	Vergadering Raad van Bestuur
Datum	19 september 2023
Agendapunt	Agendapunt 16 Nummer 23 – 333
Onderwerp	Afspraken over het gebruik van ChatGPT (Gen-AI) binnen UWV
Directeur	SBK
Opsteller	
Portefeuillehouder RvB	Nathalie van Berkel
	Onderwerp heeft instemming van
Directeur	**Toelichting**
(CIO)	
(GD)	
(BZ)	

Door Raad van bestuur te nemen besluiten

- Kennis nemen van kansen en risico's van ChatGPT (Gen-AI) en initiatieven van CIO-Office om het gebruik binnen UWV in kaart te brengen en kansen en risico's te benutten resp. beperken
- Instemmen met afspraken rondom het gebruik van ChatGPT (Gen-AI) binnen UWV:
    1. Geef opdracht aan de Commissie Data-Ethiek om op korte termijn in samenwerking met het Multifunctionele Team (MFT) een (ethisch) kader voor het gebruik van ChatGPT (Gen-AI) op te stellen
    2. Start op korte termijn een interne communicatiecampagne om collega's op te roepen mee te denken over het gebruik om onze dienstverlening te verbeteren en te wijzen op de afspraken rondom het gebruik van ChatGPT (Gen-AI)
    3. Verplicht gebruik van de proeftuin van CIO-Office voor experimenten met ChatGPT technologie door organisatieonderdelen
    4. Blokkeer tijdelijk binnen UWV-netwerk de toegang tot ChatGPT, en waar mogelijk soortgelijke toepassingen

Reden bespreking

De ontwikkeling van generatieve AI (Gen-AI) is met de lancering van ChatGPT eind 2022 in een stroomversnelling geraakt. ChatGPT heeft in een recordtijd van twee maanden 100 miljoen gebruikers aan weten te trekken: de chatbot kan namelijk de meest uiteenlopende vragen op overtuigende wijze beantwoorden, hoewel hij soms ook schrijnende fouten maakt.

Dit heeft de interesse bij UWV'ers voor de mogelijkheden van ChatGPT en soortgelijke toepassingen geprikkeld. Velen willen ermee werken en er zijn al veel ideëen. Zo bestaan er binnen UWV al plannen voor ChatGPT-experimenten met:

- Analyseren en herschrijven van teksten op de UWV-website (K&S)
- Assistentie bij medische rapportages (SMZ),
- Verbeteren van de bruikbaarheid van handboeken voor medewerkers in de uitvoering (Uitkeren),
- Verkenningen op het gebied van ontologie, matching, softwareconversie en –generatie (Werkbedrijf),
- Voorspellen of beslissingen tot bezwaren zullen leiden (Bezwaar & Beroep).

Op 17 juli 2023 is op verzoek van de raad een artikel op DWU geplaatst waarin medewerkers zijn opgeroepen om hun mooie voorbeelden van experimenten met ChatGPT te delen, maar ook is opgeroepen om terughoudend te zijn in het gebruik van ChatGPT en dat het gebruik van ChatGPT voor het verwerken van UWV-informatie, in welke vorm dan ook, niet is toegestaan.
Op 5 oktober is een innovatiebijeenkomst (webinar) georganiseerd, waarvoor grote belangstelling is (inmiddels ca. 500 aanmeldingen). Belangrijk is om daaraan voorafgaand te verduidelijken hoe we (vooralsnog) binnen UWV om willen gaan met ChatGPT en soortgelijke toepassingen. Daarvoor is deze voorlegger bedoeld.

Wat is ChatGPT en Gen-AI?

ChatGPT is momenteel een van de bekendste Gen-AI toepassingen. Gen-AI bestaat uit een algoritme dat op basis van een handmatig of automatisch ingevoerde instructie contextuele tekst-of beeldinformatie genereert (vraag-antwoord). Voor de samenstelling van die informatie wordt een stelsel van opgeslagen geclassificeerde gegevens ("large languagemodel"-LLM) gebruikt. Gen-AI bestaat een kleine vijf jaar in een operationele vorm en is een van de vele toepassingen van AI.

ChatGPT, een commerciële online Gen-AI dienst, is een merk van OpenAI. Microsoft is een sleutelpartner en gebruikt het OpenAI-LLM (GPT-4) voor zijn zoekservice Bing. Microsoft biedt de ChatGPTOpenAIservice aan via Azure. Gen-AI wordt aangeboden door een snelgroeiend aantal veelal buitenlandse partijen, niet alleen de Big-Tech bedrijven.

Afspraken voor verantwoord gebruik van Gen-AI

ChatGPT, en Gen-AI toepassingen in het algemeen, kunnen interessante mogelijkheden bieden die we goed moeten onderzoeken. Tegelijkertijd brengen ze ook risico's met zich mee, zoals:

- Fouten en desinformatie
- Bias
- Privacyschendingen
- Schendingen van auteursrecht

Om op de juiste wijze de voordelen van ChatGPT en overige Gen-AI toepassingen te kunnen benutten, is het van belang om goede interne afspraken te maken. Daarbij is het uitgangspunt dat UWV pilots en experimenten zoveel mogelijk wil stimuleren en faciliteren.

Het is goed om te weten dat op dit moment Europese regelgeving wordt voorbereid voor AI, waarin Gen-AI nu ook wordt meegenomen, maar het duurt nog zeker twee jaar voor die in werking treedt. Dat is te lang, gezien de enorme belangstelling die er nu is voor ChatGPT enerzijds en de impact en de risico's van deze technologie anderzijds.

1. Laat op korte termijn een (ethisch) kader voor ChatGPT (Gen-AI) opstellen door Commissie Data-Ethiek in samenwerking met het Multifunctionele Team (MFT) Gen-AI

Binnen UWV zijn al veel afspraken gemaakt rondom de omgang met ICT-systemen en datatoepassingen, bijvoorbeeld in de governance, UWV Beleidskaders Privacy, de Gedragscode en het Kompas Data Ethiek. Desondanks brengen ChatGPT en soortgelijke Gen-AI toepassingen nieuwe risico's met zich mee waardoor (aanvullende) interne afspraken (beleid) nodig zijn over het gebruik ervan.

We verzoeken de RvB om opdracht te geven aan de Commissie Data-Ethiek om in samenwerking met het MFT Gen-AI[1] een (ethisch) kader voor het gebruik van ChatGPT en soortgelijke Gen-AI oplossingen op te stellen. Dit kader is een aanvulling op de bestaande en in ontwikkeling zijnde kaders[2] en besteedt specifieke aandacht aan de risico's van Gen AI, zoals desinformatie en bias. Voor een dergelijk (ethisch) kader zijn inmiddels voldoende goede voorbeelden te vinden[3], maar zullen ook beelden worden opgehaald bij mede publieke dienstverleners, zoals de SVB de Belastingdienst[4], en bij het ministerie van SZW.

2. Start op korte termijn een interne communicatiecampagne om collega's op te roepen mee te denken over het gebruik om onze dienstverlening te verbeteren en te wijzen op de afspraken rondom het gebruik van ChatGPT (Gen-AI).

Een eerste berichtgeving hieromtrent is op DWU geweest. Het is belangrijk om deze communicatie te herhalen, uit te breiden en daarbij ook een uniforme boodschap uit te dragen. Daarbij kunnen collega's worden opgeroepen om zich bij het UWV-brede innovatieprogramma en het Innovatieteam van CIO-office te melden om mee te denken over potentiele Gen-AI toepassingen binnen UWV. Maak daarbij duidelijk dat de regels die de Gedragscode UWV stelt voor het omgaan met vertrouwelijke informatie (hoofdstuk 3) en de 11 richtlijnen voor het omgaan met social media (hoofdstuk 5) van toepassing zijn bij het gebruik van ChatGPT en Gen-AI toepassingen in het algemeen.

3. Stel een beveiligde proeftuin open voor organisatieonderdelen (niet voor individuele medewerkers) en maak het gebruik daarvan verplicht

CIO-Office Innovatie realiseert een ChatGPT proeftuin waarbij gebruik gemaakt wordt van door UWV ingekochte software Microsoft Azure in een afgesloten UWV Azure omgeving. Een belangrijk aspect van de beveiliging is dat gegevens en documenten bij experimenten in deze omgeving, binnen de proeftuin blijven en niet worden

[1] In het MFT zijn o.a. CIO-Office, SBK, Architectuur, Leveranciersmanagement, CISO-Office en JZ vertegenwoordigd.
[2] Zoals het model (algoritme) risico management beleid
[3] Onlangs heeft het IPO vergelijkbare kaders gepubliceerd, onder leiding van de voorzitter van onze Commissie Data-Ethiek.
[4] De SVB en de Belastingdienst geven voorzichtige reacties bij onze uitvraag en geven vooral aan nog na te denken over inzet van Gen AI zoals ChatGPT.

verwerkt in de onderliggende modellen[5] van ChatGPT (Gen-AI) die wereldwijd worden gebruikt. Op die manier worden inbreuken op de gegevensbescherming voorkomen.

In samenwerking met het innovatieprogramma (programma Dienstverlening) zal een oproep worden gedaan om kansrijke pilots te inventariseren waarvoor de proeftuin kan worden ingezet. Organisatieonderdelen die gebruik willen maken van de proeftuin kunnen casussen indienen bij het MFT Gen-AI. Een zorgvuldige toets zal bekijken of de casus aan de opgestelde (ethische) kaders voldoet, alvorens ook echt toestemming wordt verleend tot gebruik van de proeftuin. De proeftuin is niet bedoeld voor individuele medewerkers die het voor eigen werkzaamheden willen gebruiken.

Door het gebruik van de proeftuin verplicht te maken, krijgt het al bestaande MFT Gen-AI zicht op alle initiatieven binnen UWV met Gen-AI. Het is bedoeld om de relevante initiatieven in beeld te brengen, te adviseren over het gebruik van de (ethische) kaders en om de kennis en ervaring te delen die in de proeftuin wordt opgedaan.

Voordat we de beveiligde proeftuin gebruiken, leggen we op korte termijn een stevig fundament voor het (ethisch) kader dat we ontwikkelen. De ervaringen die in de proeftuin worden opgedaan gebruiken we vervolgens om het kader verder aan te scherpen en uit te diepen. Theorie leert op die manier van praktijk en vice versa.

4. Blokkeer tijdelijk binnen UWV-netwerk de toegang tot ChatGPT, en waar mogelijk soortgelijke toepassingen

De privacyrisico's van het gebruik van vrij toegankelijke Gen AI toepassingen als ChatGPT zijn aanzienlijk en de verwerking van persoonsgegevens is niet transparant. Organisaties om ons heen (zoals de SVB en de Belastingdienst) zijn dan ook terughoudend met het gebruik van ChatGPT. Ook de Autoriteit Persoonsgegevens (AP) heeft zijn zorgen uitgesproken over de omgang met persoonsgegevens bij organisaties die gebruikmaken van Gen AI en ChatGPT. De AP heeft softwareontwikkelaar OpenAI per brief om opheldering gevraagd over chatbot ChatGPT. De AP wil onder meer weten hoe OpenAI omgaat met persoonsgegevens bij het trainen van het onderliggende systeem.

De belangrijkste privacyrisico is dat data die in een openbare toepassing als ChatGPT wordt verwerkt, wordt gedeeld met het bedrijf dat erachter zit. Denk aan een verzekeringsarts die een andere kijk op een cliëntendossier verlangt of ondersteuning zoekt in het vergroten van de leesbaarheid van zijn of haar rapportages. Eventuele persoonsgegevens die worden opgenomen in de uitvraag bij ChatGPT komen bij het Amerikaanse OpenAI terecht en wordt hiermee een datalek.

Medewerkers die voorop lopen in het gebruik van tooling als ChatGPT, de innovators, hebben zich vaak verdiept in de mogelijkheden en onmogelijkheden van de technologie en zullen zich bewuster zijn van de privacyrisco's die hierbij komen kijken dan de groepen die hierop volgen. Zij worden door collega's of door artikelen gewezen op de vele voordelen en zijn zich minder bewust van de risico's. Eventueel kan voorlichting en training de kans op fouten verkleinen, maar nooit geheel wegnemen.

We adviseren de RvB om, gezien de hoge privacyrisico's, de toegang tot ChatGPT, en waar mogelijk soortgelijke toepassingen, vanuit de UWV-omgeving tijdelijk te blokkeren. Gelijktijdig werken we aan een zorgvuldige inrichting van een proeftuin met duidelijke kaders. Nadat meer ervaring is opgedaan met de kaders en voldoende zicht is op de privacyrisico's en maatregelen, kunnen de risico's op het juiste niveau vooraf worden geaccepteerd. Dit in lijn met het advies van de FG. Hierna kan, indien gewenst, besloten worden om ChatGPT, en soortgelijke toepassingen, alsnog weer toegankelijk te maken. In de tussentijd kan worden nagedacht hoe medewerkers op de hoogte kunnen worden gebracht van de ontwikkelingen, bijvoorbeeld door informatie op de blokkade-pagina's te weergeven.

Het tempo waarmee nieuwe Gen-AI diensten ontstaan en de mate waarin deze soms zijn verweven met niet Gen-AI toepassingen, zo zouden automatische suggesties in Microsoft Teams Chat ook als Gen-AI gezien kunnen worden, maakt dat blokkeren niet altijd mogelijk zal zijn. We blijven daarom goed kijken wat hier mogelijk is en welke aanvullende maatregelen we kunnen treffen, zoals het bieden van goede voorlichting en het wijzen op naleving van de Gedragscode UWV.

Gevolgen voor mensen

Het is nog te vroeg om te overzien of ChatGPT (en Gen-AI) voor betere dienstverlening gaat zorgen met zichtbare gevolgen voor cliënten en medewerkers. Het potentieel lijkt wel erg groot, daarom is het belangrijk de kansen en risico's snel beter in kaart te brengen.

Kansen en risico's voor (de opdracht van) UWV

Kansen

[5] Large Language Models

Gen-AI, zoals ChatGPT, kan snel en gemakkelijk tekst genereren, samenvatten en vertalen. Dit kan medewerkers veel tijd besparen bij het uitvoeren van taken zoals het schrijven van artikelen, het samenstellen van rapporten of het vertalen van documenten. Bovendien kan het helpen om de kwaliteit van brieven en rapportages te verbeteren. Ook kan de technologie een rol spelen in het analyseren van ongestructureerde data ten behoeve van bijvoorbeeld het beantwoorden van vragen (van clienten of medewerkers).

Risico's

Fouten en desinformatie

Een Gen-AI toepassing als ChatGPT is getraind op miljoenen teksten van het internet, de trainingsset. De herkomst daarvan is niet bekend, dus de kwaliteit is ook niet te controleren. ChatGPT beantwoordt vragen door de daarin gebruikte woorden te vergelijken met verwante woorden uit de trainingsset en die in een plausibele volgorde achter elkaar te zetten. De relatie tussen die woorden is altijd statistisch (hoe vaak komt woord A na woord B voor), nooit inhoudelijk (semantisch). Als er fouten of onwaarheden in de trainingsset staat, beantwoordt ChatGPT vragen ook met fouten of onwaarheden. Voor alle duidelijkheid: ChatGPT, en andere Gen-AI, verwerkt geen kennis en kan niet redeneren zoals mensen dat doen en dus ook geen fouten in het onderliggende model of de daaruit afgeleide antwoorden herkennen.

Bias

Op een vergelijkbare manier kan de output van Gen-AI toepassingen ook bias, vooroordelen, bevatten. Als de teksten in de trainingsset overwegend afkomstig zijn van witte, hoogopgeleide Westers-georiënteerde mannen van boven de 50, zal dat in de antwoorden gereflecteerd worden. Het is bijvoorbeeld bekend[6] dat driekwart van de teksten over AI in de trainingsset van Google zeer positief is over de mogelijkheden van AI, terwijl deskundigen daar veel terughoudender in zijn. Het zeer positieve klinkt dan door in de meeste antwoorden en vertekent op die manier de "waarheid".

Privacyschendingen

Het is niet bekend in welke mate de trainingsset persoonsgegevens bevat en dat is op zich al een probleem. Volgens de AVG zou dit controleerbaar moeten zijn (inzagerecht). Daarnaast kunnen gebruikers van Gen-AI persoonsgegevens voorleggen in hun vragen, ook van derden. Dat zijn in principe datalekken.

Schending van auteursrecht

Zolang niet bekend is welke informatie in de trainingsset is verwerkt, is niet duidelijk of daarmee auteursrechten zijn geschonden. Het is ook denkbaar dat Gen-AI antwoorden geven die bestaan uit auteursrechtelijk beschermde informatie. Dit risico is overigens groter bij AI-toepassingen die afbeeldingen samenstellen op basis van vragen van gebruikers en een trainingsset van afbeeldingen. DALL-e, van dezelfde ontwikkelaar als ChatGPT, is daar een voorbeeld van. Het is overigens twijfelachtig of schendingen van auteursrecht juridisch kunnen worden aangetoond. De eerste rechtszaken lopen nu in de Verenigde Staten.

Strategische aspecten van het besluit

Er is geen directe relatie met de UWV-strategie of -werkagenda.

Bedrijfsvoering (financieel/personeel)

C-ICT/CIO moet mensen inzetten om de voorgestelde afspraken uit te voeren (verzamelen van meldingen, laten inrichten van een proeftuin). Ook de deelnemers aan het MFT moeten hiervoor tijd krijgen.
De kosten van de proeftuin worden binnen de bestaande budgetten opgevangen.

Duurzaamheid

Het is bekend dat het ontwikkelen van Gen-AI toepassingen buitensporig veel energie vergt. Op dit moment zijn daar geen goede beheersmaatregelen voor.

Vervolgtraject besluitvorming

De voorgestelde afspraken kan UWV zelfstandig maken. Het is wel denkbaar dat de ontwikkelingen op korte termijn om nieuwe afspraken vragen, maar het is lastig die te voorspellen.

Communicatie

Het is belangrijk de afspraken snel en afdoende bekend te maken binnen de organisatie, gezien de risico's.

[6] www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/

Advies FG

Mijn advies is om generatieve AI-toepassingen, (Gen-AI) zoals ChatGPT, niet als 'niet-risicomijdend' in te zetten en in elk geval het gebruik van vrij toegankelijke AI toepassingen niet toe te staan. De privacyrisico's zijn aanzienlijk en de verwerking van persoonsgegevens is niet transparant.

Ik adviseer om op korte termijn het UWV-standpunt m.b.t. de inzet van GEN-AI-toepassingen helder te communiceren naar alle UWV-medewerkers. Gelijktijdig kan dan in rustiger tempo gewerkt worden aan een zorgvuldige inrichting van een proeftuin met duidelijke kaders.
Ik adviseer te overwegen om de toegang tot generatieve AI-toepassingen vanuit het UWV-netwerk in de tussentijd tijdelijk te blokkeren. Zie bijlage voor een uitgebreidere reactie.

Openbaarheid

Deze documenten kunnen openbaar gemaakt worden:

☒ Ja, in hun geheel
☐ Deels, markeer in de documenten wat niet openbaar gemaakt kan worden.
☐ Nee, de bijbehorende bijlage(n) niet.
☐ Nee, helemaal niet.

Metadata

Omschrijving: Om veilig met ChatGPT te kunnen experimenteren zijn een proeftuin en een ethisch kader nodig. Organisatie-onderdelen maken altijd gebruik van de proeftuin, individuele medewerkers kunnen op hun werkplek ChatGPT niet meer gebruiken.

Trefwoord(en): ChatGPT, AI, experimenteren, proeftuin, ethisch kader