner de brief. Begrijpt de inwoner deze niet, dan kan de inwoner contact opnemen met de gemeente. Bij begrip volgt eventueel actie op basis van de inhoud van de brief. Dit proces is visueel weergegeven in Figuur 2.

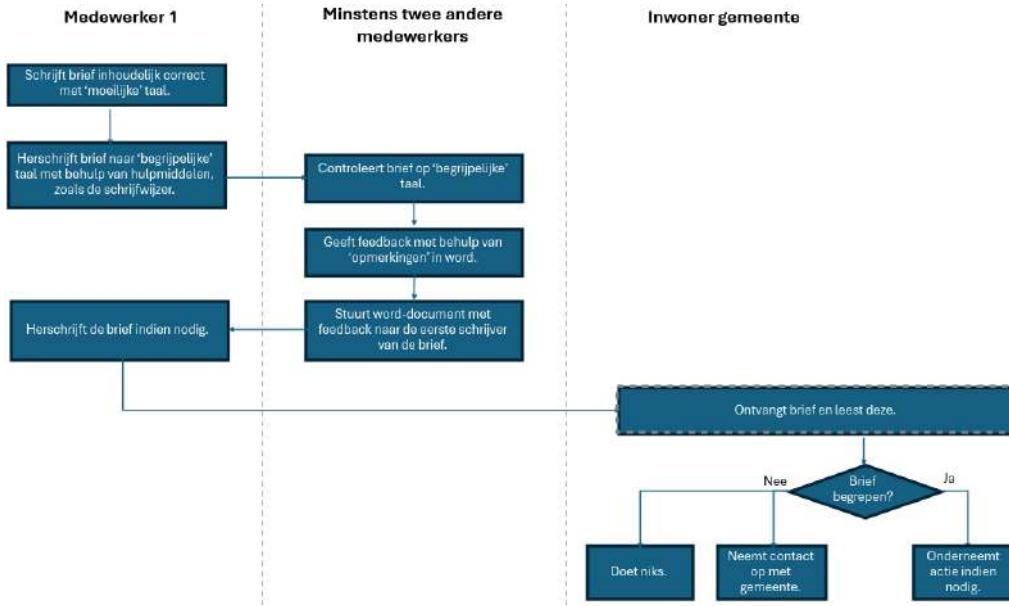


Fig. 2. Huidige situatie binnen de OVER-gemeenten

2.4 Technische, maatschappelijke en ethische overwegingen voor het inzetten van AI in dit vraagstuk

2.4.1 *Technische overwegingen.*

De inzet van AI brengt uitdagingen met zich mee. Zo zijn bestaande ATS-modellen niet specifiek afgestemd op de communicatie van Nederlandse gemeenten. Hierdoor kunnen deze modellen onnauwkeurige of juridisch onjuiste vereenvoudigingen produceren, of suggesties geven die niet overeenkomen met de schrijfwijzer van OVER-gemeenten. Vanuit privacy- en veiligheidsoogpunt (AVG) is het noodzakelijk dat persoonsgegevens niet via externe servers worden verwerkt [19]. Dit betekent technisch gezien dat de tool lokaal moet draaien en dat de data de gemeentelijke omgeving niet mag verlaten. Ook moet tijdelijke opslag van tekst geminimaliseerd of direct verwijderd worden om datalekken te voorkomen [19].

2.4.2 *Ethische overwegingen.*

Een belangrijk ethisch risico is automation bias. Dit houdt in dat medewerkers onder tijdsdruk te veel kunnen vertrouwen op AI-suggesties, waardoor foutieve suggesties alsnog worden goedgekeurd [13]. Op deze manier worden foutieve of juridisch problematische aanpassingen overgenomen zonder kritische beoordeling. Daarom moet het systeem werken volgens het human-in-the-loop-principe, waarbij medewerkers altijd zelf bepalen of suggesties correct en passend zijn voor de desbetreffende brief [40] (Bijlage E).

2.4.3 *Maatschappelijke overwegingen.*

Tot slot is het essentieel dat medewerkers worden getraind in het gebruik van het systeem. Medewerkers moeten worden geïnformeerd over de beperkingen van het systeem, zodat zij weten dat suggesties niet altijd bruikbaar zijn en handmatige aanpassingen nodig blijven. De ontwikkeling en implementatie vereisen daarom nauwe samenwerking met de stakeholders om knelpunten tijdig te signaleren en aan te pakken.

3 Requirements

Voorafgaand aan de ontwikkeling van de AI-tool zijn requirements opgesteld en geprioriteerd met behulp van een MoSCoW-analyse. Op deze manier zijn de belangrijkste eisen vastgesteld waaraan het uiteindelijke product moet voldoen (Tabel 2). De volledige lijst met requirements, inclusief eisen zonder prioriteit, is opgenomen in Bijlage C. Hierbij worden de volgende codes gebruikt: ER = ethische requirements, JR = juridische requirements, FR = functionele requirements, TR = technische requirements, OR = organisatorische requirements en DR = duurzaamheids requirements.

De organisatorische en duurzaamheidsvereisten zijn niet in dit hoofdstuk opgenomen maar uitsluitend in de bijlage vermeld, omdat het regelen van organisatorische en duurzame zaken binnen de gemeente buiten de scope van dit onderzoek valt. Dit onderzoek richt zich namelijk op het onderzoeken en ontwerpen van de AI-oplossing en niet op de interne organisatie-inrichting of implementatie binnen de OVER-gemeenten. Deze eisen zijn echter wel van belang en zullen als must-haves gelden wanneer de tool daadwerkelijk wordt geïmplementeerd binnen de OVER-gemeenten.

Table 2. Alle must-have requirements

RQ-nr	Requirement	Stakeholder	Toelichting
ER01	Gebruikers moeten de AI-tool begrijpen [21].	Medewerkers die brieven schrijven/controleren	De medewerker moet de AI-tool correct kunnen gebruiken voor het herschrijven van brieven.
ER02	Een woordsuggestie mag niet zonder toestemming van de medewerker worden toegevoegd (bijv. om bias te voorkomen). Het systeem moet een human-in-the-loop hebben; de mens blijft eindverantwoordelijk en de AI-tool ondersteunt slechts [21].	Medewerkers die brieven schrijven/controleren; Overheid	Medewerkers: De medewerker wil het herschrijven niet volledig aan de AI overlaten omdat hij/zij verantwoordelijk blijft voor de inhoud. Overheid: Wil dat inwoners correct worden geïnformeerd en dat risico op verkeerde verwoordingen wordt geminimaliseerd zodat wetgeving correct wordt vertaald.
ER03	De ontwikkeling van de AI-tool (datasets, modelkeuze, training) en de prestaties moeten gedocumenteerd worden [21].	Overheid; ICT-afdeling	Overheid: Wil inzicht in de documentatie om te controleren of het model voldoet aan wetgeving. ICT-afdeling: Moet geïnformeerd worden over de ontwikkeling zodat zij weten hoe de tool onderhouden moet worden.
ER04	Medewerkers moeten de mogelijkheid krijgen om te kiezen voor het gebruik van de AI-tool [21].	Medewerkers die brieven schrijven/controleren	Medewerkers moeten de AI-tool kunnen aan- of uitzetten wanneer zij deze niet willen gebruiken.

RQ-nr	Requirement	Stakeholder	Toelichting
ER05	De beperkingen en capaciteiten van de AI-tool moeten aan stakeholders worden gemeld; bijvoorbeeld dat de tool slechts een hulpmiddel is en niet altijd de beste suggesties geeft [21].	Medewerkers die brieven schrijven/controleren	Medewerkers moeten weten dat zij met een AI-tool werken en dat tekstsuggesties niet altijd correct zijn.
JR01	Persoonlijke informatie van burgers mag niet worden opgeslagen of gebruikt voor (verdere) training van het model. De validatie- en testdata bestaat uit anonieme brieven. [4, 19, 50]	Inwoners van OVER-gemeenten	De persoonlijke gegevens van inwoners mogen niet zonder toestemming worden verspreid.
JR02	De AI-tool mag geen persoonsgegevens delen met externe partijen of servers. [4, 48, 49]	Overheid; Inwoners van Oostzaan	Overheid: Moet controleren of dit in de praktijk wordt nageleefd. Inwoners: Hun persoonlijke gegevens mogen niet zonder toestemming worden verspreid. OVER-gemeenten: Zijn eindverantwoordelijk en mogen geen persoonlijke gegevens delen met externe partijen of servers.
FR01	De AI-tool geeft suggesties, maar de medewerker beslist wat hij of zij met deze beslissing doet [55].	Medewerkers die brieven schrijven/controleren	De medewerker moet bij alle suggesties die de AI-tool geeft, zelf bepalen of de tekst aangepast wordt.
FR02	De medewerker kan de suggestie altijd handmatig afwijzen of goedkeuren [55].	Medewerkers die brieven schrijven/controleren	De medewerker moet bij alle suggesties die de AI-tool geeft, zelf bepalen of de tekst aangepast wordt.
FR03	De AI-tool moet de medewerker informeren dat de suggesties slechts ter ondersteuning dienen en dat de medewerker de eindbeslissing neemt [55].	Medewerkers die brieven schrijven/controleren	De medewerkers mogen er niet vanuit gaan dat de AI-tool altijd de beste weerspiegeling geeft van hun eigen tekst. De AI-tool dient alleen als ondersteuning.
FR05	De AI-tool moet een lagere LiNT-II-score hebben dan het baseline model.	ICT-afdeling	Hierdoor wordt verzekerd dat de AI-tool beter presteert dan het baseline model.
FR06	Resultaten moeten betrouwbaar zijn voor alle verschillende briefcategorieën.	OVER-gemeenten	Hierdoor is de AI-tool bruikbaar voor alle brieven die geschreven worden binnen de OVER-gemeenten.
TR01	Bewonersinformatie wordt direct verwerkt en niet opgeslagen. Dit moet in de code van de AI-tool worden verwerkt [20].	Inwoners van de OVER-gemeenten; Overheden/toezichthouders	Inwoners: Willen niet dat hun persoonlijke informatie buiten de gemeente lekt. Overheden/toezichthouders: Willen verzekerd zijn dat informatie niet buiten hun netwerk lekt.

RQ-nr	Requirement	Stakeholder	Toelichting
TR02	User Interface (UI) elementen tonen aan dat de AI-tool ondersteunend is en niet beslissend [55].	Medewerkers die brieven schrijven/controleren	Medewerkers kunnen hierdoor zien dat de AI-tool een ondersteunend middel is.
TR03	De AI-tool moet gevalideerd worden ten opzichte van het baseline model en prestaties moeten worden gedocumenteerd.	ICT-afdeling	Hierdoor wordt duidelijk of de gemaakte AI-tool beter werkt dan het baseline model.
TR04	De gebruikte datasets moeten gesplitst worden in een validatie- en testset van 50%, 50%.	ICT-afdeling	Dit voorkomt data-leakage en zorgt voor correcte evaluatie.
TR05	De AI-tool moet de brieven lokaal verwerken. (Bijlage A)	Overheden/toezichthouders; ICT-afdeling	Willen niet dat persoonlijke informatie buiten de gemeente lekt.
TR06	De AI-tool moet in 2 seconden de brief controleren, fouten markeren en suggesties tonen. Hierdoor is handmatig zoeken naar suggesties in de schrijfwijzer overbodig.	Medewerkers die brieven schrijven/controleren	Hierdoor is handmatig zoeken naar suggesties in de schrijfwijzer overbodig.

De requirements worden geëvalueerd aan de hand van vooraf vastgestelde acceptatiecriteria. Voor elke requirement uit Tabel 2 is in Tabel 3 de bijbehorende evaluatiemethode opgenomen (hoe deze requirement wordt beoordeeld) en het acceptatiecriterium dat aangeeft wanneer aan de requirement is voldaan.

Table 3. Acceptatiecriteria

RQ-nr	Evaluatiemethode	Acceptatiecriterium
ER01	Paper prototype	Gebruikerstesten uitvoeren waarbij wordt onderzocht of alle instructies duidelijk zijn en of testers het systeem begrijpen.
ER02	Paper prototype	Bij het ontwerpen van het AI-systeem moet de mens eindverantwoordelijk blijven voor de beslissing of een klant alcohol mag kopen.
ER03	Documentatie	In de documentatie moeten de ontwikkeling van het AI-systeem en de nauwkeurigheid van de leeftijd voorspellingen staan.
ER04	Paper prototype	In het paper prototype moet de medewerker de optie krijgen om voor de AI-tool te kiezen.
ER05	Documentatie	Stakeholders ontvangen informatie over prestaties en beperkingen van het systeem.
JR01	Code en dataset	In de code moet staan dat de persoonlijke informatie verwijderd wordt direct na het opslaan van de brief. In de dataset moet gecontroleerd worden of de brieven anoniem zijn.

RQ-nr	Testmethode	Acceptatiecriteria
JR02	Code	In de code moet staan dat de persoonlijke informatie die opgeslagen is verwijderd wordt direct na het opslaan van de brief en dat het schrijven van de brief lokaal gebeurt.
FR01	Code	In de code moet staan dat de weergegeven suggesties eerst door de medewerker geaccepteerd moeten worden voordat de zinnen of woorden aangepast worden.
FR02	Code	In de code moet staan dat de medewerker de suggestie kan afwijzen of goedkeuren.
FR03	Code (UI)	In de code moet een stuk tekst komen dat de medewerker in zijn scherm ziet. Hierin hoort de tekst informatief te zijn, waarbij wordt weergegeven dat de AI-tool slechts ter ondersteuning dient.
FR05	Evaluatie	In de evaluatie moeten de LiNT II-scores van de AI-tool en het baseline model tegenover elkaar gezet worden.
FR06	Evaluatie	De gemiddelde LiNT-II-score van de verschillende briefsoorten moet hoger zijn dan de LiNT-II-score op het baseline model.
TR01	Code	In de code moet verwerkt worden dat bewonersinformatie uit de brieven direct verwerkt en niet opgeslagen wordt.
TR02	Code (UI)	In de code moet staan dat de AI-tool ondersteunend is voor de medewerker en niet beslissend.
TR03	Evaluatie	In de evaluatie moeten de AI-tool en het baseline model tegenover elkaar gezet worden.
TR04	Code en dataset	In de code moet een datasplitsing staan volgens een validatie- en test-set van 50% en 50%.
TR05	Verwerken	De code dient lokaal gedraaid te worden.
TR06	Evaluatie	Er dient een test uitgevoerd te worden met een medewerker, waarin de medewerker een brief verbetert door handmatig de schrijfwijzer te gebruiken en daarnaast dezelfde brief dient te verbeteren door het gebruik van de AI-tool.

4 Methodologie

4.1 Concept

4.1.1 Waarde proposities.

Het voorgestelde systeem ondersteunt medewerkers van de OVER-gemeenten bij het herschrijven van brieven naar begrijpelijke taal volgens de regels van de OVER-schrijfwijzer. Door suggesties te geven op basis van de OVER-schrijfwijzer besparen medewerkers tijd en hoeven zij minder taalkundige kennis te hebben om consistent op B1-niveau teksten te schrijven. Meer inwoners zouden hierdoor de brieven beter kunnen begrijpen, waardoor misverstanden worden verminderd. De gemeente verbetert daarmee de kwaliteit van haar dienstverlening. Aangezien de tool lokaal draait, blijven privacy en veiligheidsvereisten gewaarborgd. De volledige waardeproposities staan in Bijlage B.

4.1.2 Flow diagram en gebruikte design patterns.

In Paragraaf 2.3 is de huidige werkwijze voor het opstellen van brieven door medewerkers van de gemeente beschreven. In dit onderzoek wordt een AI-tool toegevoegd die medewerkers ondersteunt bij het herschrijven van teksten naar B1-taalniveau (Figuur 3). Het model beoordeelt of zinnen eenvoudig genoeg zijn of te complex, en genereert vervolgens suggesties voor het herschrijven van tekst met een verwijzing naar de bijbehorende regel in de schrijfwijzer. Deze suggesties worden aan de medewerker getoond, die deze naar eigen inzicht kan toepassen. Daarna verloopt het verdere proces op dezelfde manier als in de huidige situatie.

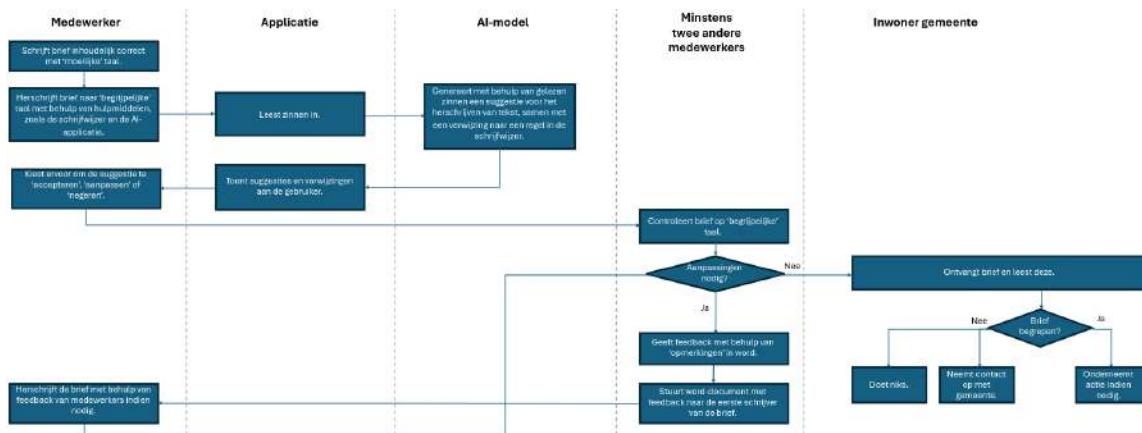


Fig. 3. Flow-diagram van de voorgestelde situatie.

Om inzicht te geven in de mate van automatisering binnen de toegevoegde stappen in het flowdiagram, is per stap van de AI-tool het bijbehorende Level of Automation (LoA) bepaald. Daarbij is gekeken naar de mate waarin de AI-tool zelfstandig handelt en in hoeverre de medewerker de inhoudelijke controle over de brief behoudt.

Binnen gemeentelijke communicatie gelden juridische en maatschappelijke eisen (FR01, FR02, ER04), waardoor volledige automatisering (LoA 6-10) niet wenselijk is. De schrijfassistent is daarom bewust gepositioneerd op een laag tot beperkt middelmatig automatiseringsniveau (LoA 1-3). Levels 4 en 5 zijn niet van toepassing, omdat de tool meer dan één of twee mogelijke suggesties kan genereren waaruit de medewerker kan kiezen. Het hoogste automatiseringsniveau binnen het systeem is LoA 3. De AI-tool ondersteunt, maar neemt geen beslissingen over het herschrijven van de brief. Tabel 4 toont per stap welk automatiseringsniveau is toegepast en waarom.

Table 4. Level of Automation per stap in de AI-tool

Stap	Beschrijving	LoA	Rol van de AI-tool / Rol van de medewerker
1	Tekst schrijven (menselijk).	1	Medewerker: schrijft de inhoud zoals in de huidige situatie in Figuur 2, zonder gebruik van AI.
2	AI-tool analyseert zinnen.	2	AI-tool: markeert moeilijke woorden, lange zinnen, passieve vorm en toonafwijkingen, etc. op basis van de schrijfwijzer.
3	AI-tool geeft herschrijfsuggesties en legt uit waarom.	3	AI-tool: genereert één of meerdere herschrijfsuggesties en legt uit waarom iets moeilijk is.
4	Medewerker ontvangt suggestie(s) en kiest ervoor om deze te accepteren, aanpassen of negeren.	3	AI-tool: past nooit automatisch wijzigingen toe. elke verandering vereist expliciete actie. Medewerker: kiest tussen <i>Accepteren</i> , <i>Aanpassen</i> of <i>Negeren</i> .

Tijdens het ontwerpproces zijn design patterns gebruikt om transparantie en controle binnen het systeem te vergroten:

- "Geef de controle terug aan de gebruiker wanneer de automatisering faalt" [34], door de medewerker de mogelijkheid te geven een suggesties te accepteren, aan te passen of te weigeren als deze de tekst niet herschreven is naar de OVER-schrijfwijzer-regels.
- "Leg uit voor begrip, niet voor volledigheid" [34], door met een verwijzing naar de schrijfwijzer toe te lichten waarom een tekst als 'niet-B1' wordt beschouwd.

4.1.3 Testopzet en evaluatiemethoden voor concept- en gebruikerstests.

Er is een paper prototype ontwikkeld om medewerkers van de OVER-gemeenten de AI-tool visueel te laten beoordelen. Deze staat weergeven in Figuur 4. Daarnaast is een testplan opgesteld (Bijlage D), waarvan de vragen zijn voorgelegd aan medewerkers van de OVER-gemeenten. De vragen richtten zich op wenselijkheid, vertrouwen en uitlegbaarheid en werden beoordeeld vanuit het perspectief van de medewerker die de brief schrijft. Enkele vragen uit het testplan staan hieronder:

- Wenselijkheid: Zijn er andere stappen in het prototype wenselijk die er op dit moment nog niet instaan?
- Uitlegbaarheid: Denk je dat jij (en je collega's) de AI-tool snel kunt leren gebruiken?
- Vertrouwen: Zou je te veel beïnvloed worden door de suggesties die de AI-tool geeft, en zou je de AI-tool daardoor te veel vertrouwen?

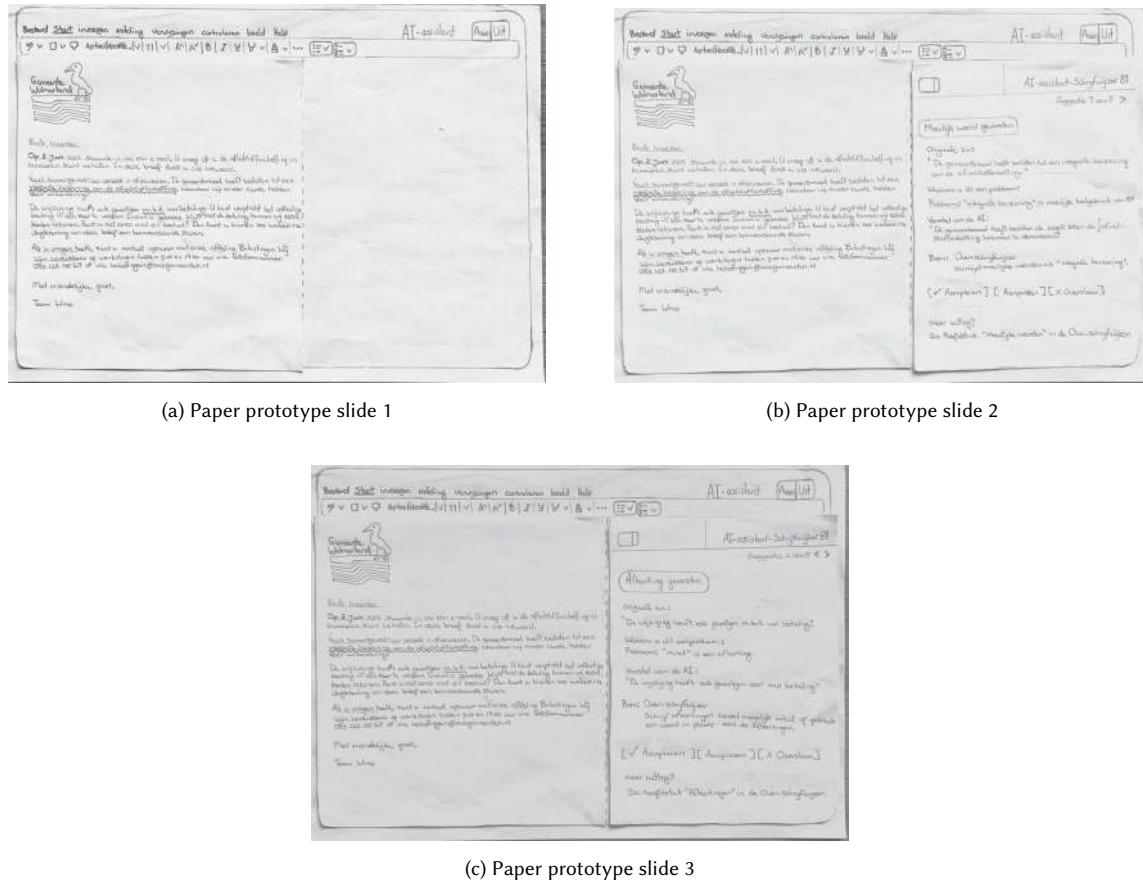


Fig. 4. Paper prototype slides

Op basis van de verkregen feedback zijn verbeteringen aangebracht aan het prototype. Onderbouwing voor deze aanpassingen wordt toegelicht in Paragraaf 6.2. Het verbeterde prototype zal in het tweede deel van dit blok opnieuw getest worden.

4.2 Data

Dit onderzoek maakt gebruik van drie typen data: de OVER-schrijfwijzer, externe woordenlijsten en bestaande brieven van de OVER-gemeenten. Samen vormen deze datasets de basis voor de ontwikkeling van de AI-tool, die gebruikmaakt van een tekstverwerkingsmodel om teksten te vereenvoudigen (Paragraaf 2.1.4).

4.2.1 OVER-schrijfwijzer.

De OVER-schrijfwijzer bevat richtlijnen voor duidelijke gemeentelijke communicatie op B1-niveau, met nadruk op korte en concrete zinnen, één kernboodschap per alinea en een logische opbouw. Medewerkers schrijven vanuit het perspectief van de inwoner, vermijden jargon en formele taal, en hanteren een actieve, positieve en adviserende toon. Daarnaast beschrijft de schrijfwijzer regels voor inclusief taalgebruik, consequent gebruik van de u-vorm en vaste notaties voor onder andere data, bedragen en adressen. De schrijfwijzer vormt de basis van de AI-tool: alle suggesties moeten hiermee in lijn zijn. Een beperking is dat de schrijfwijzer weinig concrete versimpelde synoniemen bevat voor moeilijke woorden, waardoor aanvullende externe data nodig is voor lexicale vereenvoudiging.

4.2.2 Externe data.

Ter aanvulling zijn vijf externe woordenlijsten gevonden die ondersteuning bieden bij het herkennen en vereenvoudigen van moeilijke woorden. Een overzicht van deze woordenlijsten, inclusief hun kenmerken, voor- en nadelen, is opgenomen in Tabel 5. Deze data dienen als aanvulling op de schrijfwijzer. Deze data worden pas in een tweede iteratie ingezet, nadat is onderzocht hoe goed een model functioneert op basis van uitsluitend de OVER-schrijfwijzer en nadat is vastgesteld dat de OVER-gemeenten zich daadwerkelijk aan de inhoud van deze woordenlijsten willen houden.

Table 5. Overzicht van gevonden Nederlandse woordenlijsten met uitleg, voordelen en nadelen.

Naam dataset / bron	Uitleg	Aantal woorden	Datatype	Voordeel	Nadeel
Moeilijk-woordenboek van de Gemeente Amsterdam [18]	Synoniemenlijst voor ‘moeilijke’ Nederlandse woorden.	747	String	Woorden zijn afgestemd op B1-niveau.	Geen zinnen opgenomen, waardoor geen zinsverbeteringen geleerd kunnen worden.
Helder juridisch woordenboek van de Gemeente Amsterdam [15]	Lijst met juridische termen en alternatieven op B1-niveau.	134	String	Vermijd juridisch jargon en stimuleert B1-niveau.	Geen zinnen opgenomen, waardoor geen zinsverbeteringen mogelijk.
Inclusieve woordenlijst van de Gemeente Amsterdam [17]	Lijst met termen en bijbehorende inclusieve alternatieven.	43	String	Bevordert inclusief taalgebruik.	Termen kunnen verouderen.
Synoniemen woordenlijst van B1-teksten [1], Dijk en Waard [6]	Lijst met versimpelde synonymen.	17	String	Gericht op B1-niveau en ondersteuning van eenvoudiger taalgebruik.	Beperkte lijst en geen zinnen voor zinsverbetering.

De OVER-gemeenten hebben naast de OVER-schrijfwijzer bestaande brieven aangeleverd die de huidige communicatiesituatie representeren. Omdat deze nog niet zijn herschreven volgens de schrijfwijzer, worden ze niet gebruikt als trainingsdata. Wel worden ze ingezet als validatie- en testdata om de werking van de AI-tool in een realistische context te evalueren.

4.2.3 Data-preparatie, splitsing en vector database.

De woordenlijsten waren niet beschikbaar als kant-en-klare datasets en zijn via webscraping verzameld, wat resulteerde in vier gestructureerde datasets (Tabel 5). Deze datasets zijn, zoals eerder toegelicht, in de eerste iteratie van het model nog niet gebruikt, maar zullen mogelijk in een tweede iteratie worden ingezet.

De gemeentelijke brieven zijn handmatig gegroepeerd op basis van bestandsnamen. Hierdoor ontstaan categorieën, zoals WMO en Participatiewet (Tabel 6). Vervolgens zijn de brieven met een stratified split [25] gesplitst. Door gebruik te maken van een stratified split is elke briefcategorie evenredig vertegenwoordigd, waardoor de briefcategorieën eerlijk verdeeld zijn in de datasets [25]. Vaak wordt bijvoorbeeld een 80% train-, 10% validatie- en 10% testset gebruikt, maar omdat in dit onderzoek geen gebruik wordt gemaakt van een trainset is ervoor gekozen om de brieven ook gelijk te verdelen verdeeld in 50% validatie- en 50% testset [7].

Table 6. Overzicht van de categorieën waarin de brieven verdeeld zijn.

Categorie	Uitleg	Aantal brieven
Schuldhulp	Brieven die betrekking hebben tot schuldhulp vanuit de OVER-gemeenten.	4
WMO	Brieven die betrekking hebben tot WMO en zorgvoorzieningen vanuit de OVER-gemeenten.	13
Participatiewet (PW)	Brieven die betrekking hebben tot PW en uitkeringszaken.	13
Participatiefonds (PF)	Brieven die betrekking hebben tot participatiefondsen.	4
Overig	Brieven die niet vallen binnen de bovenste vier categorieën.	2

Figuur 5 toont de verdeling van de briefcategorieën; brieven over de Participatiewet en de Wmo komen relatief vaak voor en worden daardoor vaker meegenomen in de evaluatie. Dit kan ervoor zorgen dat het onderzoek niet generaliseerbaar is naar alle categorieën brieven binnen de OVER-gemeenten.

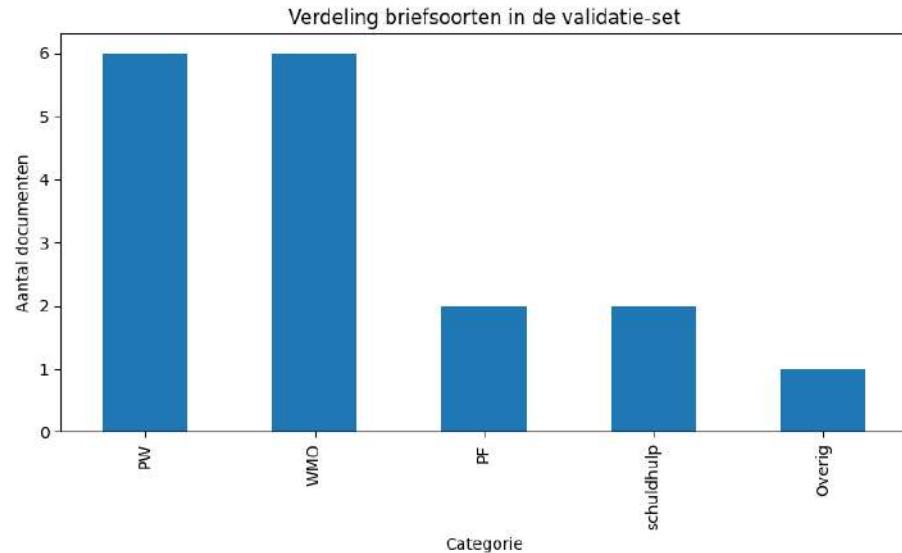


Fig. 5. Histogram van de briefcategorieën.

Om de schrijfwijzer als context beschikbaar te maken voor het taalmodel, wordt gebruikgemaakt van een vector database. Tekst wordt hierin opgeslagen als vectorrepresentaties, waardoor semantisch relevante regels kunnen worden opgehaald [11, 31, 44]. Overwogen opties van vector databases voor dit onderzoek zijn ChromaDB, LanceDB en Qdrant [3, 11, 12]. In de context van deze opdracht is het van belang dat het model lokaal kan draaien en aansluit bij de privacy- en AVG-eisen van de OVER-gemeenten, vandaar dat Qdrant minder geschikt is voor dit onderzoek (Tabel 7). Aangezien LanceDB meer filtermogelijkheden heeft dan ChromaDB is daarom gekozen voor LanceDB (Tabel 7).

Table 7. Vergelijking van vector databases

	ChromaDB [3]	LanceDB [11]	Qdrant [12]
Embedded/client-server	Hybride	Embedded	Client-server
Open-source	Ja (Apache 2.0)	Ja (Apache 2.0)	Ja Apache (2.0)
Metadata filtering	Filteren op basis van metadata en document inhoud.	Gebruikt SQL-strings en integreert met Pandas.	Geografische, tekst, en numerieke filters die in de index zijn verwerkt.

Voor opslag van de OVER-schrijfwijzer in de vector database is een chunking-strategie toegepast (Tabel 8) [11, 31, 44]. Eenvoudige methoden, zoals opsplitsen per zin of vaste lengte, zijn ongeschikt omdat richtlijnen vaak uit meerdere samenhangende zinnen en subsecties bestaan. Daarnaast is de semantische methode ongeschikt, omdat het zinnen die semantisch vergelijkbaar zijn dicht bij elkaar zet in een embedding en er hierdoor context verloren gaat. Daarom is gekozen voor een gespecialiseerde methode, “opsplitsen op structuur” (Tabel 8). Hierbij is de schrijfwijzer eerst handmatig omgezet naar het Markdown-formaat. Vervolgens is deze gebruikt om de OVER-schrijfwijzer in de vector database te plaatsen. Structureel samenhangende regels worden hierbij gezamenlijk opgeslagen, zodat voldoende context behouden blijft.

Table 8. Strategieën voor text chunking

	Strategie	Voordelen	Nadelen
Klassieke methoden	Opsplitsen op basis van een ingestelde lengte (bijvoorbeeld elke 100 karakters).	Eenvoudig te implementeren; voorospelbare chunk-grootte.	Contextverlies; kan zinnen of woorden middenin breken en negeert documentstructuur.
	Opsplitsen op zinnen.	Behoudt grammatische structuur; geen afgebroken woorden.	Gebrek aan context en zinnen variëren in lengte.
	Opsplitsen op basis van een ingestelde lengte, met een overlap, meestal rond de 10/20%.	Vermindert contextverlies op de grenzen.	Redundantie in zoekresultaten (dubbele info); hogere opslag- en indexeringskosten.
Gespecialiseerde methoden	Opsplitsen op structuur (HTML, Markdown, of LaTeX)	Behoudt semantische hiërarchie (secties, regels).	Afhankelijk van bronkwaliteit, want slechte conversie leidt tot slechte chunks.
	Semantisch	Groepeert inhoudelijk gerelateerde zinnen; optimale contextbehoud.	Rekenintensief met onvoorspelbare grenzen. Gevoelig voor hyperparameters.

4.3 Model

4.3.1 Modelarchitectuur.

Voor een eerste iteratie van het model is gekozen voor een naïeve-RAG-architectuur [41]. Een RAG-architectuur is een benadering waarbij het genereren van output actief externe, relevante informatie ophaalt uit een kennisbron en deze context gebruikt om nauwkeurigere en beter onderbouwde antwoorden te produceren [43]. Hierbij wordt een vectordatabase gebruikt om aanvullende context aan het taalmodel mee te geven. Geavanceerdere RAG-varianten zijn voor een eerste iteratie bewust buiten beschouwing gelaten om een eenvoudig en uitbreidbaar uitgangspunt te behouden. De vectordatabase van het model is gevuld met richtlijnen uit de OVER-schrijfwijzer. Relevante passages worden opgehaald en toegevoegd aan de prompt, zodat het taalmodel deze richtlijnen kan meenemen bij het herschrijven van teksten. De modelarchitectuur is weergegeven in Figuur 6.

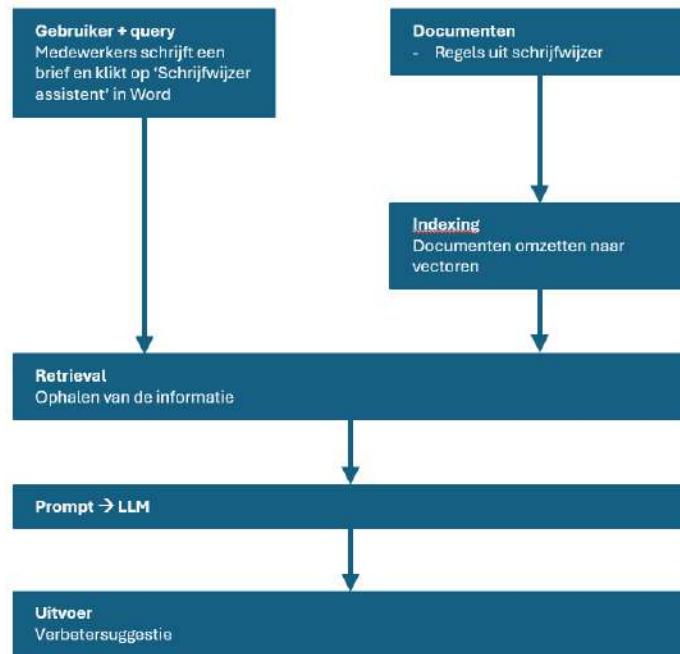


Fig. 6. Modelarchitectuur iteratie 1

4.3.2 Het gebruikte LLM.

Voor de naïve-RAG-architectuur is onderzocht welk LLM het meest geschikt is voor Nederlandstalige tekstvereenvoudiging in een gemeentelijke context. Het GEITje-model viel aanvankelijk op door zijn Nederlandstalige focus en uitgebreide training, maar is niet meer beschikbaar vanwege auteursrechtelijke bezwaren [47] en is daarom uitgesloten.

Alternatief zijn de Leesplank Noot-modellen, ontwikkeld voor Nederlandse tekstvereenvoudiging richting B1-niveau en explicet bedoeld voor publieke communicatie. Deze modellen voldoen aan eisen rondom transparantie, controleerbaarheid en veilige inzet binnen overheden. Een overzicht van de beschikbare modellen is opgenomen in Tabel 9. Het bevat een set van drie Nederlandstalige LLM's, elk gefinetuned op een samengestelde Nederlandstalige dataset voor tekstvereenvoudiging naar B1-niveau. De modellen zijn ontwikkeld voor overheids- en publieke communicatie en sluiten aan bij de EU-AI-Act: transparant, herleidbaar, controleerbaar en veilig inzetbaar in gemeentelijke communicatie. In Tabel 9 worden de drie modellen weergegeven. Op basis van Tabel 9 wordt één van deze drie modellen geselecteerd voor inzet als LLM binnen de naïve-RAG-architectuur.

Table 9. Overzicht van verschillende LLM's met bron, voordelen, nadelen en aantal parameters

LLM-model en bron	Voordelen	Nadelen	Aantal parameters
Granite-3.3-2b [9]	Biedt de hoogste kwaliteit volgens de bron, levert consistente en betrouwbare output en blijft dankzij de compacte omvang efficiënt inzetbaar op gangbare hardware.	Het model is trager dan de alternatieven en vraagt meer rekenkracht dan het lichtste model, waardoor het minder geschikt is voor situaties waar snelheid de hoogste prioriteit heeft.	2 miljard
Llama-3.2-3b [10]	Combineert een goede outputkwaliteit met hogere snelheid dan Granite en profiteert van extra modelcapaciteit die het breed inzetbaar maakt.	De grotere modelgrootte zorgt voor hogere hardware-kosten, terwijl deze extra omvang geen duidelijke kwaliteitswinst oplevert ten opzichte van Granite.	3 miljard
EuroLLM-1.7b [8]	Werkt snel, is lichtgewicht en daarmee geschikt voor toepassingen met beperkte hardware.	De lagere kwaliteit en minder stabiele verwerking van complexe context maken het minder betrouwbaar voor taken waar nauwkeurigheid belangrijk is.	1.7 miljard

Voor de naïve-RAG-architectuur wordt gekozen voor het Granite-3.3-2B-model (Tabel 9). Volgens [5] behaalt dit model de hoogste kwaliteitsscore van de drie modellen bij het vereenvoudigen van gemeentelijke teksten naar B1-niveau. Het Granite-3.3-2B-model weet teksten beter te structureren en inhoudelijk nauwkeuriger weer te geven, waarbij het model qua omvang compact genoeg blijft om efficiënt te gebruiken. De andere modellen bieden voordelen, maar sluiten minder goed aan bij de doelstelling. Llama-3.2-3B is groter, maar levert ondanks het hogere aantal parameters geen kwaliteitsverbetering. Daardoor wegen de extra rekenkosten niet op tegen de beperkte meerwaarde [10]. EuroLLM-1.7B is het snelste model, wat aantrekkelijk kan zijn voor taken die snel uitgevoerd dienen te worden. Volgens [5] behaald EuroLLM-1.7B lagere scores, wat de betrouwbaarheid van het model kan verminderen [8]. Daarom is het Granite-3.3-2B-model de meest geschikte keuze voor implementatie in het naïve-RAG-model, vanwege de hoogste kwaliteit en stabiele verwerking van teksten.

4.3.3 Baseline model.

Als baseline om het ontwikkelde naïve-RAG-model mee te vergelijken is gekozen voor een LLM dat teksten herschrijft naar eenvoudiger Nederlands op basis van een vaste prompt. Deze baseline wordt gebruikt als referentiepunt voor latere iteraties, waarin aanvullende context (zoals de OVER-schrijfwijzer en externe woordenlijsten) worden toegevoegd. Voor het baseline-model is gebruikgemaakt van hetzelfde model als in de naïve-RAG-architectuur (Paragraaf 4.3.1). De eerder geprepareerde validatieset (brieven van de OVER-gemeenten) is ingeladen en per zin herschreven met behulp van een prompt. Deze prompt is bewust kort gehouden om de prestaties van een generiek LLM te meten, zonder invloed van aanvullende richtlijnen of regels:

Je bent een assistent die teksten herschrijft naar eenvoudiger Nederlands (B1-niveau).

Herschrijf de volgende tekst:

Prompt

Geef uitsluitend de herschreven tekst, zonder toelichting.

Door deze aanpak wordt inzicht verkregen in de mate waarin een LLM, zonder domeinspecifieke kennis of expliciete schrijfregels, in staat is om gemeentelijke teksten te vereenvoudigen. De baseline vormt daarmee het uitgangspunt voor vergelijking met vervolgmodellen waarin de OVER-schrijfwijzer en human-in-the-loop-principes worden geïntegreerd.

4.3.4 Evaluatiemetrics voor het model.

De prestaties van het taalmodel zullen geëvalueerd worden met verschillende kwaliteitscriteria. Eerst is bepaald welke prestatia maat wordt gebruikt om iteraties te vergelijken. Voor dit onderzoek is een prestatia maat (performance metric) nodig die specifiek inzicht geeft in de begrijpelijkheid van herschreven teksten. Het doel is om vast te stellen of gemeentelijke brieven na vereenvoudiging daadwerkelijk leesbaar zijn en voldoen aan de regels in de OVER-schrijfwijzer. Daarom worden in dit onderzoek verschillende prestatiematen in overweging genomen voor het beoordelen van de begrijpelijkheid van gemeentelijke brieven:

- LiNT-II (Linguïstisch Instrumentarium voor Tekstleesbaarheid II) [38]
- SARI (System output Against References and against the Input) [56]
- FKBLEU (Flesch-Kincaid Bilingual Evaluation Understudy) [56]

In Tabel 10 wordt de beschrijving behorende bij deze prestatiematen gegeven.

Table 10. Overzicht van overwogen prestatiematen

Prestatiemaat	Beschrijving
LiNT-II	LiNT-II is een leesbaarheidsinstrument dat specifiek is ontwikkeld voor Nederlandstalige teksten [38]. De metric bepaalt de moeilijkheidsgraad van een tekst op basis van vier taalkundige kenmerken: woordfrequentie, syntactische afhankelijkheidslengte, het aantal inhoudswoorden per zin en het aandeel concrete zelfstandige naamwoorden [38]. Omdat LiNT-II is afgestemd op Nederlandstalige teksten en empirisch is gevalideerd met beoordelingen door docenten en scholieren [38, 42], is deze metric geschikt om te bepalen of teksten dichter bij de richtlijnen voldoen.
SARI	SARI is een veelgebruikte prestatimaat voor tekstversimpeling die zich richt op de manier waarop een vereenvoudigde tekst afwijkt van het origineel [56]. De prestatimaat meet daarmee vooral de kwaliteit van de vereenvoudiging zelf en niet de feitelijke leesbaarheid voor eindgebruikers, waardoor SARI minder geschikt is voor dit onderzoek.
FKBLEU	FKBLEU combineert een maat voor parafrasekwaliteit (iBLEU) met de leesbaarheidsindex Flesch-Kincaid Grade Level (FKGL) [56]. Hoewel FKBLEU redelijke correlaties met menselijke beoordelingen kan vertonen, laat eerder onderzoek zien dat SARI sterker samenhangt met menselijke oordelen over eenvoud en minder geneigd is om conservatieve systemen te belonen die weinig wijzigingen aanbrengen in de oorspronkelijke tekst [56].

Aangezien SARI zich niet richt op de feitelijke leesbaarheid van teksten, en FKBLEU bovendien minder goed overeenkomt met menselijke oordelen over eenvoud, worden deze prestatiematen niet gebruikt voor het beoordelen van de prestaties van de AI-tool. Op basis van de hierboven besproken overwegingen wordt LiNT-II geselecteerd als primaire prestatimaat, omdat deze prestatimaat het meest geschikt is om de leesbaarheid van Nederlandstalige gemeentelijke teksten te beoordelen. Deze prestatimaat wordt gebruikt om de gemiddelde LiNT-II-score te berekenen op de herschreven brieven uit de validatie-set. De LiNT-II-score loopt van 0 tot en met 100. De LiNT-II-score loopt van 0 tot en met 100, waarbij geldt dat hoe lager de score, hoe dichter de tekst bij het B1-niveau ligt. Een score van 100 duidt op teksten die ver boven B1-niveau zijn.

Naast LiNT-II wordt alleen het uiteindelijke, best presterende model beoordeeld op aanvullende kwaliteitscriteria: robuustheid, uitlegbaarheid, modelcomplexiteit en resource demand (Tabel 11). Het model wordt niet beoordeeld op schaalbaarheid, waarmee wordt bedoeld in hoeverre de AI-tool inzetbaar is bij andere gemeenten om ook daar brieven volgens de richtlijnen van de OVER-schrijfwijzer leesbaar te maken. Dit is in dit onderzoek niet mogelijk, omdat uitsluitend brieven van de OVER-gemeenten beschikbaar zijn en de AI-tool daardoor niet kan worden getest op brieven van andere gemeenten.

Table 11. Overzicht van aanvullende kwaliteitscriteria

Kwaliteitscriterium	Evaluatiemethode
Robuustheid	Er wordt onderzocht of de LiNT-II-score van de verbeterde brieven consistent is over verschillende briefcategorieën (Schuldhulp, WMO, Participatiewet (PW), Participatiefonds (PF) en Overig). Dit toest of het model niet alleen voor één type tekst goed presteert (FR06). De resultaten worden weergegeven als gemiddelde LiNT-II-scores per briefcategorie. Vervolgens wordt gecontroleerd of deze scores per categorie gelijk zijn.
Uitlegbaarheid	Geëvalueerd wordt of de UI-elementen duidelijk maken dat de tool ondersteunend is, fouten kunnen maken en dat de medewerker de eindbeslissing neemt (ER01, ER02, ER04, ER05, FR01–FR03, TR02). Dit wordt getest met het prototype en met testpersonen.
Modelcomplexiteit	De gebruikte LLM-architectuur (Granite-3.3-2B in de naïeve RAG) wordt beschreven in termen van het aantal parameters en de modelopbouw, om de ‘zwaarte’ van het model voor de gemeentelijke infrastructuur te beoordelen. Omdat de AI-tool lokaal moet kunnen draaien, mag het model niet te computatieloezel belastend zijn (TR05). De complexiteit wordt daarom vergeleken met die van andere tekstversimpelingsmodellen.
Resource demand	Per brief worden de verwerkingstijd en het geheugengebruik gemeten. De AI-tool moet binnen een acceptabele tijd van ongeveer twee seconden suggesties kunnen genereren, zodat de tool in de praktijk bruikbaar is zonder de werkstroom te vertragen (TR06).

5 Resultaten

5.1 Resultaten van de modeltests

De zinnen uit de brieven in de validatieset zijn herschreven met zowel het baseline model als het naïve-RAG-model. Voor beide modellen zijn de LiNT-II-scores berekend voor de oorspronkelijke brieven en de herschreven versies. Tabel 12 toont de gemiddelde LiNT-II-scores en de verandering ten opzichte van de originele brieven.

Table 12. LiNT-II-scores

Model	Gemiddelde LiNT-II-score	Gemiddelde verandering in LiNT-II-score
Geen model	47.61	-
Baseline model	45.83	1.78
naïve-RAG-model	45.75	1.86

De zinnen in de originele brieven behalen een gemiddelde LiNT-II-score van 47.61. Na vereenvoudiging met het baseline model daalt deze score naar 45.83, wat neerkomt op een gemiddelde daling van 1.78 punten. Dit laat zien dat het baseline-model teksten vereenvoudigt, maar de verbetering beperkt is. Volgens requirement FR06 moet het uiteindelijke model een lagere LiNT-II-score hebben dan het baseline model op de validatieset. Dit betekent dat het naïve-RAG-model een lagere gemiddelde LiNT-II-score moet behalen dan 45.83. Dit is behaald, aangezien het naïve-RAG-model een gemiddelde LiNT-II-score heeft van 45.75. Dit betekent dat het naïve-RAG-model tot nu toe het best presterende model is. Het verschil tussen het baseline model en het naïve-RAG-model is echter zeer klein (0.08). Het naïve-RAG-model levert nauwelijks extra verbetering op ten opzichte van het baseline model.

Een aantal herschreven voorbeeldzinnen van zowel het baseline model als het naïve-RAG-model staan in de tabellen 13 en 14. Voor beide modellen is te zien dat de voorbeeldzinnen minder formeel geschreven zijn en suggesties bevatten die geen correct Nederlands zijn, zoals ‘noemd’. Beide modellen werken dus nog niet foutloos.

Table 13. Vergelijking originele en herschreven zinnen van het baseline model

Originele zin	Versimpelde zin
Wij gaan daarna aan de slag om de ondersteuning voor u in te zetten.	We gaan daarna aan de slag om ondersteuning in te zetten voor u.
U heeft aangegeven geen voorkeur te hebben voor een zorgaanbieder.	Geen keuze gemaakt voor zorgaanbieder.
Wij gaan voor u op zoek naar een passende zorgaanbieder.	We zoeken voor u naar een geschikte zorgaanbieder.
Dit heet Zorg in Natura.	Zorg in Natura
U ontvangt hierover ons besluit met de belangrijkste informatie.	U ontvangt besluit met belangrijk info.
Tijdens het gesprek zijn enkele algemene en voorliggende voorzieningen benoemd.	Tijdens het gesprek werden enkele algemene voorzieningen genoemd.
Hiermee bedoelen wij voorzieningen, producten en diensten die voor alle inwoners van de gemeente beschikbaar zijn.	We bedoelen voorzieningen, producten en diensten die iedere burger in de gemeente heeft.
Hier is geen maatwerkvoorziening van de Wmo voor nodig.	Er is geen maatwerkvoorziening van de Wmo voor nodig.
Stichting WelzijnWonenPlus is een vrijwilligersorganisatie welke verschillende diensten biedt, onder andere tuinonderhoud, een sociaal maatje, de wijkbus/belbus en mantelzorgondersteuning.	Stichting WelzijnWonenPlus is een vrijwilligersorganisatie dat verschillende diensten aanbiedt, zoals tuinonderhoud en sociale contacten, de wijkbus/belbus, en mantelzorgondersteuning.
U kunt contact opnemen met WelzijnWonenPlus voor meer informatie.	Neem contact op met WelzijnWonenPlus voor meer details.

Table 14. Vergelijking originele en herschreven zinnen van het naïve-RAG-model

Originele zin	Versimpelde zin
Wij gaan daarna aan de slag om de ondersteuning voor u in te zetten.	Wij beginnen met inzet van ondersteuning voor u.
U heeft aangegeven geen voorkeur te hebben voor een zorgaanbieder.	U heeft geen voorkeur voor een specifieke zorgaanbieder.
Wij gaan voor u op zoek naar een passende zorgaanbieder.	Wij zoeken een geschikte zorgaanbieder voor u.
Dit heet Zorg in Natura.	Dit heet Zorg in Natura.
U ontvangt hierover ons besluit met de belangrijkste informatie.	U ontvangt het besluit met de belangrijkste informatie.
Tijdens het gesprek zijn enkele algemene en voorliggende voorzieningen benoemd.	In het gesprek werden enkele algemene voorzieningen noemd.
Hiermee bedoelen wij voorzieningen, producten en diensten die voor alle inwoners van de gemeente beschikbaar zijn.	We bieden inhoud, producten en diensten die beschikbaar zijn voor iedere burger van onze gemeente.
Hier is geen maatwerkvoorziening van de Wmo voor nodig.	Je hebt geen specifieke maatwerkvoorziening nodig uit de Wmo.
Stichting WelzijnWonenPlus is een vrijwilligersorganisatie welke verschillende diensten biedt, onder andere tuinonderhoud, een sociaal maatje, de wijkbus/belbus en mantelzorgondersteuning.	Stichting WelzijnWonenPlus is een vrijwilligersorganisatie die verschillende diensten biedt, waaronder onder andere tuinonderhoud, een sociaal maatje, de wijkbus/belbus en mantelzorgondersteuning.
U kunt contact opnemen met WelzijnWonenPlus voor meer informatie.	U kunt contact opnemen met WelzijnWonenPlus voor meer informatie.

5.2 Resultaten van concept-/gebruikerstests

5.2.1 Testen met medewerkers van de OVER-gemeenten.

Om de praktische toepasbaarheid van de AI-tool voor gemeentelijke medewerkers te toetsen, is een paper prototype getest met medewerkers van de OVER-gemeenten (teamleider WMO-afdeling, een beleidsadviseur en een WMO-consulent) en met masterstudenten Applied AI. De deelnemers kregen de opdracht om met het prototype te werken, zoals het vinden en gebruiken van knoppen, en gaven feedback op de gebruiksvriendelijkheid, wenselijkheid, uitlegbaarheid en het vertrouwen in de tool. Uit de test kwamen bevindingen naar voren die hieronder kort zijn samengevat (Bijlage G):

- De AI-tool werd over het algemeen als gebruiksvriendelijk en intuïtief ervaren. De mogelijkheid om de AI-tool aan en uit te zetten werd als positief ervaren, omdat medewerkers eerst een kladversie kunnen schrijven zonder dat direct meldingen in beeld verschijnen. Het verschil tussen de knoppen ‘accepteren’ en ‘aanpassen’ was niet voor iedereen direct duidelijk.
- Het onderstrepen van fouten en waarom een suggestie wordt gegeven, werd als leerzaam en ondersteunend ervaren.
- Voorgesteld werd om de term ‘AI-assistent’ te vervangen door ‘schrijfwijzer assistent’, vanwege de duidelijke koppeling met de OVER-schrijfwijzer en om de tool toegankelijker (minder technisch) te laten klinken.
- De bronverwijzing naar de schrijfwijzer werd als nuttig en leerzaam ervaren, maar hoeft niet twee keer te worden weergegeven.
- Sommigen gaven aan dat het overzichtelijker zou zijn om alle suggesties onder elkaar te tonen in plaats van per suggestie te navigeren. Anderen vonden dat het stap-voor-stap doorlopen van suggesties medewerkers helpt om bewust na te denken over hun keuze.
- Een knop ‘alles accepteren’ werd enerzijds als praktisch gezien, maar anderzijds als risicovol vanwege mogelijk te groot vertrouwen in de tool. Ook werd voorgesteld om een knop ‘nieuwe suggestie’ toe te voegen.
- Sommigen gaven aan behoefte te hebben aan extra ondersteuning, bijvoorbeeld in de vorm van onboarding.
- Voorgesteld werd dat een klik op een gemaakte fout de gebruiker automatisch zou moeten leiden naar de juiste pagina met uitleg en een suggestie.
- Als tip werd gegeven om bij het inschakelen van de tool een melding toe te voegen met de tekst ‘U bent zelf verantwoordelijk voor de inhoud van deze brief’, zodat medewerkers extra worden herinnerd aan hun verantwoordelijkheid.

5.2.2 Tweede testronde.

De tweede testronde is nog niet uitgevoerd in de eerste vijf weken van dit blok. Deze zal op een later moment plaatsvinden en de resultaten zullen vervolgens worden gerapporteerd.

5.3 Prestaties van het model in de context van het concept

Wanneer het huidige model in de OVER-gemeenten wordt toegepast, kan dit leiden tot ongewenste situaties. De gemiddelde LiNT-II-score is namelijk maar met 1.86 gedaald ten opzichte van de originele brieven, die niet zijn herschreven. Deze daling is zodanig klein dat het model waarschijnlijk de brief niet correct herschrijft, waardoor een brief niet voldoende voldoet aan de regels van de OVER-schrijfwijzer (Paragraaf 5.1). Hierbij bestaat het risico dat de medewerker denkt dat de brief voldoende is herschreven. In de huidige vorm van de AI-tool zal het gebruik van het systeem dus mogelijk niet direct leiden tot een compleet correct geschreven brief, maar misschien zelfs tot een verslechtering of slechts een kleine verbetering.

De medewerker fungeert hier als een human-in-the-loop en blijft eindverantwoordelijk voor de uiteindelijke beslissing. Met de huidige prestaties van het model is het echter essentieel dat de medewerker de suggesties van de AI-tool kritisch beoordeelt en zich niet te veel laat leiden door de output, aangezien deze nog kan afwijken van de regels uit de OVER-schrijfwijzer.

6 Evaluatie

6.1 Evaluatie van modelresultaten

In deze paragraaf worden de kwaliteitscriteria van het huidige best presterende model (naïve-RAG-model) geëvalueerd. Daarbij wordt ingegaan op de prestatiaaat, robuustheid, uitlegbaarheid, modelcomplexiteit en resource demand van het model.

- **Prestatiamaat**

Uit de resultaten is gebleken dat het naïve-RAG-model (gem. LiNT-II-score: 45.75) nauwelijks beter presteert dan het baseline model (gem. LiNT-II-score: 45.83), met een verschil van 0.08 punt. Daarnaast ligt deze score nog ver boven 0, wat aangeeft dat de gegenereerde teksten onvoldoende voldoen aan het B1-niveau en het model er niet in slaagt teksten effectief te vereenvoudigen.

- **Robuustheid** In Tabel 15 worden de gemiddelde LiNT-II-scores van de originele en herschreven brieven weergegeven om te onderzoeken of deze consistent zijn over de verschillende briefcategorieën. Dit toetst of het model niet alleen voor één type brief goed presteert (FR06). De resultaten worden weergegeven als gemiddelde LiNT-II-scores per briefcategorie en er wordt gecontroleerd of deze scores per categorie gelijk zijn. Uit de resultaten blijkt dat de gemiddelde LiNT-II-scores per briefcategorie van elkaar verschillen. Het model presteert dus niet consistent over de verschillende briefsoorten.

Table 15. Gemiddelde LiNT-II-scores per briefcategorie

Groep	Gem. LiNT-II-score originele brieven	Gem. LiNT-II-score herschreven brieven	Gem. verschil LiNT-II-scores
Participatiefonds (PF)	46.00	40.85	5.14
Participatiewet (PW)	45.49	41.91	4.12
Schuldhulp	57.17	50.02	7.09
WMO	47.27	45.72	1.69
Overig	43.85	39.16	4.68

- **Uitlegbaarheid**

Uit de test van het prototype met de testpersonen (Bijlage G) en de daaruit aangepaste onderdelen in de nieuwe iteratie weergegeven in Paragraaf 6.2, kan geconcludeerd worden dat de UI-elementen duidelijk zijn en de tool als ondersteunend zal worden ervaren en niet als beslissend. Hierbij is ervoor gezorgd dat de medewerker altijd de eindverantwoordelijke is en dat dit ook duidelijk wordt gemaakt aan de medewerker door gebruik van een melding in het scherm als de tool aangezet wordt.

- **Modelcomplexiteit**

De gebruikte LLM-architectuur in het naïve-RAG-model is Granite-3.3-2B. In Paragraaf 4.3.2 werd aangegeven dat dit model 2 miljard parameters bevat. Dit zijn, vergeleken met bijvoorbeeld RoBERTa die 300 miljoen parameters bevat, veel parameters [22] voor een model. Daarentegen worden bij het Taalloket ook modellen benoemd die boven de 200 miljard parameters bevatten. In vergelijking met deze modellen bevat het naïve-RAG-model weinig parameters. Als bekend is welke computers bij de OVER-gemeenten beschikbaar zijn, kan worden vastgesteld of het model daar lokaal kan draaien zonder problemen. Tijdens het vragenuur op 6 januari zullen wij navragen welke computers de medewerkers gebruiken.

- **Resource demand**

Per brief wordt de verwerkingstijd en het geheugengebruik gemeten. De AI-tool moet in staat zijn om binnen een tijd van ongeveer twee seconden suggesties te genereren (TR06). Deze metingen zijn nog niet uitgevoerd wegens tijdsgebrek, maar zullen vóór de volgende iteratie, na de deadline, alsnog worden uitgevoerd.

6.2 Onderbouwing van wijzigingen tussen concepten

Na de eerste test met het paper prototype is de ontvangen feedback van de OVER-gemeenten medewerkers en studenten (Bijlage G) verwerkt in een verbeterd paper prototype. De belangrijkste aanpassingen en toevoegingen worden hieronder toegelicht aan de hand van relevante design patterns.

6.2.1 Aanpassing 1: Terminologie.

De oorspronkelijke term 'AI-assistent' werd als technisch ervaren. Deze is vervangen voor 'Schrijfwijzer assistent', zodat de betekenis voor de medewerkers duidelijker is. Deze aanpassing is gebaseerd op de design pattern 'Voeg context toe vanuit menselijke bronnen' [34], omdat de naam van de AI-tool nu direct verwijst naar een bestaande en vertrouwde schrijfwijzer binnen de organisatie, waardoor gebruikers beter begrijpen waarvoor de tool bedoeld is.

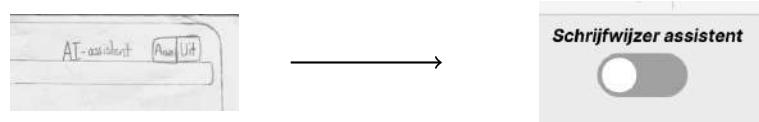


Fig. 7. Aanpassing 1

6.2.2 Aanpassing 2: Toegevoegde knop 'Nieuwe suggestie'.

Om gebruikers meer keuze en controle te geven is een extra knop 'nieuwe suggestie' toegevoegd. Hiermee kunnen zij alternatieve formuleringen opvragen wanneer de eerste niet bevalt. Dit verhoogt de gebruiksvriendelijkheid en voorkomt dat medewerkers aan één voorstel vastzitten. De aanpassing volgt het design pattern 'Laat gebruikers de automatisering begeleiden' [34], omdat gebruikers de suggesties niet alleen kunnen aanpassen, maar ook meerdere opties kunnen opvragen en daaruit kunnen kiezen.

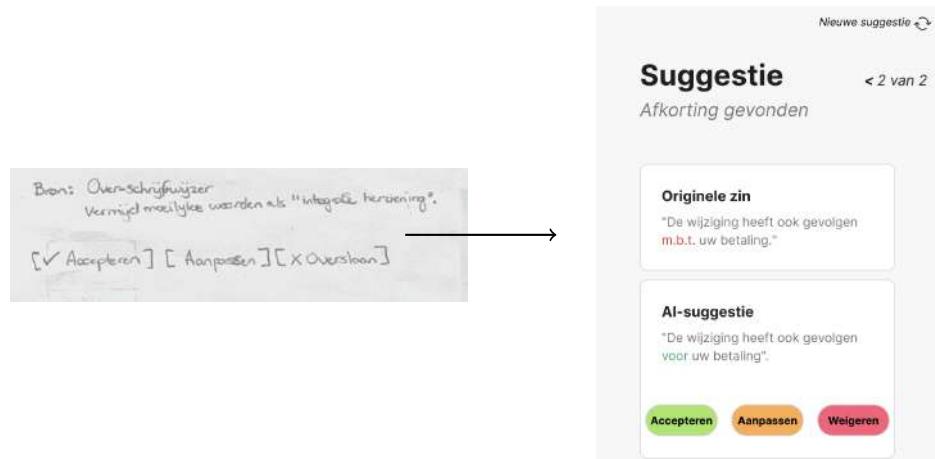


Fig. 8. Aanpassing 2

6.2.3 Aanpassing 3: Bronvermelding eenmaal tonen.

De bronverwijzing naar de schrijfwijzer werd twee keer weergegeven en wordt nu nog maar één keer weergegeven. Hoewel deze verwijzingen als nuttig en leerzaam werden ervaren, zorgen dubbele vermelding voor onnodige herhaling. Door de bron slechts eenmaal te tonen, blijft de informatie aanwezig zonder dat de interface onoverzichtelijk wordt. Deze aanpassing is gebaseerd op de design pattern 'Leg uit voor begrip, niet voor volledigheid' [34], omdat gebruikers voldoende informatie krijgen om het systeem te begrijpen zonder te worden belast met herhalende of overbodige informatie.

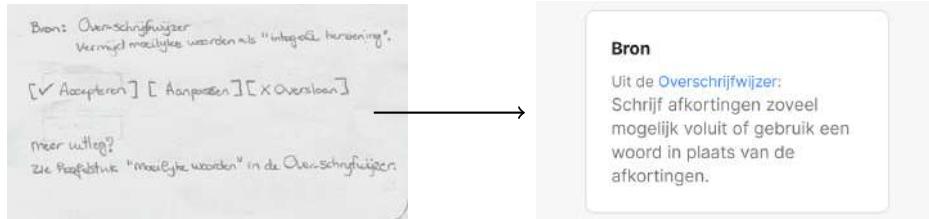


Fig. 9. Aanpassing 3

6.2.4 Toevoeging 1: Foutgerichte navigatie.

Door op een gemaakte fout te klikken, wordt de gebruiker direct doorgestuurd naar de bijbehorende uitleg en verbetersuggestie. Deze aanpassing is gebaseerd op de design pattern 'Ga verder dan uitleg op het moment zelf' [34], omdat de gebruiker direct relevante informatie te zien krijgt, zonder zelf te zoeken naar de juiste toelichting of mogelijke verbetering.

6.2.5 Toevoeging 2: Melding eigen verantwoordelijkheid.

Bij het inschakelen van de tool wordt een melding weergegeven met de tekst 'U bent zelf verantwoordelijk voor de inhoud van deze brief'. Deze melding is toegevoegd om medewerkers bewust te maken van hun eigen rol en verantwoordelijkheid bij het gebruik van de AI-tool. Deze toevoeging is gebaseerd op de design pattern 'Stel de juiste verwachtingen' [34], omdat hiermee duidelijk wordt gemaakt dat de AI slechts een hulpmiddel is en de medewerker eindverantwoordelijk blijft.

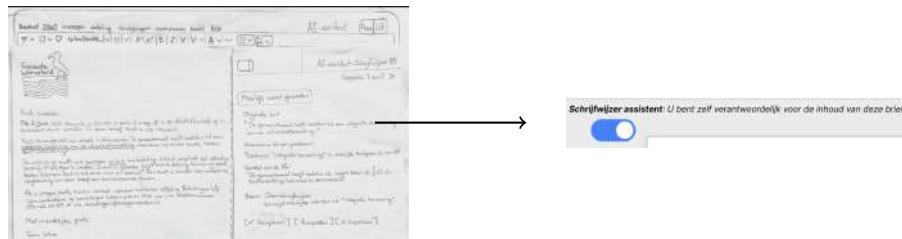


Fig. 10. Toevoeging 2

6.2.6 Toevoeging 3: Extra pagina als er geen suggesties zijn gevonden.

Wanneer de AI-tool geen moeilijke woorden of zinnen vindt, wordt een pagina getoond met de melding dat er geen verbetersuggesties zijn. Daarbij wordt aangegeven dat de brief voldoet aan de B1-richtlijnen, maar dat de medewerker de brief nog wel kritisch dient na te lezen. Dit is belangrijk omdat de medewerker verantwoordelijk blijft voor de inhoud. Deze toevoeging is gebaseerd op het design pattern ‘Stel de juiste verwachtingen’ [34], omdat duidelijk wordt gemaakt dat de AI-tool geen verbeteringen heeft gevonden, maar dat de medewerker de brief zelf nog moet controleren. Dit voorkomt dat een lege extensie (doordat er geen suggesties worden weergegeven) wordt gezien als teken dat de tekst perfect is.

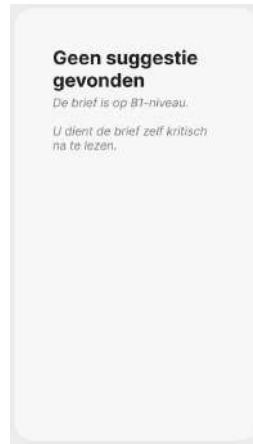


Fig. 11. Toevoeging 3

6.2.7 Aanbevelingen die niet zijn meegenomen in het verbeterde concept.

De volgende aanbevelingen die tijdens het testen door de testpersonen zijn genoemd (Paragraaf 5.2), zijn niet doorgevoerd in het verbeterde prototype:

- **Alle fouten met verbetersuggesties onder elkaar tonen in plaats van navigeren met pijlen.**
Deze suggestie is niet overgenomen, omdat de meeste testpersonen aangaven het juist prettig te vinden dat niet alle fouten tegelijk zichtbaar zijn. Hierdoor worden gebruikers gestimuleerd per fout stil te staan bij de uitleg en de voorgestelde verbetering, wat het leerproces bevordert. Dit sluit beter aan bij het doel van de tool om gebruikers te helpen hun schrijfvaardigheid te verbeteren door inzicht te geven in concrete verbeterpunten.
- **Een knop ‘alles accepteren’ toevoegen.**
Hoewel sommige testpersonen deze functie als praktisch beschouwden, is ervoor gekozen om deze niet te implementeren. Een dergelijke knop verhoogt het risico dat medewerkers suggesties automatisch overnemen zonder deze kritisch te beoordelen. Dit kan leiden tot onjuist gebruik van de AI-tool en een te groot vertrouwen in de AI, wat niet wenselijk is gezien de rol van de medewerker als eindverantwoordelijke.
- **Extra ondersteuning voor nieuwe gebruikers, bijvoorbeeld in de vorm van onboarding.**
Eén testpersoon wilde bij de eerste keer gebruiken van de tool meer uitleg. Deze aanbeveling wordt als waardevol beschouwd, maar valt buiten de scope van dit project. Het toevoegen van onboarding-functionaliteit wordt daarom gezien als een mogelijke toekomstige verbetering.

6.3 Volledig geïntegreerde evaluatie van het model binnen het concept

Het prototype is getest door medewerkers van de OVER-gemeenten en studenten. De meeste testpersonen gaven aan dat zij de tool als prettig en ondersteunend zouden ervaren in de praktijk. In het prototype was daarbij zichtbaar dat de AI-tool uitleg kon geven met een verwijzing naar de OVER-schrijfwijzer. In het huidige model is dit echter nog niet mogelijk. Naast het feit dat deze functionaliteiten nog niet beschikbaar zijn, presteert het naïve-RAG-model nauwelijks beter dan het baseline model (Paragraaf 6.1), dat eveneens onvoldoende in staat is om originele brieven naar B1-niveau te herschrijven.

Wanneer het model zich goed kan houden aan de regels van de OVER-schrijfwijzer wordt het systeem gezien als een waardevol hulpmiddel dat het werk kan vereenvoudigen en versnellen. Daarbij geven testpersonen wel aan dat er, vooral bij tijdsdruk, een gevaar bestaat dat suggesties te snel worden geaccepteerd zonder voldoende controle. Het versturen van een brief met foutieve informatie kan leiden tot ongewenste of risicovolle situaties, zoals het verkeerd informeren van inwoners. Daarom is het belangrijk dat medewerkers zich bewust blijven van hun eindverantwoordelijkheid en de inhoud altijd zorgvuldig controleren voordat een brief wordt verzonden. Hoewel bij het activeren van de AI-tool en het afronden van de suggesties meldingen worden getoond waarin wordt aangegeven dat medewerkers verantwoordelijk blijven voor de inhoud van de brief, is het noodzakelijk dat zij hier bij de implementatie ook expliciet over worden geïnformeerd.

7 Discussie

In dit hoofdstuk worden de discussiepunten besproken:

- Uit de DEDA-analyse (Bijlage F), blijkt dat de inzet van een AI-tool binnen gemeentelijke context niet alleen technische, maar ook menselijke en maatschappelijke risico's met zich meebrengt. Een belangrijk aandachtspunt is automation bias: medewerkers kunnen onder tijdsdruk geneigd zijn om AI-suggesties sneller te accepteren zonder deze kritisch te beoordelen. Dit is extra risicovol bij gemeentelijke brieven, omdat fouten gevolgen voor de inwoners kunnen hebben. Daarom zijn meldingen toegevoegd om gebruikers te wijzen op hun eindverantwoordelijkheid, en zal in een volgende testronde worden onderzocht welke aanvullende maatregelen nodig zijn om automation bias verder te beperken.
- Het naïeve-RAG-model presteert momenteel niet voor alle briefcategorieën even goed. Dit kan worden verklaard door de samenstelling van de door de OVER-gemeenten vrijgegeven data, waarin sommige briefcategorieën oververtegenwoordigd zijn. Daarnaast is het totale aantal beschikbare brieven beperkt en zijn de voorbeelden per briefcategorie ongelijk verdeeld. Hierdoor beschikt het model niet over voldoende representatieve voorbeelden, waardoor de AI-tool op dit moment nog niet als generaliseerbaar binnen de OVER-gemeenten kan worden beschouwd.
- Daarnaast is de rol van een mogelijke feedbackloop nog onvoldoende uitgewerkt (Bijlage F). In een toekomstscenario waarin het model leert van gebruikersbeslissingen, bestaat het risico dat foutieve suggesties worden versterkt wanneer deze herhaaldelijk worden goedgekeurd. Dit kan leiden tot structurele fouten in de inhoud van brieven. Daarnaast moet er bij het hertrainen opgelet worden dat privacygevoelige informatie niet wordt meegenomen. In het vervolgonderzoek van de aankomende weken moet daarom worden bekeken hoe het model kan worden getraind met behulp van het accepteren of negeren van suggesties door gebruikers, zonder dat foutieve goedkeuringen en privacygevoelige worden meegenomen.
- Hoewel de tool ondersteuning biedt bij het schrijven op B1-niveau, blijft een deel van de inwoners onvoldoende bereikt, namelijk de mensen die een lager taalniveau bezitten dan het B1-niveau. Het toepassen van B1-niveau is een eis vanuit de opdrachtgever, maar lost het bredere probleem van begrijpelijkheid in gemeentelijke teksten hiermee niet volledig op. Dit roept de vraag op of aanvullende communicatievormen nodig zijn naast schriftelijke informatie, zoals mondelinge toelichting of visuele ondersteuning.
- Als de AI-tool woordenlijsten gaat gebruiken om suggesties te geven, moeten deze lijsten worden aangepast wanneer de maatschappij verandert, zodat brieven inclusief geschreven blijven [16]. Dit komt omdat taal in de loop van de tijd verandert.

8 Conclusie

8.1 Conclusie/Terugkoppeling op de requirements

In tabel 16 wordt teruggeblickt naar de requirements en of deze behaald zijn met een bijbehorende toelichting.

Table 16. Acceptatiecriteria

RQ-nr	Acceptatiecriteria	Status	Toelichting
ER01	Gebruikerstesten uitvoeren waarbij wordt onderzocht of alle instructies duidelijk zijn en of testers het systeem begrijpen.	Behaald	De testpersonen vonden het paper prototype aantrekkelijk en konden de belangrijkste knoppen vinden. Daarnaast is de verkregen feedback meegenomen voor de volgende iteratie.
ER02	Bij het ontwerpen van het AI-systeem moet de mens eindverantwoordelijk blijven voor de beslissing of een klant alcohol mag kopen.	Behaald	Het systeem heeft een human-in-the-loop, waardoor een suggestie nooit automatisch wordt goedgekeurd.
ER03	In de documentatie moeten de ontwikkeling van het AI-systeem en de nauwkeurigheid van de leeftijd voorspellingen staan.	Behaald	Zie het verslag.
ER04	In het paper prototype moet de medewerker de optie krijgen om voor de AI-tool te kiezen.	Behaald	De AI-tool kan aan- en uitgezet worden door de medewerker.
ER05	Stakeholders ontvangen informatie over prestaties en beperkingen van het systeem.	Deels behaald	Medewerkers krijgen bij het aanzetten van de AI-tool dat ze eindverantwoordelijk blijven (omdat de AI-tool fouten kan maken). Echter is de AI-tool nog niet geïmplementeerd dus hebben de stakeholders nog geen informatie gehad over de prestaties en beperkingen van het systeem.
JR01	In de code moet staan dat de persoonlijke informatie verwijderd wordt direct na het opslaan van de brief. In de dataset moet gecontroleerd worden of de brieven anoniem zijn.	Behaald	Persoonlijke informatie van burgers is niet gebruikt bij het ontwikkelen van de AI-tool, enkel geanonimiseerde data.
JR02	In de code moet staan dat de persoonlijke informatie die opgeslagen is verwijderd wordt direct na het opslaan van de brief en dat het schrijven van de brief lokaal gebeurt.	Niet behaald	Het is nog niet duidelijk of de AI-tool lokaal kan werken, aangezien onbekend is welke computers de medewerkers gebruiken bij de OVER-gemeenten.
FR01	In de code moet staan dat de weergegeven suggesties eerst door de medewerker geaccepteerd moeten worden voordat de zinnen of woorden aangepast worden.	Behaald	Het systeem heeft een human-in-the-loop, waardoor een suggestie nooit automatisch wordt goedgekeurd.
FR02	In de code moet staan dat de medewerker de suggestie kan afwijzen of goedkeuren.	Behaald	Het systeem heeft een human-in-the-loop, waardoor een suggestie nooit automatisch wordt goedgekeurd.

RQ-nr	Acceptatiecriteria	Status	Toelichting
FR03	In de code moet een stuk tekst komen dat de medewerker in zijn scherm ziet. Hierin hoort de tekst informatief te zijn, waarbij wordt weergegeven dat de AI-tool slechts ter ondersteuning dient.	Behaald	In de AI-tool wordt de melding " <i>U bent zelf verantwoordelijk voor de inhoud van deze brief</i> " getoond als deze wordt aangezet.
FR05	In de evaluatie moeten de LiNT-II-scores van de AI-tool en het baseline model tegenover elkaar gezet worden.	Behaald	De score is met 0.08 gedaald ten opzichte van het baseline model.
FR06	De gemiddelde LiNT-II-score van de verschillende briefsoorten moet hoger zijn dan de LiNT-II-score op het baseline model.	Niet behaald	De LiNT-II-scores zijn niet voor elke briefsoort gelijk.
TR01	In de code moet verwerkt worden dat bewonersinformatie uit de brieven direct verwerkt en niet opgeslagen wordt.	Niet behaald	Het model is nog niet geïmplementeerd in de Word-extensie van de OVER-gemeenten.
TR02	In de code moet staan dat de AI-tool ondersteunend is voor de medewerker en niet beslissend.	Behaald	In de AI-tool wordt de melding " <i>U bent zelf verantwoordelijk voor de inhoud van deze brief</i> " getoond als deze wordt aangezet.
TR03	In de evaluatie moeten de AI-tool en het baseline model tegenover elkaar gezet worden.	Behaald	Zie het verslag.
TR04	In de code moet een datasplitsing staan volgens een validatie- en test-set van 50% en 50%.	Behaald	Deze exacte splitsing is toegepast.
TR05	De code dient lokaal gedraaid te worden.	Niet behaald	Het model is nog niet geïmplementeerd in de Word-extensie van de OVER-gemeenten.
TR06	Er dient een test uitgevoerd te worden met een medewerker, waarin de medewerker een brief verbetert door handmatig de schrijfwijzer te gebruiken en daarnaast dezelfde brief dient te verbeteren door het gebruik van de AI-tool.	Niet behaald	Deze exacte test is nog niet uitgevoerd.

8.2 Implicaties, toekomstig werk, aanbevelingen

De volgende punten zullen in de tweede fase uitgewerkt worden:

- We zullen de OVER-gemeenten bevragen over de computers die zij gebruiken, zodat kan worden vastgesteld of het model qua complexiteit geschikt is om lokaal op hun systemen te draaien.
- Daarnaast zullen we berekenen of het naïve-RAG-model brieven binnen twee seconden kan herschrijven, om te beoordelen of het voldoet aan de resource demand-eisen.
- Ook zullen we nagaan of de in dit rapport genoemde woordenlijsten voldoen aan de standaarden van de OVER-gemeenten en dus geschikt zijn voor gebruik in de AI-tool.
- Indien de OVER-gemeenten de woordenlijsten geschikt achten, zullen we een model ontwikkelen dat deze woordenlijsten actief gebruikt.
- We zullen onderzoeken of het mogelijk is om het model in een extensie te integreren, zodat het binnen Word kan functioneren.
- Verder onderzoeken we of het mogelijk is om bij elke suggestie de bijbehorende regel te tonen, inclusief een verwijzing naar de OVER-schrijfwijzer.
- Tot slot zal worden onderzocht op welke manier een advanced-RAG-implementatie de resultaten verder kan verbeteren.
- Wanneer het model is geïmplementeerd, kan een onboarding-functionaliteit een toegevoegde waarde bieden en wordt deze daarom gezien als een mogelijke toekomstige verbetering.

8.3 Ethische verantwoording (impact van de AI-oplossing op individu en maatschappij)

In de volgende paragrafen wordt de impact van de AI-oplossing besproken op zowel individueel niveau, voor medewerkers van de OVER-gemeenten, als op maatschappelijk niveau, voor de burgers die deze gemeenten bedienen.

8.3.1 Impact van de AI-oplossing op individu (medewerker van de OVER-gemeenten).

Als de AI-tool suggesties kan geven die voldoen aan de OVER-schrijfwijzer, kan deze medewerkers van de OVER-gemeenten ondersteunen bij het herschrijven van brieven volgens de vastgestelde richtlijnen. Medewerkers hechten er waarde aan dat bij elke suggestie de onderliggende reden en een verwijzing naar de OVER-schrijfwijzer worden weergegeven. Het huidige model voldoet hier echter nog niet aan. Implementatie in de huidige vorm zou medewerkers daarom mogelijk onvoldoende ondersteunen en zelfs tot extra werk kunnen leiden, bijvoorbeeld doordat het corrigeren van onjuiste suggesties extra tijd kost.

Desondanks is het model ethisch verantwoord, omdat de medewerker altijd eindverantwoordelijk blijft voor de inhoud. Er is daarmee sprake van een human-in-the-loop, wat ook expliciet wordt benadrukt in het tweede prototype.

8.3.2 Impact van de AI-oplossing op maatschappij (burgers van de OVER-gemeenten).

Medewerkers herschrijven brieven voor burgers in Word, waarbij de AI-tool lokaal wordt uitgevoerd. Dit betekent dat gegevens van burgers de gemeente niet verlaten en dus niet extern worden gedeeld. Op deze manier wordt voorkomen dat privacygevoelige informatie van burgers wordt verspreid naar derden.

Afhankelijk van de effectiviteit van de suggesties kunnen inwoners beter begrijpelijke brieven ontvangen. Tegelijkertijd bestaat het risico dat brieven juist minder duidelijk worden, bijvoorbeeld door automation bias of door het overnemen van onjuiste suggesties. Het is daarom belangrijk dat een medewerker de brief altijd zorgvuldig controleert op inhoudelijke en (juridische) correctheid.

References

- [1] [n. d.]. *B1-teksten in de praktijk: voorbeelden en tips / B1-teksten*. <https://b1teksten.nl/artikel/voorbeelden-van-b1-teksten>
- [2] [n. d.]. *Begrijpelijke communicatie / CommunicatieRijk*. <https://www.communicatierijk.nl/vakkennis/uitgangspunten-en-organisatie/inclusieve-communicatie/begrijpelijke-communicatie>
- [3] Chroma 2025. *Chroma-Core/Chroma*. Chroma. <https://github.com/chroma-core/chroma>
- [4] [n. d.]. *Focus op AI bij de rijksoverheid / Algemene Rekenkamer*. <https://www.rekenkamer.nl/documenten/2024/10/16/focus-op-ai-bij-de-rijksoverheid>
- [5] 2025. *nluuv/Leesplank_Noot*. https://github.com/nluuv/Leesplank_Noot original-date: 2024-10-14T07:40:41Z.
- [6] [n. d.]. *Taalniveau B1*. <https://site.dijkenwaard.nl/huisstijl/heldere-taal/direct-duidelijk-dijk-en-waard/taalniveau-b1>
- [7] [n. d.]. *Train Test Validation Split: How To & Best Practices [2024]*. <https://www.v7labs.com/blog/train-validation-test-set>
- [8] 2024. *UWV/leesplank-noot-eurollm-1.7b · Hugging Face*. <https://huggingface.co/UWV/leesplank-noot-eurollm-1.7b>
- [9] 2025. *UWV/leesplank-noot-granite-3.3-2b · Hugging Face*. <https://huggingface.co/UWV/leesplank-noot-granite-3.3-2b>
- [10] 2025. *UWV/leesplank-noot-llama-3.2-3b · Hugging Face*. <https://huggingface.co/UWV/leesplank-noot-llama-3.2-3b>
- [11] 2025. *lancedb/lancedb*. <https://github.com/lancedb/lancedb> original-date: 2023-02-28T01:15:17Z.
- [12] 2025. *qdrant/qdrant*. <https://github.com/qdrant/qdrant> original-date: 2020-05-30T21:37:01Z.
- [13] Moustafa Abdelwanis, Hamdan Khalaf Alarafati, Maram Muhanad Saleh Tammam, and Mecit Can Emre Simsekler. 2024. Exploring the risks of automation bias in healthcare artificial intelligence applications: A Bowtie analysis. 5, 4 (2024), 460–469. doi:10.1016/j.jnlssr.2024.06.001
- [14] Algemene Rekenkamer. 2016. Volwassenen met moeite met taal van rekenen: kloof tussen probleem en aanpak via riksbeleid. <https://www.rekenkamer.nl/publicaties/persberichten/2016/04/20/volwassenen-met-moeite-met-taal-van-rekenen-kloof-tussen-probleem-en-aanpak-via-riksbeleid>.
- [15] Amsterdam. [n. d.]. *Helder juridisch woordenboek*. <https://www.amsterdam.nl/schrijfwijzer/heldere-taal-basis-onze-huisstijl/helder-juridisch-schrijven/helder-juridisch-woordenboek/> Last Modified: 2025-12-02; Publisher: Gemeente Amsterdam.
- [16] Amsterdam. [n. d.]. *Inclusieve taal: richtlijnen en tips*. <https://www.amsterdam.nl/schrijfwijzer/inclusieve-taal-richtlijnen-tips/> Last Modified: 2025-12-15 Publisher: Gemeente Amsterdam.
- [17] Amsterdam. [n. d.]. *Inclusieve woordenlijst*. <https://www.amsterdam.nl/schrijfwijzer/inclusieve-taal-richtlijnen-tips/inclusieve-woordenlijst/> Last Modified: 2025-12-02; Publisher: Gemeente Amsterdam.
- [18] Amsterdam. [n. d.]. *Moeilijke-woordenboek (inclusief niet te gebruiken ambtelijke taal)*. <https://www.amsterdam.nl/schrijfwijzer/moeilijke-woorden/> Last Modified: 2025-12-07; Publisher: Gemeente Amsterdam.
- [19] Autoriteit Persoonsgegevens. 2025. *AVG-Randvoorwaarden voor generatieve AI*. Technical Report. Autoriteit Persoonsgegevens. Accessed: 2025-11-18.
- [20] Marvin van Bekkum and Frederik Zuiderveen Borgesius. 2025. De spanning tussen het non-discriminatierichtlijn en het gegevensbeschermingsrecht: heeft de AVG een nieuwe uitzondering nodig om discriminatie door kunstmatige intelligentie tegen te gaan? arXiv:2509.08836 [cs] doi:10.48550/arXiv.2509.08836
- [21] Betabit. 2018. ETHISCHE RICHTSNOEREN voor BETROUWBARE KI Deskundengroep op hoog niveau inzake kunstmatige intelligentie. Report. <https://www.betabit.nl/media/4614/ethicsguidelinesfortrustworthyai-nl.pdf>
- [22] François Chollet. 2023. *Deep Learning with Python*. Chollet. <https://deeplearningwithpython.io/> Online boek, geraadpleegd op 2025-11-21.
- [23] Mischa Corsius, Henk Pander Maat, Els van der Pool, and Wouter Sluis-Thiescheffer. 2023. Monitor Begrijpelijkheid overheidsteksten 2022. Report. <https://taalunie.org/publicaties/218/monitor-begrijpelijkheid-overheidsteksten-2022>
- [24] Mischa Corsius, Els Van Der Pool, and Wouter Sluis-Thiescheffer. 2024. Begrijp jij het? Mixed-method monitor van overheidsteksten. 45, 1 (2024), 40–65. doi:10.5117/TVT2023.3.003.CORS
- [25] Analyttica Datalab. 2021. *What is meant by ‘Stratified Split’?* <https://medium.com/@analyttica/what-is-meant-by-stratified-split-289a8a986a90>
- [26] European Union Agency for Fundamental Rights. 2025. Artikel 21 – Non-discriminatie. EU-Handvest van de grondrechten. <https://fra.europa.eu/nl/eu-charter/article/21-non-discriminatie> Geraadpleegd op 24 september 2025.
- [27] GebruikerCentraal. [n. d.]. Toolkit Taal. <https://toolkittaal.gebruikercentraal.nl/richtlijnen/>.
- [28] GebruikerCentraal. [n. d.]. Van Direct Duidelijk naar duidelijke overheidscommunicatie. <https://www.gebruikercentraal.nl/meedoen/netwerk-direct-duidelijk/duidelijke-overheidscommunicatie/> Accessed: November 13, 2025.
- [29] Gemeente Amsterdam. [n. d.]. Schrijfwijzer v4. <https://openresearch.amsterdam.nl/page/94061/schrijfwijzer-v4.pdf>.
- [30] Gemeente Oostzaan. [n. d.]. Schrijfwijzer. <https://dlo.mijnhva.nl/d2l/le/content/691830/viewContent/2743225/View>.
- [31] Hongyu Gong, Yelong Shen, Dian Yu, Jianshu Chen, and Dong Yu. 2020. Recurrent Chunking Mechanisms for Long-Text Machine Reading Comprehension. doi:10.48550/arXiv.2005.08056 arXiv:2005.08056 [cs].
- [32] Sian Gooding and Manuel Trägut. 2022. One Size Does Not Fit All: The Case for Personalised Word Complexity Models. arXiv:2205.02564 [cs] doi:10.48550/arXiv.2205.02564
- [33] Sian Gooding and Manuel Trägut. 2022. One Size Does Not Fit All: The Case for Personalised Word Complexity Models. In *Findings of the Association for Computational Linguistics: NAACL 2022*. Association for Computational Linguistics, Seattle, USA, 353–365. <https://aclanthology.org/2022.findings-naacl.27/>
- [34] Google PAIR. 2022. People + AI Guidebook: Patterns. <https://pair.withgoogle.com/guidebook/patterns> Accessed: 2025-12-02.

- [35] Grondwet voor het Koninkrijk der Nederlanden. 2023. Artikel 1: Gelijke behandeling en discriminatieverbod. DeNederlandseGrondwet.nl. https://www.denederlandsegrondwet.nl/id/vgrnb2er8avw/artikel_1_gelijke_behandeling_en Geraadpleegd op 6 december 2025.
- [36] Eliza Hobo, Charlotte Pouw, and Lisa Beinborn. 2023. “Geen makkie”: Interpretable Classification and Simplification of Dutch Text Complexity. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*. Association for Computational Linguistics, Toronto, Canada, 503–517. <https://aclanthology.org/2023.bea-1.42>
- [37] Kayleigh Hoogenboom. 2022. Duidelijke Overheidscommunicatie voor gemeenten. *Hoogenboom* (2022).
- [38] Jenia Kim and Henk Pander Maat. 2025. *LINT-II: readability assessment for Dutch*. https://github.com/vanboefer/lint_ii Python package.
- [39] Akshay Kore. 2022. *Designing Human-Centric AI Experiences: Applied UX Design for Artificial Intelligence*. Apress.
- [40] Akshay Kore. 2022. *Designing Human-Centric AI Experiences: Applied UX Design for Artificial Intelligence*. Apress. doi:10.1007/978-1-4842-8088-1
- [41] Gustavo H. Paetzold and Lucia Specia. 2017. A Survey on Lexical Simplification. 60 (2017), 549–593. doi:10.1613/jair.5526
- [42] Henk Pander Maat, Suzanne Kleijn, and Servaas Frissen. [n. d.]. LiNT: een leesbaarheidsformule en een leesbaarheids instrument. *Tijdschrift voor Taalbeheersing* 45, 1 ([n. d.]), 2–39.
- [43] Patronus AI. 2023. RAG Evaluation Metrics: Best Practices for Evaluating RAG Systems. <https://www.patronus.ai/llm-testing/rag-evaluation-metrics>. Accessed: 2025-11-28.
- [44] Renyi Qu and Ruixuan Tu. [n. d.]. Is Semantic Chunking Worth the Computational Cost? ([n. d.]).
- [45] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2022. Improving Language Understanding by Generative Pre-Training. (2022).
- [46] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2023. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. arXiv:1910.10683 [cs] doi:10.48550/arXiv.1910.10683
- [47] Edwin Rijgersberg. 2025. *Het einde van GEITje 1*. <https://goingdutch.ai/nl/posts/geitje-takedown/> Section: posts.
- [48] SecureDataService GmbH. 2023. Artikel 25 EU-AVG: Gegevensbescherming door ontwerp en door standaardinstellingen. *Privacy-Regulation.eu*. <https://www.privacy-regulation.eu/nl/artikel-25-gegevensbescherming-door-ontwerp-en-door-standaardinstellingen-EU-AVG.htm> Laatste wijziging volgens site: 04-04-2023..
- [49] SecureDataService GmbH. 2023. Artikel 32 EU-AVG: Beveiliging van de verwerking. *Privacy-Regulation.eu*. <https://www.privacy-regulation.eu/nl/artikel-32-beveiliging-van-de-verwerking-EU-AVG.htm> Laatste wijziging volgens site: 04-04-2023. Geraadpleegd op <vul-hier-je-consultatiatedatum-in>.
- [50] SecureDataService GmbH. 2023. Artikel 5 EU-AVG: Beginselen inzake verwerking van persoonsgegevens. *Privacy-Regulation.eu*. <https://www.privacy-regulation.eu/nl/artikel-5-beginselen-inzake-verwerking-van-persoonsgegevens-EU-AVG.htm> Laatste wijziging volgens site: 04-04-2023. Geraadpleegd op 6-12-2025.
- [51] KLINKENDE TAAL. [n. d.]. Is het B1? <https://ishebt1.nl/>. Accessed: 2025-12-10.
- [52] Taalunie. 2022. Monitor Begrijpelijkheid overheidsteksten 2022. <https://taalunie.org/publicaties/218/monitor-begrijpelijkheid-overheidsteksten-2022>.
- [53] Daniel Vlantis, Iva Gornishka, and Shuai Wang. 2024. Benchmarking the Simplification of Dutch Municipal Text. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. European Language Resources Association (ELRA), Torino, Italy, 2217–2226. <https://aclanthology.org/2024.lrec-main.199>
- [54] Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. 2022. A Survey of Human-in-the-loop for Machine Learning. 135 (2022), 364–381. arXiv:2108.00941 [cs] doi:10.1016/j.future.2022.05.014
- [55] Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. 2022. A Survey of Human-in-the-loop for Machine Learning. 135 (2022), 364–381. arXiv:2108.00941 [cs] doi:10.1016/j.future.2022.05.014
- [56] Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2016. Optimizing Statistical Machine Translation for Text Simplification. *Transactions of the Association for Computational Linguistics (TACL)* 4 (2016), 401–415. doi:10.1162/tacl_a_00107
- [57] Kaiyan Zhang, Yuxin Zuo, Bingxiang He, Youbang Sun, Runze Liu, Che Jiang, Yuchen Fan, Kai Tian, Guoli Jia, Pengfei Li, Yu Fu, Xingtai Lv, Yuchen Zhang, Sihang Zeng, Shang Qu, Haozhan Li, Shijie Wang, Yuru Wang, Xinwei Long, Fangfu Liu, Xiang Xu, Jiaze Ma, Xuekai Zhu, Ermo Hua, Yihao Liu, Zonglin Li, Huayu Chen, Xiaoye Qu, Yafu Li, Weize Chen, Zhenzhao Yuan, Junqi Gao, Dong Li, Zhiyuan Ma, Ganqu Cui, Zhiyuan Liu, Biqing Qi, Ning Ding, and Bowen Zhou. 2025. A Survey of Reinforcement Learning for Large Reasoning Models. arXiv:2509.08827 [cs] doi:10.48550/arXiv.2509.08827
- [58] Sanqiang Zhao, Rui Meng, Daqing He, Andi Saptono, and Bambang Parmanto. 2018. Integrating Transformer and Paraphrase Rules for Sentence Simplification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 3164–3173. doi:10.18653/v1/D18-1355
- [59] Sanqiang Zhao, Rui Meng, Daqing He, Andi Saptono, and Bambang Parmanto. 2018. Integrating Transformer and Paraphrase Rules for Sentence Simplification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun’ichi Tsujii (Eds.). Association for Computational Linguistics, Brussels, Belgium, 3164–3173. doi:10.18653/v1/D18-1355

A Informatie vanuit OVER-gemeenten

A.1 Kennismaking OVER-gemeenten (13 nov)

Op 13 november bezochten drie medewerkers van de WMO-afdeling van de OVER-gemeenten de Hogeschool van Amsterdam om het project over tekstversimpeling toe te lichten waar studenten van de Master Applied AI aan gaan werken. De OVER-gemeenten is de ambtelijke organisatie voor de Gemeente Oostzaan en Wormerland. Het doel van dit project is medewerkers te ondersteunen bij het opstellen van brieven op B1-niveau.

De bijeenkomst startte met een presentatie van Annamaria van Teeseling, waarna de studenten de gelegenheid kregen om vragen te stellen aan de drie aanwezige medewerkers. De belangrijkste notities naar aanleiding van deze kennismaking zijn hieronder vastgelegd.

A.1.1 *Introductie van Annamaria van Teeseling voer het project.*

Introductie medewerkers OVER-gemeenten

- Annamaria van Teeseling werkt voor de ambtelijke organisatie OVER-gemeenten (voor Gemeente Oostzaan en Wormerland) en is teamleider van de afdeling WMO.
- Franny van der Bijl is WMO-consultent bij OVER-gemeenten. Zij houdt zich onder andere bezig met het herschrijven van brieven volgens de interne schrijfwijzer.
- Michael is projectleider bij de ICT-afdeling. Hij bewaakt de privacyaspecten binnen het gebruik van AI en is betrokken bij het onderzoeken van mogelijkheden voor veilige inzet van AI binnen de gemeente.

Wat doet de OVER-gemeenten?

De OVER-gemeenten is verantwoordelijk voor taken zoals openbare orde, zorg, infrastructuur en sociale voorzieningen. De besluitvorming ligt bij het college van BW en de gemeenteraad, en de uitvoering bij de ambtelijke organisatie. Voorbeelden van werkzaamheden zijn:

- Ze zorgen voor huizenbouw
- Ze geven vorm aan de leefomgeving
- Ze lichten de raad in, helpen en steunen deze
- Ze schrijven beleid
- Ze behandelen WMO-aanvragen
- Ze vangen Oekraïners op
- Ze geven vergunningen af
- Ze voeren veel gesprekken met inwoners en versturen brieven

Bij brieven is het belangrijk dat de juridische inhoud begrijpelijk blijft, zonder dat deze te eenvoudig of kinderachtig overkomt.

Werken met AI in een gemeente, de uitdagingen

Herschrijven van teksten met AI is op dit moment niet toegestaan vanwege de bescherming van persoonsgegevens en andere vertrouwelijke informatie. Belangrijke aandachtspunten zijn:

- Privacy: bescherming van persoonsgegevens (AVG).

- Bedrijfsgevoelige informatie: risico op datalekken.
- Transparantie: AI-beslissingen moeten uitlegbaar zijn.
- Ethiek: gelijke behandeling en inclusie (bijvoorbeeld voorkomen van situaties zoals de toeslagenaffaire).

Hoe werkt de afdeling WMO?

De afdeling WMO voert de Wet maatschappelijke ondersteuning uit (hulp bij zelfredzaamheid). Taken zijn onder meer:

- Beoordelen van indicaties
- Bieden van ondersteuning
- Samenwerken met zorgaanbieders
- Zorgen voor begrijpelijke en toegankelijke communicatie met inwoners

Communicatie op B1-niveau

- Waarom B1? Teksten op B1-niveau zijn begrijpelijk voor ongeveer 80% van de inwoners (bron: CBS, volgens medewerkers).
- Onze ambitie: eenduidige en duidelijke communicatie.
- Hulpmiddel: interne schrijfwijzer.

De gemeente wil dat alle communicatie op dezelfde manier wordt opgesteld, zodat inwoners ervaren dat *zij* met één organisatie te maken hebben. Ter illustratie van de taalniveaus:

- A1: eenvoudig basisniveau (bijvoorbeeld taalgebruik voor op vakantie).
- A2: niveau dat vereist is voor inburgering.
- B1: helder en begrijpelijk voor een brede doelgroep.
- Etc.

Schrijfwijzer & standaard WMO-brieven

- De schrijfwijzer is opgesteld om te zorgen dat de communicatie consistent wordt.
- De standaardbrieven (van de WMO) moeten nog getoetst worden aan deze schrijfwijzer.

Opdrachten voor de studenten

- De gemeente vraagt de studenten om een slimme AI-oplossing te ontwikkelen:
 - Die medewerkers eenvoudig ondersteunt bij het schrijven van brieven volgens de schrijfwijzer,
 - Gebruikmakend van meegeleverde voorbeeldbrieven,
 - Rekening houdend met gemeentelijke kaders en privacyrichtlijnen.

A.1.2 Vragen en antwoorden met betrekking tot het onderzoek.

Stakeholders

- **Vraag:** Wat valt er onder 'OVER'-gemeenten?
- **Antwoord:** OVER-gemeenten is de ambtelijke organisatie voor de Gemeente Oostzaan en Wormerland.

Inhoud van brieven & misverstanden van burgers

- **Vraag:** Welke onderwerpen bevatten de brieven?
- **Antwoord:** De brieven gaan onder andere over schulden, uitkeringen, werk en inkomen, hulp bij het huishouden, en informatie voor Oekraïense vluchtelingen.
- **Vraag:** Welke problemen ontstaan door miscommunicatie? (Met miscommunicatie wordt bedoeld dat burgers inhoud van de brief niet begrijpen.)
- **Antwoord:**
 - Het gaat om een groot aantal brieven; binnen de WMO-afdeling alleen al ongeveer 250.
 - Daarnaast is schrijven een vak. Consulenten moeten naast hun reguliere werkzaamheden ook nog goed kunnen schrijven. De teksten moeten niet alleen juridisch dekkend zijn volgens wettelijke verplichtingen, maar ook begrijpelijk.
 - Daarnaast is de situatie van elke inwoner persoonlijk, dus daar moeten brieven op worden aangepast. (Een kanttekening bij het antwoord is dat de vragen zijn beantwoord door mensen van het team WMO met kennis in het sociaal domein.)
 - Wetgeving verplicht de gemeente tot het gebruik van begrijpelijke taal.
 - Uit data van het CBS blijkt dat veel mensen moeite hebben met het begrijpen van complexe teksten.
 - Laaggeletterden melden zich niet bij de gemeente, waardoor het probleem niet, waardoor het probleem niet direct zichtbaar is.
 - Er is een experiment gestart waarbij inwoners stickers op brieven kunnen plakken bij tekst die voor hen onduidelijk is. Dit staat nog in de kinderschoenen.
- **Vraag:** Welke doelgroepen hebben de meeste moeite met het begrijpen van brieven?
- **Antwoord:** Burgers met schulden, een uitkering en Oekraïners.

Niveau van de brieven

- **Vraag:** Kunnen brieven ook op andere niveaus dan B1 worden geschreven?
- **Antwoord:** Ja. Het taalniveau kan worden aangepast, bijvoorbeeld naar A1 als dit beter past bij de situatie van de inwoner. Soms kan het niveau juist hoger worden gezet, zoals B2.
- **Vraag:** Heeft de gemeente voorkeur voor A1 of B1, als A1 ook technisch mogelijk is?
- **Antwoord:** B1.
- **Vraag:** Welke tools worden op dit moment gebruikt?
- **Antwoord:** Medewerkers werken voornamelijk in Word, Outlook en een eigen klantsysteem genaamd Zorgnet. Word wordt het meest gebruikt.
- **Vraag:** Maken medewerkers nu gebruik van AI-tools, zoals ChatGPT?
- **Antwoord:** Nee. Het herschrijven van brieven met AI is niet toegestaan vanwege bescherming van persoonlijke gegevens.
- **Vraag:** Zijn er al conclusies uit de lopende pilot met de schrijfwijzer?
- **Antwoord:** De eerste pilotbrieven zijn ontvangen, maar over de resultaten waren ze nog niet tevreden. Franny heeft de brieven zelf nog moeten herschrijven.
- **Vraag:** Wie bepaalt nu of een tekst begrijpelijk is voor iedereen en hoe wordt de schrijfwijzer toegepast in de huidige situatie?
- **Antwoord:**
 - Medewerkers bepalen zelf of een tekst begrijpelijk is voor iedereen. Zij hebben een training gehad en toetsen elkaar stukken door brieven onderling na te lezen. Volledige garantie dat een tekst op B1-niveau is, bestaat echter niet.
 - De teamleider gaf aan dat de voorkeur uitgaat naar minder toetsing. Zij wil liever vertrouwen op de deskundigheid van de medewerkers, zonder dat onderlinge controle altijd nodig is.
- **Vraag:** Waarom is Copilot nog niet genoeg?
- **Antwoord:** Vanwege de privacy, servers, etc.

- **Vraag:** Worden er al tools gebruikt voor het begrijpelijk maken van brieven?
- **Antwoord:** Ja. Tools zoals ishetb1.nl, maar eigenlijk werkt dit niet ideaal en werken ze met hun eigen kennis, kennis van een training en de schrijfwijzer.
- **Vraag:** Moeten medewerkers nog rekening houden met andere wetgeving dan de AVG, etc.?
- **Antwoord:** Nee.

Context waarom de gemeente een hulpmiddel wil

- **Notities:**
 - Medewerkers willen niet continu moeten kijken naar de schrijfwijzer. (Deze is namelijk gigantisch en niet uit het hoofd te leren)
 - Er is geen tijd bekend over hoelang medewerkers gemiddeld bezig zijn met het schrijven van een brief. Medewerkers van de OVER-gemeenten hebben in principe geen tijdslimiet voor het schrijven van brieven. Sommige gemeenten hebben dat wel.
 - De gemeente wil voorkomen dat het proces geautomatiseerd wordt. Medewerkers moeten de controle behouden over de tekst.

Data en beschikbaarheid van brieven

- **Vraag:** Is er gebruikersdata over hoe inwoners brieven ervaren?
- **Antwoord:** Nee, er zijn geen gegevens bekend over wat mensen als onduidelijk zien.
- **Vraag:** Kunnen er brieven van andere afdelingen worden gedeeld?
- **Antwoord:** Daar gaan we achteraan.
- **Notities:**
 - De brieven die we hebben gekregen zijn standaard teksten. Tussen rode haken moet dus nog tekst geplaatst worden. Dit zijn een soort templates. Binnen de hele gemeente moeten alle afdelingen brieven schrijven op B1 niveau.
 - De brieven die we hebben gekregen zijn standaard teksten. Tussen rode haken moet dus nog tekst geplaatst worden. Dit zijn een soort templates. Binnen de hele gemeente moeten alle afdelingen brieven schrijven op B1 niveau.
 - In de beschikkingen zijn de juridische stukjes hetzelfde in elke brief. De gedeeltes over beschikkingen zijn variabel en worden per burger en per werknemer anders geschreven.

Modelkeuze (AI-technisch stuk)

- **Vraag:** Wordt het aangeraden om een model te trainen vanaf scratch?
- **Antwoord:** Nee. Dit heeft te maken met de grootte van de modellen. Ze zijn te groot om zelf te trainen. Het wordt dus aangeraden om een voorgetraind model te finetunen op eigen data. Je moet dan wel verantwoorden waarom je een bepaalde architectuur kiest. Het mag dus wel. Finetunen hoeft ook niet. Onthoud dat een NLP echt niet alleen taalmodellen zijn.
- **Notities:**
 - Het zou een nice to have zijn als het herleidbaar is waar gegenereerde content op gebaseerd is en hoe het model werkt.
 - Als tekst vereenvoudigd wordt, moet deze niet ineens 4 pagina's lang worden. Ook moet belangrijke informatie niet verdwijnen en moet het voor inwoners wel duidelijk zijn dat ze bijvoorbeeld in bezwaar kunnen gaan.
 - De tool moet door alle afdelingen gebruikt kunnen worden.

Interface

- **Vraag:** Heeft de gemeente voorkeur voor een bepaald soort interface?
- **Antwoord:** Ze werken nu voornamelijk met Firefox en sinds kort ook weer Edge (tijdelijk niet door security breach in AI-functies). Ze hebben geen algemene voorkeur voor een ideale situatie.
- **Vraag:** Willen medewerkers tijdens het schrijven al suggesties zien van bijvoorbeeld synoniemen, of pas later bij het scannen?
- **Antwoord:**
 - Ze hebben nu al een tool die dit doet. Mensen vinden het juist makkelijk dat er tussendoor al voorstellen komen (aldus Michael).
 - Michael beschreef hoe hij zelf met een tool heeft gewerkt die exact klonk als wat overleaf doet en vond het een zeer fijne tool.
 - Er moeten niet te veel meldingen tegelijkertijd in beeld komen. Niet te veel streepjes en kleurtjes tegelijkertijd (aldus Franny).
 - Het liefst zien ze dit dus al tijdens het schrijven en niet pas na het inscannen.
 - De tool moet niet te druk worden (aldus Franny).
 - Het liefst willen ze ook iets van de tool leren (aldus Franny). Ze ziet dus liever suggesties in plaats van dat automatisch een hele brief wordt herschreven.

Model lokaal draaien/technische eisen

- **Vraag:** Moet de tool lokaal kunnen draaien?
- **Antwoord:** Het liefst wel, zodat je ook met persoonlijke gegevens kan werken.
- **Notities:**
 - Het zou fijn zijn als de tool offline zou kunnen runnen.
 - Het is interessant en relevant om duurzaam te werk te gaan zover het kan, probeer hier rekening mee te houden.

Contact & beschikbaarheid

- **Vraag:** Kunnen we jullie bereiken voor vragen tijdens het project?
- **Antwoord:** Hiervoor zijn afgesproken data.

Overige vragen

- **Vraag:** Is er momenteel een feedbackloop?
- **Antwoord:** Nee, nogmaals deels door schaamte. We proberen een inclusiebeleid op te zetten. We willen meer nadruk leggen op dat iedereen de brieven begrijpt.
- **Notities:**
 - Er zijn andere gemeenten die apps gebruiken (volgens Pim). Gemeente OVER gebruikt dat niet. Ze worden getraind en dan losgelaten op de werkvloer.

A.2 Vragen en antwoorden OVER-gemeenten (21 nov)

Om meer inzicht te krijgen in de huidige situatie en context van de OVER-gemeenten konden alle projectgroepen tot 21 november vragen indienen. De docenten selecteerden vervolgens de vragen om doublures te voorkomen. Medewerkers van de OVER-gemeenten hebben deze vragen beantwoord. De antwoorden zijn hieronder te vinden.

Q1: Projectomvang – Hoe belangrijk is het om schrijvers een assistent te bieden die hen ondersteunt tijdens het schrijfproces?

Het bieden van een assistent kan voor sommige collega's helpend zijn, omdat er direct feedback wordt gegeven. Wij hopen/verwachten dat collega's hier snel van kunnen leren. Wij kunnen ons voorstellen dat dit niet voor alle collega's prettig werkt en dat zij hier niet allemaal gebruik van zullen maken.

Wij vinden daarom een oplossing die achteraf het stuk verbetert en aanpast conform de schrijfwijzer ook heel prettig. Mogelijk met optie voor functie aan- of uitzetten naar wens van de gebruiker?

Q2: Projectomvang – Hoe belangrijk is het om te beoordelen of brieven voldoen aan de vereiste leesbaarheidsnormen?

Het is heel belangrijk om te beoordelen of brieven voldoen aan de vereiste leesbaarheidsnormen. Gemeentelijke brieven gaan vaak over zaken die direct invloed hebben op het leven van inwoners, zoals belastingen, vergunningen, ondersteuning of regelingen. Als de tekst te ingewikkeld is, kunnen mensen de informatie verkeerd begrijpen of belangrijke stappen missen. Dat kan leiden tot onzekerheid, fouten of zelfs problemen.

Door te controleren of een brief op B1-niveau is geschreven, weet je zeker dat de meeste inwoners de tekst goed kunnen volgen. Dit maakt de communicatie eerlijker, duidelijker en toegankelijker voor iedereen. Bovendien scheelt het tijd en vragen voor de gemeente, omdat inwoners sneller begrijpen wat er van hen wordt verwacht.

Q3: B1-niveau & Begrijpelijke Taal – Worden er momenteel Engelse woorden in brieven gebruikt, of uitsluitend Nederlandse?

Uitsluitend Nederlands.

Q4: Huidig Schrijfproces en Workflow – Hoe verloopt de volledige procedure van het schrijven van een brief (van aanvraag tot verzending)?

Dat kan verschillen, veel afdelingen werken met standaardbrieven, waarbij veel al voorgeschreven is en de medewerker alleen kleine aanpassingen hoeft te doen die van toepassing zijn op de situatie van de inwoner.

Echter zijn er ook teams, denk aan juridische zaken, die brieven of stukken moeten schrijven die direct op de vraag aansluiten; hierbij moet de medewerker zelf een stuk schrijven.

Standaardbrieven worden vaak op voorhand door een medewerker van de afdeling gemaakt. Ook is het per team verschillend of brieven getoetst worden; daar hebben we als organisatie geen eenduidige werkwijze in.

Q5: Huidig Schrijfproces en Workflow – Welke rollen zijn betrokken bij het schrijfproces? (bijv. consulent, juridisch medewerker, kwaliteitsmedewerker, communicatie)

Alle rollen, door de hele organisatie. Ook beleid, vergunningen etc.

Q6: Huidig Schrijfproces en Workflow – Wat zijn de huidige pijnpunten of vertragingen in het proces? (bijv. juridisch taalgebruik, inconsistentie, veel correctierondes)

Niet op een eenduidige manier communiceren, onbekend met de werkwijze, veel tijd om brieven van anderen te toetsen/controleren.

Q7: Kwaliteitscontrole & Werkinstructies – Welke specifieke onderdelen van de schrijfwijzer zijn de grootste knelpunten in het huidige gebruik?

Het is een groot document, waarvan wij denken dat deze in de praktijk bij weinig mensen bekend is. Hierdoor is het gevaar dat dit een papieren werkelijkheid is en dat de medewerkers allemaal op eigen wijze door blijven communiceren.

Wij zoeken een tool waarmee we op een meer eenduidige manier, conform afspraken, kunnen communiceren.

De schrijfwijzer geeft ons, naast de schrijfstijl, ook veel detailrichtlijnen voor een eenduidige werkwijze, zoals de manier van een telefoonnummer noteren en het gebruiken van opsommingstekens. Om dit allemaal handmatig te wijzigen bij de huidige brieven zal ons enorm veel tijd kosten. De AI-tool moet ervoor zorgen dat dit voor elke brief (en dus elke afdeling) gelijk is.

Q8: AI-gebruik, Behoeften & Beslissingen – Wat is volgens jullie de ideale werkwijze voor het schrijven/evalueren van brieven met AI?

Het liefst hebben we een tool die het schrijven makkelijk maakt voor de schrijver (tijdwinst) en die de brieven zodanig toetst dat dit niet door een (tweede) collega hoeft te gebeuren.

We zoeken een tool voor zowel het herschrijven van huidige brieven, als een tool die ondersteunend is bij het schrijven van nieuwe brieven. Een tool die ons hierbij ondersteunt qua schrijfwijze (taalgebruik, zinsopbouw) en die het ons ook "leert", zodat wij zien waar we vaak "de fout" in gaan qua schrijfstijl.

Opmerking van Maikel: Mee eens, als de grootheid van het document in kleine stappen kan worden gerefereerd zal dat helpen. Ter illustratie: in de tooling voor het digitaal anonimiseren wordt er aan het einde van het document een grondslag gegeven voor de anonimisering. Zie Q6. Bijvoorbeeld: "Gegevens geanonimiseerd o.b.v. grondslag Financiële gegevens B.6.1".

Q9: AI-gebruik, Behoeften & Beslissingen – Is er binnen de gemeente of elders al onderzoek gedaan naar AI-systemen om teksten begrijpelijk te maken?

Binnen de pilot M365 Copilot zijn er zeker mogelijkheden om teksten te redigeren naar begrijpelijkere leesniveaus. Dit biedt zeker standaard al aardige mogelijkheden, maar dat is niet gekoppeld aan de schrijfwijzer. Juist die combinatie zou de tool extra zinvol maken.

A.3 Vragen en antwoorden Gemeente Oostzaan (6 jan)

Deze wordt ingevuld als de vragen zijn opgesteld en beantwoord door de OVER-gemeente in januari.

We gaan onder andere een vraag stellen met betrekking tot de gebruikte woordenlijsten van de Gemeente Amsterdam. We gaan vragen of zij deze woordenlijst correct vinden voor gebruik in hun AI-tool. Daarnaast willen we vragen op wat voor computer de medewerkers werken, zodat we kunnen concluderen of de modelcomplexiteit klein genoeg is, zodat de tool lokaal kan draaien.

B Waardeproposities per stakeholder

Stakeholder 1: OVER-gemeenten

- **Conceptnaam:**
AI-applicatie om brieven begrijpelijk te maken.
- **Door gebruik te maken van:**
Een lokaal draaiend AI-systeem dat opties geeft om brieven te vereenvoudigen.
- **Om:**
De werkdruk van medewerkers te verlagen en het schrijfproces van brieven te versnellen.
- **Wij kunnen helpen:**
Door medewerkers te ondersteunen bij het herschrijven en controleren van brieven.
- **Met een betere manier om:**
Brieven sneller begrijpelijk te maken volgens de OVER-gemeenten schrijfwijzer.
- **Zodat:**
Medewerkers minder tijd kwijt zijn aan het begrijpelijk maken van tekst (volgens de schrijfwijzer) en meer tijd overhouden voor andere werkzaamheden.
- **Met:**
Behoud van gemeentelijke kwaliteitsstandaarden, zonder afhankelijk te zijn van externe (onveilige) AI-platforms.

Stakeholder 2: Inwoners van de OVER-gemeenten

- **Conceptnaam:**
AI-applicatie om brieven begrijpelijk te maken.
- **Door gebruik te maken van:**
AI-ondersteunde taalcontrole die medewerkers suggesties geeft voor het vereenvoudigen en inclusief schrijven van tekst.
- **Om:**
Inwoners duidelijkheid te bieden over hun rechten en acties die ze eventueel moeten ondernemen.
- **Wij kunnen helpen:**
Door met behulp van de AI-applicatie medewerkers te ondersteunen bij het begrijpelijk, toegankelijker en inclusiever schrijven van brieven voor inwoners.
- **Met een betere manier om:**
Misverstanden bij inwoners te voorkomen bij belangrijke onderwerpen.
- **Zodat:**
(Alle) inwoners begrijpen wat ze moeten weten of doen.
- **Zonder:**
Moeite of stress.

Stakeholder 3: Medewerkers die brieven schrijven of controleren

- **Conceptnaam:**
AI-applicatie om brieven begrijpelijker te maken.
- **Door gebruik te maken van:**
AI-ondersteunde taalcontrole die medewerkers suggesties geeft voor het vereenvoudigen en inclusief schrijven van tekst.
- **Wij kunnen helpen:**
Door verbeteringen voor te stellen die het B1-taalniveau waarborgen.
- **Met een betere manier om:**
Sneller teksten te controleren, herschrijven en verduidelijken.
- **Zodat:**
Medewerkers hun werk efficiënter kunnen uitvoeren.
- **Met:**
Automatische verwijzingen naar de bijbehorende regel uit de schrijfwijzer, zodat medewerkers precies weten waarop de suggestie is gebaseerd.

Stakeholder 4: ICT-afdeling

- **Conceptnaam:**
AI-applicatie om brieven begrijpelijker te maken.
- **Door gebruik te maken van:**
Een lichtgewicht, lokaal te hosten AI-applicatie.
- **Om:**
De gemeente een veilige, controleerbare oplossing bieden.
- **Wij kunnen helpen:**
Door een systeem te leveren dat past binnen bestaande infrastructuur.
- **Met een betere manier om:**
AI in te zetten zonder risico op datalekken of afhankelijkheid van derde partijen.
- **Zodat:**
De ICT-afdeling in controle blijft van de gegevens en beveiliging ervan.
- **Met:**
Lokale dataverwerking.

Stakeholder 5: Overheid (wet- en regelgevers)**• Conceptnaam:**

AI-applicatie om brieven begrijpelijk te maken.

• Door gebruik te maken van:

AI-ondersteunde taalcontrole die medewerkers suggesties geeft voor het vereenvoudigen en inclusief schrijven van tekst.

• Om:

Overheidscommunicatie begrijpelijk en toegankelijker te maken.

• Wij kunnen helpen:

Door gemeenten te ondersteunen bij het naleven van richtlijnen voor begrijpelijke communicatie naar inwoners.

• Met een betere manier om:

Burgers hun rechten te laten begrijpen of acties die ze moeten ondernemen.

• Zodat:

Overheidscommunicatie transparant blijft en voldoet aan nationale richtlijnen, zoals Direct Duidelijk en de OVER-gemeente schrijfwijzer.

• Met:

Naleving van taal- en privacyrichtlijnen.

C Requirements

C.1 MoSCoW-analyse

De MoSCoW-analyse is uitgevoerd op alle ethische, juridische, organisatorische, functionele en technische vereisten. M staat voor *must have*, S staat voor *should have*, C staat voor *could have* en W staat voor *won't have*. De nummers bij een requirement (RQ) staan voor het type requirement. ER = ethische requirement, JR = juridische requirement, OR = organisatorische requirement, FR = functionele requirement, TR = technische requirement en DR = duurzaamheids requirement.

C.2 Ethische requirements

Table 17. Ethische requirements

RQ-nr	Requirement	MoSCoW	Stakeholder
ER01	Gebruikers moeten de AI-tool begrijpen [19].	M	Medewerkers die brieven schrijven/controleren
ER02	Een woordsuggestie mag niet zonder toestemming van de medewerker worden toegevoegd (bijv. om bias te voorkomen). Het systeem moet een human-in-the-loop hebben; de mens blijft eindverantwoordelijk en de AI-tool ondersteunt slechts [19].	M	Medewerkers die brieven schrijven/controleren; Overheid
ER03	De ontwikkeling van de AI-tool (datasets, modelkeuze, training) en de prestaties moeten gedocumenteerd worden [19].	M	Overheid; ICT-afdeling
ER04	Medewerkers moeten de mogelijkheid krijgen om te kiezen voor het gebruik van de AI-tool [19].	M	Medewerkers die brieven schrijven/controleren
ER05	De beperkingen en capaciteiten van de AI-tool moeten aan de stakeholder worden gemeld; bijv. dat de tool slechts een hulpmiddel is en niet altijd de beste suggesties geeft [19].	M	Medewerkers die brieven schrijven/controleren

RQ-nr	Requirement	MoSCoW	Stakeholder
ER06	De AI-tool moet worden beveiligd tegen kwetsbaarheden (bijv. hacks of technische storingen) [19].	S	ICT-afdeling; Overheid; Inwoners van Gemeente Oostzaan
ER07	De AI-tool moet een uitwijkplan hebben voor situaties met problemen zoals hacks of technische storingen [19].	S	ICT-afdeling; Medewerkers die brieven schrijven/controleren
ER08	De stakeholder moet worden geïnformeerd dat een AI-tool is gebruikt bij het schrijven van de brief [19].	C	Inwoners van Gemeente Oostzaan
ER09	Na installatie van de AI-tool moet regelmatig feedback worden gevraagd aan belanghebbenden [19].	W	Medewerkers die brieven schrijven/controleren; Inwoners van Gemeente Oostzaan
ER13	De gevolgen van het systeem op sociaal gebied moeten onderzocht worden [19].	W	Inwoners van Gemeente Oostzaan; Gemeente Oostzaan

C.3 Organisatorische requirements

Table 18. Organisatorische requirements

RQ-nr	Requirement	MoSCoW	Stakeholder
OR01	De Gemeente Oostzaan moet het gebruik van AI verwerken in hun beleid [54].	W	Gemeente Oostzaan
OR02	De Gemeente moet periodieke evaluaties inplannen om te controleren of brieven correct worden geschreven [54].	W	Gemeente Oostzaan
OR03	De medewerkers moeten een training krijgen over hoe de AI-tool werkt voor het herschrijven van brieven [54].	W	Medewerkers die brieven schrijven/controleren

RQ-nr	Requirement	MoSCoW	Stakeholder
OR04	De gemeente moet een beleid opstellen waarin wordt vastgelegd wie de eindverantwoordelijke is voor beslissingen die met behulp van de AI-tool tot stand komen [54].	W	Gemeente Oostzaan
OR05	De gemeente moet duidelijk maken aan de medewerkers dat zij niet volledig moeten vertrouwen op de tekstsuggesties van de AI-tool [54].	W	Gemeente Oostzaan; Medewerkers die brieven schrijven/controleren
OR06	De gemeente moet een feedbackmechanisme implementeren waarmee medewerkers suggesties over de AI-tool kunnen indienen [54].	W	Gemeente Oostzaan; Medewerkers die brieven schrijven/controleren
OR07	De gemeente moet een risicoanalyse opstellen waarin wordt beschreven wat de risico's zijn van de implementatie van de AI-tool bij het herschrijven van brieven [54].	W	Gemeente Oostzaan; Medewerkers die brieven schrijven/controleren

C.4 Juridische requirements

Table 19. Juridische requirements

RQ-nr	Requirement	MoSCoW	Stakeholder
JR01	Persoonlijke informatie van burgers mag niet worden opgeslagen of gebruikt voor (verdere) training van het model. De trainingsdata bestaat uit anonieme brieven [19, 50].	M	Inwoners van Gemeente Oostzaan
JR02	De AI-tool mag geen persoonsgegevens delen met externe partijen of servers [49, 50].	M	Overheden/toezichthouders; Inwoners van Gemeente Oostzaan; Gemeente Oostzaan
JR03	De AI-tool mag alleen worden gebruikt voor het herschrijven van brieven voor inwoners van de gemeente en niet voor andere doeleinden [19, 48, 50].	S	Overheden/toezichthouders

RQ-nr	Requirement	MoSCoW	Stakeholder
JR04	Er moet aangetoond kunnen worden waar de trainingsdata vandaan komt en dat er toestemming is voor gebruik hiervan [19, 20]. .	S	Overheden/toezichthouders
JR05	Het model moet worden getest op een representatieve set brieven om mogelijke fouten te identificeren. Hierbij moet specifiek worden gecontroleerd op bias (o.a. geslacht) zodat geen enkele groep wordt benadeeld en discriminatie wordt voorkomen [26, 35].	S	Overheden/toezichthouders

C.5 Functionele requirements

Table 20. Functionele requirements

RQ-nr	Requirement	MoSCoW	Stakeholder
FR01	De AI-tool geeft suggesties, maar de medewerker beslist wat hij of zij met deze beslissing doet [54].	M	Medewerkers die brieven schrijven/controleren
FR02	De medewerker kan de suggestie altijd handmatig afwijzen of goedkeuren [54].	M	Medewerkers die brieven schrijven/controleren
FR03	De AI-tool moet de medewerker informeren dat de suggesties slechts ter ondersteuning dienen en dat de medewerker de eindbeslissing neemt [54].	M	Medewerkers die brieven schrijven/controleren
FR04	De AI-tool moet na verwerking van de suggestie, de data bewaren zodat de suggestie in een soortgelijke situatie (her)gebruikt kan worden en de suggestie zo eerder wordt weergegeven [57].	M	Medewerkers die brieven schrijven/controleren
FR05	De AI-tool moet een lagere MAE-score hebben dan het nulmodel.	M	ICT-afdeling

RQ-nr	Requirement	MoSCoW	Stakeholder
FR06	Resultaten moeten betrouwbaar zijn voor alle verschillende briefcategorieën.	S	OVER-gemeenten
FR07	De AI-tool biedt mogelijkheid om door een mens gecontroleerd te worden [54].	C	ICT-afdeling
FR08	De AI-tool moet een MAE hebben die ongeveer gelijk is voor elke bevolkingsgroep.	C	Overheden/toezichthouders; ICT-afdeling

C.6 Technische requirements

Table 21. Technische requirements

RQ-nr	Requirement	MoSCoW	Stakeholder
TR01	Bewonersinformatie wordt direct verwerkt en niet opgeslagen. Dit moet in de code van de AI-tool worden verwerkt [20].	M	Inwoners van Gemeente Oostzaan; Overheden/toezichthouders
TR02	User Interface (UI) elementen tonen aan dat de AI-tool ondersteunend is en niet beslissend [54].	M	Medewerkers die brieven schrijven/controleren
TR03	De AI-tool moet gevalideerd worden ten opzichte van het nulmodel en prestaties moeten worden gedocumenteerd.	M	ICT-afdeling
TR04	De gebruikte datasets moet gesplitst worden in een train-, validatie- en testdataset van 64%, 16% en 20%.	M	ICT-afdeling
TR05	De AI-tool moet de brieven lokaal verwerken.	M	Overheden/toezichthouders; ICT-afdeling
TR06	De AI-tool moet sneller werken dan dat de medewerker handmatig de zin verbetert.	M	Medewerkers die brieven schrijven/controleren
TR07	De AI-tool moet een representatieve dataset gebruiken [20].	S	ICT-afdeling; Overheden/toezichthouders

RQ-nr	Requirement	MoSCoW	Stakeholder
TR08	Het model moet bias-mitigatie technieken gebruiken, zodat het om kan gaan met incorrect beoordeelde resultaten [20].	S	ICT-afdeling; Overheden/toezichthouders
TR09	Het systeem moet beveiligd zijn tegen hacks door middel van netwerksegmentatie.	W	ICT-afdeling; Overheden/toezichthouders

C.7 Duurzaamheids requirements

Table 22. Duurzaamheids requirements

RQ-nr	Requirement	MoSCoW	Stakeholder
DR01	Er moeten stappen worden gezet om de AI-tool zo milieuvriendelijk mogelijk te maken, van begin tot eind.	W	Overheden/toezichthouders; ICT-afdeling

D Testplan: maandag 1 december 2025

D.1 Introductie

Wij studeren met zijn vijven de Master Applied Artificial Intelligence hier op de Hogeschool van Amsterdam. Vanuit jullie, ook wel de OVER-Gemeenten, hebben wij de opdracht gekregen om een AI-tool te maken dat de gemeente kan helpen om begrijpelijk Nederlands te schrijven. In dit geval betekend begrijpelijk Nederlands ook wel het B1-niveau Nederlands, aangezien ruim 80% van de Nederlanders B1-niveau goed kan begrijpen.

Tijdens de vorige bijeenkomst gaven jullie aan dat er een schrijfwijzer is ontwikkeld (de OVER-schrijfwijzer), zodat medewerkers handvatten hebben om brieven begrijpelijk te schrijven. Daarbij werd vermeld dat deze schrijfwijzer vrij uitgebreid is en daardoor in de praktijk nog niet altijd goed wordt toegepast. Daarom is het idee ontstaan om medewerkers te ondersteunen met een AI-tool die is gekoppeld aan de OVER-schrijfwijzer, zodat teksten eenvoudiger en consistent kunnen worden herschreven.

Voor dit onderzoek willen we graag jullie expertise gebruiken om te onderzoeken of het design van ons idee naar jullie wens is. We willen jullie graag ons prototype laten zien en zullen jullie gaan vragen om feedback te geven vanuit een werknemersperspectief.

D.2 Ons idee in het kort voor dit project

In ons project willen we een AI-model ontwikkelen dat de medewerkers van OVER-gemeenten ondersteunt bij het schrijven van brieven. De medewerker krijgt tijdens het typen suggesties gericht op taalniveau en zinsopbouw, om ervoor te zorgen dat de geschreven teksten versimpeld worden.

De AI-tool gaat suggesties geven op basis van wat hij geen B1-niveau Nederlands vindt, of wat hij bijvoorbeeld te complexe zinnen vindt volgens de OVER-schrijfwijzer. Bij deze suggestie zal dan ook doorverwezen worden naar de OVER-schrijfwijzer.

Er is een knop aanwezig op het scherm, die het toelaat voor de medewerker om de AI-tool aan of uit te zetten. Hierdoor zal de medewerker niet gestoord worden tijdens het schrijven, tenzij hij of zij dat natuurlijk wel wil.

De extensie zal, als de AI-tool ingeschakeld is, rechts in beeld komen te staan. Hierbij zal per gevonden te lange zin of een te moeilijk woord, één pagina gebruikt worden. Bij 20 gevonden fouten in één document, zullen dus 20 doorklikbare pagina's beschikbaar zijn in de extensie. Zo blijven alle suggesties los van elkaar zichtbaar en kan er ook later teruggekomen worden op een suggestie.

D.3 Doel en taakverdeling

Het doel voor ons is dat wij meer te weten komen over de wenselijkheid en de uitlegbaarheid/ het vertrouwen in deze AI-tool. We willen testen of dit het gewenste prototype is, aangezien we hebben vernomen dat de medewerkers het fijn zouden vinden als de tool niet telkens in beeld springt, een overvloed aan kleurtjes heeft en over het algemeen afleidend is. We willen graag observeren wat jij van de AI-tool vindt en of je eventuele verbeterpunten ziet die zorgen voor een prettige werkervaring met deze AI-tool. Tijdens het testen wordt de rolverdeling aangehouden zoals weergegeven in de onderstaande tabel.

Naam	Taak
Amber	Aantekeningen maken
Amir	Aantekeningen maken
Kaan	Introductie geven
Terence	Het prototype uitleggen en ondersteunen tijdens de test
Bibiëne	Afwezig

Table 23. Taakverdeling tijdens de test

D.3.1 Uitleg door Terence tijdens de test.

Tijdens de test legt Terence uit hoe het prototype werkt. Hierbij gaat hij in op de volgende punten:

- Wat het prototype doet
- Het gebruik van de AI-knop (aan- en uitzetten)
- De extensie
- Verwijzingen per pagina binnen de extensie
- Verwijzingen naar de OVER-schrijfwijzer.

Vervolgens laat Terence de medewerker naar het prototype kijken en test hij of zij de aan- en uitknop van de AI-tool kan vinden. Wanneer deze knop wordt geactiveerd, verschijnt de extensie in beeld. Daarna laat Terence de testpersoon de AI-tool verder bekijken en wordt geobserveerd welke vragen of onduidelijkheden er ontstaan. Alle bevindingen worden vastgelegd en gedocumenteerd (Bijlage G).

D.4 Vragen aan de medewerkers van de OVER-gemeenten

D.4.1 Desirability (wenselijkheid).

- Voldoet de optie om de AI-tool aan en uit te zetten, aan jullie wensen?
- Zou je als medewerker prettig om kunnen gaan met de AI-tool? (*Wat zouden jullie eventueel anders willen?*)
- Is het wenselijk dat de AI-tool aangeeft dat er een suggestie is voor een bepaalde zin of een bepaald woord?
 - Zijn alle stappen wenselijk, overbodig of zijn er extra stappen nodig?
- Denk je dat je de AI-tool snel kunt leren gebruiken?
- Zijn er andere stappen in het prototype wenselijk die er op dit moment nog niet instaan?

D.4.2 Explainability & Trust (uitlegbaarheid & vertrouwen).

- Zijn de signalen van de AI-tool nuttig en voldoende uitgelegd om jouw werk te ondersteunen?
- Zou je deze AI-tool vertrouwen?
- Geeft de AI-tool informatie die jou helpt om beslissingen te nemen, of juist te veel/te weinig?
- Zou je te veel beïnvloed worden door de suggesties die de AI-tool geeft, dus zou je de AI-tool te veel vertrouwen?
- Vind je deze AI-tool vertrouwelijk in de gemeentelijke context?

E AI-breakdown & Error Flows

In	Taak	Uit	Mogelijke fouten (FP / FN, hallucinaties, meaning drift, inconsistentie, geen output)	Impact van fouten low (recommended) medium (consequential) high (life-critical)	Percentage van fouten waarbij het nog waarde heeft blijft	Mogelijke oplossingen
Tekst in brief van de gemeente	Detecteren van moeilijke woorden	Markeringen bij woorden die als 'moeilijk' geclassificeerd worden.	FP: woorden markeren als 'moeilijk' die al B1-niveau zijn (context verkeerd, eigenaardigheid/kortingen). FN: moeilijke woorden missen (samenvestigingen, vervoegingen, (arge)voorbereidende woorden, spelfouten).	Medium: FP-markeringen kunnen leiden tot onnodige aanpassingen en vertraging in het schrijfproces. High: FN kan ertoe leiden dat kempproblemen onopgemerkt blijven, waardoor B1-niveau niet wordt gehaald.	20% omdat de medewerker markeringen kan negeren.	Medewerker kan suggesties ten alle tijde afklikken/wegpikken (HTL). Blacklist/whitelist (jargonlijst), frequentie-/B1-woordlijst en (optioneel) confidence-drempel toevoegen. Toon kort waarom een woord als moeilijk is gemerkt.
Tekst in brief van de gemeente	Detecteren van zinscomplexiteit (te lang, veel bijzinnen, actief of passief geschreven)	Markering van complexe zinnen.	FP: zinnen markeren als complex pur voor lengte, terwijl ze goed leesbaar zijn (bv. opsommingen). FN: complexe zinnen missen (bijzinnen, uitdrukkende vorm, juridische constructies). Span-fout: verkeerde zinsdelen markeren (niet de echte oorzaak van complexiteit).	Medium: FP kan leiden tot onnodige herschrijfacties en verminderd vertrouwen in de tool. High: FN kan ertoe leiden dat complexe zinnen blijven staan, waardoor de brief niet op B1-niveau komt. High: Span-fouten kunnen tot verwarring leiden (medewerker begrijpt de fout niet en corrigeert verkeerd).	20%-ondel markeringen en suggesties door de medewerker genegeerd kunnen worden.	Dusdanig verminderen waarom een zin complex is (bv. tijdelijke vorm, aantal bijzinnen, lengte). Markeer (indien mogelijk) alleen het relevante zindeel en laat medewerker de markering corrigeren/afkeuren (HTL).
Tekst in brief van de gemeente	Checken of toon en inclusiviteit (o.a. vermijden van afstandelijke of stigmatiserende termen)	Feedback op toon en suggesties voor alternatieve formuleringen.	FP: neutrale woorden onterecht als 'niet-inclusief/stigmatiserend' markeren (context/taakdoeling verkeerd). FN: subtiel niet-inclusieve formuleringen missen. Generatieve fout: Alternatief voorstelt dat de betrekker/one-of voice ongewenst verschilt (te handte informant).	Low: FP is meestal herkenbaar en kan door medewerkers gecontroleerd worden. Medium: generatieve fouten of FN kan leiden tot minder passende formuleringen of gemiste verbeterkansen, wat kwaliteit en begrijpelijkheid verlaagt.	10% omdat medewerkers suggesties mogen negeren.	Randvoorwaarden toewegen (vaste lijst + contextregels) en toon kort de reden van marketing. Laat medewerker altijd kiezen tussen meerdere alternatieven of afkeuren (HTL).
Tekst in brief van de gemeente	Controleeren of tekst voldoet aan specifieke regels van de OVER-schriftwijzer.	Verwijzing naar relevante regel/verwijsgang uit de schriftwijzer waarom zinnen/woorden aangepast dienen te worden.	FP: verwijzen naar een regel die niet van toepassing is (irrelevante retrieven). FN: geen passende regel tonen terwijl er wel een overtuiging is (retreval faalt) of te algemene verwijzing gegeven. Hallucinatie: AI noemt een regel/uitvoer die niet in de schriftwijzer staat. Geen output/timeout: geen verwijzing beschikbaar door timeout.	Low: FP is vervalend maar kan genegeerd worden. Medium: FN/te algemene verwijzing kan ertoe leiden dat noodzakelijke aanpassingen niet worden gedaan. High: hallucinaties kunnen leiden tot onterechte regels en daarmee tot verkeerde bestissingen of onjuiste onderbouwing.	20% omdat de medewerker de verwijzing kan negeren en zelf kan controleren in de schriftwijzer.	Toon altijd een snippet/cta na de relevante passage + link naar sectie (verkleint hallucinatie-impact). Als er geen regel gevonden wordt: expliciete fallback-zin ("Geen passende regel gevonden in de schriftwijzer"). Reby bij timeout = loggen/monitoren van retrieval-fouten.
Tekst in brief van de gemeente	Meerdere suggesties genereren voor herschrijven volgens B1-niveau	Koop 'Nieuwe suggesties' die nieuwe zinnen/woordsuggesties per zin/woord voorstellen.	Meaning drift: betekenis verandert (juridische implicaties/verwondingen/muandes verduinen). Hallucinatie: AI introduceert info die niet in de tekst staat. Omissie: essentiële details (bedragen, data, verplichtingen) verwijderen. Geen output/timeout: geen herschrijfsuggestie terwijl dit wel mogelijk is. Inconsistentie: verschillende output per run (Nieuwe suggestie geeft tegenvallende adviezen).	High: betekenisverandering/ommixing/additie kan juridische of inhoudelijke strekking aanpassen (risico op werkende bestuurhervorming of communicatie). High: hallucinaties kunnen verwarring veroorzaken en het schrijfproces verstoren. High: geen output/timeout kan ertoe leiden dat noodzakelijke vereenvoudiging uitlijst (B1 niet gehaald). High: tegenstrijdige output vergroot de kans op een (juridisch risicovolle) verkeerde keuze en maakt resultaten minder betrouwbaar/replieverebaar.	25% omdat de medewerker de suggestie of gehele tekst kan negeren.	Medewerker kan suggesties te allen tijde afkeuren (HTL). Prompt-constraints: behoud betekenis, behoud specifieke pagina/sectie, geen nieuwe feiten. Toon verschillen (highlight wijzigingen) en bied 2-3 alternatieven.

Fig. 12. AI-breakdown (deel 1 en deel 2).

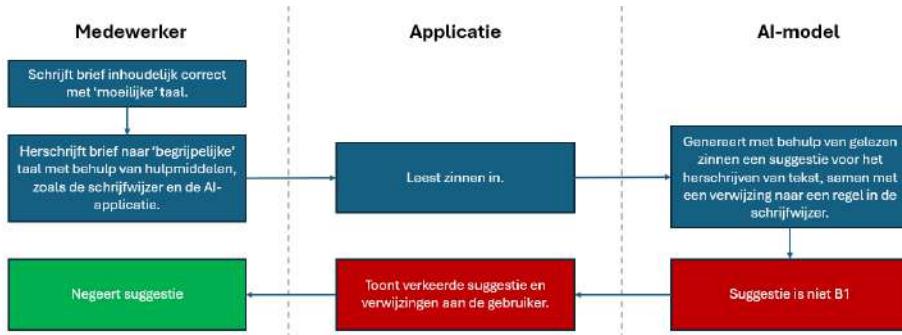


Fig. 13. Gegenereerde suggestie is niet B1

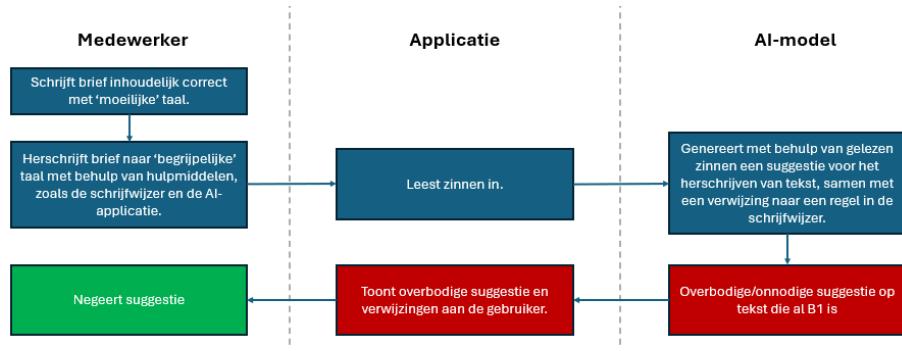


Fig. 14. De gegenereerde suggestie is overbodig of onnodig bij de huidige tekst

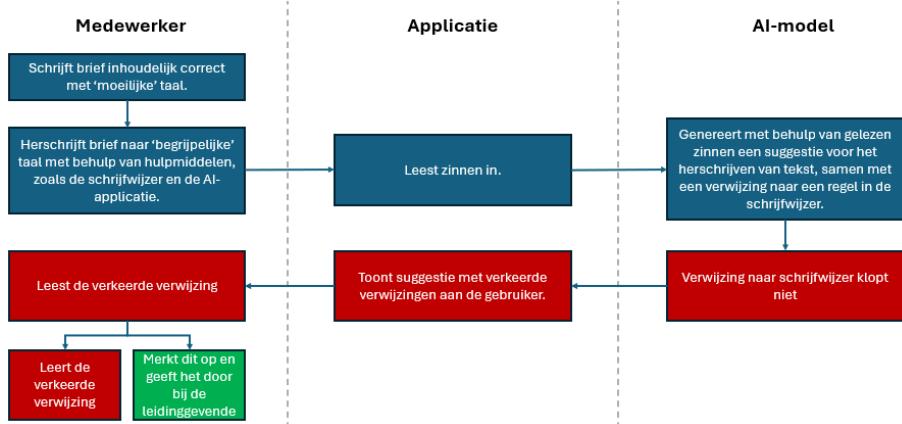


Fig. 15. De gegenereerde suggestie bevat een verwijzing die niet klopt

F DEDA-analyse

Tijdens de literatuurles op 4 december is de DEDA-tool gebruikt om te reflecteren op de ethische vraagstukken die spelen bij de ontwikkeling van de AI-tool voor de OVER-gemeente. Deze tool bestaat uit vier fasen waarin verschillende vragen worden beantwoord.

F.1 Fase één

(1) Projectnaam en datum:

AI als schrijfassistent voor heldere gemeentelijke brieven, gebaseerd op B1-niveau Nederlands, 4 december 2025.

(2) Projectteam:

Team 2.

(3) Wat houdt het project in en wat is het doel?:

Gemeentelijke brieven van de OVER-gemeenten worden herschreven volgens de schrijfwijzer, met als doel deze begrijpelijk te maken op B1-niveau.

(4) Wat zijn de eventuele maatregelen of acties op basis van de uitkomsten van dit project?:

Wanneer het project succesvol verloopt, worden brieven eenvoudiger geschreven. Dit draagt bij aan betere begrijpelijkheid voor burgers en vergroot de kans dat de inhoud daadwerkelijk wordt begrepen.

(5) Wat voor data gebruikt u? Geef een korte beschrijving van de inhoud hiervan.:

De volgende databronnen worden gebruikt:

- Brieven van de gemeente Oostzaan;
- De schrijfwijzer van de OVER-gemeenten;
- Een inclusiviteitswoordenlijst;
- Een lijst met moeilijke woorden (Gemeente Amsterdam).

(6) Wie zijn de stakeholders van dit project en op wie/wat heeft het impact?

Dit zijn de stakeholders van het project:

- De ICT-afdeling;
- Medewerkers die brieven schrijven en controleren;
- De OVER-gemeente;
- Inwoners van de OVER-gemeente;
- De overheid.

Het project heeft impact op alle stakeholders, maar vooral op de medewerkers van de OVER-gemeenten zullen te maken krijgen met de AI-tool. De burgers zullen hierdoor hopelijk de brieven beter begrijpen (Paragraaf 2.2).

(7) Wat zijn de nagestreefde voordeLEN van dit project?

- Consistenter en begrijpelijker schrijven van gemeentelijke brieven, waardoor er minder misverstanden en problemen ontstaan zoals beschreven in Paragraaf 1.2.

(8) Zijn er (mogelijke) problemEN met het project?

- Het beoordelen van B1-niveau is deels subjectief;
- Onjuiste suggesties kunnen leiden tot betekenisverlies of verkeerd woordgebruik;
- Privacygevoelige gegevens mogen de gemeente niet verlaten en mogen niet worden gebruikt voor het trainen van het model.

(9) **Welke van de onderstaande onderwerpen (algoritmen, bron, anonimiseren etc.) zijn van toepassing op uw project?**

Beslis welke er eventueel tijdens de workshop kunnen worden overgeslagen.

De volgende vragen worden tijdens de workshop overgeslagen:

- Vraag 41
- Vraag 42

F.2 Fase twee

Bij de start van de analyse is vastgesteld dat de volgende waarden belangrijk zijn voor de OVER-gemeente bij de ontwikkeling en inzet van de AI-tool:

- **Privacy:** Privacy is essentieel, omdat persoonsgegevens van inwoners niet buiten de organisatie mogen worden verwerkt of gedeeld.
- **Menselijke eindverantwoordelijkheid:** Medewerkers van de gemeente vinden het belangrijk dat de mens eindverantwoordelijk blijft voor de output van de AI-tool (Bijlage A).
- **Gebruiksvriendelijkheid:** De AI-tool moet zo worden ontworpen dat deze daadwerkelijk consistent in de praktijk gebruikt kan worden, in tegenstelling tot de huidige schrijfwijzer die als lastig in gebruik wordt ervaren.
- **Inclusiviteit:** In de schrijfwijzer staat dat teksten bedoeld zijn voor alle inwoners en dat het belangrijk is dat iedereen zich betrokken voelt. Daarom moet de AI-tool inclusief taalgebruik ondersteunen en bevorderen.

F.3 Fase drie

In deze fase worden vragen beantwoord met betrekking tot datagerelateerde overwegingen en algemene overwegingen.

F.3.1 Datagerelateerde overwegingen.

- (10) Gebruikt u een algoritme in uw project?
 - Ja
- (11) Hoe gaat u om met false positives en false negatives?
 - De medewerker is eindverantwoordelijk en keurt suggesties zelf.
- (12) Is er iemand in het team die kan uitleggen hoe het gebruikte algoritme werkt? Is het noodzakelijk dat iemand kan uitleggen wat het doet?
 - Nee, niet noodzakelijk dat iemand het exact kan uitleggen, globaal wel als het gaat om de werking van de applicatie.
- (13) Gebruikt u de uitkomsten van het model als leidend of aanvullend in uw beslissingsmodel?
 - Nee, alleen aanvullend.
- (14) Is er menselijke controle op fouten die het algoritme kan maken? Hoeveel ruimte heeft een mens om van het systeem af te wijken wanneer nodig?
 - Ja, medewerkers moeten de suggesties goed of afkeuren. De AI-tool vervangt niet meteen woorden die als 'foutief' worden gezien.
- (15) Wat is de bron van de data?
 - Gemeente Amsterdam
 - OVER-schrijfwijzer
 - Interne brieven van de OVER-gemeenten

- (16) Heeft u de kwaliteit van de dataset(s) gecontroleerd?
- Nee de brieven zijn niet gecontroleerd, omdat deze enkel voor validatie en testen wordt gebruikt. Als we de kwaliteit zouden controleren is er sprake van data leakage. Wel hebben medewerkers van de gemeente deze voor ons geselecteerd. De woordenlijsten zijn doorgenomen en lijken gebruikt te kunnen worden, maar dit moet nog nagevraagd worden bij de OVER-gemeente.
- (17) Hebben de data een houdbaarheidsdatum?
- Nee, maar bijhouden is noodzakelijk, aangezien taal in de loop der jaren verandert.
- (18) Verzamelt u de juiste informatie voor uw doel?
- Grotendeels, mogelijk aanvullende bronnen nodig.
- (19) Is het nodig om de dataset(s) te anonimiseren, pseudonimiseren of te generaliseren?
- Alleen bij gebruik van echte brieven, wat in dit geval niet zo is. In de praktijk dus wel.
- (20) In het geval van anonimiseren: is gecheckt of de data echt niet meer herleidbaar is?
- Dit is waarschijnlijk in de toekomst mogelijk, wanneer de tool geïmplementeerd wordt. Voor dit onderzoek is het in deze fase nog niet van toepassing.
- (21) In het geval van pseudonimiseren: wie heeft de sleutel om de pseudonimisering terug te draaien?
- Dit is waarschijnlijk in de toekomst mogelijk, wanneer de tool geïmplementeerd wordt. Voor dit onderzoek is het in deze fase nog niet van toepassing.
- (22) Hoe worden de uitkomsten van het project weergegeven? Worden deze gevisualiseerd?
- In een Word-extensie, zoals in het prototype is weergegeven.
- (23) Wat zou een andere manier van visualiseren zijn?
- Dat de suggesties onder elkaar worden getoond is voorgesteld door testpersonen, maar omdat dit mogelijk leidt tot minder menselijke controle en een te groot vertrouwen in de tool (automation bias), passen we dit niet toe.
- (24) Wie heeft toegang tot de dataset(s)?
- Iedereen heeft toegang tot de woordenlijsten, want deze zijn publiekelijk;
 - De mensen die brieven schrijven, docenten van de master applied AI en studenten van deze master hebben toegang tot enkele brieven van de OVER-gemeente.
 - Wanneer de applicatie echt gebruikt gaat worden, moet genoteerd worden hoe de data wordt opgeslagen en gebruikt.
- (25) Hoe wordt de toegang gemonitord?
- Wanneer de applicatie echt gebruikt gaat worden, moet later wel genoteerd worden hoe de data wordt opgeslagen en gebruikt.
- (26) Zijn de resultaten geschikt om te worden hergebruikt? Zo ja, onder welke voorwaarden?
- Er zijn nog geen resultaten, maar de manier waarop de data gebruikt zal worden, wordt vastgelegd zodat hier in de toekomst rekening mee kan worden gehouden.
- (27) Zijn (delen) van de (input)data geschikt om te worden hergebruikt? Zo ja, onder welke voorwaarden?
- Ja want deze worden nu lokaal opgeslagen en de gebruikte instellingen ook. Echter moet de data in de toekomst misschien geüpdatet worden, vanwege bijvoorbeeld verandering in de samenleving over wanneer worden worden gezien als inclusief

F.3.2 *Algemene overwegingen.*

- (28) Bestaan er binnen de organisatie beleid of richtlijnen die van toepassing zijn op dit project? Zo ja, welke?
- Nee, maar ze moeten zich houden aan algemene wetten en regelgeving, zoals artikel 50 en de AI-act.
- (29) Wie is/zijn eindverantwoordelijk voor het project?
- De mensen die besluiten het systeem te implementeren (medewerkers van de gemeente).

- (30) Zijn de verantwoordelijkheden van die persoon/personen helder?
- Nee, want die kent het systeem nog niet.
- (31) Is dit project geschikt voor samenwerking met (commerciële) partners? Zo ja, welke partijen zouden dat kunnen zijn?
- Ja, niet met commerciële partijen, maar wel met andere gemeenten.
- (32) Wat is de communicatiestrategie voor dit project? (Zowel voor de positieve als negatieve impact hiervan). In het geval van samenwerkingspartners: is deze strategie met hen afgestemd?
- Ja, er is een communicatiestrategie voor dit project. Er zijn momenten afgesproken waarop vragen gesteld kunnen worden, zowel digitaal als in persoon. Daarnaast is op 1 december 2025 een prototype getest met medewerkers van de gemeente.
- (33) Zijn er communicatiestrategieën voor het geval er iets mis gaat?
- Dat weten we niet, aangezien docenten als tussenpartij fungeren. Indien nodig kunnen wij zelf altijd nog contact zoeken.
 - Mogelijk moet aan de gemeente worden doorgegeven dat zij een communicatiestrategie opstellen voor het geval er iets misgaat, zodat er een duidelijk plan is voor hoe te handelen in zo'n situatie.
- (34) Wie is er verantwoordelijk voor deze strategieën?
- De docenten van de opleiding applied AI.
 - De communicatieafdeling en de mensen die het systeem maken (een ontwikkelteam).
- (35) Bestaat het risico op publieke verontwaardiging, nu of in de toekomst?
- Ja, aangezien burgers zich kunnen afvragen of de inhoud van de brief wel correct is en of hun persoonlijke gegevens worden gebruikt in een AI-systeem.
- (36) Hoe transparant bent u over uw project naar burgers toe?
- Tijdens het ontwikkelen van het project worden de burgers nog niet geïnformeerd over het project.
- (37) Hoe worden burgers actief betrokken?
- Aan verschillende burgers worden vragen gesteld over wat zij ervan vinden dat de gemeente AI gebruikt bij het herschrijven van brieven en over hoe transparant de gemeente hierover moet zijn.
 - Ook worden medewerkers van de gemeente betrokken bij het ontwerpproces.
- (38) Kunnen burgers bezwaar maken tegen uitkomsten van het project?
- Ja, burgers kunnen altijd klachten indienen. Zij worden geïnformeerd over de ontwikkeling van een AI-systeem dat brieven herschrijft. Daarnaast kunnen zij een klacht indienen als zij een brief niet begrijpen of het niet eens zijn met de inhoud.
- (39) Is een opt-out mogelijk voor burgers? Zo ja, op welk moment kunnen burgers ervoor kiezen om deelname te beëindigen?
- Nee, deze is er niet. Wel wordt er tijdens het ontwikkelen van het systeem rekening gehouden met de privacyrechten van burgers, etc.
- (40) Welke wetten, voorschriften en/of richtlijnen zijn van toepassing op uw project?
- Artikel 50, AI-act, etc. Zie requirements in het verslag.
- (41) Heeft u de privacyfunctionaris en/of functionaris gegevensbescherming betrokken bij het project?
- Deze vraag slaan we over.
- (42) Heeft u een DPIA (Data Protection Impact Assessment) gehanteerd?
- Deze vraag slaan we over.
- (43) Wat is uw onderbuikgevoel over dit project? Heeft u zorgen? Er zijn verschillende onderbuikgevoelens gedeeld door Bibiène Wüst, Kaan Göçay en Amir Jacobs.
- Bibiène Wüst is benieuwd of het model zich kan aanpassen naar het format van de brieven.
 - Amir Jacobs heeft zorgen over hoe het B1-niveau beoordeeld kan worden, aangezien dit subjectief lijkt om te beoordelen.
 - Kaan Göçay maakt zich zorgen over het project, omdat ongeveer 20 procent van de Nederlanders B1-niveau niet kan lezen en deze groep dus niet wordt bereikt.
- (44) Bestaat het gevaar dat bepaalde mensen of groepen gediscrimineerd zouden kunnen worden door uw project?

- Ja, mensen die onder B1-niveau lezen, worden nog steeds niet bereikt. Dat ligt niet aan de uitvoering van het project, maar aan de eis die gesteld wordt door de overheid dat teksten op B1-niveau moeten zijn.
- (45) Zijn alle verschillende groepen burgers vertegenwoordigd in de dataset(s)? Wie missen er of zijn niet zichtbaar?
- Ja, verschillende groepen burgers zijn vertegenwoordigd omdat er is gewerkt met een inclusiviteitslijst. Deze lijst zorgt ervoor dat rekening wordt gehouden met passende benamingen, zoals neutrale aanduidingen voor ouderen en het gebruik van “kinderen” in plaats van “jongens/meisjes”.
- (46) Zit er een feedback loop in het model die negatieve consequenties kan hebben?
- Ja, er kan een feedbackloop ontstaan wanneer een medewerker een suggestie accepteert die in feite onjuist is. Als het model deze acceptaties gebruikt om verder te leren, kan het verkeerde patronen gaan overnemen. Dit risico wordt groter wanneer medewerkers te veel vertrouwen op de gegenereerde tekst en minder kritisch controleren. Daarom is het belangrijk dat datasets en modeluitvoer regelmatig worden beoordeeld en bijgewerkt door mensen om fouten en vertekening te voorkomen.
- (47) Function creep: kunt u zich een toekomstig scenario voorstellen waarin de uitkomsten van dit project voor een ander doeleinde gebruikt worden?
- Ja, de uitkomsten van dit project kunnen ook gebruikt worden voor het versimpelen van brieven voor andere gemeenten.
- (48) Veranderen uw antwoorden als u de mogelijke langetermijneffecten in acht neemt? Waarom?
- Ja, mijn antwoord verandert wanneer ik rekening houd met mogelijke langetermijneffecten. Taal verandert namelijk in de loop van de tijd. Het model zal daarom regelmatig moeten worden hertraind met bijgewerkte woordenlijsten en voorbeelden om relevant en bruikbaar te blijven.
- (49) Wanneer wordt dit project geëvalueerd?
- Het project wordt geëvalueerd zodra de ontwikkelaars de vooraf vastgestelde doelen samen met de gemeente hebben behaald. Daarna kan het systeem worden ingevoerd bij de OVER-gemeenten, zodat medewerkers ermee kunnen testen. Tijdens de ontwikkeling wordt de AI-tool bovendien tussentijds beoordeeld aan de hand van verschillende kwaliteitscriteria.

F.4 Fase vier

In deze fase wordt een conclusie gevormd aan de hand van antwoorden op een aantal vragen.

- (1) Zijn de organisatiwaarden en persoonlijke waarden voldoende gewaarborgd?
- **Privacy:** Er wordt uitsluitend getraind met openbare, niet-privacygevoelige data. Brieven met privacygevoelige informatie worden alleen intern gebruikt. Vandaar dat de waarde privacy voldoende is gewaarborgd.
 - **Gebruiksvriendelijkheid:** Het systeem moet praktisch inzetbaar zijn en niet omslachtig in gebruik. Door het model te ontwikkelen als een Word-extensie wordt de gebruiksvriendelijkheid vergroot. Dit wordt geëvalueerd op basis van feedback van testgebruikers op het verbeterde prototype. Op deze manier kan deze waarde voldoende worden gewaarborgd.
 - **Menselijke eindverantwoordelijkheid:** In het ontwerp van de AI-tool is vastgelegd dat gemeentemedewerkers eindverantwoordelijk blijven voor de inhoud van hun brieven. De testpersonen hebben op 1 december 2025 bevestigd dat deze waarde in het prototype voldoende wordt gewaarborgd.
 - **Inclusiviteit:** De brieven moeten inclusief worden geschreven volgens de OVER-schrijfwijzer. Door een inclusieve woordenlijst van de Gemeente Amsterdam in het model te implementeren, wordt aan deze waarde voldaan.

(2) Wat zijn de belangrijkste ethische knelpunten?

- De belangrijkste ethische knelpunten zijn dat privacygevoelige informatie mogelijk wordt gebruikt bij het hertrainen van het model, wat risico's oplevert voor gegevensbescherming. Daarnaast is er het risico dat een deel van de bevolking, ongeveer 20 procent van de Nederlanders, ook bij teksten op B1-niveau nog steeds niet goed wordt bereikt. Tot slot bestaat de kans dat gebruikers te veel vertrouwen op het systeem en suggesties klakkeloos accepteren, waardoor fouten onopgemerkt blijven en het model op basis van onjuiste input wordt hertraind.

(3) Wat zijn nieuwe en verrassende inzichten?

- Er is nog onvoldoende nagedacht over een mogelijke feedback-loop.

(4) Onder welke voorwaarden willen we wel of niet doorgaan met dit project?

- Wanneer medewerkers het systeem te veel vertrouwen, waardoor de inhoudelijke kwaliteit van brieven achteruitgaat.

(5) Bekijk alle actiepunten die hieronder worden weergegeven en schrijf voor elk punt de actiehouder op.

- De actiepunten worden verder uitgewerkt tijdens het schrijven van het verslag, met name in het hoofdstuk 'Discussie'.

G Gebruikerstesten op 1 december

G.1 Testronde 1: Uitgevoerd door Franny van der Bijl, WMO-consulent

G.1.1 Algemene bevindingen en verbeterpunten.

De eerste indruk van de testpersoon is dat zij het leuk vond dat we een paper prototype gemaakt hadden, ondanks dat we een ICT-opleiding doen. De testpersoon heeft zelf de knop ‘AI-assistent’ gevonden en heeft deze zelf aangeklikt. Hierbij heeft teamlid Terence verteld dat het onderstreepte deel in de tekst waarschijnlijk een highlight wordt in het geel in de tweede iteratie.

De optie ‘aanpassen’ in de AI-tool vond de testpersoon onduidelijk omdat je rechts in de extensie op aanpassen klikt en vervolgens een zin in het document kan aanpassen. De testpersoon gaf aan dat ze het idee had dat niet alle collega’s zullen snappen hoe dit eruit ziet. Hierbij begreep ze wel hoe ze door kon gaan naar de volgende fout door op de pijl te klikken.

G.1.2 Vragen van de testpersoon.

- Als je op accepteren drukt, gaat hij dan automatisch door naar de volgende te verbeteren punt in het document? Je wilt namelijk door naar de volgende suggestie als je op accepteren of overslaan klikt.
- Kan het model ook leren? (Antwoord Amber: Ja dat is wel het idee, dat als medewerkers altijd op niet accepteren klikken dat het model dan gaat leren dat dit niet meer verbeterd hoeft te worden.)

G.1.3 Desirability (wenselijkheid).

- **Voldoet de optie om de AI-tool aan en uit te zetten, aan jullie wensen?**

Ze vinden het fijn dat deze knop aan- en uitgezet kan worden. Ze geeft aan dat het goed is dat je de tool aan kan zetten als je denkt dat de brief goed is.

- **Zou je als medewerker prettig om kunnen gaan met de AI-tool?**

De testpersoon vindt het fijn dat onderstreept wordt waar de fout is en dat er wordt uitgelegd waarom. Op deze manier leren de medewerkers zelf ook van hun fouten.

- **Is er te veel tekst in beeld?**

De testpersoon vindt het fijn dat de bron te zien is, maar geeft ook aan dat dit dubbelop is omdat er twee keer naar de schrijfwijzer verwezen wordt.

- **Is het wenselijk dat de AI-tool aangeeft dat er een suggestie is voor een bepaalde zin of een bepaald woord?**

De testpersoon geeft aan dat je wilt naar de volgende als je op accepteren of overslaan klikt.

- **Denk je dat je de AI-tool snel kunt gebruiken?**

De testpersoon zelf wel en denkt andere medewerkers ook. Er wordt aangeraden niet te veel kleur te gebruiken (wel duidelijk en overzichtelijk).

- **Zijn er andere stappen in het prototype wenselijk die er op dit moment nog niet instaan?**

Een knop ‘alles accepteren’ zou in sommige gevallen wel handig zijn. Soms zouden ze er door de tijd te veel op vertrouwen en op alles accepteren willen drukken en daarna de brief na willen lezen om vervolgens te controleren of deze goed werkt. Ook geeft de testpersoon aan dat je een melding krijgt dat de brief op B1-niveau is, als er niets verbeterd hoeft te worden.

G.1.4 Explainability & Trust.

- **Zijn de signalen van de AI-tool nuttig en voldoende uitgelegd om jouw werk te ondersteunen?**

Ja, misschien de bron dus dubbelop.

- **Zou je deze AI-tool vertrouwen?**

Ja, maar eerst nog een aantal keer gebruiken en dan zien of het werkt. Op alles accepteren en dan nalezen zelf.

- **Geeft de AI-tool informatie die jou helpt om beslissingen te nemen, of juist te veel/te weinig?**

Alleen de bron één keer noemen en verder is het fijn dat de tool aan- en uitgezet kan worden.

- **Zou je te veel beïnvloed worden door de suggesties die de AI-tool geeft?**

Dat weet de testpersoon niet, soms wel en soms niet. Ze zouden wel nog een vakterm moeten gebruiken, waardoor een assistent kan aangeven dat dit moeilijk is, maar deze overslaan.

- **Vind je deze AI-tool vertrouwelijk in de gemeentelijke context?**

Ja

G.2 Testronde 2: Uitgevoerd door Esmee en Pim (team 5)

G.2.1 Algemene bevindingen en verbeterpunten.

De testpersonen hebben zelf de knop ‘AI-assistent’ gevonden en hebben deze zelf aangeklikt. Hierbij heeft teamlid Terence verteld dat het onderstreepte deel in de tekst waarschijnlijk een highlight wordt in het geel in de tweede iteratie.

Esmee geeft aan dat we rechtsboven het woord ‘suggestie’ kunnen vervangen voor ‘fouten’. Ook Pim vindt, net zoals de persoon in de eerste testronde, dat twee keer de bron noemen overbodig is. Zowel Esmee als Pim geven aan deze bronnen te combineren tot één.

G.2.2 Desirability (wenselijkheid).

- **Voldoet de optie om de AI-tool aan en uit te zetten, aan jullie wensen?**
Ja, dat is juist fijn.
- **Zou je als medewerker prettig om kunnen gaan met de AI-tool?**
Ja
- **Is het wenselijk dat de AI-tool aangeeft dat er een suggestie is voor een bepaalde zin of woord?**
Ja die zijn ook wel fijn. Misschien kan er nog een knop toegevoegd worden voor ‘volgende suggestie’.
- **Zijn er andere stappen in het prototype wenselijk?**
‘suggestie’ vervangen voor ‘fouten’ rechtsboven.

G.2.3 Explainability & Trust.

- **Zou je deze AI-tool vertrouwen?**
Opzich niet.
- **Geeft de AI-tool voldoende informatie?**
Ja, er wordt voldoende informatie getoond om een suggestie te geven die onderbouwd is.
- **Zou je te veel beïnvloed worden door de suggesties?**
Nee, want doordat er stap voor stap door de suggesties wordt gelopen leer je er ook van.
- **Zou je een knop ‘alles accepteren’ willen?**
Nee want nu moet iemand er nog over nadenken.
- **Vind je deze AI-tool vertrouwelijk in de gemeentelijke context?**
Over het algemeen vinden ze hem wel nice.

G.3 Testronde 3: Uitgevoerd door Annemaria van Teeseling, teamleider WMO

G.3.1 Algemene bevindingen en verbeterpunten.

De testpersoon heeft zelf de knop ‘AI-assistent’ gevonden en heeft deze zelf aangeklikt. Hierbij heeft teamlid Terence verteld dat het onderstreepte deel in de tekst waarschijnlijk een highlight wordt in het geel in de tweede iteratie. De testpersoon kon de knop vinden, waarmee je naar de volgende fout kon gaan. Het prototype viel in de smaak.

De testpersoon gaf aan dat het beter is om het woord ‘AI-assistent’ te vervangen door ‘schrijfwijzer assistent’. Ook zou de testpersoon een optie toegevoegd willen hebben om alle suggesties te accepteren. Maar ze moeten wel echt zelf na blijven denken. Blijkt ook uit iets wat nu loopt met ‘copilot’.

G.3.2 Vragen van de testpersoon.

- Wat is het verschil tussen aanpassen en accepteren?
- Wat gebeurt er als ze hem aan het begin aanzet en aan het eind?
- Wanneer kan je hem aanzetten?

G.3.3 Desirability (wenselijkheid).

- **Voldoet de optie om de AI-tool aan en uit te zetten?**

Ja dat vindt ze heel erg fijn! Ze vindt het echt heel fijn dat deze tool ook achteraf aangezet kan worden, zodat ze eerst losjes kunnen schrijven. Ze vindt het ook heel leuk dat je fouten gemarkeerd kan krijgen, zonder de suggesties al te zien, zodat je jezelf kan uitdagen.

- **Zijn alle stappen wenselijk?**

Misschien alle suggesties onder elkaar, want dan kun je alle suggesties bekijken voordat je op alles accepteren klikt.

- **Denk je dat je de AI-tool snel kunt leren gebruiken?**

Ja (duidelijke ja), ze vindt het echt fijn dat hij aan en uit kan en dat er uitgelegd wordt waarom er een suggestie gedaan wordt. Het aanpassen vindt ze heel leuk dat dat ook rechts kan. Het verschil tussen aanpassen en accepteren moet je net even weten, maar als je het weet vindt ze het geweldig.

- **Andere wenselijke stappen?**

Een knop voor ‘andere suggestie’ zou wel heel mooi zijn.

G.3.4 Explainability & Trust.

- **Zou je te veel beïnvloed worden?**

Nee zelf niet. Het is wel goed zoals het nu is.

- **Vind je dit vertrouwelijk in gemeentelijke context?**

Ze kennen tolkje nog niet. Soms zeggen gemeenten dat ze al ver zijn, maar in de praktijk ben je al snel ver. Als je werkt met ‘copilot’. Soso vinden hun heel risicogevoelig. Misschien heeft het ook wel te maken met geld, maar dat weten ze nu niet.

G.4 Testronde 4: Uitgevoerd door Savannah Werner en Rainesh Rewat*G.4.1 Algemene bevindingen en verbeterpunten.*

De testpersonen hebben zelf de knop ‘AI-assistent’ gevonden en hebben deze zelf aangeklikt. Hierbij heeft teamlid Terence verteld dat het onderstreepte deel in de tekst waarschijnlijk een highlight wordt in het geel in de tweede iteratie. Savannah vond de inhoud erg duidelijk en de tool mooi.

Rainesh vindt dat de oorspronkelijke zin niet opnieuw getoond hoeft te worden rechts. Savannah wil dat er een onboarding komt voor als je nog nooit met de AI-tool hebt gewerkt.

G.4.2 Vragen van de testpersonen.

- Werkt de tool live?
- Hoe gaat het model om met structuur van de tekst, zoals kopjes?

G.5 Testronde 5: Uitgevoerd door Elsbeth Francisca, beleidsadviseur van de gemeente

G.5.1 Algemene bevindingen en verbeterpunten.

De testpersoon heeft zelf de knop ‘AI-assistent’ gevonden en heeft deze zelf aangeklikt. Hierbij heeft teamlid Terence verteld dat het onderstreepte deel in de tekst waarschijnlijk een highlight wordt in het geel in de tweede iteratie. Savannah vond de inhoud erg duidelijk en de tool mooi. Als het model zou leren, vind ze wel een risico als de brief in de database beland. Ze twijfelt of de term AI-gebruikt moet worden, misschien iets als ‘schrijfwijzer assistent’. Bijvoorbeeld voor collega’s die ouder zijn.

G.5.2 Desirability (wenselijkheid).

- **Voldoet de optie om de AI-tool aan en uit te zetten?**

Ja, want sommige zullen de tool misschien wel niet willen gebruiken, maar het is goed om de autonomie aan de schrijver te laten. Daarnaast kon ze de knop snel vinden.

- **Is het wenselijk dat de AI-tool aangeeft dat er een suggestie is voor een zin of woord?**

Ja. Het is wel goed dat je als schrijver nog leert en je verantwoordelijk voelt voor de inhoud. Als je in de modus van gemak zou zitten is niet goed natuurlijk. Het is goed dat een suggestie wordt gegeven en niet meteen wordt toegepast.

- **Denk je dat je de AI-tool snel kunt leren gebruiken?**

Ja en collega’s op leeftijd ook wel. Het is heel gebruiksvriendelijk.

- **Andere stappen wenselijk?**

Een melding met ‘je bent zelf verantwoordelijk voor de inhoud van deze brief’ had een andere groep en dat was erg goed. Ze zou de knop voor de volgende fout niet vinden, maar zou op het onderstreepte woord zelf klikken. Een kleur gebruiken als highlight of een strook eronder die niet zo ‘in your face is’.

G.5.3 Explainability & Trust.

- **Geeft de AI-tool informatie die helpt bij beslissingen?**

Ik zou misschien wel te veel vertrouwen op de suggesties.

- **Zou je te veel beïnvloed worden?**

Met een knop ‘Alles accepteren’ wel. Ik zou de bal bij de schrijver laten liggen.