

Advies Datafundamenten & Analytics en Klantinteractie & Services

Inhoudsopgave

1. Introductie	3
1.1. Technologische progressie versus ongewenste effecten	3
1.2. Behandeling adviesvragen	3
1.3. Managementsamenvatting advies	4
2. Adviesvraag van DF&A en KI&S	6
2.1. Adviesvraag	6
2.2. Toelichting op de adviesvraag	6
3. Advies aan DF&A en KI&S	8
3.1. Preambule - toelichting op het standpunt van de Adviescommissie	8
3.2. Randvoorwaarden en verantwoorde implementatie	8
3.3. Kernwaarden, principes en ethische uitdagingen	12
3.4. Technologische overwegingen	15
3.5. Het gebruik van de discretionaire ruimte en biaswaakzaamheid	19
3.6. Autonomie en afgebakende taken	23
3.7. Kaders en regelgeving	24
Appendix A: Overzicht vragen uit advies	26
Appendix B: Toelichting vindplaats antwoord per vraag	28
Appendix C: Technische toelichting Adviesvraag	29
Opbouw en randvoorwaarden Belbot	29
Output metriecken voor Belbot	30
Appendix D: Over de Adviescommissie Analytics	31
Procesverloop beantwoording adviesvraag	31

1. Introductie

1.1. Technologische progressie versus ongewenste effecten

Kunstmatige intelligentie (of Artificiële Intelligentie, verder: AI) vertegenwoordigt één van de grootste technologische revoluties van onze tijd. De impact ervan wordt steeds duidelijker in ons dagelijks leven, zowel op persoonlijk als professioneel niveau. Hoewel het begin van de geschiedenis van AI teruggaat tot rond de jaren 1960, werd het tot voor kort vooral gebruikt voor de analyse en verwerking van bestaande elementen: met behulp van AI is het bijvoorbeeld mogelijk om een nieuwsfeed te sorteren op basis van voorspellingen over de betrokkenheid van gebruikers. Het recent opgekomen type Generatieve AI (verder: GenAI) is daarentegen een nieuw proces dat een prompt (instructies die aan een AI-model worden gegeven om een uitvoer te genereren) verwerkt en om zet in iets compleet nieuws, zoals een afbeelding, een video of een audiobestand. De versnelde ontwikkeling en het wijdverspreide gebruik van deze technologie roept cruciale vragen op over de ethische kanten die inherent zijn aan de desbetreffende innovatie. In een wereld waarin AI een steeds centralere rol inneemt, is het essentieel om de effecten, uitdagingen en kansen ervan zorgvuldig te onderzoeken, op zoek naar een balans tussen vooruitgang en het intact houden van de menselijke maat en ethische waardenpatronen.

Een van de meeste prangende vragen bij het gebruik van steeds verdergaande technologische toepassingen is de vraag hoe je in AI een deugd als barmhartigheid inbouwt. Hierbij wordt barmhartigheid begrepen als een deugd die gekenmerkt wordt door een empathische en onbaatzuchtige betrokkenheid op mensen die hulp en ondersteuning nodig hebben en het zonder deze deugd mogelijk niet redden. Een uitvoeringsorganisatie zoals de Belastingdienst wordt geacht om de wet uit te voeren; in hoeverre is er nog ruimte voor het tonen van mededogen met mensen die het moeilijk hebben? Zal de inzet van GenAI dit streven niet bemoeilijken, omdat machines calculerend zijn? Is er als belastingbetaler nog ruimte om fouten te maken, of confronteren slimme AI-systemen ons straks onverbiddelijk met de harde gevolgen van het zijn van een feilbaar mens?

Medewerkers bij de Belastingdienst zien het potentieel van GenAI voor de automatisering en kwaliteitsverbetering van zowel eenvoudige en repetitieve als lastige en tijdrovende werkzaamheden. Hiermee wordt het een component van technologische progressie en zien ambtenaren kansen om nieuwe technieken uit te proberen ter verbetering van de interne processen om burgers en/of bedrijven beter van dienst te zijn. De vraag is nu in hoeverre de wil om deze technologische progressie te laten plaatsvinden binnen de overheid zich verhoudt tot ongewenste effecten zoals mogelijk een ongelijke behandeling, het ontbreken van de menselijke maat (het inachtnemen van de individu met diens individuele omstandigheden) en de voornoemde barmhartigheid, alsook de verlaging van de kwaliteit van maatwerk en het uit het oog verliezen van fiscaal vakmanschap.

1.2. Behandeling adviesvragen

Een adviesvraag wordt opgehangen aan de kapstok 'Willen, Mogen, Kunnen'. Een adviesvraag aan de Adviescommissie Analytics (verder: de Adviescommissie¹) is bedoeld om helderheid en richting te krijgen over de wenselijkheid ('Willen') van een bepaalde Analytics toepassing, daar waar dilemma's plaatsvinden en lastig een (ethische) afweging gemaakt kan worden. Vragen met betrekking tot 'Mogen' en 'Kunnen' dienen organisatieonderdelen eerst zelf te beantwoorden. Het gaat hier namelijk om vragen die alleen beantwoord kunnen worden wanneer men beschikt over gedetailleerde inhoudelijke kennis (over wet- en regelgeving, kaders, ICT, data science, uitvoering, etc.).

Een adviesvraag zal gaan over de maatschappelijke en morele wenselijkheid van een bepaald handelen. Het advies gaat over het afwegen van belangen vanuit een externe blik. Een advies zal dan ook eerder richtinggevend dan kaderbepalend zijn.

¹ Zie 'Appendix D: Over de Adviescommissie Analytics' voor meer informatie.

Leeswijzer

De Adviescommissie heeft ervoor gekozen om het advies in een lopend verhaal te presenteren (in plaats van vraag-antwoord).² Daardoor zijn de inleiding en de paragrafen daaronder een integraal onderdeel van de beantwoording. Omdat antwoorden op sommige vragen omgeven zijn door veel afhankelijkheden, beroept de Adviescommissie zich soms op het feit dat andere actoren (zoals bijvoorbeeld de politiek of het ministerie) in het speelveld eerst duidelijkheid dienen te verschaffen over de algemene ethische voorwaarden voor de inzet van GenAI.

1.3. Managementsamenvatting advies

De inzet van GenAI binnen de Belastingdienst, in het bijzonder via een dialoog ondersteunend systeem zoals de Belbot (een intelligente chatbot) ter ondersteuning van medewerkers bij de BelastingTelefoon, biedt zowel kansen als uitdagingen. Het potentieel voor een efficiëntere en consistentere dienstverlening staat tegenover belangrijke ethische en juridische vraagstukken. Dit adviesrapport verkent de randvoorwaarden voor een verantwoorde implementatie en geeft aanbevelingen over de wijze waarop AI kan bijdragen aan een transparante, uitlegbare en betrouwbare dienstverlening.

Kernpunten en bevindingen

1. AI als ondersteuning, geen vervanging van fiscale expertise

Het risico op een vermindering van vakinhoudelijke expertise wordt versterkt wanneer AI-systemen bepaalde patronen of aannames bevatten die niet expliciet worden herkend of gecorrigeerd door de gebruiker. De Adviescommissie onderstreept daarom het belang van een hybride werkwijze waarin AI niet de menselijke expertise vervangt, maar juist versterkt. Dit vereist niet alleen blijvende aandacht voor scholing en training, maar ook een duidelijke rolverdeling waarbij medewerkers actief betrokken blijven bij de interpretatie en verantwoording van AI-gestuurde adviezen.

2. Verloren vaardigheden voor medewerkers

Hoewel eerstelijns medewerkers bij de BetalingTelefoon geen fiscale kennis hebben en primair het belscript volgen, is het risico op (inhoudelijke) *deskilling* een aandachtspunt bij de inzet van AI-ondersteuning. Hierbij worden medewerkers door toenemende afhankelijkheid van AI-systemen geleidelijk minder vaardig in het zelfstandig interpreteren en beoordelen van fiscale vraagstukken. Wanneer AI als primair referentiepunt wordt gebruikt, kan dit leiden tot een vermindering van vakinhoudelijke expertise en een verminderd vermogen om complexe of uitzonderlijke casussen adequaat af te handelen.

3. AI-geletterdheid als randvoorwaarde

Een succesvolle en verantwoorde inzet van AI vereist voldoende AI-geletterdheid bij medewerkers. Dit betekent niet alleen kennis van de werking van AI, maar ook inzicht in juridische en ethische kaders evenals het in staat zijn tot het toepassen van een deugd als barmhartigheid na inachtneming van de individuele omstandigheden van een cliënt (de menselijke maat). De Adviescommissie adviseert structurele training en borging van deze kennis binnen de organisatie, zodat medewerkers AI-gestuurde adviezen kritisch kunnen beoordelen en, indien nodig, bijstellen.

4. Ethische en beleidsmatige overwegingen

De inzet van de Belbot moet consistent zijn met de Overheidsbrede visie op Generatieve AI, waarin principes zoals transparantie, controleerbaarheid en menselijke tussenkomst centraal staan. De Adviescommissie benadrukt het risico van wensdenken bij de implementatie van GenAI: technologische vooruitgang garandeert niet per definitie een betere dienstverlening. Kritische toetsing (met zowel kwantitatieve als kwalitatieve evaluatie) en iteratieve implementatie zijn noodzakelijk.

5. Acceptatie en randvoorwaarden voor AI-ondersteuning

De Belbot kan de dienstverlening verbeteren door medewerkers te ondersteunen bij het beantwoorden van vragen, waarbij gevalideerde kennis en (onder strikte voorwaarden) persoonsgegevens uit klantdossiers worden ingezet. Echter, de mate waarin deze technologie zich verhoudt tot de

² M.b.v. 'Appendix B: Toelichting vindplaats antwoord per vraag' kunnen vragen herleid worden naar antwoorden in de tekst.

kernwaarden van de Belastingdienst – zoals verantwoordelijkheid, geloofwaardigheid en zorgvuldigheid – hangt af van de wijze waarop AI wordt ingezet. De Adviescommissie benadrukt dat AI een hulpmiddel moet blijven en niet mag leiden tot een inperking van de beleving van de professionele discretionaire ruimte van niet-eerstelijnsmedewerkers en de hierbij in acht te nemen menselijke maat.

6. Vertrouwensbeginsel en gegevensgebruik

Het gebruik van persoonsgegevens door de Belbot raakt direct aan het vertrouwensbeginsel. De wijze van gegevensverwerking is hierbij van cruciaal belang. Een Retrieval Augmented Generation (RAG)-benadering kan de transparantie en herleidbaarheid van antwoorden verbeteren in vergelijking met een generiek Large Language Model (LLM)-gebaseerd systeem. Het gebruik van een RAG-gebaseerde oplossing kan de privacy van gegevens tot op zekere hoogte beter waarborgen, omdat sensitieve data in een aparte (lokale) database bewaard wordt en er meer controle uitgeoefend kan worden over welke data wanneer gebruikt wordt. Wel wordt de informatie tijdens het gebruik nog steeds naar het LLM model gestuurd als context voor de uiteindelijke output. Daarnaast blijft er sprake van risico's op hallucinaties en foutieve interpretaties, waarvoor adequate mitigaties (corrigerende maatregelen) noodzakelijk zijn.

7. Transparantie en keuzevrijheid voor burgers

Burgers moeten op transparante wijze worden geïnformeerd over het gebruik van AI. De Adviescommissie wijst op de noodzaak van expliciete toestemming voor het opslaan en hergebruiken van data, zoals voor kwaliteitsmonitoring of modeltraining. Daarnaast verdient het aanbeveling om bellers de mogelijkheid te bieden een schriftelijke bevestiging te ontvangen van de verstrekte informatie, inclusief de onderliggende gegevens en fiscale onderbouwing.

8. Governance en externe afstemming

Gezien het geïntensiveerde toezicht door de Autoriteit Persoonsgegevens (verder: AP) is het noodzakelijk om vroegtijdig afstemming te zoeken over de verwerking van persoonsgegevens binnen AI-toepassingen. Daarnaast moet worden voorkomen dat de Belbot leidt tot ongewenste afhankelijkheid van externe technologieaanbieders. Voor een intern gehost AI-model wordt aanbevolen om de controle over data en privacywaarborgen te versterken.

Conclusie en aanbevelingen

De Adviescommissie erkent het potentieel van de Belbot om bij te dragen aan een efficiëntere en effectievere dienstverlening. Tegelijkertijd vraagt een verantwoorde implementatie om een zorgvuldige afweging van ethische, juridische en organisatorische aspecten. De Belbot moet een ondersteunend instrument blijven, waarbij menselijke controle, uitlegbaarheid en privacybescherming gewaarborgd zijn. De Adviescommissie adviseert om de implementatie gefaseerd uit te rollen, met duidelijke evaluatiemomenten, en roept op tot voortdurende monitoring en bijsturing waar nodig.

2. Adviesvraag van DF&A en KI&S

Deze adviesvraag komt vanuit de Belastingdienst, Corporate Dienst Datafundamenten & Analytics (DF&A) met als mede-indiener Directie Klantinteractie & - services (KI&S). Deze adviesvraag heeft betrekking op de inzet van GenAI bij de Belastingtelefoon. Eigenaar van de adviesvraag is

Persoonsgegevens

Persoonsgegevens

2.1. Adviesvraag

Het doel van de adviesvraag is om een verkennend gesprek met de Adviescommissie tot stand te brengen. De adviesvrager is benieuwd of de Adviescommissie hen bepaalde overwegingen mee wil geven met betrekking tot de verantwoorde vormgeving en inbedding van de Belbot en of ze specifieke criteria, methodieken of publicaties aanbeveelt. Op deze wijze hoopt de adviesvrager nog beter zicht te krijgen op de voorwaarden voor verantwoorde inzet van GenAI voor dialoogondersteuning van de servicemedewerker bij de Belastingtelefoon. De adviesvrager stelt deze adviesvraag als aanvulling op eigen analyses en kennisuitwisseling in interne en interdepartementale werkgroepen rond de visievorming op GenAI.

De hoofdvraag luidt: Is het acceptabel om servicemedewerkers bij de Belastingtelefoon te laten bijstaan door een op GenAI gebaseerd systeem voor dialoogondersteuning?

Deze hoofdvraag wordt genuanceerd door een aantal deelvragen³ die toezien op verschillende mogelijke implementaties van een op GenAI gebaseerd systeem voor dialoogondersteuning en op een aantal beginselen en waarden die de overheid, en specifiek de Belastingdienst, nastreeft. Daarnaast zijn aanvullende vragen gesteld voor die elementen waarop de implementatie van een op GenAI gebaseerd systeem voor dialoogondersteuning kan variëren en waarin overige ethische aspecten zijn verdisconteerd. In dit advies worden deze vragen per thema beantwoord. In 'Appendix B: Toelichting vindplaats antwoord per vraag' staat een overzicht in welke paragraaf welke (deel)vraag terug te vinden is.

2.2. Toelichting op de adviesvraag

Uit een innovatief project (de Drie Dagen Challenge, verder: 3DC) binnen de Belastingdienst hebben DF&A en KI&S de vraag naar voren gebracht of GenAI mogelijkheden biedt om servicemedewerkers bij de Belastingtelefoon te faciliteren door een deel van de informatieverwerking m.b.v. GenAI te automatiseren.

De adviesvrager denkt dat deze toepassing goed past in de Overheidsbrede visie GenAI⁴, omdat het kan bijdragen aan het welzijn van zowel de bellers als de medewerkers door een adequatere dienstverlening met oog voor de menselijke maat. Wel moeten er voldoende waarborgen zijn ingebouwd om het systeem veilig en betrouwbaar in te zetten, zonder dat er sprake is van privacy-aantasting en hallucinaties.⁵ Bij de ontwikkeling van de Belbot wordt IAMA⁶ als algemeen toetsingskader ingezet, terwijl ook rekening wordt gehouden met de uitgangspunten van de Overheidsbrede visie GenAI. Recente onderzoeksrapporten laten zien dat zorgvuldigheid is geboden.

Aannemend dat het systeem voor dialoogondersteuning voldoende waarborgen kent om veilig en betrouwbaar ingezet te kunnen worden bij de werkzaamheden van medewerkers, zijn de adviesvragers benieuwd of er nog ethische risico's zijn verbonden aan de voorgestelde toepassing die onvoldoende in beeld zijn, maar wel meegenomen moeten worden in het ontwerp van het systeem zelf en de processen waarbinnen het zou moeten functioneren. Hierbij wordt ook de mogelijkheid opengehouden dat de

3 Voor een overzicht van de deelvragen, zie 'Appendix A: Overzicht vragen uit advies'

4 <https://www.rijksoverheid.nl/documenten/rapporten/2024/01/01/overheidsbrede-visie-generatieve-ai>

5 Bij AI is een hallucinatie een zelfverzekerde reactie van een AI die niet lijkt gerechtvaardigd te worden door zijn trainingsgegevens (bron: [https://nl.wikipedia.org/wiki/Hallucinatie_\(kunstmatige_intelligentie\)](https://nl.wikipedia.org/wiki/Hallucinatie_(kunstmatige_intelligentie))).

6 <https://www.rijksoverheid.nl/documenten/rapporten/2021/02/25/impact-assessment-mensenrechten-en-algoritmes>

Belbot een oplossing kan bieden voor tekortkomingen in de bestaande processen die juist door deze vorm van automatisering goed zijn op te lossen, zoals het vastleggen en valideren van gegeven antwoorden.

Uit een analyse van KI&S van de oorzaken achter de verminderde bereikbaarheid van de BelastingTelefoon is blijkens de aangeleverde adviesvraag naar voren gekomen dat het aantal telefoontjes toeneemt, terwijl het aantal afgehandelde gesprekken daalt en de formatieve pool van medewerkers instabiel is (door een hoge doorstroom met lastige werving en lange inwerktijden). Technologische ondersteuning, zoals een intelligente chatbot (de Belbot), is een mogelijke oplossingsrichting om de dienstverlening op peil te houden. Het basisidee is dat de Belbot de servicemedewerker tijdens het gesprek op de volgende manieren kan ondersteunen:

1. Met het opzoeken en klaarzetten van informatie. De Belbot kan zo bijdragen aan een snellere afhandeling (reductie van de zoektijd), betere kwaliteit (juiste, consistente en begrijpelijke antwoorden) en een betere verantwoording (reproduceerbare antwoorden met een heldere onderbouwing).
2. Door middel van ondersteuning voor het opleidingstraject van nieuwe medewerkers, bijvoorbeeld door bij concrete vragen aan te geven welke informatie nodig is, waar die informatie staat en hoe je tot een antwoord kunt komen.

De adviesvragers verwachten dat de hulpvragen aan de BelastingTelefoon door de assistentie van de Belbot efficiënter en beter kunnen worden afgehandeld. Hierdoor kunnen meer belastingplichtigen de juiste hulp krijgen, met inachtneming van hun persoonlijke omstandigheden, hetgeen een positief effect zal hebben op de compliantie.

Aanvullend zal voor daadwerkelijke inzet in het werkproces aanvullende privacytoetsing worden uitgevoerd om te bepalen welke brondata wel en niet ingezet mag worden in dit proces. Idealiter verwerkt het LLM zelf (anders dan de vooranalyse door de RAG-component) geen directe persoonsgegevens (zoals BSN of naam), maar alleen fiscaal relevante gegevens (vanuit de data en documentatie die de RAG-component voor het LLM klaarzet). Bijkomend voordeel is dat op deze manier mogelijk een kleinere kans op bias ontstaat en verdere informatieverwerking anoniemer is. Het is een open vraag in hoeverre de informatie-uitwisseling in echte gesprekken via opnames, transcripties en gesprekslogs mag worden ingezet voor het evalueren en zelfs bijtrainen van het AI-systeem, en in welke component(en) de traceerbare persoonsgegevens mogen worden verwerkt.

Om meer zicht te krijgen op ethische risico's is in februari 2024 een ethische casusdialoog (rollenspel) georganiseerd met een ethische klankbordgroep bij DF&A. Hierbij is het voorgestelde product vanuit verschillende invalshoeken bekeken, waaronder de directie, de interne gebruiker, de burger, de politiek en de media. De belangrijkste aandachtspunten die hierbij naar voren kwamen zijn dat de medewerker graag van tevoren duidelijk wil hebben wie verantwoordelijk is voor het gegeven antwoord en dat zowel de medewerker als de burger zorgen hebben over de acceptatie van het AI-systeem als onbekende derde in het gesprek. Beide zorgen kunnen worden geadresseerd door de burger te informeren dat de servicemedewerker hulp krijgt van een AI-systeem en dat de burger kan aangeven of de burger hiermee akkoord gaat.

3. Advies aan DF&A en KI&S

De beantwoording van de hoofdvraag ‘of het acceptabel is om servicemedewerkers bij de Belastingtelefoon te laten bijstaan door een op GenAI gebaseerd systeem voor dialoogondersteuning’ wordt door de Adviescommissie **positief** beantwoord met een ‘ja’. De Adviescommissie geeft middels dit rapport meerdere (ethische) overwegingen die meegenomen moeten worden voor verantwoorde implementatie.

3.1. Preambule - toelichting op het standpunt van de Adviescommissie

Alvorens over te gaan tot het advies wil de Adviescommissie graag een preambule uitbrengen om haar standpunten duidelijk te maken ten opzichte van diverse AI gerelateerde onderwerpen en thema's. Deze preambule is belangrijk om in het achterhoofd te houden tijdens het lezen van onderstaand advies.

Centraal in de preambule is het standpunt van de Adviescommissie dat niet alle vragen beantwoord kunnen worden zonder deze los te zien van het standpunt van de politiek over AI. Het ministerie van Financiën (of de bewindspersoon) dient een helder en duidelijk standpunt in te nemen over het gebruik en de inzet van GenAI alvorens bepaalde vragen beantwoord kunnen worden in de context van een uitvoeringsdienst zoals de Belastingdienst. De Belastingdienst is namelijk ook afhankelijk van de politieke en maatschappelijke afwegingen hierin. Deze keuzes kan de Adviescommissie niet maken; wel kan de commissie richting geven en aangeven waar men rekening mee dient te houden.

Daarnaast constateert de Adviescommissie dat niet alle vragen expliciet over de Belbot gaan, maar de vraagstukken betreffende Belbot ingebed zijn in een veel bredere (maatschappelijk) discussie. Dat maakt het solistisch beantwoorden van dergelijke vragen uitdagend omdat de context essentieel onderdeel is van de vraag. De Adviescommissie is zich ervan bewust dat de inzet van GenAI gepaard gaat met keuzes die niet alleen te maken zijn door één organisatie maar plaatsvinden in en mede bepaald worden door een samenleving als geheel c.q. door de volksvertegenwoordiging.

Hierdoor zal een deel van de adviesvraag beantwoord worden door het politieke proces dat wij in Nederland kennen en bij het uitspreken welke waarden, deugden of belangen belangrijker zijn dan andere. Ook wordt er in de adviesvraag gevraagd naar keuzes met betrekking tot de implementatie van de Belbot. Deze implementatie is uiterst afhankelijk van de context en vragen die hierop betrekking hebben kunnen dus niet altijd beantwoord worden door de Adviescommissie. Zo is er bijvoorbeeld een voorlopige richtlijn die de inzet van GenAI inperkt, onder meer door ambtenaren te verbieden om gebruik te maken van niet-gecontracteerde clouddiensten die gebruik maken van generatieve AI, in afwachting van verdere ethische visievorming. Deze richtlijn is relevant voor het advies omtrent de inzet van GenAI bij de Belbot.

De Adviescommissie wordt middels het stellen van bepaalde vragen uitgenodigd om een principiële positie ten opzichte van de operationele uitvoering te bepalen, terwijl de Adviescommissie onvoldoende in staat is om de impact van de applicatie op deze operationele uitvoering in te schatten. De Adviescommissie staat met enige afstand van de organisatie die de adviesvraag indient, mede om het onafhankelijke karakter van de commissie te waarborgen.

3.2. Randvoorwaarden en verantwoorde implementatie

Deze paragraaf beschrijft de randvoorwaarden en implicaties van de ondersteuning van servicemedewerkers door de Belbot. Het bevat concrete voorstellen waarop gelet zou moeten worden bij een verantwoorde implementatie. Een kernvraag is of en onder welke voorwaarden de Belbot acceptabel kan zijn. De technologie zou de dienstverlening kunnen verbeteren door antwoorden consistent en meer persoonlijk te maken, maar dit roept direct vragen op over de waardering van het vertrouwensbeginsel en de mate waarin AI gebruik mag maken van persoonsgegevens. De manier waarop gegevens worden verwerkt – bijvoorbeeld via de RAG-methode in plaats van een generiek LLM-gebaseerd dialoogsysteem – kan hierbij een cruciale rol spelen.

Een andere overweging is de mate van transparantie en keuzevrijheid voor de burger. Moet de beller bijvoorbeeld vooraf worden geïnformeerd over het gebruik van AI en de mogelijke voordelen hiervan? En in hoeverre moet de Belbot standaard een schriftelijke bevestiging verstrekken van het verstrekte advies, inclusief de gebruikte gegevens en juridische onderbouwing?

Visie vanuit de Adviescommissie

De Adviescommissie erkent – net zoals de adviesvrager – dat een toepassing van de Belbot goed zou passen in de Overheidsbrede visie Generatieve AI, zoals door het vergroten van het welzijn van zowel de bellers als de medewerkers. De Adviescommissie is benieuwd naar de specifiekere motivatie om een op GenAI gebaseerd systeem voor dialoogondersteuning in te zetten bij de BelastingTelefoon. In de adviesvraag wordt aangegeven dat de Belbot kan bijdragen aan kernwaarden, zoals het vergroten van het wederzijds begrip en dus het welzijn van zowel de bellers als de medewerkers en dat beide groepen baat hebben bij een systeem dat zorgt voor een adequatere dienstverlening met oog voor de menselijke maat. Echter, uit de vragen die volgen blijkt dat er vooral gekeken wordt vanuit een efficiency-oogpunt met een sterk technologische insteek en minder naar de subjectieve toetsingscriteria zoals: vindt de belastingplichtige aan de telefoon dat het probleem is opgelost? Is belastingplichtige aan de telefoon tevreden met het antwoord? Is er sprake van geweest dat diens persoonlijke omstandigheden voldoende in acht genomen zijn? De Adviescommissie is van mening dat deze zaken ook van belang zijn als je wilt bijdragen aan de implementatie van de kernwaarden van de Overheidsbrede visie Generatieve AI.

Rapportcijfers van de BelastingTelefoon

De adviesvrager geeft aan dat de Belbot een oplossing kan bieden voor tekortkomingen in de bestaande processen die juist door deze vorm van automatisering goed zijn op te lossen, zoals het vastleggen en valideren van de gegeven antwoorden. De Adviescommissie erkent dit maar wil ook stil staan bij het feit dat de BelastingTelefoon voor het eerst in 15 jaar een voldoende heeft gescoord in een jaarlijks kwaliteitsonderzoek van de Consumentenbond. De Belastingdienst krijgt een 6,1. Toch is de Consumentenbond nog niet tevreden, want er gaat nog te veel fout en de wachttijden zijn erg lang. Dit is op basis van de huidige vorm van automatisering (zonder GenAI) en met een grote menselijke factor.

Technologie (zoals GenAI) draagt vaak een belofte met zich mee dat voorgaande problemen opgelost kunnen worden door nieuwe vormen van automatisering. De Adviescommissie wil de adviesvrager erop attent maken dat er sprake kan zijn van een bepaalde mate van wensdenken. De kwaliteit van het systeem wordt niet alleen bepaald door de werking van een LLM, maar is ook afhankelijk van andere factoren, zoals kwaliteiten van de medewerker aan de telefoon. **De introductie van een LLM-aangedreven ondersteunende chatbot zal invloed op de kwaliteit hebben, echter met het onderzoek van de Consumentenbond in gedachten dient nader onderzocht te worden op welke aspecten het om een positieve invloed gaat.**

Externe stakeholders

Het gebruik van de Belbot door de Belastingdienst raakt ook externe stakeholders zoals belastingadviseurs, intermediairs, notarissen en andere partijen die zorgdragen voor de fiscale belangen van hun cliënten. Omdat de Belbot naar alle waarschijnlijkheid ook gebruikt gaat worden voor meer complexe fiscale vragen (de zogeheten tweedelijns vragen), zullen ook fiscale professionals te maken krijgen met de Belbot en de impact die dit zal hebben op hun werk. Hier kan een onderscheid gemaakt worden in gebruikerscategorieën. Zo zijn er drie soorten stakeholders te onderscheiden: primaire (belastingmedewerkers, burgers en bedrijven), secundaire (belastingadviseurs) en tertiaire (de hele sector: alle Belastingplichtigen in Nederland). **Er dient per categorie goed in kaart gebracht te worden wat de inzet van de Belbot voor elk van deze categorieën betekent, welke belangen deze stakeholders hebben en hoe die het best gediend kunnen worden met het gebruik van een Belbot.**

Een betrouwbare overheid begint met vertrouwen

De adviesvrager stelt de vraag wat de inzet van Belbot voor het vertrouwensbeginsel betekent als de Belbot ook gebruik mag maken van persoonsgegevens. De Adviescommissie heeft bij de beantwoording een aanname gedaan over wat precies onder het vertrouwensbeginsel wordt verstaan door de adviesvrager en heeft zich beperkt tot het formuleren van een advies vanuit een technisch perspectief.

De voorgestelde RAG-methodiek (Retrieval Augmented Generation), gebaseerd op interne databronnen en kennisbanken, biedt een meer gestructureerde en meer transparante manier om persoonsgegevens te benutten voor het genereren van antwoorden dan alleen een externe LLM of een platform-gebaseerd systeem zoals ChatGPT. Externe opties zijn vaak minder transparant in de manier waarop antwoorden worden gegenereerd en gegevens worden verwerkt. Het RAG-systeem maakt gebruik van interne kennisbanken, waardoor er minder risico is op het lekken van gevoelige gegevens naar externe partijen. Dat draagt bij aan de bescherming van persoonsgegevens en de versterking van het vertrouwensbeginsel. De RAG-methodiek zorgt ervoor dat antwoorden gebaseerd zijn op gevalideerde bronnen en dat de redeneerlijn voor de servicemedewerker inzichtelijk blijft. **De RAG-methode maakt het voor medewerkers eenvoudiger om de gebruikte bronnen te controleren, zodat ze kunnen nagaan hoe een bepaald antwoord tot stand is gekomen en meer transparantie kunnen bieden richting de beller.**

Er blijft echter, ook wanneer gekozen wordt voor een RAG-opzet, kans op ‘hallucinaties’ (onjuiste informatie). Daarom ook blijft het van belang om de overige eisen van de AVG niet uit het oog te verliezen. Wanneer Belbot gebruik gaat maken van persoonsgegevens dient een Data Protection Impact Assessment (DPIA) te worden uitgevoerd. Aandachtspunten daarbij zullen in ieder geval de grondslag voor de gegevensverwerking en het doelbindingsbeginsel zijn. **Om daadwerkelijk bij te kunnen dragen aan een mogelijke versterking van het vertrouwensbeginsel is het onder meer essentieel dat de verwerkte persoonsgegevens juist zijn (zie art. 5 lid 1 onder d AVG) en dat burgers adequaat over de gegevensverwerking worden geïnformeerd (zie art. 12-14 AVG).**

Gelet op het feit dat de Belastingdienst vijf jaar onder geïntensiveerd toezicht van de AP staat, raadt de Adviescommissie aan dat indien de Belastingdienst overgaat tot het verwerken van persoonsgegevens via Belbot, afstemming met de AP te laten plaatsvinden om te waarborgen dat de gegevensverwerking AVG-compliant plaatsvindt en de burger adequaat beschermd wordt.

De adviesvrager vraagt of de beller geïnformeerd moet worden over de voor- en nadelen van het ondersteund worden door een GenAI-gebaseerd systeem voor dialoogondersteuning, met de mogelijkheid dat de keuze voor de Belbot door voordelen wordt gestimuleerd.

De Adviescommissie is van mening dat de technische validatie van de Belbot voor- en nadelen expliciet zou moeten maken. Het streven zou moeten zijn dat de technologie het werk van de medewerkers effectief kan ondersteunen en de dienstverlening verbetert. **Uitleg over Belbot op de website en registratie in een Algoritmenregister draagt zeker bij aan transparantie voor de gebruiker.**

Informed consent

Een vraag die gesteld kan worden is of de beller expliciet toestemming kan geven voor het hergebruik van zijn informatie voor het geanonimiseerd bijtrainen van het systeem. Is het mogelijk om *informed consent* te faciliteren en de nodige informatie te verstrekken of leidt dit tot een lange uitleg aan de belastingtelefoon? Ook hier is het van belang of de training intern of door een commerciële partij wordt gedaan. Moet de beller vooraf instemmen dat het gesprek wordt ondersteund door een op GenAI gebaseerd systeem voor dialoogondersteuning? En in het geval dat de beller niet instemt, wordt de mogelijkheid gegeven om voor een terugbelafsprak te kiezen?

De vraag is hoe ver de dienstverlening door de Belbot gaat. Is de Belbot diegene die antwoord geeft op vragen van de burger of ondersteunt de Belbot een competente, opgeleide medewerker van de Belastingdienst om haar taken effectiever en meer ethisch verantwoord te kunnen uitvoeren, met inachtneming van de menselijke maat? In het geval dat de expertise verschuift naar de Belbot kunnen zich structureel verschillen voordoen in de kwaliteit van het advies gegeven door een minder competente medewerker plus Belbot en het advies van een competente medewerker zonder Belbot. In dit geval zal toestemming nodig zijn. **Omdat de Belbot niet direct met de burger communiceert en ook nadrukkelijk een ondersteunende rol heeft in de informatieprocessen, meent de Adviescommissie dat informed consent door de beller voor gebruik van de Belbot niet expliciet nodig is. Wél moet toestemming gevraagd worden voor de registratie/opname van het gesprek door de Belbot, een recording device of anderszins, inclusief een opt-out optie zodat bellers altijd een keuze hebben.**

De Adviescommissie is van mening dat het een goede service zou zijn als de beller een schriftelijke bevestiging zou kunnen opvragen van het verstrekte advies, inclusief gebruikte gegevens en fiscale onderbouwing, indien er sprake is van een fiscaal gepersonaliseerd antwoord. Omdat dit in het nadeel zou kunnen zijn van de Belastingdienst met het oog op een eventuele stijging van het aantal bezwaar- en beroepsprocedures, is de vraag hoeveel meerwerk dit met zich mee zou brengen en in hoeverre dit eigenlijk raakt aan het domein van belastingadviseurs en accountants. Een reden om dit wel door de voeren is de transparante onderbouwing van het antwoord (gebaseerd op duidelijk gespecificeerde bronnen en gegevens) wat in de ogen van de Adviescommissie een grote bijdrage geeft aan een betrouwbare overheid. Er moet dus een afweging worden gemaakt m.b.t. de verhouding van de kosten van het eventuele ‘meerwerk’ tot de voornoemde goede service.

Objectieve toetsingscriteria bij een verantwoorde implementatie

De adviesvrager geeft aan een op GenAI gebaseerd systeem voor dialoogondersteuning (verder: een GenAI ondersteuningssysteem) te willen inzetten voor het consistentere en meer gepersonaliseerd beantwoorden van vragen die binnenkomen bij de Belastingtelefoon. Er zijn verschillende toetsingscriteria mogelijk. Welke drempelwaarden gebruikt worden hangt af van de sensitiviteit en impact van de context. **De Adviescommissie adviseert voor consistentie en reproduceerbaarheid van een Belbot-toepassing om naar metriecken uit het veld te kijken, zoals:**

- De mate waarin de output van de BelBot hetzelfde is voor dezelfde vraag en context, bijv. door te kijken naar cosine similarity.⁷ Het Eureka framework⁸ is een open-source framework dat de mogelijkheid biedt om voor taalmodellen te kijken naar verschillen in uitkomst voor identieke input.
- De mate waarin de uitkomst van verschillende medewerkers met ondersteuning van de Belbot hetzelfde is voor dezelfde vraag en context.
- De mate waarin de uitkomst van verschillende medewerkers zonder ondersteuning van de Belbot hetzelfde is voor dezelfde vraag en context.

Ook voor nauwkeurigheid (juistheid) zijn er een aantal objectieve toetsingscriteria bij een verantwoorde implementatie aan te raden. De Belbot-nauwkeurigheid moet getest worden op een *testset*, die bijvoorbeeld geannoteerde voorbeelden bevat van de juiste uitkomst gegeven een vraag en context. Hierbij kunnen verschillende criteria worden onderscheiden:

- Performance op een testset die bestaat uit data met een vergelijkbare distributie als de trainingset.
- Performance op een testset die een andere distributie heeft.
- Performance op een testset die veel outliers of bijzondere gevallen bevat.

Ook bij het annoteren van data bestaan verschillen tussen mensen onderling en dit kan mede de werking en prestaties van het algoritme of de inzet van AI bepalen. **De Adviescommissie wil benadrukken dat er geïnvesteerd dient te worden in (service)medewerkers om de kwaliteit van de antwoorden te testen en blijvend op zoek te gaan naar verbeteringen.** Het is menskracht die de kwaliteitscontrole uitvoert.

Na de ingebruikname van het systeem kan de nauwkeurigheid gemonitord worden door de medewerker een optie te geven om terug te koppelen in hoeverre de uitkomst van het systeem juist of helpend was (bijvoorbeeld met een ‘thumbs up’/‘thumbs down’ functionaliteit). **De Adviescommissie adviseert om voor personalisatie en begrijpelijkheid wederom naast objectieve criteria ook subjectieve criteria te beoordelen.** Bijvoorbeeld: vindt de belastingplichtige aan de telefoon dat het probleem is opgelost; of is belastingplichtige aan de telefoon tevreden met het antwoord; voelt de belastingplichtige zich als individu met diens persoonlijke omstandigheden begrepen en zou de belastingplichtige het antwoord kunnen uitleggen aan een andere burger? Dit zou bijvoorbeeld standaard opgenomen kunnen worden met één en dezelfde benadering van de ‘thumbs up’/‘thumbs down’ functionaliteit, zodat burgers direct kunnen aangeven of het gesprek behulpzaam was.

⁷ Dit is een maat om aan te geven in hoeverre twee stukken tekst een gelijke tekstuele inhoud hebben.

⁸ <https://www.microsoft.com/en-us/research/blog/eureka-evaluating-and-understanding-progress-in-ai/>

3.3. Kernwaarden, principes en ethische uitdagingen

Deze paragraaf beschrijft welke randvoorwaarden cruciaal zijn om een verantwoorde en betrouwbare toepassing te waarborgen. Naast de kernwaarden van de Belastingdienst moet de inzet van AI ook worden beoordeeld binnen het bredere kader van de Uitvoerings- en handavingsstrategie van de Belastingdienst. Hierin staan principes als doelmatigheid, doeltreffendheid, rechtszekerheid en rechtsgelijkheid centraal, evenals het inachtnemen van de menselijke maat in de uitvoering. Een cruciale vraag is of en hoe de Belbot kan bijdragen aan deze uitgangspunten en of een concrete casus, zoals het beoordelen van een verzoek tot kortlopend uitstel van betaling, kan dienen als illustratie van de praktische toepasbaarheid van AI in de belastingdienstverlening.

Daarnaast speelt de bredere ethische afweging een belangrijke rol. De Overheidsbrede visie op GenAI stelt eisen op het gebied van transparantie, controleerbaarheid en publieke verantwoording. Dit roept de vraag op welke ethische argumenten voor of tegen de inzet van de Belbot pleiten en hoe deze techniek in overeenstemming kan worden gebracht met de bredere maatschappelijke verantwoordelijkheid van de overheid.

Visie vanuit de Adviescommissie

De Adviescommissie is van mening dat de Belbot een uitstekende aanleiding vormt om te onderzoeken hoe de normatieve ambities van de organisatie zich moeten manifesteren in een operationeel instrument. De Adviescommissie wil benadrukken dat deze vraag niet uitsluitend bij de Adviescommissie belegd kan worden. Veel van de deelvragen lijken slechts operationeel van aard te zijn maar raken het strategische niveau. De Adviescommissie kan hier wel adviseren, maar het moet duidelijk zijn dat het handelingsvermogen van de Adviescommissie hier (terecht) beperkt blijft. Het gaat dan om vragen die op bestuurlijk niveau moeten worden bepaald (zie: preambule).⁹ De Adviescommissie kan voor alle verschillende opties in deze vraag wel onderbouwde voor- en nadelen noemen. In het geval dat deze vragen uitsluitend worden voorgelegd aan de Adviescommissie dreigt bovendien een zogenaamde *accountability sink* (Davies 2024)¹⁰. Hiervan is sprake als de aanbevelingen als een *carte blanche* zouden worden ingezet ook al vraagt de specifieke context andere praktijken.

In het geval van deze adviesvraag moet de Adviescommissie benadrukken dat het gebruik van de Belbot een reeks van deelvragen impliceert die direct verband houden met beslissingen op het niveau van de bewindspersoon. Deze moet vanuit een politiek en bestuurlijk perspectief keuzes maken uit een reeks van verschillende mogelijkheden. Dit geldt voor vragen zoals de keuze tussen een LLM dat door een commerciële partij wordt aangeboden en gehost, of een eigen model op eigen servers. Dit geldt evenzeer voor ethische vragen; er dient een keuze gemaakt te worden welke morele waarden in de AI moeten worden gemanifesteerd (zie ook de opmerking over barmhartigheid)¹¹.

Cruciaal bij het voornemen van een verantwoorde implementatie van een GenAI

ondersteuningssysteem, is om hier al over na te denken in de ontwerpfase. De Adviescommissie wil haar waardering uitspreken voor het feit dat bij de ontwikkeling van de Belbot het IAMA *framework* als algemeen toetsingskader wordt gehanteerd. Daarnaast geeft de adviesvrager aan dat er rekening gehouden wordt met de uitgangspunten van de Overheidsbrede visie Generatieve AI. **Dit zijn keuzes die passen bij een verantwoorde implementatie. De Adviescommissie roept dan ook op om deze aspecten die centraal zijn gesteld ook gedurende de gehele technische implementatiefase te respecteren en hierin geen concessies te doen.**

Kernwaarden in relatie tot de Belbot en de menselijke maat

De menselijke maat wordt vaak gezien als het vermogen en de wil om rekening te houden met specifieke omstandigheden of eigenschappen van een individu in een bepaald geval. Dit betreft vragen van

9 Om een afweging van waarden te maken, in lijn met de bestaande Uitvoerings- en handavingsstrategie of in lijn met een nieuwe politieke lijn, kan het bestuurlijk niveau gebruik maken van een instrument als CODIO dat de waarde keuzen expliciet maakt (Meijer & Ruijter 2021).

10 Davies, Dan. *The Unaccountability Machine: Why Big Systems Make Terrible Decisions-and How The World Lost its Mind*. Profile Books, 2024.

11 Meijer, A., and E. Ruijter. "Code Goed Digitaal Openbaar Bestuur (CODIO): Borgen van waarden bij de digitalisering van het openbaar bestuur." (2021).

belastingplichtigen over hun belastingverplichtingen en kan ook van toepassing zijn bij de BelastingTelefoon bij de beantwoording van fiscale vragen door de medewerkers. Bij de beoordelaar moeten de omstandigheden, middelen, de bereidheid en motivatie (en de communicatie modaliteiten/vaardigheden die daarmee gepaard gaan) zodanig zijn dat begrip wordt opgebracht voor de concrete en individuele relevante eigenschappen van het individu of het geval: de situatie van de individuele burger. Helpt de inzet van een op GenAI gebaseerd systeem voor dialoogondersteuning dat we de menselijke maat in acht blijven nemen? Hoe zorgt GenAI ervoor dat we extra voorzichtig zijn bij het gebruik ervan en de menselijke maat niet uit het oog te verliezen? Bij welke kernwaarden dient extra aandacht te komen bij een verantwoorde implementatie?

Om deze vragen te beantwoorden moet onder andere gekeken worden in hoeverre de menselijke dispositie verdisconteerd is. Daarmee wordt bedoeld de grondhouding waarin de medewerker van de Belastingdienst een reflexieve ruimte heeft (ontwikkeld) om de uitkomsten van de Belbot op waarde te schatten met betrekking tot de vraag van een individu. Anders gezegd, er dient altijd rekening mee gehouden te worden of iemand bereid of in staat is om de uitkomsten van een Belbot te ontvangen. En hoe staat de servicemedewerker tegenover het gebruik van een Belbot? Is een servicemedewerker bereid en in staat om samen te werken met de Belbot of ziet deze medewerker de Belbot voornamelijk als een vervanging van zijn of haar eigen werk? Tenslotte: er dient sprake te zijn van enige vorm van vrijwilligheid met betrekking tot het gebruik van de Belbot door de servicemedewerker. Dat valt te betwijfelen bij de inzet van de Belbot in de context van de BelastingTelefoon, omdat de verschaft middelen vaak verplicht zijn om mee te werken.

De Adviescommissie wil een aantal observaties meegeven aan de adviesvrager:

- De uitgangssituatie voor de Belastingdienst bij het invoeren van GenAI in telefonisch contact met cliënten/burgers is problematisch. De perceptie van externen op nieuwe initiatieven is al snel negatief en dit soort innovaties zou gezien kunnen worden als het zoveelste bewijs van een verstoorde verhouding tussen burger en overheid i.c. de Belastingdienst.¹²
- Er is een spanning tussen het gebruik van GenAI en het inachtnemen van de menselijke maat in het telefonisch contact met cliënten, in de zin dat GenAI een techniek is waarbij de meest waarschijnlijke reactietekst wordt voorspeld gegeven een vraag (prompt). GenAI kan dus per definitie minder goed rekening houden met meer specifieke, minder voorspelbare situaties van cliënten. Om het nodige maatwerk te leveren zijn menselijke tussenkomst en menselijk contact essentieel. Vanuit het perspectief van menselijke maat is het geïndiceerd om een implementatievorm te kiezen waarbij een menselijke medewerker uiteindelijk altijd beslist welke informatie voor de cliënt relevant is, waarbij GenAI een ondersteunende rol zou kunnen spelen.
- Als het gaat om verantwoordelijkheid zou spanning met GenAI kunnen ontstaan wanneer medewerkers of burgers zich in hun advies of handelen in sterke mate baseren op de output van het GenAI systeem. Wanneer het systeem een fout maakt kan een zogeheten *responsibility gap* ontstaan: wie is er verantwoordelijk? Hier moet binnen de organisatie duidelijk over gecommuniceerd worden.
- Er kan bij servicemedewerkers afhankelijkheid ontstaan van de geleverde informatie van GenAI. Deze is echter niet in alle gevallen betrouwbaar en er kunnen mogelijke *biases* ontstaan waarbij niet alle bellers dezelfde kwaliteit van advies krijgen. Een Nederlandse bank had bijvoorbeeld bij hun interne klantenservice-chatbot last van een LLM die aannam dat alle klanten van de bank mannen zijn.
- Differentiatie (via een keuzemenu) naar verschillende situaties zoals extreem sensitieve casussen, simpele informatieverstrekking, routineus advies, of FAQ zou overwogen kunnen worden. Met een mogelijkheid om af te breken, of een medewerker te kunnen spreken.

12 Ter illustratie: de regeling die Toeslagen-ouders €5.000,- in het vooruitzicht stelde in ruil voor hun afzien van een klacht, werd kritisch, met argwaan of morele verontwaardiging ontvangen.

Ethische principes in relatie tot de Belbot

De Adviescommissie constateert dat er een breed scala aan ethische principes en vanuit de Belastingdienst gedefinieerde waarden een rol spelen bij de inzet van GenAI bij de BelastingTelefoon. De vraag vanuit de adviesvrager omvatte ook de vraag of er nog andere ethische aspecten zijn die voor of tegen de inzet van de Belbot pleiten, ook gezien de uitgangspunten van de Overheidsbrede visie GenAI.

In vele publicaties waarin richtlijnen zijn opgenomen met betrekking tot ethische en technische waarden van AI-systemen, zijn acht algemene principes te vinden die de ontwikkeling en implementatie van op AI gebaseerde systemen moeten leiden: mensenrechten, welzijn, gegevensagentschap, efficiëntie, transparantie, verantwoordingsplicht, bewustzijn van misbruik en competentie. De meest genoemde en bestudeerde principes met betrekking tot de ontwikkeling van op AI gebaseerde systemen zijn: transparantie, rechtvaardig- en eerlijkheid, niet-schade principe, verantwoordelijkheid en privacy¹³. Omdat de adviesvrager expliciet vraagt naar met welke ethische principes (nog meer) rekening gehouden dient te worden, volgt hieronder een toelichting van de principes die de Adviescommissie in deze context zeker van belang acht:

- **Transparantie.** Transparantie van handelingen is een van de belangrijkste aandachtspunten als het gaat om AI. Het gaat erom hoe en waarom de AI een bepaalde beslissing neemt. Transparantie moet niet alleen worden overwogen bij de werking van het systeem, maar moet worden geïntegreerd in de technologische ontwikkeling van AI om de besluitvorming transparanter en betrouwbaarder te maken.
- **Privacy.** Nu er steeds meer aandacht is voor privacy en veilig gegevensbeheer, moet AI de bescherming van gebruikersgegevens gedurende de hele levenscyclus van het systeem waarborgen. In het algemeen kan het privacy-principe worden gedefinieerd als 'het recht om controle te hebben over de eigen informatie en gegevens', maar in AI-gestuurde omgevingen is de kwestie veel complexer. Het systeem verwerkt immers gegevens van gebruikers en voert daarop opschonings-, samenvoegings- en interpretatiebewerkingen uit.
- **Verantwoordingsplicht.** Betrokken actoren moeten verantwoordelijk zijn voor de beslissingen en acties van het systeem om mogelijke aansprakelijkheidsproblemen tot een minimum te beperken en hebben de plicht om zowel technische als sociale verantwoording af te leggen voor en na de ontwikkeling, implementatie en werking van het systeem.
- **Rechtvaardigheid.** De discriminatie die in sommige beslissingen van AI-systemen wordt aangetroffen, leidt tot ethische vragen omdat deze discriminatie gemeenschappelijke waarden aantasten, waaronder die van de intrinsieke menselijke waardigheid van elke individu waar ook ter wereld. Het verwijderen van vooroordelen uit AI-systemen kan de sociale rechtvaardigheid bevorderen. Dit kan bijvoorbeeld worden bereikt door het besluitvormingsproces expliciet te maken, de transparantie te vergroten en verantwoordelijke entiteiten aan te wijzen.

AI is een methodiek die niet beperkt is tot technologie: hierop gebaseerde systemen zijn noodzakelijkerwijs afhankelijk van menselijke tussenkomst om ze te ontwikkelen, bij te werken, aan te passen en te beheren. Het zijn professionals die ervoor zorgen dat deze systemen goed functioneren, die de datapijplijnen ontwikkelen die nodig zijn om de algoritmen te trainen, en die beveiliging van de informatieomgeving waarborgen. Daarom is het noodzakelijk om te erkennen dat AI niet alleen 'uit technologie bestaat', maar ook een dynamisch geheel is van mensen en actoren die samenwerken aan de werking en ontwikkeling ervan.

De ethische uitdagingen bij het gebruik van de Belbot

De volgende problemen zijn evenwel te constateren met betrekking tot het ethisch gebruik en de ontwikkeling van op AI gebaseerde systemen:

- **Gebrek aan ethische kennis.** Het ontbreekt technisch personeel vaak aan de expertise om de morele en ethische implicaties van AI-systemen te beoordelen.

13 Artificial Intelligence: the global landscape of ethics guidelines, Cornell University: <https://arxiv.org/abs/1906.11668>

- Geen eenduidige ethiek of geen eenduidig mensbeeld. Er zijn geen universeel overeengekomen ethische principes bij medewerkers die een AI-systeem als de Belbot gebruiken en er is geen eenduidig mensbeeld dat aan hun handelen ten grondslag ligt.
- Conflicten in de praktijk. Bedrijven, commissies en groepen die betrokken zijn bij de ontwikkeling van richtlijnen en principes voor AI-ethiek hebben verschillende visies op de praktische implementatie van AI-ethiek. Het Britse Hogerhuis heeft bijvoorbeeld voorgesteld dat robots niet autonoom kunnen werken, maar door mensen moeten worden geleid.
- Gebrek aan technisch inzicht en AI-geletterdheid. Beleidsmakers, filosofen en theologen die zich met AI bemoeien, missen meestal technische expertise: dit kan de ethiek van AI in de praktijk tot een uitdaging maken. Aan de andere kant hebben degenen die AI-systemen ontwikkelen vaak weinig interesse in het aanpakken van ethische aspecten. Het gebrek aan technisch inzicht creëert een kloof tussen systeemontwerp en ethisch denken, waardoor het erg moeilijk wordt om een 'ethics by design' aanpak voor technologische innovatie door te voeren.

Kortom: omdat AI-systemen in staat zijn om sommige menselijke capaciteiten te evenaren of zelfs te overtreffen, is een innovatieve ethische benadering nodig om ervoor te zorgen dat AI op verantwoorde wijze wordt ontwikkeld en gebruikt, met respect voor fundamentele waarden en principes.

Bij de inzet van AI binnen de publieke dienstverlening is het essentieel dat de technologie blijft functioneren binnen de kaders van menselijke waarden en maatschappelijke normen. Dit geldt evenzeer voor de Belbot. AI zelf bezit geen intrinsieke moraliteit; de beslisprincipes op basis waarvan het opereert, zijn numeriek en afhankelijk van de data en het ontwerp van de modellen. Dit betekent echter niet dat AI zonder ethische richting moet functioneren. Door middel van zorgvuldig ontworpen algoritmische waarborgen kunnen ethische kaders worden ingebouwd die ervoor zorgen dat AI-systemen binnen verantwoorde grenzen opereren. Hierover moet in een vroeg stadium al overleg zijn tussen de technici en de ethici en dit overleg moet gaande het proces blijven.

Cruciaal hierbij is het primaat van de mens: AI moet ondersteunend zijn en mag niet de autonomie en het beoordelingsvermogen van de mens ondermijnen. Waar de technologie in staat is om op basis van geavanceerde modellen antwoorden te genereren, blijft het de verantwoordelijkheid van de gebruiker om deze kritisch te wegen. Zoals de theoloog en AI-adviseur van de Verenigde Naties Paolo Benanti stelt, is het vermogen om te twijfelen en de juistheid van gegenereerde antwoorden te beoordelen een fundamentele menselijke eigenschap – een eigenschap die AI zelf ontbeert¹⁴. Dit roept bredere vragen op over de wijze waarop AI binnen een rechtvaardig en inclusief kader kan opereren. De ontwikkeling en implementatie van AI is niet louter een technologisch vraagstuk, maar raakt direct aan machtsverhoudingen, governance en maatschappelijke verantwoordelijkheid. De Belastingdienst moet zich bewust blijven van de implicaties van AI-gebruik binnen de dienstverlening en waarborgen dat deze technologie bijdraagt aan publieke waarden in plaats van deze te ondermijnen.

3.4. Technologische overwegingen

Om te beoordelen of en hoe de Belbot verantwoord kan worden ingezet, zullen in deze paragraaf de technologische overwegingen en de vraag welke objectieve toetsingscriteria gehanteerd kunnen worden centraal staan. Hierbij wordt gekeken naar de specifieke functionaliteiten van de Belbot, waaronder het efficiënter verwerken van inkomende gesprekken en het automatisch genereren van schriftelijke samenvattingen met juridische onderbouwing. Daarnaast wordt stilgestaan bij de technologische keuzes die gemaakt kunnen worden voor een verantwoorde implementatie van de Belbot, zoals het onderscheid tussen een intern ontwikkeld model en een cloudgebaseerde oplossing, en de mogelijke implicaties hiervan voor privacybescherming en informatiebeveiliging. Een ander belangrijk aandachtspunt is de vraag of en in hoeverre de inhoud van AI-ondersteunde gesprekken bewaard dient te worden. Hierbij spelen verschillende belangen een rol, zoals de wens om de dienstverlening te verbeteren en de kwaliteit van adviezen te bewaken, versus de noodzaak om persoonsgegevens te beschermen en de impact van langdurige gegevensopslag te beperken. Daarnaast worden er in dit

¹⁴ <https://www.paolobenanti.com/>

hoofdstuk enkele vragen behandeld waarbij ethische aspecten centraal staan in relatie tot deze technologisch keuzes.

Visie vanuit de Adviescommissie

De Adviescommissie erkent de behoefte bij overheidsdiensten om (meer) geavanceerde technologie in te zetten ter verbetering van de publiek dienstverlening. Publieke belangen moeten leidend zijn bij de implementatie hiervan. De Adviescommissie stelt vast dat het gebruik van grote databestanden, 'machine learning' en algoritmes voor complexe analyses niet weg te denken is uit het overheidshandelen. De technologie is hard nodig om de complexe opgaven van deze tijd aan te kunnen. Tegelijkertijd vraagt die technologie om goed in het vizier te houden wat het doel ervan is: ondersteuning van de principes van goed bestuur. Dat betekent in de eerste plaats dat burgers via de techniek zo goed mogelijk worden bediend en rechtvaardig worden behandeld. In de tweede plaats zijn de processen van overheidsorganisaties van belang. Foutieve modellen, 'long term errors' en verkeerde indicatoren raken niet alleen burgers, maar ook organisaties. Dat kan er bijvoorbeeld toe leiden dat de belastinginning stukt. Tenslotte is het zaak de werkwijze in het vizier te houden volgens welke de technologie wordt ingezet. Ook deze moet beantwoorden aan de regels van behoorlijk bestuur.

Gefaseerd implementeren

De adviesvrager geeft in de adviesvraag aan de Belbot graag te willen inzetten om tekortkomingen in de bestaande processen op te lossen, zoals het valideren en vastleggen van de gegeven antwoorden. Dat is een goed streven. Het blijkt dat rekenkracht een belangrijk criterium is als het aankomt op een goede werking van een GenAI ondersteuningssysteem. De adviesvrager stelt dat bij DF&A momenteel de haalbaarheid van het Belbot idee wordt beproefd in de vorm van een Conceptlab dat als doel heeft om tot een verantwoorde implementatie te komen. Echter zijn de technische middelen die bij de eerste fase van Conceptlab voorhanden waren onvoldoende om tot een gedegen technische analyse te komen. Er is simpelweg te weinig rekenkracht om alle technische functionaliteiten te testen, zeker niet zoals deze geformuleerd zijn in de adviesvraag. Daarnaast lijkt de zoektocht naar een perfecte balans bij het afstellen van het model in de praktijk gepaard te gaan met een hoge mate van complexiteit. Zonder diep in te gaan op de interne werking van een op GenAI gebaseerd systeem wil de Adviescommissie benadrukken dat er sprake is van een fragiele balans in de afstelling ervan. De Belastingdienst wil een zo efficiënt mogelijke dialoogondersteuning en tegelijkertijd de kwaliteit ervan bewaken.

De Adviescommissie adviseert om de implementatie van een dergelijk system gefaseerd te laten plaatsvinden, inclusief evaluatie van de fasen in het proces vanuit technische en ethische invalshoek. Concreet betekent dit om te beginnen met vragen van algemene aard (standaardvragen) zonder het gebruik van persoonsgegevens en in een volgend stadium de koppeling te maken met het gebruik van persoonsgegevens uit verschillende systemen. De praktijk leert dat de balans tussen efficiency en kwaliteit soms lastig te bewaken is en dat dit door middel van veel testen en reflectie tot verbetering kan leiden. Dit proces verdient tijd en dient zorgvuldig te worden doorlopen, waarbij de inhoudelijke fiscale en technische experts van alle domeinen ruimte krijgen om hier een waardevolle bijdrage aan te leveren.

Eén van de aspecten bij het implementeren van een GenAI ondersteuningssysteem is dat hier ook veel rekenkracht bij komt kijken. De keuze voor een bepaalde LLM en de locatie van hosting (intern of extern) heeft gevolgen voor de benodigde capaciteit. Zo zijn er beperkingen aanwezig op het Nederlandse stroomnetwerk en zijn de benodigde technologische componenten (GPU's) misschien niet altijd even tijdig voorhanden.

Objectieve toetsingscriteria bij een verantwoorde implementatie

Bij de vraag welke objectieve toetsingscriteria er zijn om te bepalen of het acceptabel is om servicemedewerkers bij de Belastingtelefoon te laten bijstaan, is de Adviescommissie van mening dat **hoewel meetbare, kwantificeerbare toetsingscriteria een objectieve interpretatie bevorderen, meer subjectieve toetsingscriteria ook heel relevant zijn om een holistisch beeld te geven van de acceptatie.**

De adviesvrager geeft aan dat zij wil voldoen aan randvoorwaarden van betrouwbaarheid, kwaliteitsbewaking, privacywaarborging, personalisatie, uitlegbaarheid & reproduceerbaarheid, open source en energieminimalisatie. De Adviescommissie verwacht dat ook voor de randvoorwaarden objectieve toetsingscriteria gehandhaafd worden, wat niet wordt afgedekt in de beschreven vier Belbot

‘toepassingen’ in de adviesvraag. Daarnaast zou het goed zijn om de randvoorwaarden te bekijken niet alleen als van toepassing op AI, maar op het ‘team’ van mens plus AI. Op deze manier zou worden erkend dat bij het evalueren van de randvoorwaarden ook de rol van de gebruiker (medewerker die de BelBot gebruikt) in ogenschouw genomen moet worden. Bijvoorbeeld: hoe betrouwbaar is de medewerker die de AI opereert? Hoe consistent is het gebruik van de Belbot door de medewerker? Dit zal invloed hebben op de betrouwbaarheid van het totale systeem (AI en medewerker). Zo moet er ook aandacht zijn voor de kwaliteit van de LLM-toepassingen: waar komt het antwoord vandaan, wat is de kwaliteit van het antwoord (hoe zeker kun je zijn van de bruikbaarheid?) en krijgt iedereen hetzelfde terug c.q. krijg je hetzelfde antwoord als je de vraag meerdere keren stelt?

Belbot-toepassing: efficiënter afhandelen

Bij de beantwoording van de adviesvraag heeft de Adviescommissie efficiëntie geïnterpreteerd als de tijdsduur waarmee de medewerker bezig is vragen van belastingplichtigen te beantwoorden. Bijvoorbeeld met het zoeken naar de relevante informatie, waarbij aangenomen wordt dat de hier gewonnen tijd zou leiden tot meer aandacht voor en kwaliteit van service aan de beller. Indien dit overeenkomt met de intentie van de adviesindiener dan zouden de volgende toetsingscriteria toegepast kunnen worden:

- Time to Find: de tijd die de medewerker bezig is met informatie opzoeken tijdens of voorafgaand aan de interactie. Hiervoor moet er een loggingsmechanisme aanwezig zijn in de gebruikte tools.
- Time to Resolve: de totale interactietijd met de gebruiker, onder de aanname dat als het gesprek korter is, er eerder een antwoord op de vraag gegeven kon worden (zie ook laatste alinea).
- Aantal keer dat medewerkers gemiddeld moet escaleren naar een andere/meer ervaren medewerker om een vraag te beantwoorden.
- De gemiddelde wachttijd voor de beller.
- Of een beller meerdere malen belt over hetzelfde probleem.

Belbot-toepassing: samenvatten

Samenvatten is een functionaliteit die de mogelijkheid biedt om de adviezen die de bellers van de Belastingtelefoon ontvangen na het gesprek geautomatiseerd in schriftelijke vorm beschikbaar te stellen, inclusief de juridische onderbouwing van deze adviezen. Voor de helderheid gebruikt de Adviescommissie de term ‘rapport’ om te verwijzen naar de gegenereerde informatie die uit Belbot gegenereerd wordt. Dit rapport wordt nog geëvalueerd en eventueel aangevuld door de medewerker voor definitieve verzending. Indien de Belbot-toepassing *samenvatten* meeneemt zouden de overstaande toetsingscriteria relevant kunnen zijn:

- Tijd die de medewerker kwijt is om het rapport te controleren/herschrijven.
- Vergelijking in tijd en kwaliteit van door medewerkers vervaardigde rapporten en door Belbot en medewerker samen gegenereerde rapporten (neem hierbij bijv. ook vertaalwerk mee).
- Kosten per gegenereerd advies (uitgesplitst in tijd van de medewerker, kosten van analoge of digitale post, energiekosten van het runnen van de BelBot).

Daarnaast zijn er specifieke metrieken die relateren aan de output¹⁵ van het model.¹⁶

Interne of externe hosting

Het maakt uit of de Belbot op een intern LLM wordt gebaseerd of op een cloudoplossing, ook als dit goede waarborgen op het vlak van privacybescherming biedt. Echter, **de Adviescommissie is van mening dat niet alle vragen in deze beantwoord kunnen worden zonder deze los te zien van het standpunt van de politiek over AI gerelateerde zaken en dat er bovendien een duidelijk standpunt van het ministerie van Financiën nodig is.** De inzet van een intern LLM wordt geprefereerd, zoals dat ook tot uiting komt in de Kamerbrief van de staatssecretaris van het ministerie BZK d.d. 11 december 2023 met het ‘*Voorlopig standpunt voor Rijksorganisaties bij het gebruik van generatieve AI*’. Het toepassen van interne LLM-modellen wordt geprefereerd omdat hier betere waarborgen zijn te geven m.b.t. data-privacy en transparantie m.b.t. de gebruikte databronnen, maar definitief is deze keuze nog altijd niet. De

15 Output: in dit geval het resultaat gegenereerd door de BelBot, wat gebruikt wordt ter ondersteuning van de medewerker.

16 Zie verder bij ‘Appendix C: Technische toelichting Adviesvraag’

mogelijkheid tot een externe of een *on premise* hosting van een LLM behoort voorsnog tot de mogelijkheden zolang de politiek hier geen definitief standpunt over heeft ingenomen.

Het is volgens de Adviescommissie een opgave voor het ministerie van Financiën om te bepalen of de soevereiniteit en de doelstellingen van de organisatie beter gediend zijn bij een LLM in de cloud (van een externe aanbieder) of bij een intern LLM. Ook relevant is om na te denken over de data-soevereiniteit en de privacy van de burgers als data (in dit geval conversaties over belastingaangiften) gedeeld worden met een externe partij. Welke afspraken worden er gemaakt? Worden de data door de commerciële partij gebruikt om AI te trainen (denk aan vergelijkbare voorbeelden zoals het trainen van AI met behulp van wetenschappelijke artikelen zonder enige toestemming van de auteurs). Als een cloudgebaseerde oplossing toch wordt overwogen, moeten zeer strenge waarborgen aanwezig zijn, waaronder aantoonbare naleving van de AVG, onder meer door een strikt toegangsbeheer, volledige transparantie over de gegevensverwerking en specifieke afspraken over de dataretentie. Onafhankelijk van een externe of interne cloud-oplossing, is het relevant om na te denken hoe afhankelijk medewerkers van de Belbot worden in het geval dat dit systeem uitvalt. Recent heeft de AP in haar Sectorbeeld Overheid¹⁷ expliciet gewezen op het risico van (te) grote afhankelijkheid van een specifieke IT-dienstleverancier. De keuze voor een intern LLM is meer in lijn met de Overheidsbrede visie op de inzet van een GenAI ondersteuningssysteem dat veiligheid, rechtvaardigheid, menselijke autonomie en privacy centraal stelt. De Belbot kan hierdoor meer verantwoord en effectiever worden ingezet, zonder concessies te doen aan de kernwaarden van de dienstverlening door de Belastingdienst. Het verdient daarbij de voorkeur dat de cloudoplossing op Europese servers wordt gehost.

Door bij de verdere implementatie en ontwikkeling van een GenAI ondersteuningssysteem te kiezen voor een intern en open source LLM, houdt de adviesvrager volgens de Adviescommissie meer controle op de data- en privacybescherming, omdat alle gegevens binnen de eigen infrastructuur van de Belastingdienst blijven. Het is een keuze die goed uit te leggen valt en past in de kaders en richtlijnen die opgesteld zijn voor een verantwoord gebruik van AI. Het gebruik van een interne (gehost) LLM vermindert het risico op datalekken en biedt betere mogelijkheden voor compliance met de AVG en het IAMA-framework. Interne modellen kunnen specifiek worden afgestemd op de behoeften van de Belastingdienst, met gerichte training op domeinspecifieke kennis en regels, hetgeen de betrouwbaarheid en uitlegbaarheid vergroot. Hoewel een gelicenseerde cloudoplossing privacy waarborgen kan hebben, blijft er een grotere afhankelijkheid van de derde partij die het model beheert. Dit kan leiden tot risico's zoals onvoorziene datatoegang door externe partijen, complexere naleving van wetgeving en beperktere controle over de gebruikte algoritmen en trainingsdata.

Bewaren van historische gespreksinformatie

Bij de beoordeling van de wenselijkheid om de inhoud van de AI-ondersteunde gesprekken te bewaren voor analyse doelen is de Adviescommissie van mening dat het recente onderzoek naar de hallucinaties van Whisper heeft aangetoond hoe cruciaal het bewaren van de originele gesprekken kan zijn alsmede dat dit de transparantie van een betrouwbare overheid kan bevorderen. De Adviescommissie staat positief tegenover de mogelijke keuze dat de inhoud van de AI-ondersteunde gesprekken bewaard blijft.

AI-geletterdheid

De Adviescommissie ziet het niveau van AI-geletterdheid als één van de hoekstenen van een verantwoord gebruik van een GenAI ondersteuningssysteem. De recente AI Act van het Europees Parlement en het advies van de AP omtrent AI-geletterdheid ondersteunen dit¹⁸. De AI Act stelt duidelijk dat afhankelijk van de rol van de organisatie en het gebruik van het soort AI er per risiconiveau de juiste geletterdheid aanwezig dient te zijn bij de gebruikers. De opkomst van AI en het gebruik van chatbots binnen de publieke sector brengt niet alleen nieuwe mogelijkheden met zich mee, maar ook een toenemende verantwoordelijkheid voor organisaties en medewerkers om AI-systemen op een verantwoorde en doordachte manier te benutten. AI-geletterdheid is daarbij een essentiële voorwaarde. Het gaat niet alleen om kennis van de technologie zelf, maar ook om inzicht in de ethische en juridische

17 Autoriteit Persoonsgegevens, Sectorbeeld Overheid, oktober 2024, p. 5.

18 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689> en <https://www.autoriteitpersoonsgegevens.nl/documenten/aan-de-slag-met-ai-geletterdheid>

kaders, de risico's van bias en de impact op burgers. Overheidsorganisaties die AI inzetten, moeten ervoor zorgen dat medewerkers in staat zijn om AI-ondersteunde systemen kritisch te beoordelen, resultaten op waarde te schatten en bij te sturen waar nodig. Zonder deze geletterdheid bestaat het risico dat AI wordt ingezet zonder voldoende begrip van de implicaties, wat kan leiden tot onbedoelde effecten op besluitvorming en dienstverlening.

Het belang van AI-geletterdheid strekt zich verder uit dan individuele medewerkers; het raakt aan bredere governance-vraagstukken en de borging van publieke waarden. Een goed begrip van AI draagt bij aan transparantie, neemt de menselijke maat in acht, zodat de hulp vragende burger zich werkelijk gekend weet en zorgt ervoor dat AI-toepassingen binnen de kaders van behoorlijk bestuur blijven. Dit vraagt om continue ontwikkeling van kennis en vaardigheden, niet alleen op technisch niveau, maar ook op juridisch en ethisch vlak.

De Adviescommissie is van mening dat elke medewerker, in dienst dan wel inhuur, die een geavanceerde datatoepassing inzet, zich ervan bewust moet zijn dat daar risico's aan kleven.

Geavanceerde systemen vragen om aangepaste werkprocessen waarin ruimte is voor reflectie en toetsing. Voor momenten dus waarop medewerkers zichzelf en elkaar vragen stellen als: waarom doen we wat we doen, wat zijn onze drijfveren? Hoe verhoudt onze doelstelling om de burger de verschuldigde belasting te laten betalen tot het in acht nemen van diens unieke persoonlijke omstandigheden en van de factoren die het deze burger moeilijk maakt de belasting te voldoen? Valt wat we doen binnen wettelijke kaders, is het ethisch en zijn de uitkomsten gewenst? Hoe borgen we het toezicht erop?

Eén van de beoogde effecten van de adviezen van de Adviescommissie is om een cultuurverandering mede tot stand te brengen, zodat medewerkers meer handelingsbekwaam worden in relatie tot geavanceerde datatoepassingen en zich vrij voelen om aan de bel te trekken als er toch iets misgaat. En dat het vervolgens mogelijk is om de uitkomsten of werkprocessen aan te passen. Hiervoor is enerzijds voldoende kennis van zaken bij de medewerkers nodig. Anderzijds vraagt het om een veilige sfeer, gecreëerd door leidinggevenden die niet bang zijn voor statusverlies, die besluitvormers durven tegen te spreken en die hun medewerkers vertrouwen en beleidsvrijheid geven. Een manier om dit te realiseren is het trainen van de servicemedewerkers die met een AI-systeem als de Belbot moeten werken. Het is de verantwoordelijkheid van de organisatie om de juiste trainings- en opleidingsmiddelen aan de servicemedewerkers aan te reiken en ervoor te zorgen dat het kennisniveau op pijl blijft.

Belbot-toepassing: trainen

Een andere Belbot-toepassing is het trainen van nieuwe servicemedewerkers, die via het dialoogsysteem leren welke informatie ze moeten uitvragen en hoe ze hiermee tot een juist antwoord kunnen komen. Daarnaast kan de Belbot worden gebruikt voor de verbreding van de inzetbaarheid van bestaande servicemedewerkers.

De Adviescommissie komt tot het volgende advies met betrekking tot welke objectieve toetsingscriteria er zijn om bij te dragen aan de implementatie van deze toepassing:

- Percentage van medewerkers dat de training heeft gevolgd en succesvol heeft afgerond.
- Bijhouden van een classificatie van de problemen waar bellers mee bellen door middel van 'tags'. Gemeten kan worden of gemiddeld genomen medewerkers in staat zijn om vragen van diverse categorieën te beantwoorden.
- Aantal keer dat medewerkers gemiddeld moet escaleren naar een meer ervaren medewerker om een vraag te beantwoorden. In welke scenario's is ervaring belangrijk?
- Voor het meten van AI *literacy* zou een gevalideerde schaal gebruikt kunnen worden.

3.5. Het gebruik van de discretionaire ruimte en biaswaakzaamheid

De integratie van (Gen)AI binnen de dienstverlening van de Belastingtelefoon roept fundamentele vragen op over verantwoordelijkheid, autonomie en gegevensverwerking. Hoewel AI-ondersteunde systemen zoals de Belbot de efficiëntie en nauwkeurigheid van de dienstverlening kunnen verbeteren, is het cruciaal om de ethische, juridische en praktische implicaties zorgvuldig af te wegen. Deze paragraaf richt zich op de kernvragen die bepalen onder welke voorwaarden de inzet van AI bij de

Belastingtelefoon als verantwoord en acceptabel kan worden beschouwd. Het biedt een analyse van ethische en beleidsmatige vraagstukken.

Een belangrijk uitgangspunt hierbij is de verantwoordelijkheid voor de antwoorden die aan burgers worden verstrekt. In een traditionele situatie ligt deze verantwoordelijkheid bij de servicemedewerker, maar in een AI-ondersteund gesprek kan deze dynamiek veranderen. De vraag is dan ook of en hoe deze verschuiving in verantwoordelijkheden invloed heeft op de betrouwbaarheid en juridische houdbaarheid van de verstrekte informatie.

Daarnaast spelen keuzevrijheid en autonomie van medewerkers een rol. Moet het gebruik van AI-ondersteuning verplicht worden gesteld of mag de servicemedewerker zelf bepalen of hij hiervan gebruikmaakt, zelfs als dit ten koste gaat van de efficiëntie? Ook de reikwijdte van de Belbot is een essentieel punt van overweging: moet de AI uitsluitend algemene fiscale vragen beantwoorden, of mag het systeem ook persoonlijke vragen behandelen op basis van eerder geregistreerde klantgegevens?

Ten slotte komt het raadplegen en verwerken van persoonsgegevens aan bod. In hoeverre mag de Belbot servicemedewerkers toegang geven tot persoonlijke gegevens van de beller om een beter antwoord te formuleren? En onder welke voorwaarden zou de Belbot gegevens van een andere belastingplichtige mogen inzien, bijvoorbeeld in het geval van een nabestaande of een belastingadviseur die namens een cliënt belt?

Visie vanuit de Adviescommissie

Discretionaire ruimte verwijst naar de beoordelingsvrijheid die ambtenaren binnen de Belastingdienst hebben bij de toepassing van wet- en regelgeving in individuele gevallen. Deze ruimte is van belang om recht te doen aan de menselijke maat en om besluiten aan te laten sluiten bij de specifieke omstandigheden van burgers en bedrijven. Tegelijkertijd vraagt het om zorgvuldige afwegingen binnen duidelijke kaders, zodat beslissingen transparant, uitlegbaar en consistent blijven.

De inzet van AI en data-analyse kan ondersteunend zijn bij het structureren van keuzes binnen deze ruimte, maar de Adviescommissie is van mening dat dit nooit de plaats mag innemen van een inhoudelijke en proportionele afweging door de servicemedewerker van de Belbot. Dit vraagt om een goed doordachte balans tussen automatisering en maatwerk, waarbij technologische ondersteuning niet leidt tot een verschraving van discretionaire bevoegdheden, maar juist bijdraagt aan een rechtvaardige en evenwichtige uitvoering van beleid. Cruciaal is ook de beleving van de professionele ruimte bij de servicemedewerkers van de Belbot. Kunnen zij bijvoorbeeld afwijken van de gegeven antwoorden door de Belbot?

Verantwoordelijkheden bij de inzet van de Belbot

Wanneer er wordt gekeken naar wie er uiteindelijk verantwoordelijk zou moeten zijn voor de gegeven antwoorden en of het hierbij uitmaakt of het om een traditioneel gesprek gaat of een AI-ondersteund gesprek, is het belangrijk om onderscheid te maken tussen verschillende types verantwoordelijkheden. Uiteindelijk werken de Belastingdienst en haar ambtenaren onder ministeriële verantwoordelijkheid. De ministers zijn (politiek) verantwoordelijk voor wetgeving en beleid. Het handelen van ambtenaren moet daarom te allen tijde legitiem zijn. Met andere woorden, het handelen moet gefundeerd zijn in wet -en regelgeving, waarbij de beginselen van behoorlijk bestuur een belangrijke leidraad vormen. Aangezien ambtenaren de uitvoering van wet -en regelgeving mede vormgeven, hebben zij in die zin ook een medeverantwoordelijkheid voor het realiseren van een goed bestuur. Deze verschillende maar complementaire verantwoordelijkheidsvormen veranderen niet door het inzetten van een AI-systeem.

Wat wel verandert is de wijze waarop deze verschillende verantwoordelijkheden betekenisvol zijn en blijven wanneer AI een rol speelt in de uitvoeringsprocessen. Zowel de minister als de ambtenaren moeten voor alles het principe 'uitlegbaarheid' respecteren en toepassen. Het parlement kan alleen haar taak uitoefenen en grip op het openbaar bestuur behouden als de minister kan uitleggen hoe dat bestuur onder zijn/haar verantwoordelijkheid tot stand komt. **Voor ambtenaren, bijvoorbeeld zij die met de Belbot werken, is die uitlegbaarheid van belang in relatie tot de burger, die het recht heeft om een toelichting te verkrijgen bij een beslissing of advies dat die burger aangaat** (zie bijvoorbeeld Art. 22

AVG). AI-systemen kunnen deze beide vormen van uitlegbaarheid bemoeilijken, omdat door technische complexiteit niet (voldoende) te achterhalen valt hoe bepaalde uitkomsten tot stand zijn gekomen.

Betekenis geven aan medeverantwoordelijkheid van de ambtenaar

Als we ervan uitgaan dat uitlegbaarheid een cruciaal aspect is van de invulling van betekenisvolle medeverantwoordelijkheid van de ambtenaar die werkt met de Belbot, waar dient die uitlegbaarheid dan aan te voldoen?

- Ten eerste dient de uitlegbaarheid het loutere descriptieve te overstijgen, dus niet “dit is wat er uit het systeem komt”. Om iets uit te kunnen leggen, moet je begrijpen hoe het tot stand komt. In het geval van de Belbot: er moeten redenen gegeven kunnen worden waarom tot een bepaald advies is gekomen op basis van de geanalyseerde data. Dit vergt van ambtenaren dat zij naast vakinhoudelijke kennis, ook kennis hebben van het functioneren van de Belbot.
- Ten tweede dienen die redenen inzichtelijk te zijn voor de burger aan de andere kant van de lijn. Dit vraagt van de ambtenaar het vermogen om te verklaren waarom en hoe de algemene regels van toepassing zijn op de specifieke omstandigheden van de burger in een voor de burger begrijpelijke taal.
- Ten derde moet de uitlegbaarheid de burger in staat stellen, wanneer zij het niet eens is met het advies of de beslissing, er (naderhand) bezwaar tegen aan te tekenen. Met andere woorden, in de gevallen “computer says no” of wanneer de medewerker een antwoord van de Belbot niet overneemt, dan blijft de servicemedewerker (gedelegeerd) verantwoordelijk voor het gegeven antwoord en voert deze verantwoordelijkheid via de organisatie terug naar de minister. Het antwoord dat uit de Belbot komt, maar ook de onderbouwing van dit antwoord, dient voor burgers opvraagbaar te zijn als het bepalend is voor de fiscale keuzes van die burger. De ambtenaar die met de Belbot werkt of een andere ambtenaar aan wie deze taak is toebedeeld, moet de expertise en het mandaat hebben af te wijken van wat de Belbot als antwoord adviseert, mits daar goede redenen voor zijn. Deze redenen kunnen, naast bijvoorbeeld het detecteren van een fout in het antwoord van de Belbot, ook gebaseerd zijn op nieuwe informatie aangedragen door de burger.

Ontwikkel een Belbot die begrijpelijk is voor de medewerker die ermee moet werken

Welke maatregelen dienen getroffen te worden om de uitlegbaarheid die nodig is voor betekenisvolle medeverantwoordelijkheid te borgen? In de AI Act wordt veel aandacht besteed aan de ontwikkelfase van de AI-applicatie. In deze ontwikkelfase wordt namelijk in grote mate bepaald of de gebruiker (in dit geval de servicemedewerker die met de Belbot werkt) in staat wordt gesteld betekenisvolle medeverantwoordelijkheid te dragen. Zo moet de aanbieder van een AI-systeem ervoor zorgen dat gebruikers toezicht kunnen houden op het systeem en interveniëren wanneer nodig. Er moet voldoende informatie voorzien worden over de wijze waarop het systeem functioneert, over waar het systeem sterk in is en onder welke omstandigheden het mogelijk de mist ingaat.

Aangezien de Belbot als in-house applicatie wordt ontwikkeld, is er zowel technische als vakinhoudelijke kennis aanwezig om de applicatie te ontwerpen op zo een manier dat aan deze eisen voldaan kan worden. Begin hiervoor niet helemaal bij nul, maar kijk naar de huidige praktijk en zet als doel de Belbot zoveel mogelijk te laten aansluiten bij de huidige manier van werken. Zo kan je gebruik maken van de bestaande protocollen en *checks and balances* die hun nut reeds hebben bewezen. Alleen waar dat niet mogelijk is, bijvoorbeeld omdat er nieuwe risico's ontstaan of omdat mitigerende maatregelen niet naar de Belbot-situatie omgezet kunnen worden, moeten nieuwe checks en balances ontwikkeld worden.

Zorg voor een adequate governance structuur

Lazcoz en de Hert (2023)¹⁹ beargumenteren dat om op een verantwoorde wijze gebruik te maken van AI-applicaties, betekenisvol toezicht van mensen noodzakelijk maar niet voldoende is. Zelfs goed onderwezen en ervaren ambtenaren kunnen ten prooi vallen aan *confirmation bias* (het interpreteren van informatie op zodanige wijze dat het de eigen overtuiging ondersteunt) en *automation bias* (de neiging om de output van een systeem als waarheid aan te nemen, zelfs als er goede redenen zijn om

19 Lazcoz, G., & De Hert, P. (2023). Humans in the GDPR and AIA governance of automated and algorithmic systems. Essential prerequisites against abdicating responsibilities. *Computer Law & Security Review*, 50, 105833.

dat niet te doen). Daarnaast kan het systeem natuurlijk ook onderpresteren, bijvoorbeeld omdat het in de praktijk minder goed werkt bij bepaalde groepen burgers of omdat het niet voldoende is getraind op veranderende omstandigheden in de buitenwereld.

Er moet dan ook een governance structuur zijn die toezicht houdt op de processen waarin mens en AI samenwerken. Hierbij wordt aangeraden om te kijken hoe dit toezicht zoveel mogelijk opgenomen kan worden in de bestaande *governance* sfeer. Schuw hierbij niet gedetailleerd vast te leggen hoe de taakverdeling tussen gebruiker en systeem is geregeld en wanneer en hoe medewerkers bij gebruik van de Belbot hun discretionaire ruimte in kunnen vullen. Het uitvoeren van een DPIA kan een goed middel zijn om hier beter inzicht in te verkrijgen (Lazcoz en de Hert 2023).

Geïntensiveerd toezicht Autoriteit Persoonsgegevens

Er is bekendgemaakt dat de Belastingdienst vijf jaar onder geïntensiveerd toezicht van de AP zal staan, mede naar aanleiding van recente berichten over de inzet van Risico Analyse Modellen. Dit toezicht houdt onder meer in dat begeleiding wordt geboden bij het duurzaam verbeteren van de bescherming van persoonsgegevens. **Nu de AP ook aangewezen is als de algoritmetoezicht-houder raden wij de adviesvrager aan om dit toezicht te zien als een kans om actief met de AP in contact te treden over onder meer de juridische aspecten van de Belbot.** De Belastingdienst kan gebruik maken van het toezichtpalet dat middels dit geïntensiveerde toezicht aangeboden wordt om in gesprek te gaan over aspecten waar nu nog onduidelijkheid over bestaat.

Het is de mening van de Adviescommissie dat het aan de Belastingdienst zelf is om te bepalen hoe hun medewerkers technologie als de Belbot inzetten. Omdat de Belbot wellicht ook goed kan documenteren lijkt het zinvol om de technologie breed in te zetten en zo te voorkomen dat er kwalitatief verschillende dossiers ontstaan. Relevant is dat de dienst voldoende inspanning verricht om medewerkers effectief te trainen in het gebruik van de technologie. De vraag raakt de kop “Engagement” in het WRR Rapport Opgave AI. De organisatie zou de medewerker moeten meenemen bij de ontwikkeling en implementatie van de Belbot en op die manier ook een community van expert users opbouwen die de Belbot helpen verbeteren en ontwikkelen. Dat kan de medewerkers ondersteunen in het effectief gebruik maken van de technologie.

De vraag of de Belbot alleen algemene vragen mag beantwoorden of ook persoonlijke fiscale vragen, valt terug op het advies dat eerder gegeven is in dit adviesrapport, namelijk om voorzichtig te werk te gaan en de implementatie van een GenAI ondersteuningssysteem gefaseerd te laten plaatsvinden. Concreet betekent dit om eerst te beginnen met vragen van algemene aard (standaardvragen) zonder het gebruik van persoonsgegevens en pas daarna de koppeling te maken met het gebruik van persoonsgegevens uit verschillende systemen, indien de resultaten van de inzet naar verwachting zijn. In een later stadium van de implementatie zou geëxperimenteerd moeten worden met de inzet van de Belbot zodat servicemedewerker de mogelijkheid krijgt om via de Belbot persoonlijke gegevens van de beller te raadplegen die volgens het systeem nodig zijn voor de beantwoording van de vraag. Het verwerken van persoonlijke gegevens is inherent aan de processen die de medewerker uitvoert. Maar hierbij moet altijd gelden dat de juiste waarborgen genomen zijn (zoals het controleren van een juiste werking van de Belbot).

Of het systeem ook de mogelijkheid onder voorwaarden moet bieden om informatie van een andere belastingplichtige te raadplegen (bv. bij nabestaande of belastingadviseur) kan *a priori* moeilijk worden beantwoord omdat dit vooruit loopt op toekomstige operationele mogelijkheden.

(No) Human in the loop

Uit de adviesvraag blijkt dat de adviesvrager overweegt om een GenAI ondersteuningssysteem autonoom te laten opereren bij goed afgebakende taken. Dit betreft functionaliteiten waarbij de Belbot zelfstandig voorbereidende handelingen uitvoert, zoals het opzoeken van relevante gegevens op basis van een ingevoerd BSN-nummer of het geautomatiseerd doorverbinden naar de juiste medewerker op basis van een ingesproken vraag. Het autonoom opereren van de Belbot roept echter verschillende ethische en praktische vragen op. Enerzijds kan een dergelijke toepassing de dienstverlening versnellen en de werklast voor servicemedewerkers verlichten door hen beter voorbereid het gesprek te laten starten. Anderzijds moeten er duidelijke kaders worden vastgesteld om te waarborgen dat de chatbot

geen beslissingen neemt die buiten zijn mandaat vallen en dat de privacy- en gegevensbescherming van burgers gewaarborgd blijven.

Een belangrijke overweging is de mate van autonomie die de chatbot mag hebben bij het verzamelen en interpreteren van gegevens. In hoeverre kan en mag een op GenAI gebaseerd systeem voor dialoogondersteuning zelfstandig gegevens ophalen en verwerken voordat een menselijke medewerker betrokken wordt? Daarnaast speelt de betrouwbaarheid van spraakherkenning en kennisbanken een cruciale rol. Het correct koppelen van vragen aan onderwerpen en het doorverbinden naar de juiste afdeling vereist een nauwkeurige werking van de onderliggende technologie om misinterpretatie en foutieve doorverwijzingen te voorkomen. Deze technologie is (nog) niet onfeilbaar.

3.6. Autonomie en afgebakende taken

Wat zijn de randvoorwaarden en ethische afwegingen rondom het autonoom opereren van de Belbot binnen goed gedefinieerde taken? Onderzocht wordt onder welke voorwaarden een dergelijke toepassing verantwoord kan worden ingezet, welke waarborgen nodig zijn om de betrouwbaarheid te garanderen en hoe de interactie tussen AI en menselijke medewerkers optimaal kan worden vormgegeven.

Visie van de Adviescommissie

De Adviescommissie acht het acceptabel als de Belbot autonoom opereert bij goed afgebakende taken. Hierbij moet wel worden nagedacht over bijvoorbeeld een controlevraag door de medewerker (e.g., laatste cijfer of korte herhaling). Het zou kunnen dat de Belbot-functionaliteit minder goed werkt voor een bepaalde groep bellers (bijvoorbeeld oudere mensen die minder verstaanbaar praten, of mensen die in een luidruchtige omgeving zijn of wonen). Daarnaast is het van belang dat er sprake is van domeinexpertise bij de medewerkers van de Belastingdienst, zodat zij in staat zijn om de beoordeling te maken wanneer je wel of niet een GenAI-systeem inzet. Als gesteld moet er ruimte blijven om af te wijken van het advies van de Belbot en dat kan het beste bereikt worden door de medewerkers voldoende professionele ruimte te bieden. **Het is belangrijk om van tevoren over deze kwetsbaarheden na te denken en mitigaties te bedenken (bijvoorbeeld door deze groepen al vroeg in het menu een controlevraag te stellen, waardoor bijvoorbeeld de audiokwaliteit vastgesteld kan worden).**

Autonomie

Hierbij dient een noot geplaatst te worden dat dergelijke inzet een opmaat zou kunnen zijn tot een toekomst waarin de Belbot bepaalde vragen compleet autonoom afhandelt, zonder menselijke tussenkomst, of dat er alleen dan geëscaleerd wordt naar een medewerker wanneer de Belbot geen of geen passend antwoord kan geven. In een dergelijke situatie kan worden voorzien dat er een kloof ontstaat waarbij alleen de meer digitaal geletterden of volhardende bellers menselijke service zouden kunnen krijgen. Er moet te allen tijde een duidelijk pad zijn om een menselijke medewerker te kunnen spreken, zodat het nooit zo kan zijn dat menselijke hulp en service slechts voor bepaalde groepen is weggelegd.

Er is bij het gebruik van de Belbot door de medewerkers van de Belastingdienst een bepaalde scepsis nodig om de resultaten (output) van de Belbot niet als “foutloos” en “zonder vooroordeel” te lezen. Net als bij het advies dat de Adviescommissie heeft uitgebracht rondom de bias- en fairnesstoetsing aan DF&A, is hier ook essentieel dat de “human in the loop” niet slechts symbolisch is, maar competent genoeg om mogelijke (door het algoritme veroorzaakte) discriminatie of onjuistheden in de beantwoording van de vragen te herkennen. Daarnaast dient deze ‘human in the loop’ over de capaciteit en het mandaat te beschikken om adequaat te kunnen reageren en moet deze ondersteund worden door processen om dit mogelijk (en bij voorkeur makkelijk) te maken. Algemene vaardigheden op het gebied van data en AI literacy, ethiek, en aandacht voor de sociale impact zijn hiervoor uiterst relevant.

Afgebakende taken

Er zijn diverse ontwikkelingen op het gebied van AI, zoals onder andere de inzet van *agentic* AI of Chain-of-Retrieval Augmented Generation²⁰ waarbij AI *agents* een goed afgebakende taak kunnen uitvoeren.

20 Chain-of-Retrieval Augmented Generation, Cornell University: <https://arxiv.org/abs/2501.14342>

Dat zegt weinig over of het daadwerkelijk een goed afgebakende taak is en het zegt weinig over de kwaliteit van de uitkomst. Het gevaar is dat deze prestaties van een GenAI ondersteuningssysteem die afgebakende taken naar alle waarschijnlijkheid goed uitvoert, tot een generieke goedkeuring van praktijken kan leiden, waarbij de kwaliteit voor bredere toepassingen niet gegarandeerd zijn. Als alleen de individuele en afgebakende taken worden beoordeeld op prestaties en juistheid, hoe zit het dan met het complete advies dat afgegeven wordt door de servicemedewerker die gebruik maakt van een GenAI ondersteuningssysteem? Er dient daarom niet alleen op goed afgebakende taken, maar ook op het geheel bereidheid te zijn bij de Belastingdienst om het functioneren van de Belbot te monitoren en te corrigeren indien nodig.

De Adviescommissie is kritisch over het woord ‘afgebakend’ omdat hierbij de mogelijkheid kan bestaan dat dit zou kunnen leiden tot een *accountability sink*. *Accountability sinks* zijn structuren die de gevolgen van een beslissing absorberen of verhullen, zodat niemand er rechtstreeks verantwoordelijk voor kan worden gehouden. Op het moment dat er sprake is van een goed functionerende Belbot, beoordeeld op losse en afgebakende taken, waar de fiscale dienstverlening niet als geheel als product gezien wordt, kan er sprake zijn van het verschuiven van de verantwoordelijkheid van het afgegeven fiscale advies. De crux in deze vraag is het woord ‘afgebakend’ dat een flexibele en rekbare definitie kan hebben. Een slechts beknopt en positief antwoord van de Adviescommissie zou hier de zorgvuldigheid kunnen ondermijnen. Het principiële antwoord van de Adviescommissie is dan ook dat de Belbot zeker ingezet kan worden voor afgebakende taken, maar dat deze inzet wel goed gewaarborgd moet zijn en dat duidelijk gedefinieerd is welke vragen de Belbot zelfstandig zou mogen beantwoorden en welke vragen de servicemedewerker voor zijn of haar rekening neemt. De Adviescommissie adviseert daarom om het begrip ‘afgebakend’ nader te definiëren met oog voor de bewezen functionaliteit van de Belbot, het iteratieve proces van validatie en mogelijk nodige correcties.

3.7. Kaders en regelgeving

Deze paragraaf schetst de belangrijkste kaders en regelgeving die relevant zijn bij de inzet van chatbots in de publieke sector. Daarbij wordt ingegaan op de juridische verplichtingen, ethische randvoorwaarden en de stappen die nodig zijn om AI op een verantwoorde manier te implementeren binnen een overheidsorganisatie. Hoewel deze technologie de dienstverlening efficiënter en toegankelijker kan maken, moet de implementatie zorgvuldig worden afgestemd op bestaande wettelijke en ethische kaders. Overheidsorganisaties dienen bij de introductie van AI-gestuurde chatbots niet alleen rekening te houden met technologische en operationele aspecten, maar ook met de geldende regelgeving en maatschappelijke waarden.

Een van de belangrijkste juridische kaders op dit gebied is de Verordening Kunstmatige Intelligentie (AI Act) van de Europese Commissie. Deze wetgeving classificeert AI-systemen op basis van risiconiveaus en stelt strikte eisen aan systemen die in de publieke sector worden ingezet. Chatbots die in direct contact staan met burgers kunnen onder deze regelgeving vallen, wat betekent dat transparantie, uitlegbaarheid en toezicht op de werking van het systeem gewaarborgd moeten worden.

Daarnaast is de Algemene verordening gegevensbescherming (AVG) een cruciale leidraad bij het ontwikkelen en implementeren van chatbots. Omdat dergelijke systemen vaak persoonsgegevens verwerken – zoals naam, BSN of fiscale gegevens – moeten overheidsinstanties zorgen voor een rechtmatige grondslag voor verwerking, minimale gegevensverwerking en adequate beveiligingsmaatregelen. Het principe van privacy-by-design en by default moet worden toegepast om de rechten van burgers te beschermen.

Naast juridische normen spelen ook ethische uitgangspunten een grote rol. Het gebruik van AI binnen de overheid moet transparant, controleerbaar en inclusief zijn. Dit betekent dat burgers moeten weten wanneer ze met een AI-gestuurde chatbot communiceren, dat er altijd een mogelijkheid moet zijn om over te schakelen naar menselijke ondersteuning en dat er mechanismen aanwezig zijn om eventuele fouten of bias in de chatbot tijdig te signaleren en corrigeren.

De Adviescommissie is positief over de initiatieven die al door de adviesvrager zijn ondernomen en het voornemen om bij overheidstoepassingen van GenAI, zoals de Belbot, een toetsing te doen op basis van de geldende wetgeving en beschikbare instrumenten. Hiervoor is het gebruik van het IAMA-framework

en de toetsing aan de grondrechten van betrokkenen een goed middel. Aanvullend is er een voorlopige richtlijn die richting geeft aan de inzet van GenAI bij overheidsdiensten en zijn binnen twee jaar bedrijven en overheidsinstellingen ook verplicht om high-risk AI-toepassingen aan de AI Act te toetsen. Ter aanvulling op bovenstaande voert de adviesvrager ook eigen analyses uit en wisselt actief kennis uit in interne en interdepartementale werkgroepen rond de visievorming op generatieve AI.

De Adviescommissie moedigt al deze initiatieven ten zeerste aan. Daarbij wil de Adviescommissie de adviesvrager ook attenderen op het feit dat er gebruik gemaakt kan worden van De Ethische Data Assistent (DEDA²¹) om in onderlinge samenwerking met fiscale specialisten, techneuten en andere belanghebbende partijen ethische problemen in de Belbot te herkennen. Bij DF&A is er intern ook een Ethisch Assessment ontwikkeld dat kan helpen bij het beoordelen van de inzet van algoritme in relatie tot de Belbot. De Adviescommissie adviseert daarom ook om gebruik te maken van interne toetsingen om tot een goed afgewogen kader te komen bij de implementatie van de Belbot.

21 <https://deda.dataschool.nl/>

Appendix A: Overzicht vragen uit advies

1. Hoofdvraag

Is het acceptabel om servicemedewerkers bij de Belastingtelefoon te laten bijstaan door een op Generatieve AI gebaseerd systeem voor dialoogondersteuning?

- a) Welke objectieve toetsingscriteria zijn bij deze afweging van belang? We verzoeken de adviescommissie om dit voor onderstaande vier Belbot-toepassingen toe te lichten.
 - i. Het efficiënter afhandelen van een inkomend gesprek door automatisch opzoeken, combineren en interpreteren van gegevens en fiscale wetten en regelingen, zodat de servicemedewerker minder tijd kwijt is aan informatie verzamelen en meer aandacht kan besteden aan de duiding van de vraag en de interactie met de beller.
 - ii. Het consistentere en meer gepersonaliseerd beantwoorden van de vragen die binnenkomen bij de Belastingtelefoon, doordat de Belbot gebruik maakt van een gevalideerde kennisbank en (desgewenst) de al beschikbare gegevens uit het klantdossier, en in staat zou zijn om op basis hiervan juiste, begrijpelijke en gepersonaliseerde antwoorden te produceren.
 - iii. Het realiseren van de mogelijkheid om de adviezen die de bellers van de Belastingtelefoon van een servicemedewerker ontvangen na het gesprek geautomatiseerd in schriftelijke vorm beschikbaar te stellen, inclusief de juridische onderbouwing van deze adviezen (door de gebruikte gegevens en fiscale regels te vermelden).
 - iv. Het trainen van nieuwe servicemedewerkers, die via het dialoogsysteem leren welke informatie ze moeten uitvragen en hoe ze hiermee tot een juist antwoord kunnen komen, maar ook voor de verbreding van de inzetbaarheid van bestaande servicemedewerkers.
- b) Hoe verhoudt de inzet van Generatieve AI, en meer specifiek ondersteuning van servicemedewerkers door de Belbot, zich tot de kernwaarden bij de dienstverlening door medewerkers van de Belastingdienst (verantwoordelijkheid, geloofwaardigheid, zorgvuldigheid)? Welke voorwaarden moeten specifiek aandacht krijgen om deze techniek te kunnen benutten?²²
- c) Wat betekent de inzet van Belbot voor het vertrouwensbeginsel als Belbot ook gebruik mag maken van persoonsgegevens? Hangt dit af van de implementatiewijze, en zo ja, is de voorgestelde RAG-methodiek voor de benutting van de al bekende en extra verstrekte persoons- en dossiergegevens een betere aanpak dan verwerking van de persoonsgegevens via een LLM-gebaseerde chat-dialoog (waar een platform als ChatGPT toe uitnodigt)?
- d) Hoe verhoudt de inzet van Generatieve AI bij de Belastingtelefoon zich tot de centrale waarden uit de Uitvoerings- en handhavingsstrategie van de Belastingdienst (doelmatigheid, doeltreffendheid, rechtszekerheid, rechtsgelijkheid en menselijke maat c.q. recht doen aan de belangen van burgers en bedrijven)? Is het mogelijk om dit te illustreren aan de hand van de Belbot-casus om te reageren op een verzoek tot kortlopend uitstel van betaling of verlenging van de betalingsregeling?
- e) In hoeverre passen de eerder beschreven Belbot-toepassingen bij de uitgangspunten van de Overheidsbrede Visie Generatieve AI?

Bijkomende vragen

- a) Maakt het voor de beantwoording van de hoofdvraag uit of de Belbot op een intern LLM wordt gebaseerd of dat er sprake is van een gelicenseerde cloudoplossing met goede waarborgen op het vlak van privacybescherming?
- b) Wie zou uiteindelijk verantwoordelijk moeten zijn voor de gegeven antwoorden en maakt het hierbij uit of het om een traditioneel gesprek gaat of een AI-ondersteund gesprek?
- c) In hoeverre zijn onderstaande voorwaarden van invloed op het antwoord op vraag 1?

²² Voor algemene achtergrondinformatie, zie o.a. de Uitvoerings- en Handhavingsstrategie (UHS).

- i. Of de medewerker zelf mag beslissen of hij gebruik maakt van de Belbot, ook als dat betekent dat hij minder efficiënt werkt
 - ii. Of de beller er vooraf mee moet instemmen dat het gesprek wordt ondersteund door een digitale assistent die gebruik maakt van Generatieve AI en hierbij de mogelijkheid heeft om voor een terugbelafspraak te kiezen.
 - iii. Of de beller geïnformeerd wordt over de voor- en nadelen van zijn keuze, met de mogelijkheid dat de keuze voor Belbot met extra voordelen wordt gestimuleerd (zoals een snellere en betrouwbaardere beantwoording van zijn vragen).
 - iv. Of Belbot alleen algemene vragen beantwoordt of ook persoonlijke fiscale vragen (waar mogelijk op basis van eerdere geregistreerde klantgegevens).
 - v. Of de medewerker de mogelijkheid krijgt om via de Belbot persoonlijke gegevens van de beller te raadplegen die volgens het systeem nodig zijn voor de beantwoording van de vraag.
 - vi. Of het systeem ook de mogelijkheid biedt onder voorwaarden informatie van een andere belastingplichtige te raadplegen (bv. bij nabestaande of belastingadviseur).
- d) Is het acceptabel als de Belbot autonoom opereert bij goed afgebakende taken (bijvoorbeeld alvast gegevens opzoeken bij het BSN-nummer dat de beller via het belmenu heeft ingevoerd, als basis voor het gesprek met de servicemedewerker), of het doorverbinden naar de juiste persoon (bijv. op basis van een ingesproken vraag die via spraakherkenning wordt gekoppeld aan een kennisbank voor het koppelen van vragen aan onderwerpen)?
- e) Is het wenselijk om de inhoud van de AI-ondersteunde gesprekken te bewaren voor analysedoelen, en hoe wegen de volgende punten mee in het antwoord op deze vraag?
- i. Of alle gesprekken worden vastgelegd ten behoeve van kwaliteitsbewaking en trainingsdoeleinden, zodat de antwoorden kunnen worden gevalideerd/verbeterd.
 - ii. Of de beller expliciet toestemming kan geven voor het hergebruik van zijn informatie voor het geanonimiseerd bijtrainen van het systeem (wat een verdergaande toepassing is dan het bijtrainen van de servicemedewerker).
 - iii. Of de beller standaard een schriftelijke bevestiging kan opvragen van het verstrekte advies, inclusief gebruikte gegevens en fiscale onderbouwing.
 - iv. Dat de transparante onderbouwing van het antwoord (gebaseerd op duidelijk gespecificeerde bronnen en gegevens) een reden kan zijn om de Belbot in te voeren.
- f) Zijn er nog andere ethische aspecten die voor of tegen de inzet van de Belbot pleiten, ook gezien de uitgangspunten van de Overheidsbrede visie Generatieve AI?
- g) Komt er een moment dat we deze adviesvraag aan Belbot of een andere door Generatieve AI ondersteunde Chatbot kunnen stellen als representant van de huidige adviescommissie?

Appendix B: Toelichting vindplaats antwoord per vraag

Vindplaats antwoorden per vraag

Paragraaf	Vragen
3.2. Randvoorwaarden en verantwoorde implementatie	1a ii, 1c, 2c iii, 2 e ii, 2 e iii, 2e iv, 2f
3.3. Kernwaarden, principes en ethische uitdagingen	1b, 1d, 2f
3.4. Technologische overwegingen	1a, 1a i, 1a iii, 2a, 2e i, 1a iv
3.5. Het gebruik van de discretionaire ruimte en biaswaakzaamheid	2b, 2c i, 2c iv, 2c v, 2c vi
3.6. Autonomie en afgebakende taken	2d
3.7. Kaders en regelgeving	2c ii, 2c vi, 2f

De vraag (1e)

“In hoeverre passen de eerder beschreven Belbot-toepassingen bij de uitgangspunten van de Overheidsbrede Visie GenAI?”

De commissie van mening is dat deze vraag een toetsingsvraag is. De Adviescommissie geeft geen advies over ethische aspecten en is geen toetsingscommissie.

De vraag (2g)

“Komt er een moment dat we deze adviesvraag aan Belbot of een andere door Generatieve AI ondersteunde Chatbot kunnen stellen als representant van de huidige adviescommissie?”

Deze vraag is bewust niet beantwoord in het advies zelf. De dag dat adviesvragen aan een Belbot of een andere door een op GenAI gebaseerd systeem voor dialoogondersteuning gesteld kunnen worden, waarbij deze kan dienen als een soort representant – en hiermee feitelijk als een soort van vervanging zou kunnen optreden - van de huidige Adviescommissie is een dag waarbij de Adviescommissie vaststelt dat er sprake is van een technologische singulariteit of ook wel het bereiken van *Artificial General Intelligence* (AGI). Tot dan zullen de leden van de Adviescommissie zich blijven inzetten om van toevoegde waarde te zijn.

Appendix C: Technische toelichting

Adviesvraag

Opbouw en randvoorwaarden Belbot

Het basisidee van de Belbot is dat deze de servicemedewerker tijdens het gesprek kan ondersteunen met het opzoeken en klaarzetten van informatie en zo bijdraagt aan een snellere afhandeling en een betere verantwoording.

Belbot moet in het algemeen kunnen helpen bij de beantwoording van vragen die betrekking hebben op fiscale processen en toegestuurde brieven, bijvoorbeeld welke mogelijkheden de belastingplichtige heeft. Hierbij moet de servicemedewerker zelf beoordelen wat hij mag meedelen en hoe hij dit formuleert. In een conceptlab wordt een demo-versie van de Belbot ontwikkeld met deze beoogde kenmerken:

- Binnenkomende vragen classificeren en routeren;
- Uit actuele procesdocumentatie afleiden wat de benodigde gegevens voor het antwoord;
- Deze gegevens ophalen uit het klantdossier en kan delen met de servicemedewerker;
- De ontbrekende informatie opvragen via de servicemedewerker of een belmenu;
- Een correcte conclusie trekken uit deze gegevens op basis van een redenering (met expliciete regels) die controleerbaar is voor de servicedeskmedewerker (en de beller);
- Een passend antwoord formuleren;
- Het via de telefoon verstrekte antwoord aansluitend ook (onder nader te bepalen; voorwaarden) in schriftelijke vorm beschikbaar stellen.

De Belbot is opgebouwd uit twee geïntegreerde componenten, waartussen gegevensverkeer mogelijk is:

1. Een op een LLM (Large Language Model) gebaseerd RAG-systeem (Retrieval Augmented Generation, voor het opsporen van relevante contextinformatie in een interne tekstgebaseerde kennisbank) specifiek voor het beantwoorden van vragen die binnenkomen bij de Belastingtelefoon. De kern bestaat uit het Embedding Model, dat ervoor zorgt dat de kennis uit interne documenten (Eigen Data) gestructureerd (als informatievektor) in de Vector Database wordt vastgelegd. De gebruikersvragen worden op dezelfde wijze in informatievektoren omgezet, waarna via een similarity-analyse gekeken wordt welke interne documenten het beste bij de vraag passen. Die bronnen moet het LLM als basis nemen voor de aanmaak van passende antwoorden voor de servicemedewerker, zodat die de beller juist te woord kan staan.
2. Een dashboard waarin de juiste klantgegevens en informatiebronnen gekoppeld worden om de dialoogondersteuning persoonsgebonden te maken. Dit Dashboard communiceert met de gebruikersinterface voor de RAG voor input en het tonen van de antwoorden van het systeem.

Omdat de adviesvrager het van groot belang vindt een verantwoorde Belbot neer te zetten, wordt onder meer met de volgende randvoorwaarden rekening gehouden:

1. Betrouwbaarheid: gebruik van alleen interne bronnen, vermelding van de gebruikte bronnen, altijd in interactie met de servicemedewerker; medewerker houdt de leiding.
2. Kwaliteitsbewaking: door casusgebaseerd te werken kan de respons van het systeem goed worden gevalideerd en wordt hallucinatie voorkomen; via logging van de chatdialoog en onderliggende systeemstappen is doorlopende kwaliteitsbewaking mogelijk.
3. Privacy: geen data-uitwisseling met de cloud; er wordt gebruik gemaakt van een lokale installatie van het taalmodel (LLM), te gebruiken binnen een afgeschermd omgeving.
4. Persoonlijke adviezen: koppeling met klantdossier, er wordt niet meer informatie opgehaald dan nodig.
5. Uitlegbaar en reproduceerbaar door vastlegging van chathistorie en onderliggende stappen; klant kan desgewenst een afschrift van het advies ontvangen.

6. We kiezen een LLM dat zich aan open source principes houdt en voor zover bekend zonder uitbuiting (i.e. inzet onderbetaald personeel) en copyrightscheiding tot stand is gekomen.
7. Energieminimalisatie: het ingezette LLM gebruikt niet meer energie dan noodzakelijk om de voorgelegde vraag te kunnen beantwoorden (o.a. klein LLM, hergebruik eerdere query's).

Output metrieke voor Belbot

Er zijn specifieke metrieke die relateren aan de output²³ van de Belbot. Er zijn open-source *frameworks* zoals Holistic Evaluation for Language Models (HELM)²⁴ en Ragas²⁵ die helpen om een volledige evaluatie te doen van een model en metrieke geven voor onderstaande concepten:

- Source transparency: percentage van het rapport dat onderbouwd wordt door bronnen.
- Relevance & Recall (information retention): hoeveel van de geregistreeerde informatie uit het gesprek (zowel gebruikersinformatie als gegeven adviezen) komt terug in het rapport?
- Faithfulness/Diversity: hoeveel variatie zit er in het rapport? In hoeverre overlapt het met de gegevens in het systeem?
- Fluency & Coherence: hoe grammaticaal correct en natuurlijk is het rapport? Is er een logische flow van het rapport en is het consistent?
- Juistheid van vertalingen, bijvoorbeeld met de BLEU-metrik²⁶.

²³ Output: in dit geval het resultaat gegenereerd door de BelBot, wat gebruikt wordt ter ondersteuning van de medewerker.

²⁴ <https://crfm.stanford.edu/helm/>

²⁵ <https://docs.ragas.io/en/stable/>

²⁶ Voor een overzicht van verschillende toetsingscriteria en toepassingen – waaronder BLEU, zie <https://cloud.google.com/vertex-ai/generative-ai/docs/models/determine-eval#bleu>

Appendix D: Over de Adviescommissie Analytics

Het ministerie van Financiën heeft de onafhankelijke Adviescommissie Analytics (verder: de Adviescommissie) ingesteld voor de Belastingdienst, Dienst Toeslagen, Douane en het kerndepartement om haar kritische blik te versterken en bij te dragen aan een lerende organisatie. Het doel van de Adviescommissie is bijdragen aan een meer verantwoorde omgang met data-analyse, algoritmes, risicomodellen en artificiële intelligentie. De Adviescommissie adviseert in brede zin vanuit de vijf perspectieven die de Algemene Rekenkamer (AR) in haar rapport 'Aandacht voor algoritmes' als toetsingskader heeft meegegeven.

Voordat adviesvragen worden gesteld aan de Adviescommissie, worden deze intern voorbereid door een brede expertgroep, afkomstig van het ministerie van Financiën, de Belastingdienst, Dienst Toeslagen en de Douane. Een adviesvraag kan gericht zijn op hoe data en *Analytics* ingezet kan worden in een nieuwe context met daarbij de nadruk op ethische en sociale aspecten.

Voor wat betreft de werkwijze staat het voeren van dialoog centraal. De Adviescommissie is nadrukkelijk geen onderzoeks- of toetsingscommissie.

De leden van de Adviescommissie betreffen:

- prof. dr. E.H.L. Aarts (Emile) - voorzitter
- prof. dr. S. Bhulai (Sandjai)
- prof. dr. P.J.J. van Geest (Paul)
- prof. dr. M.J. van den Hoven (Jeroen)
- mr. F.C. van der Jagt (Friederike, lid tot 28-02-2025)
- prof. dr. E.L.O. Keymolen (Esther)
- dr. P. Prüfer (Patricia)
- dr. M.T. Schäfer (Mirko)
- dr. A. van Wissen (Arlette)

Procesverloop beantwoording adviesvraag

Begin augustus 2024 heeft de adviesvrager de adviesvraag gepresenteerd en toegelicht aan de Adviescommissie met ruimte voor verklarende en verdiepende vragen van de Adviescommissie. Ter vergadering is voor de behandeling van de adviesvraag een subcommissie ingesteld, bestaande uit Sandjai Bhulai, Paul van Geest, Mirko Schäfer en Arlette van Wissen. De subcommissie is op 27 september online bijeengekomen voor een eerste inhoudelijke bespreking van het ingediende advies. Daarna is de Adviescommissie Analytics fysiek bijegekomen op 7 oktober tijdens een openbare vergadering te 's-Hertogenbosch en hierbij is het advies ook mondeling besproken. Er hebben daarna drie subcommissie-vergaderingen online plaatsgevonden om de inhoud van het advies vorm te geven, te weten op 6 november, 4 december en 22 januari. Op maandag 10 februari stond de volgende openbare vergadering te 's-Hertogenbosch van de voltalige Adviescommissie gepland en is het conceptadvies *GenAI* Belbot besproken met alle leden. Beide Rapporteurs zijn vanaf het begin betrokken geweest bij het opstellen van het advies en hebben goed naar de leden van de Adviescommissie geluisterd en op basis daarvan een advies geschreven. Met de input is de rapporteur aan de slag gegaan en op 26 maart is de subcommissie online bijeen geweest om te bespreken. Na een schriftelijke ronde en een feitelijke check door de adviesvrager is het definitieve concept geagendeerd in de vergadering van de Adviescommissie op 14 april.

