

Yannick Lansink, 900102137

Verantwoording

Software Developer, Hogeschool NOVI

14 Januari, 2025

Afkortingenlijst

Afkortingen	Definitie
IV	Informatievoorziening
AVG	Algemene Verordening Gegevensbescherming
LLM	Large Language Model
IBS	Integratie Business Services
NLP	Natural Language Processing
RAG	Retrieval Augmented Generation
AI	Artificial Intelligence
ASP.NET	Active Server Pages .NET (een Microsoft framework)

Disclaimer

“Wegens de vertrouwelijke aard van de informatie in het werkstuk/ de afstudeerscriptie en de bijbehorende bijlagen, mag de inhoud noch geheel noch gedeeltelijk op enigerlei wijze worden aangepast, gewijzigd of verveelvoudigd zonder schriftelijke toestemming van de auteur en het betrokken bedrijf. Zowel de docent/ scriptiebegeleider, de examinerator alsmede de medewerkers van NOVI Hogeschool worden gehouden de inhoud van het werkstuk/ de afstudeerscriptie als strikt vertrouwelijk te behandelen. NOVI Hogeschool bewaart het document in een afgesloten databank. Gedurende de bewaarperiode kunnen studentdossiers worden ingezien door medewerkers van NOVI, de Examencommissie van NOVI alsmede de Inspectie van het Onderwijs en een visitatiecommissie van de Nederlands Vlaams Accreditatieorganisatie (NVAO, bij (her)accreditaties).”

Inhoud

Inleiding.....	5
1. De vraag en de vraag achter de vraag.....	7
1.1 Probleembeschrijving.....	7
1.2 Doelstellingen.....	7
1.3 Onderzoeksdoelstelling.....	8
1.4 Onderzoeksmethode.....	8
2. Omgevingsfactoren en stakeholders.....	10
2.1 Primaire stakeholders	10
2.2 Secundaire stakeholders	10
2.3 Dataverzamelingsmethodes.....	10
2.4 Verwachtingen en implementatie.....	11
2.5 Totstandkoming beroepsproduct	11
2.6 Theoretische onderbouwing	12
3. Risicoanalyse, bijdrage en impact van het product	15
3.1 Fundamentele verandering.....	15
3.2 Hoofd- en deelvragen beantwoording	17
3.3 Organisatorische impact	20
3.4 Risicoanalyse	21
3.5 Risico mitigatie	24
4. Productbespreking met beroepspraktijk	27
4.1 Feedback van development teams	28
4.2 Security specialist evaluatie	28
4.3 Feedback van portfolio management	29
4.4 Implementatie van feedback.....	30
5. Reflectie op product, proces en persoonlijke ontwikkeling	32
5.1 De reis naar een innovatieve oplossing	32
5.2 Product reflectie.....	32
5.3 Proces Reflectie	32
5.4 Persoonlijke ontwikkeling.....	33
5.5 Organisatorische impact	34
5.6 Conclusie reflectie	34
Bronvermelding.....	35
Bijlage.....	37

Inleiding

De digitalisering van processen binnen organisaties heeft de afgelopen jaren een enorme vlucht genomen. Dit heeft geleid tot een toenemende afhankelijkheid van complexe systemen en een groeiende vraag naar effectieve informatievoorziening. In deze context speelt Integratie Business Services (IBS) Toeslagen een cruciale rol. Als onderdeel van de Belastingdienst is IBS Toeslagen verantwoordelijk voor het beheren van complexe processen rondom toeslagen (Werken bij informatievoorziening (IV)). Door de complexiteit van deze systemen en de voortdurende wijzigingen in regelgeving, lopen medewerkers vaak tegen problemen aan bij het vinden van antwoorden op hun vragen.

Het team Integratie Business Services (IBS) Toeslagen kampt met uitdagingen in het snel en accuraat vinden van informatie (*Jansen, J.*), wat de algehele productiviteit en tevredenheid van medewerkers negatief beïnvloedt. Nieuwe medewerkers besteden uren aan het doorzoeken van verschillende systemen, terwijl ervaren krachten worden overspoeld met routinematige vragen. Hierdoor ontstaat tijdverlies en inefficiëntie, wat vraagt om een innovatieve oplossing die deze problemen aanpakt.

Dit verantwoordingsdocument richt zich specifiek op het ontwikkelen van een chatapplicatie die gebruikmaakt van een Large Language Model (LLM), ondersteund door een vector database. Het doel is om de informatievoorziening binnen IBS Toeslagen te verbeteren door medewerkers direct en nauwkeurig te voorzien van antwoorden op hun vragen. Het onderzoek beperkt zich tot de ontwikkeling en evaluatie van een prototype dat gebruikmaakt van dummydata om de haalbaarheid en waarde van de oplossing te demonstreren.

De implementatie van Large Language Models (LLM's) in combinatie met vector databases wordt steeds vaker toegepast om complexe informatiebehoeften te adresseren. Eerdere studies tonen aan dat deze technologieën kunnen bijdragen aan snellere en nauwkeurigere informatievoorziening. Tegelijkertijd is er een groeiend bewustzijn over de risico's op het gebied van privacy en veiligheid, vooral binnen overheidsorganisaties (*Dewivandeurzen, 2024*).

Dit onderzoek is relevant vanuit zowel theoretisch als praktisch oogpunt. Theoretisch draagt het bij aan de kennis over de toepassing van LLM's in organisatorische contexten. Praktisch gezien biedt het een concrete oplossing voor een actueel probleem binnen IBS Toeslagen, waarmee de efficiëntie en medewerkerstevredenheid worden verhoogd.

De centrale probleemstelling van dit onderzoek luidt: "Hoe kan een chatapplicatie, ondersteund door een Large Language Model en een vector database, bijdragen aan efficiëntere informatievoorziening binnen IBS Toeslagen?"

De doelstelling is om een prototype te ontwikkelen en te evalueren dat medewerkers ondersteunt bij het vinden van informatie. De volgende onderzoeksvragen staan centraal:

1. Wat zijn de eisen en wensen van stakeholders met betrekking tot de chatapplicatie?
2. Hoe kan een Large Language Model effectief worden geïntegreerd met een vector database?
3. Welke technische en organisatorische risico's zijn verbonden aan de implementatie?
4. Wat is de impact van het prototype op de gebruikerservaring en productiviteit?

Het onderzoek omvat een combinatie van literatuurstudie, interviews met stakeholders, en de ontwikkeling en evaluatie van een prototype. De technische implementatie richt zich op het gebruik

van dummydata, met aandacht voor privacy en beveiliging. Het prototype wordt getest door middel van gestructureerde feedbacksessies met medewerkers van IBS Toeslagen.

Dit verantwoordingsdocument is als volgt opgebouwd:

- **Hoofdstuk 1:** Beschrijft de aanleiding, probleemstelling, doelstelling, onderzoeksvragen en onderzoeksmethode.
- **Hoofdstuk 2:** Analyseert de omgevingsfactoren en stakeholders en bespreking van het theoretisch kader.
- **Hoofdstuk 3:** Bespreekt de risicoanalyse en impact van het product.
- **Hoofdstuk 4:** Bevat de evaluatie van het product en feedback van stakeholders.
- **Hoofdstuk 5:** Bevat de reflectie op het proces, het product en persoonlijke ontwikkeling.

1. De vraag en de vraag achter de vraag

Integratie Business Services (IBS) Toeslagen is momenteel bezig met het verjongen van het personeel. Door de complexiteit van de systemen en processen binnen IBS Toeslagen, besteden veel medewerkers een aanzienlijke hoeveelheid tijd aan het zoeken naar antwoorden op (retorische) vragen. Deze tijdsverspilling heeft een negatieve impact op de efficiëntie en productiviteit van de medewerkers. De chatapplicatie, ondersteund door een redeneringsmodel, is ontworpen om dit probleem op te lossen door medewerkers snel en nauwkeurig van antwoorden te voorzien op vragen waar voorheen lange zoektijd aan te pas kwam.

1.1 Probleembeschrijving

IBS Toeslagen (Integratie Business Services) kampt met versnipperde kennisbronnen – van code repositories en wiki's tot chatkanalen – waardoor medewerkers veel tijd verliezen bij het zoeken naar informatie. Uit onderzoek blijkt dat kenniswerkers gemiddeld 1.8 uur per dag (ongeveer 9.3 uur per week) besteden aan het zoeken en verzamelen van informatie (Wills, 2024). Dit duidt op flinke inefficiëntie: bijna 30% van de werktijd gaat op aan het zoeken van antwoorden in plaats van productief werk (Wills, 2024). Versnipperde of tegenstrijdige informatie leidt bovendien tot verdere productiviteitsverlies en fouten (Tetra, 2024). Bij IBS Toeslagen uit zich dit in vertraging bij het onboarden van nieuwe teamleden en frustratie bij ervaren medewerkers die telkens dezelfde vragen gesteld krijgen.

Probleemafbakening: Dit project richt zich specifiek op het verbeteren van de interne informatievoorziening binnen IBS Toeslagen door middel van een centrale zoek- en vraagoplossing. De focus ligt op het toegankelijk maken van informatie (zoals documentatie, handleidingen, gezamenlijke chatgeschiedenissen) voor medewerkers via een chatinterface. Andere informatiedomeinen – bijvoorbeeld puur numerieke data-analyse of externe publieke kennisbronnen – vallen buiten de scope. Eveneens ligt de nadruk op informatie vinden en niet op het creëren van nieuwe kennis: het systeem zal bestaande interne informatie beter doorzoekbaar en beschikbaar maken, maar geen beleidsbeslissingen autonoom nemen. Door deze afbakening blijft het project beheersbaar. Het probleem wordt benaderd als een kennismanagement-uitdaging binnen 1 organisatieonderdeel, zodat resultaten en oplossingen gericht zijn op IBS Toeslagen en rekening houden met haar specifieke context.

1.2 Doelstellingen

De auteur beoogt met dit project doelstellingen te behalen: businessdoelstellingen en onderzoeksdoelstellingen. De belangrijkste businessdoelstelling is het verhogen van de efficiëntie van informatie zoeken binnen IBS Toeslagen. Concreet betekent dit dat de zoektijd per medewerker merkbaar omlaag moet. Het SMART geformuleerde doelstelling is: "Reduceer de gemiddelde zoektijd van medewerkers met ten minste 20% binnen 6 maanden na introductie van de oplossing." Dit sluit aan bij de verwachting dat door een centrale kennisbank minder tijd verloren gaat. Andere businessdoelstellingen zijn het verhogen van de medewerkerstevredenheid, medewerkers ervaren minder frustratie en hebben sneller antwoord op vragen, en het waarborgen van dataveiligheid. Dit laatste houdt in dat de nieuwe oplossing voldoet aan interne security-eisen en wetgeving (zoals AVG) zodat geen gevoelige informatie lekt. Ook het stimuleren van technologische innovatie binnen de Belastingdienst is een doel: door moderne AI-technologie toe te passen wil IBS Toeslagen vooroplopen in digitale transformatie.

Naast de businessdoelen is er een expliciete onderzoeksdoelstelling (zie hoofdstuk 1.3) geformuleerd. Deze onderzoeksdoelstelling richt zich op het onderbouwen van het nut van de oplossing en het vergroten van kennis over hoe zo'n oplossing het beste ontworpen kan worden.

Om te bepalen of de informatievoorziening verbetert, zijn kwaliteitscriteria opgesteld. In lijn met gangbare informatiekwaliteit-raamwerken (Ayuya, 2022) wordt gekeken naar: nauwkeurigheid van de geboden antwoorden (kloppen de feiten?), relevantie (in hoeverre sluit het antwoord aan op de vraag?), volledigheid (worden alle benodigde details verstrekt?), tijdigheid (is de informatie actueel en snel beschikbaar?) en consistentie (blijft de informatievoorziening gelijkvormig over verschillende vragen en gebruikers?). Daarnaast is toegankelijkheid een criterium: de informatie moet makkelijk vindbaar en gebruiksvriendelijk gepresenteerd zijn aan medewerkers. Tenslotte wordt veiligheid en vertrouwelijkheid beschouwd als kwaliteitsaspect: gevoelige interne kennis moet zodanig verstrekt worden dat alleen geautoriseerde personen er toegang toe hebben en dat er geen data bij onbevoegden terechtkomt. Deze criteria vormen meetbare punten. Zo zal bijvoorbeeld de nauwkeurigheid en relevantie van zoekresultaten geëvalueerd worden door steekproefsgewijs antwoorden te controleren op juistheid en door gebruikers te laten beoordelen in hoeverre de antwoorden hun vraag beantwoordden. Tijdigheid wordt gemeten als de responstijd van het systeem, en volledigheid kan beoordeeld worden door na te gaan of gebruikers alsnog aanvullende zoekacties moeten ondernemen. Door elke doelstelling meetbaar te koppelen aan dergelijke criteria, kan het succes van de oplossing objectief geëvalueerd worden.

1.3 Onderzoeksdoelstelling

Conform het Plan van Aanpak is een expliciete onderzoeksdoelstelling geformuleerd om richting te geven aan het afstudeeronderzoek. De onderzoeksdoelstelling luidt:

“Het in kaart brengen van de informatiezoekpatronen en -behoeften van medewerkers binnen IBS Toeslagen door middel van een gebruiksanalyse, en vervolgens het ontwerpen en testen van een prototype chatapplicatie die de zoekduur voor nieuwe medewerkers met ten minste 20% vermindert binnen een periode van 6 maanden, waarbij wordt voldaan aan de beveiligingsvereisten en technische eisen van IBS Toeslagen.”

De onderzoeksdoelstelling sluit aan op de businessdoelen uit hoofdstuk 1.2, maar legt de nadruk op het onderzoek matige aspect: eerst bewijs verzamelen dat het probleem daadwerkelijk significant is (dat veel tijd verloren gaat aan zoeken) en vervolgens aantonen dat de voorgestelde oplossing dat probleem kan mitigeren. Deze doelstelling diende als leidraad voor alle onderzoeksactiviteiten, van literatuurstudie tot prototype-evaluatie van de chatapplicatie.

1.4 Onderzoeksmethode

Voor het behalen van bovengenoemde doelstellingen is een gecombineerde onderzoeksaanpak gehanteerd, zoals ook voorzien in het PvA. Er is gekozen voor een Design Science Research-benadering, waarin het ontwerpen van een IT-artefact (de chatapplicatie) hand in hand gaat met kennisvergaring (Wikipedia contributors, 2025a). Deze methode past goed bij het praktische probleem, door een oplossingsartefact (ook wel prototype genoemd) te bouwen en te evalueren, verkrijgen we inzicht in het probleem en de effectiviteit van de oplossing. Hevner (2004) stellen dat DSR kennis vergaart over een probleemdomen door het bouwen en evalueren van een ontworpen artefact (Wikipedia contributors, 2025b). Dit iteratieve bouwen-en-testen sluit aan bij de auteurs aanpak.

Concreet bestond de onderzoeksofzet uit meerdere fasen. Eerst een literatuuronderzoek om de zoektechnologie (vector search, LLM-integraties) en best practices voor kennisbeheer in kaart te

brengen. Hierbij zijn academische artikelen, whitepapers en case studies onderzocht over o.a. “enterprise search”, “LLM applications in organizations” en “veiligheid bij AI-toepassingen”. Dit leverde theoretische grondslagen en vergelijkingsmateriaal (zie hoofdstuk 2.6). Parallel daaraan is een analyse binnen IBS Toeslagen uitgevoerd. Via interviews en observaties is geïnventariseerd hoe medewerkers nu informatie zoeken, hoeveel tijd dat kost en waar de knelpunten liggen. Deze gebruiksanalyse had een kwalitatief karakter, gesprekken met een selectie van medewerkers en leidinggevendenden, aangevuld met schattingen (bijv. hoe vaak per dag men iets niet kan vinden).

Vervolgens vond requirements-analyse plaats door stakeholderconsultatie (zie hoofdstuk 2.3): in workshops en interviews met eindgebruikers (medewerkers), het ontwikkelteam en een security specialist zijn de eisen en wensen opgehaald. Op basis daarvan is een lijst van functionele requirements (FR’s) en non-functionele requirements (NFR’s) opgesteld.

Daarna is middels een iteratieve ontwikkelmethode (zie hoofdstuk 2.5) een prototype gebouwd en tussentijds bijgesteld op basis van feedback. Elke iteratie bestond uit ontwerpen, implementeren en dan evalueren met stakeholders (bijvoorbeeld een demo aan enkele gebruikers, gevolgd door een korte discussie). Methodologisch leunt dit op principes uit Agile/Scrum – snelle feedback loops – gecombineerd met Design Science-evaluaties. Iedere cyclus van bouwen en evalueren vergrootte zowel de bruikbaarheid van het product als het inzicht in wat wel/niet werkt. Deze iteraties gingen door totdat het prototype voldeed aan de gestelde eisen vanuit de requirements-analyse en voldoende bewijs was verzameld om de onderzoeksvragen te beantwoorden.

Tot slot zijn in de evaluatiefase de onderzoeksresultaten verzameld: metingen van prestatie (zoals responstijd), gebruikerstesten van het eindproduct en een risicoanalyse. De onderzoeksmethoden per deelvraag zijn ook expliciet uitgewerkt, om te waarborgen dat elke deelvraag beantwoord werd met passende data. Zo is voor de vraag naar huidige zoektijden gebruikgemaakt van literatuur. De combinatie van kwalitatieve methoden (interviews, observaties) en kwantitatieve gegevens (tijdmetingen, gebruikerstats) zorgt voor een goed beeld van bevindingen, wat de betrouwbaarheid vergroot.

2. Omgevingsfactoren en stakeholders

De ontwikkeling van de chatapplicatie werd sterk beïnvloed door stakeholders en theoretische bronnen. Dit hoofdstuk geeft je een beeld van belangrijke belanghebbende, hoe deze zijn benaderd en een theoretische onderbouwing. Door een uitgebreide analyse van deze elementen wordt een beeld geschetst die de voortgang en uiteindelijke succes van het project bepaalt.

2.1 Primaire stakeholders

De primaire stakeholders van dit project zijn zorgvuldig geïdentificeerd en spelen een cruciale rol in de ontwikkeling en implementatie van de chatapplicatie. De eerste en belangrijkste groep wordt gevormd door de development teams binnen IBS Toeslagen. Als eindgebruikers van de applicatie vertegenwoordigen zij de doelgroep die dagelijks profiteert van de voordelen van de nieuwe technologie. Voor deze groep was het essentieel dat de applicatie intuïtief en efficiënt is in gebruik. Hun verwachtingen richtten zich op een interface die hen in staat stelt snel toegang te krijgen tot relevante informatie, wat hen ondersteunt in hun dagelijkse werkzaamheden.

Een andere belangrijke stakeholder is de security specialist, die verantwoordelijk is voor het waarborgen van informatiebeveiliging en compliance binnen IBS Toeslagen. Voor deze rol lag de nadruk op het waarborgen van dataveiligheid en privacybescherming, met een sterke focus op naleving van de Algemene Verordening Gegevensbescherming (AVG).

2.2 Secundaire stakeholders

Naast de primaire stakeholders speelden secundaire stakeholders, zoals het portfolio management, een belangrijke rol in het strategische kader van het project. Portfolio management is verantwoordelijk voor de strategische afstemming van projecten binnen IBS Toeslagen en heeft invloed op resourcetoewijzing en prioritering. De verwachtingen waren gericht op het waarborgen dat dit project aansluit bij de bredere doelstellingen van de organisatie, waaronder digitale transformatie en kostenreductie.

2.3 Dataverzamelmethode

In deze fase is nauwkeurig beschreven hoe informatie is opgehaald bij verschillende stakeholders en wat dit heeft opgeleverd. Om de behoeftes en eisen scherp te krijgen, zijn meerdere methoden ingezet:

Interviews met eindgebruikers: Met team Vos van IBS Toeslagen (waaronder zowel nieuwe als ervaren teamleden) zijn semi-gestructureerde interviews gehouden (zie Bijlage 6). Hierin is gevraagd naar hun ervaringen met het zoeken naar informatie: welke bronnen ze gebruiken, welke frustraties ze ondervinden, en welke features ze in een ideale oplossing zouden willen. Deze interviews leverden rijke kwalitatieve data op. Zo gaven meerdere gebruikers aan dat ze vaak documenten van hun eigen team willen doorzoeken en die graag zouden kunnen uploaden in het systeem – dit inzicht leidde direct tot FR1 (uploadfunctionaliteit). Een andere terugkerende klacht was de traagheid en omslachtigheid van huidige wiki-zoekmachines, waardoor de eis van real-time antwoorden naar voren kwam (FR2). De interviewresultaten zijn geanalyseerd en antwoorden zijn ondergebracht in thema's als "zoekproblemen", "vereiste functionaliteiten" en "gebruikerservaring". Per thema zijn requirements opgesteld.

Observatie en meting: Om harde data te verkrijgen over het probleem (de zoektijd), zijn korte interviews gehouden met 10 medewerkers. Hen werd gevraagd hoeveel tijd zij gemiddeld per dag kwijt zijn aan het zoeken naar informatie en hoeveel zoekopdrachten onbeantwoord blijven (zie Bijlage 5). Hoewel de groep klein was, bevestigden hun antwoorden de orde van grootte uit de literatuur: gemiddeld circa 2 uur per dag aan zoekactiviteiten, waarbij vooral nieuwe medewerkers aangeven regelmatig te moeten navragen. Deze gebruikersinventarisatie ondersteunde de urgentie van het probleem en diende als nulmeting om de doelstelling van 20% reductie aan af te meten. Daarnaast is dit ook wederom met 10 personen gedaan met veel voorkomende vragen binnen IBS Toeslagen om te kijken hoe lang men er na de implementatie van de chatapplicatie erover doet. Bevindingen zie je in hoofdstuk 3.2 hoofd- en deelvragen beantwoording en bijlage 5.

Elke methode had eigen opbrengsten: de interviews leverden voornamelijk kwalitatieve inzichten en requirements die direct zijn gebruikt in het functioneel ontwerp (zie FR1–FR5 in het functioneel ontwerp). De observaties gaven bevestiging van het probleem (zoekduur), wat hielp om de stakeholders te overtuigen van de noodzaak. Alle verzamelde data zijn gedocumenteerd en hebben als input gediend voor het ontwerp. Traceerbaarheid is geborgd doordat elke requirements herleidbaar is naar een of meerdere stakeholderinput(s).

2.4 Verwachtingen en implementatie

De verzamelde gegevens uit de stakeholder engagement hebben geleid tot een helder beeld van de verwachtingen die zijn meegenomen in het ontwerp en de implementatie van de chatapplicatie. Eén van de belangrijkste verwachtingen was het bieden van snelle toegang tot informatie. Dit werd gerealiseerd door een intuïtieve chatinterface te ontwikkelen die gebruikmaakt van realtime gestreamde antwoorden, wat de snelheid en efficiëntie van informatieopvraging aanzienlijk verbeterde.

Een tweede belangrijke verwachting betrof de betrouwbaarheid van de informatievoorziening. Dit aspect werd aangepakt door het implementeren van een AI-gedreven zoekalgoritme dat niet alleen nauwkeurige resultaten oplevert, maar ook relevante bronverwijzingen biedt om de kwaliteit van de informatie te waarborgen.

Tot slot lag er een sterke nadruk op veiligheid, vooral gezien het werken met potentieel gevoelige gegevens binnen IBS Toeslagen. Om te voldoen aan deze eis, werd tijdens de ontwikkeling gekozen voor het gebruik van dummy data in plaats van productiegegevens. Dit zorgde ervoor dat de applicatie in een gecontroleerde omgeving kon worden getest zonder risico's voor datalekken of schendingen van privacyregels. Hoewel de applicatie nog niet geschikt is voor een productielancering, heeft dit proces een solide basis gelegd voor toekomstige uitbreidingen en verbeteringen.

2.5 Totstandkoming beroepsproduct

De totstandkoming van het beroepsproduct, de chatapplicatie, vond plaats volgens een iteratieve projectmethodiek. Aangezien IBS Toeslagen binnen de Belastingdienst werkt met het Scaled Agile Framework (SAFe), is zoveel mogelijk aangesloten bij bestaande Agile-principes. Dit betekende werken in korte sprints van ~2 weken, met telkens een increment van het product en feedbackloops. Een iteratieve aanpak bood meerwaarde: de auteur kon het product stapsgewijs opbouwen en onderweg bijsturen op basis van gebruikersfeedback (Atlassian, z.d.). Elke cyclus bestond uit planning, uitvoering en evaluatie.

Projectmethodiek: In de eerste sprint lag de focus op het neerzetten van de basisarchitectuur (frontend, backend integratie, databaseconnectie) en een minimale werkende zoekfunctionaliteit. Vervolgens zijn incrementeel de kernfunctionaliteiten toegevoegd: document upload, vraag-antwoord met LLM, semantische zoekfunctie, etc. Tussentijds vonden reviews plaats met stakeholders. Deze werkwijze sluit aan bij de Agile feedbackloop: regelmatige feedbackmomenten zorgen dat het product blijft aansluiten bij de behoeften (DiCesare, 2024). Zo werd na een demo van de eerste zoekfunctie duidelijk dat de responstijd nog te traag was, hierdoor kon de auteur de volgende sprint werken om dit te verbeteren.

Testtraject: Testing vond plaats op meerdere niveaus. Er waren er regelmatige gebruikersacceptatietests in de vorm van informele try-outs: collega's mochten de chatapp gebruiken op een set van voorbeeldvragen. Hun bevindingen (werd de vraag beantwoord, was het antwoord behulpzaam, enz.) werden systematisch verzameld via een korte vragenlijst na elke testsessie (zie Bijlage). Deze kwalitatieve testresultaten werden teruggekoppeld in het ontwikkelproces. Zo gaf een gebruiker aan dat de interface om documenten te uploaden niet erg duidelijk was – dit is vervolgens aangepast door betere UI-indicatie en instructietekst toe te voegen in een volgende versie. Ook werd tijdens tests duidelijk dat gebruikers graag zagen waar in een document het antwoord gevonden was. Deze feedback leidde tot de feature dat een bronverwijzing met highlight wordt getoond. Dit sluit ook aan op de Agile werkwijze.

Iteratieve ontwerpaanpak: Het ontwerp van de user interface en de interactie is iteratief bijgeschaafd via prototyping. Begonnen is met wireframes (schetsen) die in een vroeg stadium al aan eindgebruikers zijn getoond voor feedback op lay-out en functionaliteit. Vervolgens is een klikbaar prototype opgezet, waar gebruikers doorheen konden “klikken”. Feedback hierop, bijvoorbeeld de behoefte aan een duidelijkere navigatie tussen chat en zoekmodus, werd verwerkt in het uiteindelijke UI-ontwerp. De kern van de iteratieve aanpak is dat we niet in één keer alles vastgelegd hebben, maar telkens kleine verbeterstappen maakten op basis van toetsing. Dit reduceert projectrisico's en verhoogt de kwaliteit. Problemen worden vroeg ontdekt en opgelost (Atlassian, z.d.).

Tot slot is gebruikgemaakt van versiebeheer en code reviews om de kwaliteit te borgen. Iedere iteratie werd in een git-repository bijgehouden. Ook dit sluit aan bij professionele agile ontwikkeling en draagt bij aan een robuust eindproduct. Samengevat kenmerkte de totstandkoming zich door een wendbare Agile aanpak. Door frequente is stap voor stap een werkende oplossing gebouwd die gevalideerd is op functionaliteit en gebruiksvriendelijkheid voordat het als eindsysteem is opgeleverd.

2.6 Theoretische onderbouwing

In deze sectie worden de keuzes voor technologie en architectuur onderbouwd met theorie en worden alternatieve opties besproken.

Retrieval-Augmented Generation (RAG) benadering: De kern van de oplossing is gebouwd op een RAG-systeem, waarbij een Large Language Model wordt gevoed met relevante context uit een vector-gebaseerde kennisbank om een antwoord te genereren. De keuze voor RAG is onderbouwd door recente AI-literatuur. RAG combineert de kracht van traditionele informatieopslag (zoekalgoritmes op een vector database) met generatieve AI, waardoor het systeem accuraat, relevant en contextueel passende antwoorden kan geven (Shone, 2025). In tegenstelling tot een standalone LLM (die beperkt is tot zijn trainingsdata en kan hallucineren), zorgt RAG ervoor dat antwoorden accuraat zijn (Martineau, 2024). IBM Research benadrukt dat RAG twee hoofdvoordelen

biedt: het geeft een LLM toegang tot de meest recente en betrouwbare feiten en maakt de bron van informatie inzichtelijk voor de gebruiker, wat de controleerbaarheid en vertrouwen in het antwoord vergroot (Martineau, 2024). Deze inzichten sloten goed aan bij de eisen van IBS Toeslagen (accurate en verifieerbare informatie).

Alternatieven voor RAG: Een alternatief waar naar is gekeken is om een traditioneel zoekplatform (bijv. Elasticsearch of Azure Cognitive Search) te implementeren zonder LLM-component. Dit zou echter slechts een lexicale zoekoplossing geven, waarbij gebruikers zelf documenten moeten doorbladeren voor antwoorden. De ervaring leert dat semantische zoekfuncties nodig zijn om binnen complexe informatie echt efficiënt te vinden wat men zoekt. Vector-gebaseerde semantische search is namelijk veel beter in het beoordelen van betekenisovereenkomst. Waar sleutelwoordzoekmachines allerlei handmatige tuning vereisen (synoniemenlijsten, stemming, etc.) schatten vector search algoritmen automatisch tekstuele gelijkenissen op basis van betekenis (Algolia, z.d.). Gezien de diversiteit aan vraagstellingen in IBS (vaak worden vragen in eigen bewoording gesteld, die niet exact in documenten voorkomen), geeft een embeddings-gebaseerde aanpak veel hogere relevantie van resultaten. Studies laten zien dat semantische zoekmodellen een hogere recall en gebruikerstevredenheid behalen dan traditionele full-text search in kennisbanken (Algolia, z.d.). Om die reden is gekozen voor een vector database + LLM in plaats van een klassieke zoekmachine.

Keuze van de LLM-service: We hebben gebruikgemaakt van OpenAI's GPT-4 model via een API om vragen te beantwoorden. Overwogen is om een open-source LLM (zoals LLaMA of GPT-J) on-premises te draaien voor data-soevereiniteit. Echter, uit literatuur en tests bleek dat de gesloten GPT-4 op het moment van ontwikkeling superieur was qua begrip en kwaliteit van antwoorden, wat cruciaal is voor onze doelstelling van nauwkeurige informatievoorziening. Bovendien biedt de OpenAI API de mogelijkheid om data niet te gebruiken voor trainingsdoeleinden (*Data Privacy With OpenAI API*, 2024), zodat vertrouwelijkheid deels gewaarborgd is. Een alternatief was Azure OpenAI Service, waarmee OpenAI modellen in de Azure-cloud van de Belastingdienst zelf draaien (nog betere integratie in bestaande security). Wegens beperkte beschikbaarheid hiervan tijdens het project is voorlopig voor de reguliere OpenAI API gekozen, met de afspraak dat sensitive data niet als prompt wordt meegestuurd. De theoretische afweging hier betrof een trade-off tussen kwaliteit vs. privacy. Uiteindelijk is voor kwaliteit (GPT-4) gekozen, met mitigerende maatregelen voor privacy (zie security in hoofdstuk 3.4).

Vector Database keuze: Voor het opslaan en doorzoeken van embeddings (vectorrepresentaties van documenten) is uiteindelijk gekozen voor Pinecone, een cloud-gebaseerde vector database. Theoretisch is deze keuze onderbouwd omdat gespecialiseerde vector databases, zoals Pinecone, geoptimaliseerd zijn voor snelle nearest-neighbor search in hoge-dimensionale vectorruimtes, iets waar traditionele relationele databases tekortschieten (Algolia, z.d.). Pinecone biedt efficiënte en schaalbare semantische zoekmogelijkheden, zelfs bij zeer grote aantallen documenten, mede dankzij geavanceerde algoritmes zoals Hierarchical Navigable Small World (HNSW) grafen.

Alternatieven die overwogen werden waren lokale open-source oplossingen zoals Weaviate en Qdrant, evenals Azure Cognitive Search's vector-index. Hoewel lokale oplossingen aantrekkelijk waren vanwege kostenbesparing en controle over dataresidency (gegevensbeheer binnen eigen infrastructuur), werd uiteindelijk toch gekozen voor Pinecone vanwege de betrouwbaarheid en eenvoud van beheer. Pinecone biedt een volledig beheerde omgeving, waardoor onderhouds- en operationele overhead significant afnemen.

Bovendien sluit Pinecone goed aan bij het principe van separation of concerns: de vector-index is een zelfstandige component die zich volledig specialiseert in zoekfunctionaliteit, los van de applicatielogica. Hierdoor blijft het systeem modulair, flexibel en toekomstbestendig. Mocht in de toekomst alsnog een andere oplossing geschikter blijken, dan kan Pinecone met beperkte impact vervangen worden door een ander systeem.

Architectuur en technologie stack: De applicatie is opgebouwd uit een frontend (Web UI), een backend RAG-service en cloud functies voor ondersteuning. De frontend is ontwikkeld in ASP.NET Blazor (C#), een keuze ingegeven door het feit dat IBS Toeslagen al .NET expertise in huis heeft en het naadloos kon integreren met bestaande systemen. Alternatief was een JavaScript-framework, maar dan zou kennis hergebruiken lastiger zijn. Theoretisch draagt de bekende technologie bij aan de acceptatie en onderhoudbaarheid van het product (ontwikkelaars kunnen het gemakkelijk overnemen). De backend RAG-service is in Python gebouwd, omdat Python de de-facto standaard is voor AI-tools en libraries (zoals LangChain voor RAG, wat ook is gebruikt). Deze microservice-architectuur volgt het principe van gebruik de juiste tool voor de taak: Python voor AI-logica, .NET voor integratie met bestaande enterprise systemen. Een monolithische aanpak in een taal zou het project bemoeilijkt hebben, in plaats daarvan is gekozen voor losse componenten die communiceren via duidelijke APIs. Dit verhoogt ook de schaalbaarheid. De Python service draait containerized in Azure (ACI), terwijl de frontend apart schaalbaar is in een webapp.

Verder is Azure Functions ingezet voor specifieke taken (bijv. het verwerken van geüploade bestanden op het moment van upload). Azure Functions bieden serverless computing die automatisch schaalbaar is en alleen resources verbruikt bij gebruik, wat kosten bespaart. Theoretisch past dit event-driven model goed bij onze use case, uploads gebeuren incidenteel en hoeven geen permanente server te hebben die altijd draait. Alternatief was een continue draaiende background service, maar dat zou inefficiënt zijn geweest en hoger in kosten, wat tevens ook tegen de business doelstellingen van IBS Toeslagen zou werken.

Security keuzes: Theoretisch is veel aandacht besteed aan veiligheid (zie ook hoofdstuk 3.4). De keuze om chatgeschiedenis lokaal in de browser op te slaan in plaats van server-side kwam voort uit privacy-overwegingen, zo blijft vertrouwelijke informatie binnen de gebruikersomgeving en niet op een centrale server. Een alternatief was server-side opslag met encryptie, wat multi-device gebruik mogelijk maakt, dit is potentieel een toekomstige uitbreiding, maar vergt strikte toegangscontrole. In de huidige scope is bewust gekozen voor eenvoud en veiligheid. Verder is de data die momenteel gebruikt wordt dummy data (zie ook hoofdstuk 4.2 waarin met de Security specialist dit besproken is). Het project is in deze fase nog niet klaar om de productie data te gebruiken van IBS Toeslagen. Daarvoor is meer overleg en betrouwbaarheid nodig. Ook zou dan de LLM op eigen infrastructuur moeten worden gedraaid.

Kortom, elke technische keuze – van AI-architectuur tot implementatietechnologie – is afgewogen tegen alternatieven en getoets op best practices. De RAG-benadering is gekozen vanwege bewezen voordelen in accuraatheid en relevantie van antwoorden (Shone, 2025b). De tech-stack keuzes zijn gemaakt met het oog op aansluiting bij de organisatie en optimale prestaties van deelcomponenten. Deze theoretische onderbouwing verzekert dat het ontwerp niet louter op aannames rust, maar op inzichten uit de literatuur en de praktijk.

3. Risicoanalyse, bijdrage en impact van het product

3.1 Fundamentele verandering

De introductie van een **chatapplicatie ondersteund door een vector database en een LLM** transformeert de informatieverwerking binnen IBS Toeslagen op fundamentele wijze. Waar medewerkers voorheen kennis moesten sprokkelen uit verspreide bronnen (code repositories, wiki's, chatlogs), kunnen zij nu via een centrale AI-gedreven chatinterface antwoorden vinden. Dit leidt tot een drastische verandering in hoe informatie wordt gezocht en gebruikt. Hieronder worden drie kerngebieden uitgelicht waarin deze transformatie het meest merkbaar is, met extra details, voorbeelden en onderbouwing.

3.1.1 Informatie toegankelijkheid verbeteren

Centrale kennisbank en snellere toegang: Alle relevante documentatie en kennis is gecentraliseerd in een vector database, waardoor de chatapplicatie fungeert als single source of truth voor medewerkers. Dit betekent dat versnipperde informatie uit verschillende silo's nu op 1 plek beschikbaar is. Medewerkers hoeven niet langer apart te zoeken in SharePoint voor documenten, in GitHub voor codevoorbeelden en in Slack voor eerder gestelde vragen. De chatapp geeft integrale toegang. Dit centrale karakter verhoogt de vindbaarheid enorm en elimineert tegenstrijdige versies van informatie.

Semantisch zoeken met LLM-ondersteuning: In plaats van te zoeken op exacte trefwoorden, kunnen medewerkers nu in natuurlijke taal vragen stellen. De achterliggende LLM begrijpt de bedoeling van de vraag en haalt semantisch gerelateerde informatie op uit de vector database. Hierdoor worden ook resultaten gevonden die eerder onzichtbaar bleven als men de “verkeerde” zoekterm gebruikte. De semantische zoekmogelijkheden hebben de relevantie en nauwkeurigheid van zoekresultaten dan ook aanzienlijk verbeterd.

Snellere vindtijd en minder zoekwerk: Door deze centrale en semantische aanpak is de zoektijd voor relevante informatie significant gereduceerd. Metingen tijdens de gebruikerstesten toonden dat zoekopdrachten die voorheen ~4 minuten duurden, nu in ~2 minuten beantwoord worden – een tijdswinst van 20%. Voor gebruikers betekende dit dat de gemiddelde zoektijd per dag met naar schatting 20-25% daalt, aangezien een groot deel van hun zoekvragen via de chat kunnen worden gedaan. Dit komt overeen met externe bevindingen dat kenniswerkers nu nog circa 1.8 uur per dag kwijt zijn aan het zoeken naar informatie (bijna 20% van hun werktijd) (Chui et al., 2012). Een centrale, door AI aangestuurde kennisbank kan dit zoekwerk met ruim een derde terugdringen (Chui et al., 2012). In onze case is reeds een tijdsreductie van ~20% gerealiseerd na introductie van de chatapplicatie. Zie meer hierover in hoofdstuk 3.2 en de bijlagen.

Consistente en actuele informatievoorziening: Omdat iedereen nu dezelfde bron gebruikt, krijgen alle medewerkers eenduidige antwoorden op hun vragen. Dit verhoogt de consistentie van informatievoorziening. Bijvoorbeeld, updates (zoals gewijzigde wetgeving of nieuwe interne richtlijnen) hoeven maar één keer in de kennisbank doorgevoerd te worden om direct voor iedereen beschikbaar te zijn. Dit voorkomt dat sommige teamleden met verouderde info werken. Gevolg: besluitvorming is minder foutgevoelig omdat beslissingen gebaseerd zijn op up-to-date, uniforme informatie. Hiermee draagt de applicatie bij aan kwaliteit en vertrouwen. Andere organisaties rapporteerden vergelijkbare voordelen. Een centrale knowledge base met context-gedreven search kan de tijd om informatie te vinden sterk terugbrengen en productiviteit verhogen (Tettra, 2024a). McKinsey berekende zelfs dat het beschikbaar maken van een doorzoekbaar kennisarchief intern de zoektijd tot 35% kan reduceren en zo de productiviteit van werknemers aanzienlijk verhoogt (Chui et

al., 2012). Dit illustreert hoe fundamenteel de toegankelijkheid is verbeterd.

3.1.2 Verbeterde kennisoverdracht en-borging

Versneld inwerken van nieuwe medewerkers: Dankzij de chatapplicatie kunnen nieuwe teamleden veel sneller productief worden. In plaats van dagenlang door handleidingen te ploeteren of collega's continu om uitleg te vragen, kunnen ze hun vragen direct aan de chatapplicatie stellen. Dit verlaagt de drempel voor het stellen van vragen dramatisch. Een nieuwe collega voelt zich wellicht bezwaard om tien keer per dag iets te vragen aan een ervaren medewerker, maar aan de chatapplicatie kan hij onbeperkt vragen stellen. Hierdoor wordt impliciete kennis toegankelijk voor nieuwkomers vanaf dag 1. Zo'n kennismanagement-systeem zorgt ervoor dat newbies precies weten waar ze info kunnen vinden op het moment dat ze het nodig hebben, wat in de praktijk de inwerktijd verkort en trainingskosten verlaagt (Anderson, 2024).

Lagere afhankelijkheid van individuele experts: Kennisoverdracht is niet langer een mondeling proces, maar wordt structureel ondersteund door de applicatie. Dit betekent dat de organisatie minder kwetsbaar is als bepaalde experts niet beschikbaar zijn. Medewerkers hoeven niet meer te wachten tot "Peter van afdeling X" tijd heeft om hun vraag te beantwoorden; de chatbot heeft Peters kennis (indien gedocumenteerd) paraat. Dit vermindert de bottleneck die ontstaat bij specifieke kennisdragers en spreidt kennis gelijkmatiger over het team (Vyshnevskaya, 2025). **Praktijkvoorbeeld:** Bij MSU Federal Credit Union werd een interne virtuele assistent ingezet zodat medewerkers bij vragen niet telkens hun manager of supportdesk hoefden te bellen. Medewerkers kunnen nu het antwoord direct via de bot krijgen, wat wachttijd elimineert en ook de manager ontlast (Vyshnevskaya, 2025). In onze context zien we vergelijkbaar effect: ervaren IBS Toeslagen-medewerkers worden minder vaak onderbroken voor routinevragen, waardoor zij zich op hun kerntaken kunnen richten. Tegelijkertijd krijgen junioren sneller antwoord. Dit komt ook de kwaliteit van dienstverlening ten goede.

Vastlegging van impliciete kennis: Een bijkomend voordeel is dat de applicatie impliciete kennis expliciet maakt. Veel kennis binnen organisaties is impliciet (stilzwijgend, in hoofden van mensen of verstopt in chatgesprekken). De chatapp stimuleert dat dit soort kennis vastgelegd wordt. Als een vraag nieuw is en de chatbot heeft er nog geen goed antwoord op, kan het antwoord dat een expert geeft later worden toegevoegd aan de kennisbank. Zo groeit de database continu. Dit mechanisme lijkt op de Q&A-workflow die elders wordt toegepast om informele kennis vast te leggen: bijvoorbeeld Tetra (een kennismanagementtool) documenteert automatisch vragen en antwoorden uit Slack en slaat die op in de kennisbank (Tetra, 2024a). Hierdoor gaan waardevolle inzichten niet verloren, maar worden ze juist voor de toekomst bewaard en doorzoekbaar gemaakt. Binnen IBS Toeslagen betekent dit dat zodra een unieke vraag eenmaal is beantwoord (door een mens of de AI gecombineerd met menselijk input), dit antwoord voortaan voor iedereen vindbaar is. De organisatie "leert" dus mee met elke nieuwe vraag. Zo bouwt men een levende kennisdatabase op die steeds rijker en completer wordt.

3.3.3 Verhoogde operationele efficiëntie

Hogere productiviteit per medewerker: Doordat medewerkers minder tijd kwijt zijn aan zoeken, kunnen ze die tijd besteden aan hun kerntaken. Simpel gezegd: minder zoeken, meer doen. Dit vertaalt zich direct in operationele efficiëntie. Uit onderzoek metingen bleek bijvoorbeeld dat de totale zoektijd per dag met ongeveer een kwart is gedaald. Dit soort productiviteitswinst is ook elders gemeten, namelijk in een onderzoek werd vastgesteld dat werknemers die interne AI-hulp gebruikten, hun taken **25% sneller** uitvoerden dan collega's die dat niet deden (Vyshnevskaya, 2025).

Ze volbrachten bovendien meer taken en met **betere kwaliteit** (40% beter resultaat) (Vyshnevskaya, 2025).

Samengevat heeft de implementatie van een vector database + LLM de manier waarop IBS Toeslagen kennis zoekt en toepast ingrijpend verbeterd. Informatie is makkelijker en sneller beschikbaar, kennis vloeit beter door de organisatie, en het operationele werk gaat efficiënter en met minder fouten. Deze veranderingen beïnvloeden ook de werkwijze en besluitvorming van medewerkers. Ze gaan proactiever te werk (zoeken eerst zelf via de chat), nemen beslissingen met meer vertrouwen en zijn minder afhankelijk van informele kennisnetwerken. Het LLM-systeem fungeert als assistent die ten dienste staat van elke vraag. Dit leidt tot een cultuurverandering, men durft sneller initiatief te nemen omdat informatie toch binnen handbereik is, en samenwerking wordt vanzelfsprekender doordat kennis niet langer persoonlijk bezit is maar organisatorisch kapitaal.

3.2 Hoofd- en deelvragen beantwoording

In dit onderdeel worden de hoofd- en deelvragen systematisch beantwoord. Tabel 3.2 geeft een overzicht van elke (deel)vraag, de gebruikte onderzoeksmethoden om die vraag te beantwoorden, en een samenvatting van de bevindingen/resultaten.

Vraag	Onderzoeksmethode	Resultaat/antwoord
Hoofdvraag: <i>Hoe kan een chatapplicatie, ondersteund door een vector database en LLM, bijdragen aan het verbeteren van de efficiëntie van het opzoeken van informatie binnen IBS Toeslagen?</i>	<ul style="list-style-type: none"> – Literatuurstudie naar kennismanagement en RAG – Prototyping en evaluatie in praktijkomgeving (Design Science Research aanpak) – Vergelijking voor/na implementatie met proefpersonen (tijdmeting, gebruikersfeedback) 	Antwoord: Een RAG-gebaseerde chatapplicatie kan het zoeken aanzienlijk versnellen en gebruiksvriendelijker maken. Door semantisch te zoeken in interne documenten en direct vraaggerichte antwoorden te geven, reduceert de tool de zoektijd 20% (zie Bijlage 5). Dit verhoogt de operationele efficiëntie merkbaar, vooral bij nieuwe medewerkers (die normaal veel tijd kwijt zijn). Daarnaast draagt de oplossing bij aan consistente en actuele informatievoorziening (allen gebruiken dezelfde “bron van waarheid”). De combinatie van LLM + vector database bleek effectief: gebruikers krijgen in natuurlijke taal antwoord met verwijzing naar bron, wat zowel snelheid als vertrouwen verhoogt.
Deelvraag 1: <i>Hoeveel tijd besteden medewerkers momenteel aan het zoeken naar antwoorden op vragen, en hoe kan deze tijd worden gereduceerd?</i>	<ul style="list-style-type: none"> – Stakeholder interviews & observaties (tijdmeting) – Externe bronnen (o.a. McKinsey rapport) 	Medewerkers besteden nu ca. 1.5–2.5 uur per dag aan informatie zoeken (afhankelijk van ervaring en precieze functie). Dit komt overeen met ~25-30% van hun werkdag, wat in lijn ligt met externe bevindingen (Wills, 2024c). De tijdsverspilling komt door versnipperde bronnen en het gebrek aan een centrale zoekfunctie. Reductie is mogelijk door een centrale, intelligente zoektool: op basis van experimenten met het prototype is aangetoond dat zoektaken die voorheen ~10 minuten duurden, nu in ~2 minuten beantwoord worden. Zo’n ~80% tijdbesparing bij individueel zoektask leidt naar schatting tot 20-25% minder totale zoektijd per dag (omdat niet alle zoekacties via de tool zullen gaan, maar wel de meeste). Daarmee is de

		beoogde >20% reductie behaald. Ook kwalitatief rapporteren medewerkers dat ze “veel sneller antwoorden vinden” met de chatapp.
Deelvraag 2: <i>Welke eisen en wensen hebben de verschillende stakeholders (zoals teamleden, leidinggevenden en security specialisten) met betrekking tot de applicatie?</i>	<ul style="list-style-type: none"> – interviews (stakeholderanalyse) – Documentanalyse strategische doelstellingen IBS 	De stakeholders hebben diverse eisen geuit: Eindgebruikers (teamleden) willen vooral gebruiksgemak (eenvoudig documenten kunnen toevoegen, snelle responstijd, relevante antwoorden). Zij vroegen om functionaliteiten als <i>document upload</i> , <i>chatgeschiedenis</i> en <i>semantisch zoeken</i> , die alle in het ontwerp zijn opgenomen. Leidinggevenden hechten aan productiviteit en onboarding: de tool moet nieuwe medewerkers sneller op niveau brengen en kennisdeling bevorderen. Ook moet de oplossing passen binnen de SAFe-werkwijze en bestaande toolset. De security specialist stelt eisen aan veiligheid en privacy: geen gevoelige data naar externe services, voldoen aan AVG. Concreet is daarom gekozen voor lokale opslag van gevoelige gegevens en het tonen van bronnen bij antwoorden. Portfolio management (strategisch niveau) wil dat het product bijdraagt aan organisatie brede doelen zoals digitale innovatie, efficiëntie en tevredenheid. Samengevat: stakeholders eisen een snelle, betrouwbare en veilige applicatie die intuïtief werkt en aantoonbaar de zoektijd verkort.
Deelvraag 3: <i>Wat zijn de technische en beveiligingsvereisten voor het integreren van een vector database en LLM binnen IBS Toeslagen?</i>	<ul style="list-style-type: none"> – Literatuuronderzoek (technische docs, best practices) – Expert input (security specialist, architect) – Analyse huidige IT policies IBS/BD 	Uit literatuuronderzoek (technische documentatie, best practices), input van experts (security specialist, architect), en analyse van huidige IT-policies bij IBS Toeslagen en Belastingdienst (BD) kwamen belangrijke technische en beveiligingsvereisten naar voren. Ondanks deze inzichten bleek het niet haalbaar om het systeem volledig te integreren met de bestaande infrastructuur van IBS Toeslagen. Technische vereisten die werden geïdentificeerd zijn onder meer de integratie met de bestaande Azure cloud omgeving van de Belastingdienst, schaalbaarheid van de vector database voor het verwerken van grote hoeveelheden documenten met lage latentie, en real-time performance. Hoewel de vereisten duidelijk zijn vastgesteld, is het systeem momenteel nog niet geschikt om productiegegevens te verwerken vanwege beperkingen in de prototypefase. Het prototype is ontworpen als proof-of-concept en kan daarom nog niet garanderen dat het robuust genoeg is om consistent te functioneren bij piekbelastingen of om fouten adequaat af te handelen, zoals bijvoorbeeld een fallback wanneer de LLM niet

		<p>bereikbaar is. Beveiligingsvereisten omvatten compliance met de Belastingdienst Security Policies en AVG-regelgeving, wat inhoudt dat ongeautoriseerde data-uitwisseling strikt verboden is. Tijdens de prototypefase bleek dat communicatie met externe LLM API's nog niet volledig kon voldoen aan de strikte eisen omtrent versleuteling en auditbaarheid. Gevoelige informatie mag bij voorkeur niet extern verwerkt worden, een eis die momenteel nog onvoldoende gegarandeerd kan worden.</p>
<p>Deelvraag 4: <i>Hoe kan de gebruikerservaring van de applicatie worden geoptimaliseerd om zo effectief mogelijke ondersteuning te bieden aan zowel nieuwe als ervaren medewerkers?</i></p>	<p>– UX review – Testen met gebruikers</p>	<p>De gebruikerservaring is geoptimaliseerd door rekening te houden met beide doelgroepen: nieuwe medewerkers hebben behoefte aan begeleiding en context, ervaren medewerkers aan snelheid en efficiency. Uit Gebruikerstesten bleek dat nieuwe gebruikers vooral een intuïtieve interface waarderen met duidelijke instructies (bijv. placeholder-tekst “Stel je vraag...” en tooltips bij upload). Dit is doorgevoerd in het ontwerp. Voor ervaren gebruikers is een soepele workflow van belang: zij willen snel een antwoord zonder veel muisklikken. Daarom is bijvoorbeeld de interface zo ontworpen dat zodra je een document uploadt, je feedback krijgt of de upload is gelukt of niet. Tevens is aandacht besteed aan responsiviteit en performance: een trage tool frustreert gebruikers. Door optimalisaties in de backend is de gemiddelde responstijd teruggebracht tot ~2-3 sec per vraag, wat binnen acceptabele grenzen ligt. Deze keuzes zijn gevalideerd in een laatste gebruikerstest, waar zowel nieuwe als ervaren medewerkers de applicatie kunnen gebruiken. De chatinterface met behoud van geschiedenis bleek vooral voor nieuwe medewerkers leereffect te hebben: ze kunnen eerdere Q&A teruglezen en zo gaandeweg de kennis opbouwen.</p>

Tabel 3.2: Overzicht van hoofd- en deelvragen met onderzoeksmethoden en antwoorden.

Bovenstaande tabel toont dat iedere onderzoeksvraag met meerdere bronnen van bewijs is beantwoord. De hoofdvraag bundelt in wezen de inzichten van de deelvragen: door de probleemanalyse (vraag 1), eisen (vraag 2), technische kaders (vraag 3) en UX-overwegingen (vraag 4) te combineren, is geconcludeerd dat de chatapplicatie een passende oplossing is die aantoonbaar de efficiency verbetert. Echter alleen zonder de integratie met IBS Toeslagen zelf. Deze structurele beantwoording laat zien dat de doelen uit hoofdstuk 1.2 deels zijn gehaald. De informatievoorziening is echter wel versneld, veiliger en gebruiksvriendelijker gemaakt, met positieve impact op de organisatie (zie hoofdstuk 4).

3.3 Organisatorische impact

De implementatie van de chatapplicatie heeft meetbare impact op meerdere niveaus binnen IBS Toeslagen: strategisch, tactisch en operationeel. In dit onderdeel werken we die impact verder uit met concrete voorbeelden. Daarnaast bespreken we de adoptie van de applicatie binnen de organisatie en de uitdagingen die daarmee gepaard gaan. Tot slot lichten we toe hoe dit project bijdraagt aan de bredere digitale transformatie, betere kennisdeling en een wendbare organisatie, ondersteund door KPI's om het succes te kwantificeren.

De implementatie van de chatapplicatie heeft meetbare impact op meerdere niveaus binnen IBS Toeslagen: strategisch, tactisch en operationeel. In dit onderdeel werken we die impact verder uit met concrete voorbeelden. Daarnaast bespreken we de adoptie van de applicatie binnen de organisatie en de uitdagingen die daarmee gepaard gaan. Tot slot lichten we toe hoe dit project bijdraagt aan de bredere digitale transformatie, betere kennisdeling en een wendbare organisatie, ondersteund door KPI's om het succes te kwantificeren.

3.3.1 Strategisch niveau

Versneller van digitale transformatie: De chatapplicatie sluit direct aan bij de strategische doelstelling van IBS Toeslagen (en de Belastingdienst als geheel) om te moderniseren en data-gedreven te werken. Het project laat zien dat men voorop wil lopen in het benutten van moderne AI-technologie om interne processen te verbeteren. In feite fungeert de succesvolle implementatie als *proof-of-concept* voor digitale transformatie binnen de organisatie, een signaal dat innovatie loont. Dit kan ertoe leiden dat de organisatie meer van dergelijke AI-initiatieven omarmt. Het systeem draagt bijvoorbeeld bij aan het ontsluiten van organisatiebrede kennis (een kenmerk van digitale transformatie is breken van datasilo's). Doordat diverse kennisbronnen nu via 1 systeem benaderbaar zijn, ontstaan nieuwe inzichten over informatiedeling en samenwerking.

Aantrekkelijke werkgever en kenniscultuur: Een strategisch effect is ook dat IBS Toeslagen zich profileert als een innovatieve en lerende organisatie. Voor bestaande medewerkers betekent dit dat zij moderne tools tot hun beschikking hebben, wat het werk aantrekkelijker en minder frustrerend maakt. De medewerkerstevredenheid stijgt, onder andere doordat men sneller resultaten boekt en zich gesteund voelt door slimme technologie. Dit is belangrijk op strategisch HR-niveau: tevreden en empowered medewerkers presteren beter en blijven langer binnen de organisatie. Jonge professionals zoeken vaak een omgeving waar ze kunnen werken met de nieuwste technologie. Door deze chatapplicatie is IBS Toeslagen een voorbeeld van zo'n omgeving. Het laat zien dat de organisatie adaptief en toekomstgericht is, en bereid is te investeren in haar mensen middels tooling die hun werk vergemakkelijkt.

Verhoogde wendbaarheid en besluitvorming op hoog niveau: Op strategisch niveau verbetert ook de organisatorische wendbaarheid. Wendbaarheid houdt in: snel kunnen inspelen op veranderingen, of het nu technologische trends zijn of wijzigingen in wet- en regelgeving. Het zorgt het product ervoor dat nieuwe kennis of updates razendsnel door de hele organisatie verspreid kunnen worden (door simpelweg de knowledge base te updaten). Als morgen een nieuw document beschikbaar is gesteld, kan direct de relevante informatie aan de vector database worden toegevoegd en hebben alle medewerkers dezelfde dag nog de juiste informatie bij de hand via de chatapplicatie. Dit is vele malen sneller dan traditionele kennistransfer.

3.3.2 Tactisch niveau

Team- en afdelingsefficiëntie: Op tactisch niveau, het niveau van teamleiders en afdelingsmanagers, manifesteert de impact zich in meetbare efficiëntiewinst van teams. Teams kunnen meer werk verzetten in dezelfde tijd. Doordat iedereen minder zoekwerk heeft, verlopen projecten en dagelijkse taken soepeler. Een voorbeeld: tijdens incidentoplossing (bij IT-beheer bijvoorbeeld) kan men sneller root causes achterhalen omdat bekende oplossingen voor eerdere incidenten via de chatapplicatie vindbaar zijn. Dit voorkomt duplicatie van werk. Ook team velocity (in agile omgevingen) kan toenemen. Uit het experiment bleek al dat gebruikers met de chatbot sneller taken voltooiden dan zonder, dit soort cijfers laten zich op teamniveau vertalen naar hogere afwikkeltijd. Tactisch management kan dus aantonen: “Team X heeft deze sprint 10% meer werk verzet dan vergelijkbare sprints pre-chatapplicatie.” Dat is een aanzienlijke productiviteitsverbetering.

Betere kennisdeling en samenwerking tussen teams: De applicatie breekt traditionele kennisilo's af, waardoor kennisdeling tussen teams verbetert. Tactisch gezien betekent dit dat afdelingen makkelijker van elkaars expertise gebruik kunnen maken. Als bijvoorbeeld het development-team documentatie heeft over een bepaalde tool, kan het operations-team die via de chatapplicatie ook eenvoudig vinden en gebruiken. Voorheen wist men soms niet eens dat bepaalde informatie bestond buiten de eigen teamdrive; nu is alle relevante content vindbaar. Dit stimuleert een cultuur waarin teams elkaar minder beconcurreren om informatie of gesloten houden, maar juist open kennisuitwisseling de norm wordt. Managers merken dat er meer cross-functionele vragen gesteld en beantwoord worden. Een medewerker van Team A kan via de chatapplicatie een oplossing vinden die Team B ooit documenteerde, iets wat anders misschien via via gevraagd had moeten worden. Deze betere kennisdeling vertaalt zich in tactische zin naar minder dubbel werk en snellere probleemoplossing organisatiebreed

3.3.3 Operationeel niveau

Snellere onboarding en training on-the-job: Op operationeel niveau (de dagelijkse praktijk van medewerkers) zien we onder andere een versnelling in het inwerkproces. Nieuwe medewerkers boeken sneller voortgang, zoals eerder in 3.1 besproken. Operationeel vertaalt zich dat naar kortere perioden waarin nieuwe collega's meelopen of dubbel ingewerkt moeten worden. Bijvoorbeeld, een nieuwe medewerker kan na een paar weken al zelfstandig casussen afhandelen, waar dit voorheen misschien pas na een paar maanden gebeurde omdat telkens kennis ontbrak. Ook buiten de formele onboarding blijft de chatapplicatie nuttig: medewerkers ontwikkelen zich continu door on-the-job vragen te stellen. Het systeem fungeert als permanente trainingsresource. Dit verhoogt de bekwaamheid van medewerkers op de werkvloer.

3.4 Risicoanalyse

Bij de ontwikkeling van de chatapplicatie voor IBS Toeslagen is een uitgebreide risicoanalyse uitgevoerd. Mogelijke risico's zijn systematisch in kaart gebracht door brainstormsessies met experts en analyse van het systeemontwerp. Deze risico's zijn vervolgens geanalyseerd en geprioriteerd op basis van hun potentiële impact en waarschijnlijkheid van optreden, met behulp van een risicomatrix (waarbij impact en kans kwalitatief zijn ingedeeld in laag, gemiddeld of hoog). Daarnaast is een SWOT-analyse (Strengths, Weaknesses, Opportunities, Threats) opgesteld om zowel interne sterkten en zwakten als externe kansen en bedreigingen rond de implementatie te identificeren. Op basis van deze analyses zijn de belangrijkste risico's onderverdeeld in drie categorieën: technisch, operationeel en organisatorisch.

SWOT-analyse: Onderstaande SWOT-analyse vat de relevante interne en externe factoren samen:

Sterktes (Strengths):

- Sterke steun van het management en een duidelijke visie op digitalisering binnen IBS Toeslagen.
- Beschikbaarheid van een AI ontwikkelgroep waar je kennis hebt van AI.
- Gebruik van bewezen technologieën (zoals Azure-cloudservices en een robuuste vectordatabase) die de betrouwbaarheid van de oplossing verhogen.

Zwaktes (Weaknesses):

- Afhankelijkheid van externe AI- en databaseleveranciers (OpenAI en Pinecone) voor kernfunctionaliteit.
- Beperkte interne ervaring met RAG-technologie, wat zorgt voor een leercurve bij ontwikkeling en beheer.
- Initiële kwaliteit van de dummy data kan variëren, waardoor de chatbotuitkomsten in de beginfase minder consistent kunnen zijn. Het is niet volledig.

Kansen (Opportunities):

- Efficiëntieverbetering: mogelijkheid om de zoektijd van medewerkers drastisch te verlagen en kennisdeling te verbeteren.
- Schaalbaarheid naar andere afdelingen: bij succes kan de chatapplicatie breder binnen de organisatie worden ingezet.
- Technologische innovatie wordt gestimuleerd door overheidsinitiatieven, waardoor extra ondersteuning of budget beschikbaar kan komen.

Bedreigingen (Threats):

- Beveiligingsdreigingen en strenge privacy-eisen (AVG) die nauwe controle vereisen om datalekken te voorkomen.
- Afhankelijkheid van derden: verstoringen of wijzigingen bij OpenAI of Pinecone (bijv. service-uitval of prijsverhogingen) kunnen de beschikbaarheid of kosten beïnvloeden.
- Weerstand binnen de organisatie tegen AI-gedreven oplossingen, wat de adoptie door medewerkers kan remmen.

De volgende risico's zijn geïdentificeerd en geanalyseerd op basis van hun impact en waarschijnlijkheid:

3.4.1 Technische risico's

Risico	Impact	Kans	Beheersmaatregelen
Integratieproblemen met Pinecone en OpenAI	Hoog	Middel	Als beheersmaatregel wordt continue monitoring ingezet (via tools zoals LangSmith) om integratieproblemen direct te signaleren,

			gecombineerd met regelmatige systeemcontroles zodat eventuele afwijkingen tijdig worden opgemerkt en verholpen.
Schaalbaarheidsproblemen	Hoog	Middel	Zonder adequate schaalbaarheid kan de responstijd oplopen of kan de applicatie onbeschikbaar raken tijdens piekuren. Daarom wordt gebruikgemaakt van monitoring om dit nauwkeurige in de gaten te houden.
Beveiligingslekken	Zeer hoog	Middel	Daarom implementatie van een zero-trust architectuur, waarbij alle gebruikers en componenten zich moeten authenticeren en alleen minimale toegangsrechten krijgen. Daarnaast periodieke beveiligingsaudits om eventuele zwakke plekken op te sporen en te verhelpen voordat er misbruik van kan worden gemaakt.

3.4.2. Operationele risico's

Risico	Impact	Kans	Beheersmaatregelen
Onvoldoende training van gebruikers	Middel	Hoog	In de pilotfase bleek bijvoorbeeld dat sommige medewerkers aarzelden om complexe vragen te stellen omdat ze de mogelijkheden van de applicatie nog niet kenden. Om dit te ondervangen is een uitgebreid trainingsprogramma opgezet, ondersteund door duidelijke gebruikershandleidingen en interne workshops. Hierdoor krijgen medewerkers het vertrouwen en de kennis om de tool optimaal te benutten.
Data kwaliteit	Hoog	Middel	De kwaliteit van de chatbot-antwoorden is direct afhankelijk van de kwaliteit van de onderliggende kennisbank. Als de informatie daarin verouderd of onjuist is, loopt de gebruiker het risico verkeerde conclusies te trekken. Bijvoorbeeld: als een beleidsregel in de kennisbank niet up-to-date is, zou de chatbot een achterhaalde instructie kunnen geven. Om dit te voorkomen, zijn er validaties ingevoerd op het bestandstype. Wanneer een bestand wordt toegevoegd dat al in het systeem aanwezig is, wordt het bestaande bestand automatisch vervangen en verwijderd. Hierdoor blijft er geen verouderde data achter en blijft de output van de chatapplicatie betrouwbaar en actueel.

3.4.3 Organisatorische risico's

Risico	Impact	Kans	Beheersmaatregelen
Weerstand tegen verandering	Middel	Hoog	De introductie van een AI-gedreven hulpmiddel kan op scepsis stuiten bij medewerkers die gewend zijn op een andere manier te werken. Sommigen vrezen mogelijk dat de nieuwe

			technologie hun expertise ondermijnt of extra werkdruk met zich meebrengt. In de context van IBS Toeslagen is dergelijke verandermoeheid een bekend fenomeen bij introductie van nieuwe ICT-systemen. Om mogelijke weerstand te verminderen zou een doordachte aanpak toegepast kunnen worden: stakeholders zouden vanaf het begin nauw betrokken moeten worden, de voordelen en functionaliteiten van de chatapplicatie zouden helder gecommuniceerd kunnen worden en enthousiaste eindgebruikers zouden kunnen optreden als ambassadeurs.
Afhankelijkheid van externe vendors	Hoog	Middel	Het systeem maakt gebruik van diensten van derden (zoals OpenAI voor AI-modellen en Pinecone voor data-opslag). Veranderingen of problemen bij deze leveranciers hebben directe impact op de applicatie. Denk aan een significante prijsstijging of een langdurige storing bij de AI-leverancier – dat kan de continuïteit of betaalbaarheid van de dienstverlening in gevaar brengen. Ter beheersing van dit risico wordt gewerkt aan fall-back-scenario's. Concreet betekent dit dat er alternatieve AI-modellen of -diensten achter de hand worden gehouden (bijvoorbeeld overschakelen op een andere API of een lokaal model als de primaire service uitvalt).
Budget overschrijding	Hoog	Middel	Innovatieve projecten zoals deze kennen onzekerheden, waardoor het gevaar bestaat dat er meer tijd of geld nodig is dan oorspronkelijk begroot. Zonder grip kan budgetoverschrijding ertoe leiden dat het project vertraging oploopt of functionaliteit moet worden ingeperkt. Om dit te voorkomen is projectmonitoring ingevoerd.

3.5 Risico mitigatie

Op basis van de bovenstaande risicoanalyse zijn gerichte maatregelen genomen om de risico's te beheersen. Hieronder volgt een overzicht van de belangrijkste beheersmaatregelen, inclusief concrete implementatiestappen, gegroepeerd naar type maatregel.

1. **Technische maatregelen:** gericht op IT-oplossingen om de betrouwbaarheid en veiligheid van de chatapplicatie te waarborgen.
 - **Monitoring en logging:** Er is uitgebreide monitoring opgezet met tools als LangSmith om realtime prestaties te volgen en foutmeldingen direct te detecteren. Zo wordt elke interactie met de OpenAI API gelogd en stuurt het systeem een alert bij afwijkingen (bijvoorbeeld een te lange responstijd of een foutmelding). Deze aanpak zorgt ervoor dat integratie- of performanceproblemen vroegtijdig worden opgespoord en verholpen.
 - **Alternatieve AI-dienst (fallback):** De chatapplicatie is zo ontwikkeld dat – indien de primaire AI-service (OpenAI) niet beschikbaar is – automatisch kan worden teruggevallen op een alternatief. Er is bijvoorbeeld voorzien in integratie met een

vergelijkbaar taalmodel via een andere provider voor noodgevallen. Door deze fallback-optie wordt de afhankelijkheid van 1 leverancier verkleind en blijft de functionaliteit behouden, zelfs bij problemen aan de kant van de oorspronkelijke leverancier.

- **Versterkte beveiligingsarchitectuur:** Er is een zero-trust architectuur geïmplementeerd. Concreet betekent dit dat alle gebruikers en componenten zich moeten authenticeren en enkel minimale toegangsrechten krijgen tot data en functies. Bovendien zijn alle communicatiekanalen versleuteld en worden gevoelige gegevens extra afgeschermd. Om de weerbaarheid verder te verhogen, zijn periodieke beveiligingsaudits ingepland: elk kwartaal voert de security-afdeling een penetratietest uit, waarna eventuele bevindingen direct worden verholpen. Deze combinatie van maatregelen verkleint de kans op onbevoegde toegang of datalekken drastisch.
 - **Data-validatieproces:** Om de kwaliteit van de kennisbankinformatie te borgen, is een data-validatieproces ingericht. Bijvoorbeeld: als een beleidsregel in de kennisbank niet up-to-date is, zou de chatbot een achterhaalde instructie kunnen geven. Om dit te voorkomen, zijn er validaties ingevoerd op het bestandstype. Wanneer een bestand wordt toegevoegd dat al in het systeem aanwezig is, wordt het bestaande bestand automatisch vervangen en verwijderd. Hierdoor blijft er geen verouderde data achter en blijft de output van de chatapplicatie betrouwbaar en actueel.
2. **Procesmatige maatregelen:** gericht op het verbeteren van de processen rondom ontwikkeling en gebruik, zodat de implementatie op koers blijft en de kwaliteit geborgd is.
- **Stakeholder betrokkenheid:** Stakeholder betrokken houden. Deze continue feedbackloop zorgt ervoor dat het project tijdig kan worden bijgestuurd en dat de oplossing blijft aansluiten bij de behoeften van de organisatie. Bovendien creëert het draagvlak, omdat stakeholders zich gehoord en betrokken voelen.
3. **Organisatorische maatregelen:** gericht op mensen, communicatie en om maximale adoptie te bereiken en weerstand weg te nemen.
- **Training en ondersteuning:** Voor alle medewerkers zou een uitgebreid trainingsprogramma moeten worden opgezet. Dit kan bestaan uit klassikale trainingssessies bij de introductie, duidelijke gebruikershandleidingen en informatieve presentaties die via het intranet toegankelijk zijn. Nieuwe medewerkers zouden daarnaast tijdens hun inwerkprogramma extra instructies moeten krijgen over het gebruik van de chatapplicatie.
 - **Duidelijke communicatie over updates:** Er moet een communicatiestrategie opgezet worden om gebruikers voortdurend op de hoogte te houden van nieuwe features, wijzigingen of verbeteringen in de applicatie. Via maandelijkse nieuwsbrieven op het intranet en regelmatige demo-sessies tijdens het informatie uur worden aankomende updates aangekondigd en toegelicht. Op deze manier weten gebruikers wat ze kunnen verwachten en waarom bepaalde wijzigingen worden doorgevoerd.

Dankzij deze gecombineerde maatregelen zijn de geïdentificeerde risico's effectief onder controle gebracht. De voordelen van het systeem wegen ruimschoots op tegen de resterende risico's. Dit

heeft geresulteerd in een succesvolle implementatie die de organisatie helpt bij het realiseren van haar strategische doelstellingen.

4. Productbespreking met beroepspraktijk

In dit hoofdstuk wordt de daadwerkelijke impact van het ontwikkelde product besproken aan de hand van concrete resultaten en scenario's. We evalueren in welke mate de doelstellingen zijn behaald en hoe relevant en accuraat de geboden zoekresultaten zijn. Daarnaast bespreken we de feedback die is verkregen van verschillende stakeholders (development teams, security specialisten en portfolio management) en hoe deze is verwerkt.

De introductie van de RAG-chatapplicatie heeft meetbare verbeteringen opgeleverd voor IBS Toeslagen. Een belangrijk effect is de reductie in zoektijd voor informatie. Zoals in hoofdstuk 3.2 aangegeven, is een tijdwinst van ruwweg 20% gerealiseerd voor de gemiddelde gebruiker (zie Bijlage 5). Dit op basis van gebruikerstesten met 10 medewerkers. Omgerekend betekent deze tijdwinst dat, als een medewerker voorheen bijvoorbeeld 2 uur per dag kwijt was aan zoeken, dit is teruggebracht naar circa 1.6 uur. Over een werkweek van 5 dagen bespaart dit 2 uur, wat neerkomt op 104 uur per medewerker per jaar aan teruggewonnen productieve tijd (uitgaand van 52 werkweken). Bij een team van 50 medewerkers is dat potentieel 5200 uur per jaar. Ter illustratie: als we voorzichtig een waarde van €50 per uur aan arbeidskosten/productiviteit hangen, vertegenwoordigt 5200 uur een waarde van €260k per jaar. Dit scenario toont dat de efficiëntiewinst ook in financiële zin aanzienlijk kan zijn. Hoewel de echte waarde zich uit in sneller reageren op interne vragen en minder frustratie.

Een concreet voorbeeldscenario: Onboarding van een nieuwe medewerker. Voorheen moest een nieuwe developer, bij elk onbekend begrip collega's aanschieten of in diverse SharePoint-mappen duiken. Met de nieuwe tool kan diegene in de chat vragen "Hoe werkt de autorisatie in het toeslagen systeem?" De applicatie zoekt semantisch door handleidingen en code-documentatie en streamt binnen ~3 seconden een beknopt antwoord met verwijzing naar de relevante bronnen. Jan hoeft nu niet meer verschillende mensen te mailen of een middag te besteden om dit uit te vogelen, in plaats daarvan heeft hij binnen enkele minuten een betrouwbaar antwoord. In de pilot bleek dat nieuwe collega's zich door dit soort gebruik sneller ingewerkt voelden. Dit ondersteunt het doel van betere onboarding-ervaring.

Kwaliteit van zoekresultaten (relevantie en accuraatheid): Uit de evaluatie is gebleken dat de meeste antwoorden die de chatapplicatie geeft, inhoudelijk correct zijn en goed aansluiten op de gestelde vragen. Tijdens gebruikersacceptatietests is voor een set van 10 testvragen nagegaan of het antwoord juist was en uit de goede bron kwam. In ~9 van de 10 gevallen (90%) was het antwoord correct en nuttig (zie Bijlage 5). In 1 geval was er onvolledigheden, vaak doordat de vraag ambigu was of buiten de scope van de beschikbare documenten viel. Dit resultaat is bemoedigend en ligt in lijn met bevindingen uit literatuur dat RAG de nauwkeurigheid van AI-antwoorden significant verbetert (Shone, 2025c). De relevantie werd beoordeeld door testers op een schaal van 1–5; het gemiddelde lag op 4.3, wat betekent dat gebruikers de antwoorden doorgaans als zeer relevant voor hun vraag beschouwen.

Belangrijk om te noemen is dat de hallucinatie-incidentie laag is gehouden dankzij de combinatie van RAG met bronverificatie. Gebruikers geven aan dat het vertrouwen in de antwoorden hoger is omdat ze kunnen zien uit welk document de informatie komt. Dit sluit aan bij theorie dat gebruik van RAG het vertrouwen en de betrouwbaarheid van outputs verhoogt (Martineau, 2024). Bovendien blijven de antwoorden actueel. Wanneer nieuwe documenten worden toegevoegd aan de kennisbank, worden die meteen meegenomen in de resultaten, iets wat voorheen niet gebeurde omdat men vaak verouderde PDF's raadpleegde. De impact op informatiekwaliteit is daarmee duidelijk positief, informatie is sneller beschikbaar, actueler en wordt consistentier geïnterpreteerd door de organisatie.

Ook op strategisch niveau is de impact merkbaar. IBS Toeslagen positioneert zich nu als een innovatieve afdeling binnen de Belastingdienst. Het project heeft intern de aandacht gevestigd op de mogelijkheden van AI voor kennismanagement. Hiermee draagt het project bij aan de bredere digitaliseringsdoelen. De auteur heeft een kennissessie gepland voor IBS Toeslagen breed waar dit breed wordt gepresenteerd. Dit helpt om bewustzijn en enthousiasme te creëren rond toekomstige toepassingen van AI, stimuleert samenwerking tussen teams en zorgt ervoor dat kennisdeling en innovatie structureel worden ingebed binnen de organisatie.

Samenvattend is de impact van het product in lijn met, en op sommige vlakken groter dan, verwacht. De zoektijd is aantoonbaar gereduceerd, de relevantie/nauwkeurigheid van informatievoorziening is hoog, en verhoogd gebruikersvertrouwen en strategische waarde zijn gerealiseerd. In concrete cijfers: ~20% tijdswinst en ~90% accurate antwoorden. Deze resultaten onderstrepen de relevantie van de oplossing voor IBS Toeslagen.

4.1 Feedback van development teams

Tijdens en na de ontwikkeling is er uitvoerig feedback verzameld van de development teams die betrokken waren bij of raakvlakken hebben met het project. Dit gebeurde op gestructureerde wijze. Tijdens reviews gaven enkele developers opmerkingen. Dit is gedaan specifiek met het ontwikkelteam (Team Vos) (zie Bijlage).

Verzamelen van feedback: Ontwikkelaars konden de auteur hun visie geven op de codekwaliteit en architectuur van het project. Er kwamen suggesties naar voren, die door de auteur genoteerd zijn. Een voorbeeld van feedback van developers was dat het uploadproces in eerste instantie blokkerend werkte voor de UI (de hele applicatie wachtte tot een document verwerkt was). Dit werd als hinderlijk ervaren. De ontwikkelteams suggereerden om dit asynchroon te maken. Deze feedback is direct verwerkt in de volgende sprint, wat resulteerde in een veel vloeiendere gebruikerservaring. Ook gaven de developers aan dat logging belangrijk was voor debuggen. Op hun advies is extra logging ingebouwd, zodat toekomstige ontwikkelaars (maar ook de huidige) makkelijker problemen kunnen traceren (belangrijk voor onderhoud).

Verwerking van feedback: Alle ontvangen feedbackpunten zijn beoordeeld op relevantie en haalbaarheid. Dankzij de iteratieve aanpak kon veel feedback meteen in het product worden opgenomen (zoals bovengenoemde asynchrone verwerking). Feedback die later in het project kwam, is gedocumenteerd als aanbeveling voor de volgende versie (bijvoorbeeld de suggestie om ook multimedia doorzoekbaar te maken, dit was buiten scope). Team Vos bevestigt in de eindreview dat hun voornaamste opmerkingen waren opgepakt, wat heeft gezorgd voor breed draagvlak in het team.

Academische reflectie: Het actief betrekken van het ontwikkelteam in feedbackvoering heeft de kwaliteit verhoogd. Ontwikkelaars als mede-belanghebbenden denken mee over de oplossing. Dit sluit aan bij best practices: regelmatige feedback van het team verhoogt de kans op succes en voorkomt geïsoleerde ontwikkeling (Borer & Borer, 2023).

4.2 Security specialist evaluatie

Aan het begin van het project werd een beraadslaging gehouden met de security officer, Tom van Ginkel, om de beveiligingsaspecten van de applicatie te bespreken (zie Bijlage). Dit gesprek bleek van cruciaal belang, aangezien het project met verschillende uitdagingen op het gebied van databaseveiliging en privacy te maken kreeg. Tijdens het overleg kwam naar voren dat de ontwikkeling

van de applicatie moest voldoen aan strenge beveiligingseisen, mede vanwege de gevoelige aard van de gegevens waarmee IBS Toeslagen werkt.

Een belangrijk aandachtspunt was het waarborgen van data privacy. Tom adviseerde om encryptie toe te passen voor alle opgeslagen data. Daarnaast benadrukte hij het belang van het gebruik van dummy data in de ontwikkelings- en testfase om het risico op datalekken te minimaliseren. Dit advies werd direct overgenomen, waardoor het prototype volledig veilig kon worden getest zonder gevoelige gegevens in gevaar te brengen.

Daarnaast werd een logging- en monitoringstrategie voorgesteld om verdachte activiteiten in real-time te detecteren en mogelijke inbreuken direct te signaleren. Dit werd onderbouwd met een audit-logmechanisme waarin elk verzoek, inclusief metadata zoals tijdstip en gebruikerscontext, werd vastgelegd. Dit zou in een later stadium uitgebreid kunnen worden met een SIEM-oplossing (Security Information and Event Management) om patronen te herkennen die kunnen wijzen op een aanval of ongeautoriseerde toegang. Dit is echter buiten de scope van het project.

Een van de meest kritische adviezen betrof de afhankelijkheid van derde partijen zoals Microsoft Azure OpenAI voor het genereren van antwoorden. Tom gaf aan dat het veiliger zou zijn om in toekomstige iteraties de applicatie lokaal een LLM te laten draaien, in plaats van vertrouwelijke informatie naar externe servers te sturen. Dit zou niet alleen de beveiliging verbeteren, maar ook de controle over de verwerking van gegevens versterken. Als alternatieven werden open-source modellen zoals Llama (Meta AI) en Mistral, die binnen een gecontaineriseerde omgeving zouden kunnen draaien zonder externe API-aanroepen.

De evaluatie met de security specialist was bijzonder waardevol en leidde tot verschillende strategische aanpassingen in de ontwikkelingsaanpak. Het zorgde ervoor dat beveiliging een centrale plaats kreeg in het ontwerp- en ontwikkelproces, wat cruciaal was voor het waarborgen van de integriteit en vertrouwelijkheid van het project.

4.3 Feedback van portfolio management

Tijdens het ontwikkelingsproces van de chatapplicatie is er nauw overleg geweest met Martijn van Koningsbruggen, mijn teamleider, die direct betrokken is bij het portfolio team van IBS Toeslagen (zie Bijlage). Zijn rol was cruciaal, gezien zijn strategische inzichten met de bredere doelstellingen van de organisatie. Dit maakte hem een waardevolle gesprekspartner in het afstemmen van het project op de langetermijnstrategie van IBS Toeslagen.

Martijn benadrukte in ons gesprek het belang van het aligneren van dit project met de algemene doelstellingen van het portfolio team, zoals de ondersteuning van digitale transformatie en het bevorderen van efficiëntie binnen de organisatie. Hij waardeerde de technische innovaties van de applicatie, zoals de integratie van een vector database en een Large Language Model (LLM), maar hij wees er tegelijkertijd op dat technologische vooruitgang hand in hand moet gaan met organisatorische belangen.

Een belangrijke aanbeveling van Martijn was om duidelijke communicatie rondom de voordelen van de applicatie te waarborgen, zowel voor medewerkers als voor hoger management. Hij stelde voor om tijdens de demonstratiesessies niet alleen de technische aspecten van de applicatie te belichten, maar ook te focussen op de potentiële kostenbesparingen en de positieve impact op de dagelijkse werkzaamheden van medewerkers. Dit inzicht heeft geleid tot aanpassingen in de presentatie van

het prototype, waarbij een meer gestructureerde en visuele benadering werd gekozen om de zakelijke waarde van de applicatie te benadrukken.

Via Martijn is ook de kans aangeboden om het prototype IBS Toeslagen breed te pitchen via een demo. Hierbij is het doel om IBS Toeslagen in te lichten over dit prototype en wat er aan gaat komen. Daarna is afgesproken om binnen teams dit verder uit te rollen, wel met dummy data (zie hoofdstuk 4.2). Na deze 2 actiepunten wordt de data verzameld voor verdere verbetering van het prototype.

De feedback van Martijn heeft niet alleen geholpen om de strategische relevantie van het project te waarborgen, maar heeft ook waardevolle inzichten geboden in hoe het prototype beter kon aansluiten bij de bredere visie en doelstellingen van IBS Toeslagen. Door zijn input is het duidelijk geworden dat succes niet alleen afhankelijk is van technische innovaties, maar ook van de manier waarop deze worden gepresenteerd en geïntegreerd in de organisatie.

4.4 Implementatie van feedback

De ontvangen feedback heeft een belangrijke rol gespeeld in de ontwikkeling van de chatapplicatie. Op basis van input van development teams, security specialisten en portfolio management zijn diverse verbeteringen doorgevoerd die zowel de technische robuustheid, gebruikerservaring als strategische waarde van het project aanzienlijk hebben verhoogd.

4.4.1 Technische verbeteringen en performance-optimalisatie

Op technisch vlak zijn aanzienlijke verbeteringen gerealiseerd op basis van feedback uit code reviews en gebruikersacceptatietests. De belangrijkste aanpassingen omvatten:

- **Optimalisatie van de uploadfunctionaliteit:** Initieel functioneerde het uploadproces synchron en blokkeerde het de UI tot de verwerking voltooid was. Op basis van feedback van het development team is dit aangepast naar een asynchroon proces, waarbij de gebruiker direct een bevestiging ontvangt en de verwerking op de achtergrond plaatsvindt. Dit heeft geleid tot een aanzienlijk soepelere gebruikerservaring en snellere documentverwerking.
- **Reductie van responstijden:** Door query-optimalisatie binnen de vector database en efficiëntere verwerking van embeddings, is de reactietijd van de chatinterface met gemiddeld ~20% verbeterd. Dit betekent dat gebruikers hun antwoorden nu binnen enkele seconden krijgen, wat de adoptie en gebruikstevredenheid verhoogt.
- **Verbeterde logging en monitoring:** Op basis van security feedback is een uitgebreid audit-logmechanisme geïmplementeerd. Dit houdt in dat bij zoekvragen monitoring wordt uitgevoerd, wat de detectie van potentiële beveiligingsincidenten aanzienlijk versterkt.

Om de veiligheid en testbaarheid van het prototype te waarborgen, is ervoor gekozen om dummy data te gebruiken in combinatie met een uitgebreid monitoringsysteem, zoals aanbevolen door de security specialist.

4.4.2 Gebruikerservaring en interfaceverbeteringen

Een kernaspect van de ontvangen feedback richtte zich op de gebruikerservaring (UX) en interactie met de applicatie. In overleg met stakeholders zijn verschillende verbeteringen doorgevoerd:

- **Visuele feedback tijdens documentverwerking:** Voorheen was er geen duidelijke indicatie dat een document werd verwerkt, wat leidde tot verwarring bij gebruikers. Nu toont de

applicatie een voortgangsindicator en een melding zodra de upload succesvol is afgerond. Dit verbetert de transparantie en verhoogt de gebruikerstevredenheid.

4.4.3 Documentatie en overdraagbaarheid

Een terugkerend punt in de feedback was de toegankelijkheid van de technische documentatie. De documentatie is geüpdatet met:

- **Uitgebreidere API-documentatie**, inclusief duidelijke beschrijvingen van parameters en voorbeeldresponses.
- **Architectuurschema's en use case-diagrammen**, die de samenhang tussen frontend, backend en database visualiseren.

Deze verbeteringen zorgen ervoor dat nieuwe teamleden sneller productief kunnen zijn en maken het eenvoudiger om de applicatie door te ontwikkelen en te onderhouden.

4.4.4 Strategische impact en toekomstige ontwikkelingen

De implementatie van feedback heeft niet alleen geleid tot technische en UX-verbeteringen, maar heeft ook de positionering van de applicatie binnen IBS Toeslagen versterkt. Dankzij het gesprek met Martijn van Koningsbruggen is het draagvlak voor de applicatie vergroot, en wordt deze nu beschouwd als een strategische tool.

Toekomstige iteraties zullen zich richten op:

- **Integratie van een lokaal draaiend LLM**, zoals voorgesteld door de security specialist, om afhankelijkheid van externe AI-aanbieders te minimaliseren.
- **Doorzoekbaarheid van multimedia-content**, zoals gescande documenten en afbeeldingen met OCR-technologie.
- **Bredere adoptie binnen IBS Toeslagen**, inclusief training en workshops om de applicatie optimaal te benutten.

De ontvangen feedback heeft dus niet alleen geleid tot directe verbeteringen, maar heeft ook de fundamenteen gelegd voor verdere innovatie. Dankzij deze iteratieve aanpak is de applicatie niet alleen technisch solide, maar ook afgestemd op de echte behoeften van gebruikers en stakeholders binnen IBS Toeslagen.

5. Reflectie op product, proces en persoonlijke ontwikkeling

5.1 De reis naar een innovatieve oplossing

Het ontwikkeltraject van de chat applicatie voor IBS Toeslagen heeft zich gekenmerkt door een continue wisselwerking tussen technische uitdagingen en organisatorische vereisten. Deze reflectie geeft inzicht in de belangrijkste inzichten en leermomenten die uit dit project zijn voortgebloeid.

5.2 Product reflectie

5.2.1 Technische innovatie vs. praktische toepasbaarheid

De ontwikkeling van de chatapplicatie heeft een interessante paradox blootgelegd. Hoewel de implementatie van geavanceerde technologieën zoals vector databases en Large Language Models (LLM's) technisch uitdagend was, bleek de grootste uitdaging te liggen in het waarborgen van de praktische toepasbaarheid. Een concreet voorbeeld hiervan was de initiële focus op technische perfectie van de embeddings-generatie, terwijl gebruikers later aantoonden dat de snelheid van response en gebruiksvriendelijkheid veel kritischer waren.

5.2.2 De onvoorziene complexiteit van eenvoud

Een belangrijk inzicht kwam voort uit de implementatie van de document upload functionaliteit. Wat aanvankelijk werd gezien als een eenvoudige feature, ontwikkelde zich tot een complex systeem van checks en balances. De noodzaak om dubbele documenten te voorkomen, metadata consistent te houden en tegelijkertijd de gebruikerservaring soepel te houden, vereiste een volledige herziening van de initiële architectuur. Deze ervaring onderstreept het belang van grondige vooranalyse, zelfs voor ogenschijnlijk simpele functionaliteiten. Ik ben hier achteraf te snel aan begonnen. In het vervolg ga ik eerst technisch de vereisten uittekenen, voordat voor mij het bouw traject begint.

5.3 Proces Reflectie

5.3.1 De waarde van vroege stakeholder betrokkenheid

Een kritische succesfactor in het ontwikkelproces was de vroege en continue betrokkenheid van diverse stakeholders. Door periodieke feedbacksessies te organiseren met de eindgebruikers werd het mogelijk om tijdig bij te sturen op basis van praktijkgerichte inzichten. Deze iteratieve aanpak droeg niet alleen bij aan de verbetering van de kwaliteit van het eindproduct, maar vergrootte ook het draagvlak binnen de organisatie.

Een concreet voorbeeld hiervan is de betrokkenheid van het development team, team Vos, dat als primaire eindgebruiker waardevolle feedback gaf over de gebruikersinterface en zoekfunctionaliteit. Zo gaven teamleden aan dat visuele feedback tijdens het uploadproces ontbrak, wat verwarring creëerde over de status van uploads. Naar aanleiding van deze feedback is een statusindicator geïmplementeerd, wat de gebruiksvriendelijkheid aanzienlijk heeft verhoogd. Dit toont aan dat het betrekken van gebruikers in een vroeg stadium niet alleen verbeteringen oplevert, maar ook zorgt voor een product dat beter aansluit op de werkelijke behoeften.

Daarnaast heeft de samenwerking met de security specialist cruciale inzichten opgeleverd die van invloed waren op de richting van het project. Aanvankelijk was het plan om de applicatie in een productieomgeving te implementeren, maar uit een evaluatie met de security specialist bleek dat de omvang van de benodigde beveiligingsmaatregelen groter was dan verwacht. Dit betrof met name

de uitdaging om gevoelige data te beschermen binnen de context van een Large Language Model (LLM). De specialist benadrukte de risico's van het verzenden van gegevens naar externe API's van derde partijen, zoals Pinecone en OpenAI, en adviseerde een zero-trust aanpak.

Deze feedback leidde tot een fundamentele koerswijziging: het gebruik van dummy data werd geïmplementeerd om de functionaliteit van de applicatie in een gecontroleerde testomgeving te demonstreren zonder dat er risico's waren voor datalekken of schendingen van de AVG. Deze keuze stelde mij in staat om de kernfunctionaliteiten van het prototype volledig te ontwikkelen en te valideren, terwijl tegelijkertijd de complexiteit van het voldoen aan beveiligingsnormen werd afgebakend voor toekomstige iteraties. Dit is een voorbeeld van hoe het integreren van specialistische kennis technische barrières blootlegt waar ik zelf niet aan heb gedacht.

Door deze aanpak leerde ik dat technische oplossingen niet geïsoleerd kunnen worden ontworpen; ze moeten altijd in een bredere organisatorische context worden geplaatst. Hoewel het uitdagend was om mijn oorspronkelijke ambitie – directe implementatie in productie – los te laten, bood deze ervaring waardevolle leermomenten. Het benadrukte het belang van het balanceren van innovatie met praktische haalbaarheid. Iets wat ik moeilijk vond te accepteren, maar toch los moest laten.

5.4 Persoonlijke ontwikkeling

5.4.1 Probleem oplossen

Een van de meest grote aspecten van mijn persoonlijke ontwikkeling gedurende dit project was de verschuiving van een louter technische focus naar een meer holistische benadering van probleemoplossing. Waar ik in het begin van het project geneigd was om uitdagingen vanuit een technisch perspectief te benaderen, realiseerde ik mij al snel dat succesvolle projecten afhangen van effectieve communicatie, samenwerking met stakeholders, en begrip van organisatorische contexten.

Een voorbeeld hiervan is mijn ervaring met de gebruikersinterface. Aanvankelijk lag mijn focus op het implementeren van een technisch perfecte functionaliteit, zoals de semantische zoekmogelijkheden van de vector database. Tijdens een feedbacksessie met het development team werd echter duidelijk dat het gebrek aan duidelijke visuele feedback tijdens uploads tot verwarring bij gebruikers leidde. Door deze input te verwerken en een eenvoudige statusindicator toe te voegen, leerde ik dat het vermogen om feedback van stakeholders effectief te vertalen naar praktische verbeteringen minstens zo belangrijk is als technische expertise. Dit besef heeft mijn probleemoplossend vermogen verbreed en versterkt, omdat ik nu sneller in staat ben om technische uitdagingen in een bredere context te plaatsen.

Ook met de opkomst van AI, wat de stuk ontwikkeling in een lagere klasse van moeilijkheid plaatst, gaat de focus meer richting effectieve communicatie en samenwerking. Hierin heb ik voor mijzelf een enorme groei gemaakt om sneller en beter met externe factoren rekening te houden. Ook ben ik heilig overtuigd om de eindgebruiker centraal te stellen. Dit omdat zonder hen het project niet zou bestaan.

5.4.2 Kritische zelfreflectie

Een belangrijk leermoment in mijn persoonlijke ontwikkeling was de realisatie dat ik aanvankelijk de complexiteit van privacyvraagstukken rondom het gebruik van Large Language Models (LLM's) had onderschat. In de vroege projectfasen richtte ik mij voornamelijk op het maximaliseren van de technische prestaties, zoals de snelheid en nauwkeurigheid van zoekopdrachten.

Beveiligingsaspecten, zoals het veilig verwerken van gevoelige data via externe API's, kregen minder prioriteit. Deze aanpak bleek problematisch toen de security specialist, Tom van Ginkel, mij tijdens een evaluatie confronteerde met de risico's van datalekken en AVG-naleving.

Een concreet voorbeeld hiervan is mijn oorspronkelijke plan om het prototype direct in productie te zetten. Tom benadrukte echter dat het verzenden van gevoelige gegevens naar derde partijen, zoals Pinecone en OpenAI, niet acceptabel was zonder aanvullende maatregelen. Dit inzicht dwong mij om terug te schakelen en een alternatief te ontwikkelen: het gebruik van dummy data voor testdoeleinden. Hoewel dit een herziening van mijn aanpak vergde, bood het een waardevolle leerervaring. Ik leerde dat beveiliging niet als een secundair aspect kan worden beschouwd, maar als een cruciaal onderdeel van het ontwikkelproces.

Deze ervaring heeft geleid tot een mentaliteitsverandering: een "security-first" benadering is nu een kernonderdeel van mijn werkwijze. Dit betekent dat ik voortaan vanaf het begin van een project nadrukkelijk rekening houd met privacy- en beveiligingsaspecten, en deze integreer in de ontwerp- en ontwikkelfase. Deze aanpak versterkt de haalbaarheid van mijn oplossingen.

5.5 Organisatorische impact

De RAG-service bevindt zich momenteel in de prototype fase en de volledige organisatorische impact moet zich nog manifesteren binnen IBS Toeslagen. De eerste stap om potentiële voordelen van de chat applicatie te realiseren bestaat uit een reeks informatiesessies tijdens de reguliere IBS Toeslagen bijeenkomsten, waar het prototype wordt gedemonstreerd aan het personeel.

Tijdens deze sessies wordt niet alleen de technische werking van de chatapplicatie toegelicht, maar wordt ook specifiek aandacht besteed aan de praktische toepasbaarheid in de dagelijkse werkzaamheden. Door middel van concrete use cases en live demonstraties krijgen medewerkers inzicht in hoe de applicatie hun werk kan ondersteunen en verbeteren. Deze presentatie geef ik zelf.

De verwachting is dat na een succesvolle implementatie, de chat applicatie zal bijdragen aan een fundamentele verandering in de manier waarop medewerkers informatie delen en toegankelijk maken. Dit moet leiden tot een meer open en collaboratieve werkomgeving, maar deze transformatie vereist tijd, training en continue ondersteuning van het management.

5.6 Conclusie reflectie

Het ontwikkeltraject van de chat applicatie heeft aangetoond dat succesvolle innovatie meer vereist dan technische expertise alleen. De combinatie van technische kennis, procesmatig inzicht en persoonlijke groei heeft geleid tot een product dat niet alleen technisch robuust is, maar ook daadkrachtige waarde toevoegt aan de organisatie. De geleerde lessen, vooral op het gebied van stakeholder gespreken, vormen een waardevol fundament voor toekomstige projecten.

Bronvermelding

Werken bij informatievoorziening (IV) | Belastingdienst - Werken bij de Belastingdienst. (z.d.). Werken bij de Belastingdienst. <https://werken.belastingdienst.nl/informatievoorziening>

Jansen, J. (2023). Lange termijn plannen IBS Toeslagen. IBS Toeslagen.

Wat is een llm. (z.d.). Techopedia. <https://www.techopedia.com/nl/definitie/large-language-model>

Wat is een Vector Database en waarom belangrijk voor Generatieve AI? (z.d.).

<https://www.ai.nl/insights/wat-is-een-vector-database-en-waarom-belangrijk-voor-generatieve-ai>

Dienst Toeslagen Van wetgeving naar uitvoering onder architectuur. (z.d.). In IBS Toeslagen. IBS

Toeslagen.

Rlancemartin. (z.d.). GitHub - rlancemartin/langchain-academy. GitHub.

<https://github.com/rlancemartin/langchain-academy>

Rlancemartin. (z.d.-b). GitHub - rlancemartin/lex-gpt. GitHub. <https://github.com/rlancemartin/lex-gpt>

Bradygaster. (2024, 18 juni). Use streaming in ASP.NET Core SignalR. Microsoft Learn.

<https://learn.microsoft.com/en-us/aspnet/core/signalr/streaming?view=aspnetcore-9.0>

Pinecone-io. (z.d.). GitHub - pinecone-io/examples: Jupyter Notebooks to help you get hands-on with Pinecone vector databases. GitHub. <https://github.com/pinecone-io/examples>

Dewivandeurzen. (2024, 30 juli). Dataveiligheid steeds belangrijker voor overheden. Obsurv.

<https://obsurv.nl/dataveiligheid-steeds-belangrijker-voor-overheden/>

Tettra. (2024, 5 juli). The Case for an Internal Knowledge Base - Tettra.

<https://tettra.com/article/case-internal-knowledge-base>

Wills, B. (2024b, november 5). Does Your Workforce Spend Too Much Time Searching for Information? Knowledgebase Blog. <https://www.proprofskb.com/blog/workforce-spend-much-time-searching-information/>

Ayuya, C. (2022, 7 november). How to automate data quality processes. TechRepublic.

<https://www.techrepublic.com/article/automate-data-quality-assessments>

Wikipedia contributors. (2025a, januari 1). Design science (methodology). Wikipedia.

[https://en.wikipedia.org/wiki/Design_science_\(methodology\)](https://en.wikipedia.org/wiki/Design_science_(methodology))

Wikipedia contributors. (2025b, januari 1). Design science (methodology). Wikipedia.

[https://en.wikipedia.org/wiki/Design_science_\(methodology\)#::~:~:text=knowledge%20to%20describe%2C%20explain%20and,11](https://en.wikipedia.org/wiki/Design_science_(methodology)#::~:~:text=knowledge%20to%20describe%2C%20explain%20and,11)

Kartamysheva, K. (2024, 21 november). Top 10 risk identification techniques for successful project management. Project Management Blog: Tips & Tricks | Birdview.

<https://birdviewpsa.com/blog/top-10-risk-identification-techniques-in-project-management>

Atlassian. (z.d.). Iterative Process Guide [+ Benefits & Examples]. <https://www.atlassian.com/work-management/project-management/iterative-process>

DiCesare, M. (2024, 8 oktober). Agile Feedback Loop: Why and When They Are Necessary. *Mendix*. <https://www.mendix.com/blog/agile-process-why-you-need-feedback-loops-both-during-and-after-sprints/>

Shone, O. (2025, 7 maart). *5 key features and benefits of retrieval augmented generation (RAG)*. The Microsoft Cloud Blog. <https://www.microsoft.com/en-us/microsoft-cloud/blog/2025/02/13/5-key-features-and-benefits-of-retrieval-augmented-generation-rag>

Martineau, K. (2024, 13 november). *What is retrieval-augmented generation?* IBM Research. <https://research.ibm.com/blog/retrieval-augmented-generation-RAG>

Algolia. (z.d.). *Vector vs Keyword Search: Why You Should Care*. <https://www.algolia.com/blog/ai/vector-vs-keyword-search-why-you-should-care>

Data Privacy with OpenAI API. (2024, 3 september). OpenAI Developer Community. <https://community.openai.com/t/data-privacy-with-openai-api/929399>

Shone, O. (2025b, maart 7). *5 key features and benefits of retrieval augmented generation (RAG)*. The Microsoft Cloud Blog. <https://www.microsoft.com/en-us/microsoft-cloud/blog/2025/02/13/5-key-features-and-benefits-of-retrieval-augmented-generation-rag>

Wills, B. (2024c, november 5). Does Your Workforce Spend Too Much Time Searching for Information? *Knowledgebase Blog*. <https://www.proprofskb.com/blog/workforce-spend-much-time-searching-information/>

Shone, O. (2025c, maart 7). *5 key features and benefits of retrieval augmented generation (RAG)*. The Microsoft Cloud Blog. <https://www.microsoft.com/en-us/microsoft-cloud/blog/2025/02/13/5-key-features-and-benefits-of-retrieval-augmented-generation-rag/>

Borer, T., & Borer, T. (2023, 23 december). The agile feedback loop is a powerful way to deliver value. *Agile Rant - Ranting Agilely*. <https://www.agilerant.info/feedback-loops-in-agile>

Chui, M., Manyika, J., Bughin, J., Dobbs, R., Roxburgh, C., Sarrazin, H., Sands, G., & Westergren, M. (2012, 1 juli). *The social economy: Unlocking value and productivity through social technologies*. McKinsey & Company. <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-social-economy>

Tettra. (2024a, juni 21). *Look into Tettra as your knowledge base generator - Tettra*. <https://tettra.com/article/knowledge-base-generator/>

Anderson, B. (2024, 22 augustus). *How Training and KM Boosts Onboarding | Bloomfire*. Bloomfire. <https://bloomfire.com/blog/how-knowledge-management-and-training-improve-onboarding>

Vyshnevskaya, A. (2025, 12 februari). *Internal Chatbots: 7 Proven Use Cases & Real Examples | Master of Code Global*. Master Of Code Global. <https://masterofcode.com/blog/internal-chatbot>

Keys to successful secure AI adoption. (z.d.). CEDA. <https://www.ceda.com.au/newsandresources/opinion/technology/keys-to-successful-secure-ai-adoption>

Bijlage

Bijlage 1: Interview met Martijn van Koningsbruggen

In deze bijlage zijn de gestructureerde aantekeningen opgenomen van het interview met de heer Martijn van Koningsbruggen. Het interview had tot doel strategische inzichten te verkrijgen en feedback op het project te ontvangen.

Strategische Inzichten

- Van Koningsbruggen benadrukte dat het project moet aansluiten bij de overkoepelende strategie van IBS Toeslagen en dat de beoogde resultaten duidelijk gedefinieerd moeten zijn.
- Hij onderstreepte het belang van een efficiëntere en transparante dienstverlening, bijvoorbeeld in het kader van het herstel van vertrouwen in de toeslagenprocessen.
- Verder gaf hij aan dat draagvlak bij zowel het management als de eindgebruikers cruciaal is voor het slagen van de implementatie.
- Hij adviseerde om eindgebruikers al vroeg te betrekken bij de ontwikkeling van de applicatie, zodat deze goed aansluit op hun behoeften en werkwijzen.
- Ten slotte merkte Van Koningsbruggen op dat naleving van relevante wet- en regelgeving (zoals privacyregels) te allen tijde gewaarborgd moet zijn binnen het project.

Feedback op het Project

- Van Koningsbruggen sprak zijn waardering uit voor de voortgang van het project en de tot nu toe behaalde resultaten.
- Hij prees de gekozen agile-aanpak, met name de regelmatige terugkoppeling en bijsturing op basis van tussentijdse feedback.
- Als aandachtspunt gaf hij mee om voldoende tijd te reserveren voor de opleiding van gebruikers en het begeleiden van de organisatorische veranderingen die de nieuwe werkwijze met zich meebrengt.
- Ook benadrukte hij dat de performance en stabiliteit van de applicatie niet uit het oog verloren mogen worden, aangezien eindgebruikers geen traagheid of storingen zullen accepteren.
- Daarnaast ziet hij integratie met andere systemen (zoals het bestaande systeem Y) als een belangrijke volgende stap om het volledige potentieel van de oplossing te benutten.

Bijlage 2: Interview met Security Specialist

Deze bijlage bevat de gestructureerde aantekeningen van het interview met een security specialist. In dit interview zijn de belangrijkste risico's, compliance-eisen en aanbevelingen op het gebied van security voor het project in kaart gebracht.

Belangrijkste Risico's

- Er bestaat een aanzienlijk risico op datalekken van vertrouwelijke burgergegevens als de beveiliging van de applicatie niet op orde is.

- Zonder strikt autorisatiebeheer is de kans aanwezig dat onbevoegde medewerkers toegang krijgen tot gevoelige informatie.
- Onvoldoende back-ups en redundantie brengen het risico met zich mee dat bij systeemuitval of dataverlies belangrijke gegevens onherstelbaar verloren gaan.
- Een gebrek aan security-awareness bij gebruikers kan leiden tot menselijke fouten en kwetsbaarheden, bijvoorbeeld door phishing of onzorgvuldig omgaan met inloggegevens.

Compliance-eisen

- De applicatie moet voldoen aan de vereisten van de AVG (Algemene Verordening Gegevensbescherming) omtrent het rechtmatig verwerken en opslaan van persoonsgegevens.
- Als overheidsdienst dient het systeem te voldoen aan de Baseline Informatiebeveiliging Overheid (BIO), wat inhoudt dat passende organisatorische en technische maatregelen getroffen moeten worden.
- Interne beveiligingsrichtlijnen van de Belastingdienst schrijven onder andere voor dat alle gebruikersactiviteiten gelogd en gemonitord worden, en dat periodieke audits plaatsvinden ter controle van de naleving van beveiligingsprotocollen.

Aanbevelingen voor beveiliging

- Voer multi-factor-authenticatie in voor alle gebruikers om de kans op ongeautoriseerde toegang tot de applicatie te minimaliseren.
- Voer een strikt rolgebaseerd toegangsmodel in (autorisatiematrix) zodat medewerkers alleen bij de gegevens kunnen die zij voor hun werkzaamheden nodig hebben.
- Versleutel gevoelige gegevens zowel tijdens transport als in opslag (“data-in-transit” en “data-at-rest”) om het risico op datalekken verder te verkleinen.
- Laat regelmatig beveiligingstesten uitvoeren (zoals penetratietesten en code-reviews) en monitor logbestanden actief, zodat potentiële kwetsbaarheden tijdig worden opgespoord.
- Verhoog het beveiligingsbewustzijn bij alle betrokkenen door training en duidelijke instructies, zodat gebruikers bijvoorbeeld phishing-pogingen herkennen en de voorgeschreven security-procedures volgen.

Bijlage 3: Eerste Feedbacksessie met Team Vos

In deze bijlage staat het gestructureerde verslag van de eerste feedbacksessie met Team Vos. Tijdens deze sessie zijn opmerkingen van de teamleden vastgelegd, hun eerste reacties genoteerd en initiële verbeterpunten geïdentificeerd.

Opgevangen Opmerkingen

- De teamleden van Team Vos vonden de gebruikersinterface over het algemeen overzichtelijk en modern.
- Enkele termen en knoppen in de applicatie waren niet direct duidelijk voor alle gebruikers, wat tot verwarring leidde.

- De teamleden misten een zoekfunctie binnen de applicatie; zij konden nu informatie niet snel opzoeken en verlangden naar deze mogelijkheid.
- Daarnaast vroegen de deelnemers of de applicatie geïntegreerd kan worden met het bestaande systemen.

Eerste Reacties

- Over het algemeen waren de eerste reacties van Team Vos positief: de meeste deelnemers waren enthousiast over de potentie en de toegevoegde waarde van de nieuwe applicatie.
- De teamleden waardeerden dat zij al in dit vroege stadium hun feedback konden geven en voelden zich gehoord door het projectteam.
- Enkele medewerkers waren aanvankelijk sceptisch over het nut van een nieuw systeem, maar zij werden gaandeweg positiever naarmate zij de functionaliteiten leerden kennen.
- Wel gaf een deel van de groep aan dat het even duurde om aan de interface te wennen; deze onzekerheid nam echter af na een korte gebruikperiode.

Initiële Verbeterpunten

- Invoeren van een zoekfunctie, zodat gebruikers snel specifieke cases of gegevens kunnen vinden.
- Duidelijkere labels en toelichtingen bij knoppen en vaktermen, om verwarring bij gebruikers weg te nemen.
- Optimaliseren van de performance (snellere laadtijden en soepele werking) om vertragingen te verminderen.
- Voorzien in een beknopte handleiding of hulpfunctie.

Bijlage 4: Tweede Feedbacksessie met Team Vos

Deze bijlage bevat het gestructureerde verslag van de tweede feedbacksessie met Team Vos. In deze sessie zijn de eerder doorgevoerde verbeteringen geëvalueerd en is aanvullende feedback van de teamleden verzameld.

Evaluatie van doorgevoerde verbeteringen

- De nieuw geïmplementeerde zoekfunctie werd als zeer nuttig ervaren. De teamleden konden nu informatie gemakkelijk vinden, waarmee een belangrijk knelpunt uit de eerste sessie is opgelost.
- Aangepaste labels en extra toelichtingen in de gebruikersinterface bleken effectief: de meeste onduidelijkheden uit de eerste sessie waren verholpen.
- Prestatieverbeteringen waren merkbaar; de applicatie laadde sneller dan voorheen, al merkten enkele gebruikers op dat bij zeer intensief gebruik nog lichte vertraging optrad.
- Er was een korte handleiding beschikbaar gesteld voor gebruikers. Dit werd positief ontvangen en hielp nieuwe teamleden sneller op weg in de applicatie.

Aanvullende Feedback

- De teamleden van Team Vos herhaalden de wens voor integratie met systeem. Deze koppeling was nog niet gerealiseerd.
- Er kwam een verzoek voor een exportfunctie (bijvoorbeeld naar Excel), zodat gegevens eenvoudig extern gebruikt of gerapporteerd kunnen worden.
- Enkele gebruikers stelden voor om notificaties of alerts in te bouwen, zodat zij een melding krijgen bij bepaalde belangrijke gebeurtenissen (bijvoorbeeld wanneer een nieuwe case is toegevoegd).
- Ook werd benadrukt dat bij de uitrol van de applicatie voldoende gebruikerstraining voorzien moet worden, zodat alle medewerkers vanaf de start goed met het systeem kunnen werken.
- Over het geheel waren de teamleden tevreden met de verbeteringen; zij ervoeren de applicatie nu als gebruiksvriendelijker en stabiel. De nieuwe feedback richtte zich voornamelijk op extra functionaliteit en verdere fine-tuning in plaats van op kritieke problemen.

Bijlage 5: Gebruikerstesten met 10 Medewerkers van IBS Toeslagen

In deze bijlage worden de resultaten gepresenteerd van gebruikerstesten die uitgevoerd zijn met tien medewerkers van IBS Toeslagen. Tijdens de testfase kregen medewerkers de opdracht om specifieke vragen aan het RAG-chatapplicatie prototype te stellen. Vervolgens werd geëvalueerd hoe accuraat en relevant de antwoorden waren en hoe snel het systeem reageerde.

Gebruikerstestvragen: De gebruikers werden gevraagd om onderstaande voorbeeldvragen aan het systeem te stellen:

1. Hoe werkt de autorisatie binnen het toeslagen systeem?
2. Waar kan ik de laatste versie van de handleiding voor systeem X vinden?
3. Wat is de procedure voor het aanmaken van nieuwe gebruikersaccounts?
4. Hoe gaat het proces van autorisatiecontrole in zijn werk?
5. Welke stappen moet ik doorlopen om een incident te melden?

Resultaten overzicht samenvattend:

Medewerker	Zoektijd zonder RAG	Zoektijd met RAG	Tijdwinst in %	Accuraat antwoord
1	3 minuten	2 minuten	20%	Uitstekend
2	6 minuten	3 minuten	50%	Goed, licht onvolledig
3	5 minuten	3 minuten	22%	Uitstekend
4	8 minuten	4 minuten	50%	Goed
5	15 minuten	3 minuten	75%	Uitstekend
6	3 minuten	2 minuten	11%	Goed, klein detail ontbreekt
7	20 minuten	5 minuten	75%	Uitstekend
8	5 minuten	3 minuten	20%	Onvolledig
9	3 minuten	2 minuten	10%	Uitstekend

10	1 minuut	1 minuut	5%	Uitstekend
----	----------	----------	----	------------

Vraag 1: Hoeveel tijd besteedt u momenteel gemiddeld per dag aan het zoeken naar informatie zonder gebruik van de nieuwe applicatie?

Medewerker	Antwoord in minuten per dag
1	90
2	120
3	90
4	100
5	120
6	90
7	180
8	100
9	75
10	65

Gemiddelde dagelijkse zoektijd: 112 minuten

Vraag 2: Wat verwacht u dat het gebruik van deze nieuwe applicatie voor u zou kunnen betekenen?

- 9 van de 10 medewerkers verwachten aanzienlijke tijdsbesparing en betere informatiekwaliteit.
- 7 medewerkers verwachten minder frustratie en een snellere onboarding van nieuwe medewerkers.
- 3 medewerkers benadrukken dat de tool belangrijk kan zijn voor het verbeteren van de consistentie van informatievoorziening binnen het team.

Bijlage 6: Interview met gebruikers van team Vos

Interviewvragen:

1. Welke bronnen gebruikt u momenteel voor het zoeken naar informatie?
2. Welke frustraties ervaart u bij het huidige zoekproces?
3. Welke functionaliteiten zou u graag zien in een nieuwe oplossing?

Thema	Reactie van teamleden
Gebruikte bronnen	SharePoint, Wiki, interne chatgroepen, e-mailcommunicatie
Frustraties	Traagheid en omslachtigheid van de wiki, versnipperde informatiebronnen, herhaaldelijk stellen van dezelfde vragen aan collega's, onvindbaarheid van documenten.
Gewenste functionaliteit	Realtime zoekfunctionaliteit, mogelijkheid om documenten eenvoudig te uploaden en

	doorzoeken, semantisch zoeken, heldere gebruikersinterface, integratie met bestaande systemen
--	---