

Bestuurlijk standpunt m.b.t. het advies aan DF&A en KI&S inzake inzet Generatieve AI voor dialoogondersteuning bij de BelastingTelefoon.

01-09-2025, Persoonsgegevens namens de dienstonderdelen KI&S en DF&A

Managementsamenvatting

De directie Klantinteractie en -services (KI&S) en Corporate Dienst Datafundamenten en Analytics (DF&A) hebben met interesse kennisgenomen van het uitgebreide advies van de Adviescommissie Analytics voor het Ministerie van Financiën over de wenselijkheid en randvoorwaarden van een mogelijke nieuwe toepassing van Generatieve AI (GenAI), te weten een interne chatbot (aangeduid als *BelBot*) voor dialoogondersteuning aan servicemedewerkers van de BelastingTelefoon.

De aanvragers van dit advies zijn verheugd dat de Adviescommissie zich positief uitsprekt over de gekozen opzet van deze interne chatbot (in ontwikkeling bij DF&A), die tijdswinst en kwaliteitsverbetering mogelijk moet maken bij het beantwoorden van vragen aan de BelastingTelefoon. Wel verbindt de Adviescommissie een flinke reeks voorwaarden aan dit positieve advies, maar naar het oordeel van de dienstonderdelen die de adviesvraag hebben ingediend, gaat het om redelijke voorwaarden die grotendeels al een plaats kregen in de interne visie op verantwoord omgaan met AI bij de Belastingdienst in het beleidsprogramma 'Regie op AI'. Parallel zijn veel van deze voorwaarden ook al in beeld bij de afdeling KI&S bij de uitwerking en afstemming van plannen voor de ontwikkeling of inkoop van AI-toepassingen.

De genoemde dienstonderdelen zien het advies dan ook als een bevestiging van de ingeslagen koers met de ontwikkeling van kaders en richtlijnen voor GenAI-toepassingen en de invulling van interne innovatietrajecten rond GenAI. Zo wordt al rekening gehouden met de aanbevelingen van de Overheidsbrede visie op Generatieve AI (met waarden als het welzijn van medewerkers en burgers, menselijke maat en duurzaamheid) en principes zoals transparantie, controleerbaarheid en menselijke tussenkomst. Dat betekent niet dat deze dienstonderdelen niets meer kunnen leren van het uitgebrachte advies. Specifiek met betrekking tot de GenAI-toepassing waar de adviesvraag betrekking op had, vermeldt het advies vele behartenswaardige overwegingen, aandachtspunten, metrieken en mitigerende maatregelen.

De kern van het advies bestaat uit aanbevelingen m.b.t. de onderstaande acht punten:

1. AI als ondersteuning, geen vervanging van fiscale expertise
2. Verloren vaardigheden voor medewerkers
3. AI-geletterdheid als randvoorwaarde

4. Ethische en beleidsmatige overwegingen
5. Acceptatie en randvoorwaarden voor AI-ondersteuning
6. Vertrouwensbeginsel en gegevensgebruik
7. Transparantie en keuzevrijheid voor burgers
8. Governance en externe afstemming

Zowel de gehanteerde acht beoordelingscriteria als de aanbevelingen m.b.t. de keuze tussen interne en externe hosting, toetsing van kernwaarden, ethische waarden en kwaliteitscriteria (c.q. metrieken) zijn zeer bruikbaar voor de verdere uitwerking van kaders en richtlijnen voor verantwoorde implementatie van GenAI-toepassingen, en in het verlengde hiervan ook voor de specificatie van tenders voor inkooptrajecten. In deze bestuurlijke standpuntbepaling wordt per invalshoek aangegeven hoe de dienstonderdelen dit mee zullen nemen in het vervolg van het Belbot-project, of het nu intern wordt ontwikkeld of via een aanbesteding wordt neergezet. De concrete aanbevelingen m.b.t. hosting, kernwaarden, ethische waarden en kwaliteitsmetrieken zijn samengebracht in de appendix van deze reactie, zodat ze echt in beeld blijven.

Uit het advies blijkt dat het lopende Belbot-innovatietraject een goede opzet heeft, maar nog tekortschiet wat betreft actieve toepassing van het klantperspectief: tot nu toe lag de focus meer op de technische uitdagingen rond de daadwerkelijke inzet van GenAI.

De Adviescommissie adviseert bijvoorbeeld:

- Om na te denken over de vraag hoe het systeem de beoogde waarden kan realiseren, zoals het bevorderen van de menselijke maat, omdat de scope blijkens de projectbeschrijving nu nog sterk op efficiency-verbetering zit.
- Om de focus te verbreden van de puur technische uitdaging om een betrouwbaar systeem neer te zetten naar inzicht krijgen in wat de gebruiker wil en vindt. Wat is er nodig om vertrouwen en waardering van de gebruikers te krijgen; draagt het systeem echt bij aan een snellere en meer betrouwbare beantwoording van de vragen aan de BelastingTelefoon en is de beller ook blij met het verloop van de gesprekken die met hulp van de Belbot worden gevoerd?
- Om daadwerkelijk te meten hoe het met de kwaliteit en de waardering is gesteld aan de hand van concrete metrieken (zie de appendix).
- Om zorg te dragen voor de tijdige inrichting van governance m.b.t. de ethische vraagstukken die kunnen ontstaan m.b.t. de dataverwerking via GenAI.

Door deze perspectieven in een vroeg stadium samen te brengen neemt de kans toe dat er ook echt een bruikbare en verantwoorde AI-toepassing ontstaat.

Bestuurlijk standpunt in reactie op het uitgebrachte advies

Inleiding

De directie Klantinteractie en Services (KI&S) en Corporate Dienst Datafundamenten en Analytics (DF&A) danken de Adviescommissie Analytics voor het uitgebreide en rijke advies in reactie op de adviesvraag over de inzet van een toekomstig op Generatieve AI gebaseerd systeem (GenAI-systeem) voor dialoogondersteuning voor servicemedewerkers van de BelastingTelefoon, met de werknaam BelBot. Een advies dat door zijn intellectuele diepte en accuraatheid beslist niet voorspelbaar was met de inzet van de huidige GenAI-platforms. Met de adviesvraag nodigden KI&S en DF&A de Adviescommissie uit te reflecteren op de wenselijkheid van de inzet van de BelBot. In het verlengde werd de Adviescommissie ook gevraagd verdiepend advies te geven over de vormgeving en implementatie van een dergelijk systeem.

Deze bestuurlijke reactie is primair gericht op de weging van de acht punten uit de samenvatting van dit advies (verder aan te halen als *het BelBot-advies*) op praktische haalbaarheid en conformiteit met bestaande kaders en inzichten. Hieronder lopen we deze punten een voor een langs, waarbij we steeds een appreciatie geven. Aanvullend zullen we ook stilstaan bij concrete, meer specifieke adviezen uit de verdere notitie, voor zover nog niet gedekt door onze reactie op de samenvatting.

Specifiek erkennen we de relevantie van de kritiekpunten op de scope van het huidige ontwikkeltraject bij DF&A, dat (los van de uitwerking van deze adviesvraag) sterk technisch is gericht en als primaire doel heeft om een betrouwbaar en robuust platform neer te zetten voor efficiency-bevorderende dialoogondersteuning bij scriptgebaseerde telefoongesprekken van servicemedewerkers. In de volgende projectfasen besteden we graag aandacht aan de klantwaarden. Zie de managementsamenvatting voor de hiermee samenhangende verbeterpunten.

Over het geheel genomen kunnen we ons goed vinden in de voorgestelde aandachtspunten en maatregelen. Ook de geraadpleegde adviseurs uit de Expertgroep Analytics (waarin alle DG's van Financiën en het beleidsdepartement zijn vertegenwoordigd) steunen deze lijn.

1. AI als ondersteuning, geen vervanging van fiscale expertise

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. 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.

Reactie: De inrichting en doorontwikkeling van deze toepassing moet zeker getoetst worden op inhoudelijke, dienstverlenende en ethische criteria. Die laatste twee zullen in de praktijk ook overlappen. We hebben hier ook AI-waarden geformuleerd. De servicemedewerker is en blijft zelf verantwoordelijk voor zijn antwoord. Waarbij we streven naar een uniform antwoord binnen de persoonlijke context van burger/ondernemer. Dit is in feite niet anders dan met de huidige ondersteuning. Echter, ook wij zien het risico dat de systematische ondersteuning door AI de actieve inzet van vakinhoudelijke expertise van medewerkers kan belemmeren. Anderzijds maakt een medewerker die gebruik maakt van belscripts nu ook afwegingen bij de formulering van zijn antwoord. Een combinatie van juiste training, juiste coaching en monitoring van enerzijds de gesprekken maar ook de werking en inrichting (bijv. guardrails en monitoring) van de BelBot moeten hierbij helpen.

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.

Reactie: We zijn het volledig eens dat medewerker moet snappen hoe een BelBot werkt. Dit kan via een gespecialiseerde AI-cursus worden bereikt die voortbouwt op de basiscursus AI-geletterdheid. Verder moeten medewerker hetzelfde kennisniveau als nu hebben wat betreft de besproken materie om het gegeven antwoord te kunnen toelichten. Daarbij is het de verwachting dat het aandeel meer complexe vragen toeneemt door het afvangen van makkelijk te beantwoorden vragen, onder andere door inzet van kanaalsturing en verdere ontwikkeling van portalen.

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.

Reactie: AI-geletterdheid is een must, dit onderschrijven wij volledig. Dit zal dan ook actief worden meegenomen bij de implementatie van een GenAI-systeem als BelBot. Voor zowel gebruikers als de ondersteunende rollen, mede gebaseerd op een risicoprofiel. Dit gebeurt in aanvulling op de sterk geadviseerde algemene cursus AI-geletterdheid i.v.m. de beschikbaarstelling van Copilot Chat als hulpmiddel in het dagelijkse werk. Ethische kaders (wij noemen het AI-Waarden, denk bijvoorbeeld aan rechtvaardigheid, maar ook duurzaamheid) worden ingezet bij ontwerp, testfase en implementatie van de BelBot. Daarnaast nemen we het mee in een toekomstige training. Onderdeel is daarbij ook effectief gebruik van de BelBot. Net zoals medewerkers nu worden getraind en gecoacht op het effectief gebruik van systemen, zal dit straks ook gebeuren bij de ingebruikname van een BelBot. Het lijkt ons zinvol om de bevordering van data-geletterdheid en AI-geletterdheid in de toekomst te combineren. Dit raakt ook aan de omgang met kernwaarden, in de zin dat medewerkers een goede balans moeten vinden tussen de harde data en empathie, waarbij steeds de vraag aan de orde is of er sprake is van gelijke behandeling van gelijke gevallen.

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.

Reactie: We hebben geen onbeperkte keuzeruimte waar het gaat om welke technologie gebruikt kan worden voor onze processen, dus er zijn soms suboptimale keuzes nodig. Uiteraard volgen we Rijksbrede richtlijnen en kaders. Ten aanzien van kritische toetsing: ook onze medewerkers en daarmee huidige processen zijn niet volledig voorspelbaar. Daarom hebben we meerdere processen om de gewenste kwaliteitsstandaard te monitoren en waar nodig bij te sturen. Zoals:

- Actieve monitoring op kwaliteitsniveau: inhoudelijke steekproeven, bij start massaler.
- Feedback servicemedewerkers: kwalitatief en kwantitatief, bijvoorbeeld in eerste fase sessies/enquêtes. Duimpjes meten voor AI-gebaseerde antwoorden vinden we een prima idee, al is dit wel afhankelijk van de technische mogelijkheden.
- Feedback burgers/ondernemers, zeker een wens, speelt ook breder. Interessant om te zien of we dit mee kunnen nemen bij implementatie als a/b test of iets dergelijks. Er is al een regulier klanttevredenheidsonderzoek vanuit de contactkanalen.
- Reguliere kwaliteitsmonitoring: hierbij kijkt men zowel naar het inhoudelijk aspect als naar het dienstverleningsaspect (sterke link met de kernwaarden).

We nemen bij dit thema graag ook de concrete suggesties voor toetscriteria mee. Als de inzet van GenAI tot gelijkwaardige resultaten leidt als bij serviceverlening zonder inzet van GenAI (onze baseline), zien we dit in principe als een positieve usecase. Er wordt dan tijdsbesparing mogelijk zonder kwaliteitsverlies.

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 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.

Reactie: Het BelBot-concept is nog niet uitgedacht op werkinstructie-niveau voor medewerkers. Wel lijkt het logisch om medewerkers altijd de ruimte te laten om de AI-assistent te negeren (na geven feedback) en handmatig de kennisbank te raadplegen en/of een hulplijn in te zetten. Dit is in lijn met hoe wij nu ook al werken. Dit gaat om inhoudelijke afwegingen, maar ook zeker om de menselijke maat/barmhartigheid.

Deze specifieke casus gaat om ondersteuning van medewerkers. Bij de implementatie zal hier uitgebreid bij stil worden gestaan. Onze servicemedewerkers houden de leiding over het antwoordproces en blijven dus ook verantwoordelijk voor het antwoord. We onderkennen de geschetste risico's, maar dit is wat ons betreft niet fundamenteel anders dan het nu is met de ondersteuning van een kennisbank (c.q. kennis-managementsysteem) en bronsystemen.

Wij erkennen de noodzaak van de toegang tot menselijke hulp in aanvulling op AI-gebaseerde systemen. Vooralsnog is de BelBot ook alleen een intern hulpsysteem voor de dialoog tussen de menselijke servicemedewerker en de beller. Uit het advies maken we op dat het (in lijn met onze eigen verwachting) niet nodig is om hier expliciet toestemming voor te vragen aan de beller. Praktisch gezien zou dat ook erg ingewikkeld zijn en de effectiviteit van de dienstverlening via de BelastingTelefoon kunnen aantasten.

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.

Reactie: Het vertrouwensbeginsel is voor ons essentieel. Transparantie en herleidbaarheid zijn hier inderdaad belangrijke kernwoorden. En een belangrijk aandachtspunt is om hallucinaties te minimaliseren. We onderschrijven dat keuzes t.a.v. bronnen, RAG's, LLM's en databases gevolgen hebben als je kijkt naar risico's en benodigde mitigerende maatregelen. De exacte technische invulling en aanvullende maatregelen (technisch en procesmatig) moeten nog worden bepaald.

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.

Reactie: De wijze waarop wij omgaan met data en inzet van AI zullen wij conform de richtlijnen expliciet vermelden. Denk daarbij aan de IVR (keuzemenu), maar ook de

privacyverklaring. Dit is nu ook al het geval bij het gebruik van opnames van interacties voor coaching en analyse doeleinden (nog geen AI toepassing). Dit is ook beschreven en getoetst door middel van een DPIA. Indien er bezwaar is tegen een opname is er een werkproces voor verwijdering van de opname en transcript. Mocht blijken dat er aanvullende maatregelen nodig zijn bij de inzet van de BelBot, dan geven we hier een invulling aan. De genoemde nieuwe toepassingen vergen nog veel afstemming.

8. Governance en externe afstemming

Er hoort een governance structuur te 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. 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.

Reactie: We onderschrijven het geschetste belang. De uitwerking van de governance is een Belastingdienst-breed vraagstuk, waar ook aandacht voor is bij de strategievorming m.b.t. de inzet van AI. Anderzijds ontslaat dit de afnemer van een AI-toepassing niet van eigen risico mitigerende maatregelen in een robuuste samenhang (met elkaar en de overkoepelende kaders en governance). De standaard werkwijze is om een DPIA uit te voeren bij invoering van nieuwe, analytics-gebaseerde oplossingen. Voor AI-gebaseerde oplossingen zal dit niet anders zijn. Hiernaast vindt ook een ethische toetsing plaats aan de hand van het IAMA-kader, in de vorm van een ethisch-juridisch assessment.

De verantwoorde verwerking van persoonsgegevens en privacygevoelige informatie wordt standaard meegenomen bij de juridische toetsing van onze applicaties en analytics-toepassingen. Verder staan we in nauw contact met de AP. We zullen waar logisch zeker gebruik maken van de kennis aldaar. Tot slot is de zogenaamde exit-strategie een wezenlijk onderdeel van onze architectuur en we onderkennen zeker het belang van zogenaamde vendor lock ins en data soevereiniteit.

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.

Reactie: Het is prettig dat de vraag over het ‘willen’ van de door ons voorgelegde AI-toepassing positief wordt beantwoord in het advies, ervan uitgaande dat de gestelde randvoorwaarden in orde worden gemaakt. Het ‘mogen’ en ‘kunnen’ is een vraagstuk dat we de komende periode met elkaar verder moeten uitwerken. Waarbij we te maken hebben met afhankelijkheden van nog in ontwikkeling zijnde richtlijnen en kaders, maar ook de voortdurende technische ontwikkelingen.

De context van massale dienstverlening

Bij de afweging van keuzes rond de inzet van AI-oplossingen is het van belang om de massaliteit van onze dienstverlening in relatie te zien tot het belang van continuïteit en schaalbaarheid van de oplossing. Met circa 10 miljoen interacties op jaarbasis en circa 1000 gelijktijdige users vraagt onze dienstverlening om een zeer robuuste oplossing. Momenteel lopen hier ook verschillende gesprekken over dit soort toepassingen. Risico's en maatregelen t.a.v. (nog ontbrekende) richtlijnen, wetgeving, security, soevereiniteit en technische (on)mogelijkheden die de Adviescommissie ook zelf aanraakt in het advies, zijn hierbij onderdeel van gesprek.

Visie op dienstverlening

Uitgangspunt voor de Belastingdienst is dat de inzet van AI-ondersteuning voor onze servicemedewerkers, zoals een BelBot, moet helpen bij het einddoel: het bevorderen van compliantie. Vanuit het perspectief van de belastingplichtige bezien moet het bijdragen aan het kunnen voldoen aan plichten en het gebruik maken van rechten, zo makkelijk mogelijk en waar nodig met hulp. Dit willen we doen op een congruente, efficiënte en effectieve manier. Uiteraard vanuit een dienstverlenende houding. De BelBot moet hier een aanvulling op zijn en daarmee passen binnen de bestaande kwaliteitscriteria.

Gefaseerde implementatie

We onderkennen dat de implementatie van een AI-gebaseerde innovatie als de BelBot een grote verantwoordelijkheid met zich meebrengt, ook vanwege inhoudelijke uitdagingen en het technologische vertrekpunt. We zien dan ook het belang van een gefaseerde implementatie met evaluatiemomenten. Qua toetscriteria zullen we zoveel mogelijk aansluiten bij de huidige methodiek, genaamd kwaliteitsmonitoring. Dit proces geeft enerzijds een uniform beeld over het kwaliteitsniveau (inhoudelijk en klantvriendelijkheid) en anderzijds is het een middel om de medewerker te helpen in zijn/haar ontwikkeling. Daarnaast zal het proces en dus ook het toekomstig systeem

moeten voldoen aan kaders en richtlijnen rond privacy, security en inzet AI. Dit vraagt om een lerende aanpak, met passende ondersteuning op inhoud, technologie en kaders in een veilige omgeving.

Human in the loop

Het scenario van autonome GenAI-toepassingen voor “afgebakende taken” is nu nog verder weg. We zien het echter wel als een relevante ontwikkeling in onze visie op dienstverlening. We onderschrijven daarbij dat het nooit dwingend kan zijn of een digitale weg zonder afslag naar menselijke ondersteuning. Het belang van een effectieve *Human in the loop* onderkennen wij volledig. Dit zien we op twee manieren: 1) goed opgeleide en geïnformeerde medewerkers (op inhoud en AI-geletterdheid) en 2) monitoring van de werking van het systeem door geëquipeerde medewerkers.

We onderkennen volledig de relevante van de ethische en praktische vragen bij de inzet van een autonome GenAI op afgebakende taken. Mochten wij dit overwegen, dan zullen wij ook extra training en monitoring inzetten om de bijkomende risico's te mitigeren. Hierbij zal een escape richting een menselijke medewerker er één van zijn. Zowel bij ontwikkeling als in de run moet aandacht zijn voor continue verbetering. De betekenis van “afgebakende taak” zullen we nader concretiseren per usecase.

Metrieken

We zijn van plan de in het advies aangereikte metrieken voor het evalueren en monitoren van de output van zowel de BelBot zelf als die van de servicemedewerker die met de BelBot werkt (zie appendix) goed te bestuderen en (deels) te beproeven. Niet alleen voor de technische evaluatie, maar ook voor het monitoren van de tevredenheid van de gebruikers (servicemedewerkers) en de bellers (belastingplichtigen). De aangedragen metrieken kunnen ook breder doorwerken door ze mee te nemen bij de uitwerking van kwaliteitsrichtlijnen voor alle AI/GenAI-toepassingen. Ook met dit voornemen willen we aan de slag.

Tot slot

We onderkennen dat de BelBot niet gewoon een nieuwe extra tool is, maar een meer ingrijpende impact heeft. De overwegingen, handreikingen en aanbevelingen in het uitgebrachte advies helpen ons zeer om richting te geven aan een verantwoord ontwikkel- en implementatietraject. Niet alle aanbevelingen uit het advies zijn in onze bestuurlijke reactie meegenomen. Zo doet de adviescommissie hele concrete suggesties m.b.t. ethische principes, evaluatiemetrieken en kwaliteitscriteria. Die raken aan de intern ontwikkelde kaders en richtlijnen m.b.t. technologie en waarden. Onze inzet is om het rapport van de Adviescommissie standaard mee te laten wegen bij de verdere uitwerking en implementatie van de AI-kaders.