Pensioenfederatie
Prinses Margrietplantsoen 90
2595 BR Den Haag
Postbus 93158
2509 AD Den Haag
T +31 (0)70 76 20 220
info@pensioenfederatie.nl
www.pensioenfederatie.nl

KvK Haaglanden 52988368
IBAN: NL20TRIO0212407368
BIC: TRIONL2U

CONSULTATIEVERSIE

KENMERK: B/2024/226/OH DATUM: 12-06-2024
ONDERWERP: Gedragslijn AI en Ethiek voor de pensioensector

Gedragslijn 'AI en Ethiek voor de Pensioensector'

Consultatieversie

Deze versie (B/2024/226/OH, datum 12-06-2024) wordt aan de leden van de Pensioenfederatie aangeboden op de Algemene Ledenvergadering van 27 juni 2024. Na de presentatie van deze gedragslijn, wordt een consultatieperiode geopend. Wij roepen leden op om hun feedback op deze consultatieversie te sturen naar de Pensioenfederatie, t.a.v. Otto Hulst (otto.hulst@pensioenfederatie.nl). Reacties kunnen tot en met 11 oktober 2024 ingezonden worden.
Alle ontvangen feedback wordt verwerkt, en de definitieve versie van de gedragslijn zal op de Algemene Ledenvergadering van de Pensioenfederatie op 12 december 2024 ter stemming worden gebracht.

Inleiding

Artificial Intelligence (verder AI) of Kunstmatige Intelligentie is wereldwijd bezig met een opmars in de financiële sector. AI brengt naast substantiële kansen ook risico's met zich. Enerzijds kan AI leiden tot betere dienstverlening voor deelnemers en meer gepersonaliseerde producten, en biedt het Pensioenfondsen de kans om kosten te verlagen. Anderzijds zijn er risico's, onder andere op het gebied van datakwaliteit, gegevensbescherming, uitlegbaarheid, incorrecte resultaten, discriminatie en uitsluiting, en een hogere mate van afhankelijkheid van derde partijen.

Van pensioenfondsen mag dan ook verwacht worden dat zij AI verantwoord inzetten, om deze risico's te beheersen. Dit wordt ook verlangd vanuit de wetgeving, waarop vervolgens de toezichthouders AFM en DNB ook toezicht houden. Echter, de geldende wetgeving voor Pensioenfondsen op het gebied van AI is nog niet heel concreet. Vandaar dat de Pensioenfederatie een Gedragslijn AI en Ethiek voor de Pensioensector heeft ontwikkeld, zodat Pensioenfondsen handvatten krijgen hoe zij verantwoord om kunnen gaan met het toepassen van AI.

Het ethisch kader richt zich niet alleen op de techniek. Dat zou ook een te beperkte uitleg zijn van AI en te weinig recht doen aan de potentie van AI. Er wordt een koppeling gemaakt tussen governance, technologie en cultuur. Vanwege dit brede kader en de aankomende AI-verordening van de EU is ook gekozen voor een gedragslijn. Een gedragslijn voor de pensioensector zorgt voor uniforme uitleg en het geeft een duidelijk signaal af: de pensioensector vindt AI een belangrijk onderwerp en neemt dit serieus.

Door het grote belang van AI, is het niet gebruiken van deze technologie ook een risico. Wanneer niet voldoende geëxperimenteerd wordt, zal dat nadelig uitwerken op de concurrentiepositie ten opzichte van (buitenlandse) financiële instellingen en niet-financiële partijen die zich ook op de Nederlandse (pensioen)markt begeven. Deze gedragslijn geeft pensioenfondsen de kaders waarbinnen het kan experimenteren met AI.

Ten vierde reden is gekozen voor een gedragslijn omdat de Europese wetgever financiële instellingen en brancheorganisaties stimuleert om gedragscodes op te stellen voor AI dat geclassificeerd kan worden als ‘”lage of minimale risico's”. Elke sector heeft zijn eigen unieke karakter, waarbij een gedragslijn op dat unieke karakter kan inspelen.

Gezien de snelle technologische en maatschappelijke ontwikkelingen zal deze gedragslijn een levend document zijn waarbij jaarlijks gekeken zal worden of er aanpassingen nodig zijn. De gedragslijn heeft tot doel om binnen de pensioensector een verantwoord gebruik van AI-toepassingen te bewerkstelligen. De gedragslijn is technologisch neutraal opgesteld en is principle-based. Uitgangspunten voor de gedragslijn zijn eerlijkheid en gelijkheid, transparantie en uitlegbaarheid rondom het gebruik van AI. De gedragslijn geeft niet aan op welke wijze (HOE) het Pensioenfonds en/of zijn uitbestedingspartijen invulling moeten

geven aan de principes. Daarvoor is de pensioensector te divers. Belangrijk uitgangspunt hierbij is dat de (strategische) doelstellingen van Pensioenfondsen grotendeels overeenkomen, hetgeen zich vertaalt in het uitgangspunt dat Pensioenfondsen samenwerking en kennisdeling op het gebied van inzet van AI binnen de pensioensector stimuleren.

Binnen dit kader wordt AI als volgt gedefinieerd: *een machinaal systeem dat is ontworpen om met verschillende niveaus van autonomie te werken en dat na de uitrol aanpassingsvermogen kan vertonen en dat, voor expliciete of impliciete doelstellingen, uit de ontvangen input afleidt hoe output te genereren zoals voorspellingen, inhoud, aanbevelingen of beslissingen die van invloed kunnen zijn op fysieke of virtuele omgevingen.*

In de bijlage is een definitielijst opgenomen met de belangrijkste definities. Voor alle overige definities sluiten we aan bij de definitielijst in het rapport ‘Verkennen kansen AI’ van SIVI.

Scope gedragslijn

De Gedragslijn AI en Ethiek voor de Pensioensector is gericht op alle toepassingen van AI die door de pensioensector wordt ingezet. Processen die zien op klantcommunicatie of het selecteren van financiële instrumenten, het handhaven van de verplichtingstelling en optimalisatie van repeterende werkzaamheden zijn logische aandachtsgebieden die binnen de scope van de gedragslijn vallen. Het startpunt is een proces en of daar AI voor wordt ingezet. De definitie van AI is hierboven weergegeven.

De concept AI-verordening is gericht op de bescherming van belangrijke waarden en grondrechten van de Europese Unie, zoals veiligheid en non-discriminatie. Deze concept Verordening maakt daardoor onderscheid tussen AI-toepassingen met de risicoclassificatie onacceptabele risico's (verboden AI-praktijken), hoge risico’s en lage of minimale risico's. AI dat valt onder de classificatie onacceptabele risico's, mag niet gebruikt worden door Pensioenfondsen.

Ethisch kader

Dit is een ethisch kader voor het inzetten van AI-toepassingen bij Pensioenfondsen die lid zijn van de Pensioenfederatie. Het geeft weer waar Pensioenfondsen en hun uitbestedingspartijen in de keten voor staan bij het gebruik van AI en andere datagedreven producten en processen. Het kader vormt een bovenwettelijk uitgangspunt bij de inzet van moderne technieken: als een bepaalde techniek wettelijk toegestaan is, doch strijdig met deze principes, zullen Pensioenfondsen en hun uitbestedingspartijen deze niet toepassen dan expliciet en gemotiveerd aangeven waarom het Pensioenfonds deze techniek toch inzet.

Het kader is gebaseerd op de aanbevelingen van de High-Level Expert Group on Artificial Intelligence en de AI-verordening van de Europese Unie.

Dit adviesorgaan van de Europese Commissie bepaalde dat er bij ethisch gebruik van AI zeven vereisten voor verantwoordgebruik in acht moeten worden genomen:

1. Verantwoording
2. Menselijke autonomie en controle
3. Technische robuustheid en veiligheid
4. Privacy en data governance
5. Transparantie
6. Diversiteit, non-discriminatie en rechtvaardigheid
7. Maatschappelijk welzijn

In de gedragslijn zijn per vereiste de principes voor Pensioenfondsen uitgewerkt die van invloed zijn op het deelnemersvertrouwen[1]. Denk hierbij tenminste aan kernprocessen zoals verwerken deelnemersmutaties, innen van premies, berekenen van de aanspraken/individuele pensioenpotjes na invaren, uitkeren van pensioenen, inkomende en uitgaande waarde-overdrachten, interne adviseringen klantendienst (inclusief deelnemersportaal). Maar ook binnen processen van Vermogensbeheer worden naar verwachting steeds meer AI-toepassingen ingezet, bijvoorbeeld besluitvorming rondom beleggingsstrategieën.

Naarmate het (reputatie)risico hoger wordt bij de bewuste AI-toepassing (denk aan berekenen aanspraak, pensionering, uitkeren) zal een striktere aantoonbare naleving van de principes van dit kader meer gewenst zijn. De principes zijn ook

[1] Denk aan de inzet van chatbots, straight through processing, gebruik van biometrische data, social media data etc.

van toepassing bij externe inkoop van data en/of technieken waarbij AI wordt ingezet.

NB! Daar waar in het principe Pensioenfonds staat opgenomen wordt voor deze gedragslijn het Pensioenfonds en zijn uitbestedingspartijen bedoeld. Hierbij is het niet relevant of het een kritieke dan wel niet-kritieke uitbesteding betreft.

Onderstaand worden per vereiste de principes opgenomen. In totaal bestaat het ethisch kader uit 21 principes. We willen de lezer meegeven dat enkele principes ook onder meerdere vereisten geplaatst zouden kunnen worden.

Verantwoording

Er moeten mechanismen worden ingevoerd om de verantwoordelijkheid en verantwoordingsplicht voor AI-systemen en de resultaten ervan te waarborgen, zowel voor als na de uitvoering ervan. Controleerbaarheid van AI-systemen is in dit verband van cruciaal belang, aangezien de beoordeling van AI-systemen door interne en externe auditors en de beschikbaarheid van dergelijke evaluatieverslagen sterk bijdragen aan de betrouwbaarheid van de technologie. Externe controleerbaarheid moet met name worden gewaarborgd bij toepassingen die van invloed zijn op de grondrechten, met inbegrip van veiligheidskritieke toepassingen. Potentiële negatieve effecten van AI-systemen moeten worden geïdentificeerd, beoordeeld, gedocumenteerd en tot een minimum worden beperkt. Het gebruik van effectbeoordelingen vergemakkelijkt dit proces. Deze beoordelingen moeten in verhouding staan tot de omvang van de risico's die de AI-systemen met zich meebrengen. Afwegingen tussen de vereisten – die vaak onvermijdelijk zijn – moeten op een rationele en methodologische manier worden aangepakt en er moet rekening mee worden gehouden. Ten slotte moet er in geval van een onrechtvaardig negatief effect worden voorzien in toegankelijke mechanismen die een adequaat rechtsmiddel waarborgen.

Een Pensioenfonds hanteert met betrekking tot verantwoording de volgende principes:

Principe 1: Beleid
Een Pensioenfonds stelt beleidsregels op voor de inzet van AI-toepassingen binnen zijn processen. In deze beleidsregels wordt ten minste aandacht besteed aan de risicoanalyse en factoren die bij de risicoanalyse van belang zijn. Ook de naleving van deze gedragslijn moet in het Fondsbeleid zijn opgenomen.

Principe 2: Beheersbaarheid, aantoonbaarheid en controleerbaarheid
Een Pensioenfonds zorgt voor een intern controle- en verantwoordingsmechanisme voor het gebruik van AI-systemen, de resultaten en de gebruikte databronnen.

Principe 3: Deskundigheid
Een Pensioenfonds bevordert de kennis van bestuurders, interne toezichthouders en medewerkers ten aanzien van AI-toepassingen.

Principe 4: Klachten
Een Pensioenfonds informeert gebruikers, in lijn met de generieke klachtenprocedure, ook bij AI-toepassingen over de mogelijkheden om klachten in verband met het gebruik van deze AI-toepassingen kenbaar te maken.

Menselijke autonomie en controle

AI-systemen moeten individuen ondersteunen bij het maken van betere, beter geïnformeerde keuzes in overeenstemming met hun doelen. Ze moeten fungeren als katalysatoren voor een bloeiende en rechtvaardige samenleving door menselijke keuzevrijheid en grondrechten te ondersteunen, en niet de menselijke autonomie te verminderen, te beperken of te misleiden. Het algehele welzijn van de gebruiker moet centraal staan in de functionaliteit van het systeem.
Menselijk toezicht helpt ervoor te zorgen dat een AI-systeem de menselijke autonomie niet ondermijnt of andere nadelige effecten veroorzaakt. Afhankelijk van het specifieke op AI gebaseerde systeem en het toepassingsgebied ervan, moet worden gezorgd voor de passende mate van controlemaatregelen, met inbegrip van het aanpassingsvermogen, de nauwkeurigheid en de verklaarbaarheid van op AI gebaseerde systemen.
Toezicht kan worden uitgeoefend door middel van governance-mechanismen, zoals het waarborgen van een human-in-the-loop-, human-on-the-loop- of human-in-command-benadering. Er moet voor worden gezorgd dat overheidsinstanties in staat zijn hun toezichtsbevoegdheden uit te oefenen in overeenstemming met hun mandaat. Als alle andere dingen gelijk blijven, hoe minder toezicht een mens kan uitoefenen op een AI-systeem, hoe uitgebreider testen en strenger bestuur vereist is.

Een Pensioenfonds hanteert met betrekking tot de menselijke autonomie en controle de volgende principes:

Principe 5: Risicobereidheid
Een Pensioenfonds stelt zijn risicobereidheid met betrekking tot de inzet van AI-toepassingen vast en beoordeeld vervolgens minimaal één keer per jaar of de risicobereidheid nog voldoet.

Principe 6: Risicoclassificatie AI-toepassingen
Een Pensioenfonds classificeert zijn AI-toepassingen conform de risicoclassificatie in de concept-AI verordening van de EU. Een Pensioenfonds zet geen AI-toepassingen in die overeenkomstig de concept-AI-verordening van de EU classificeren als onacceptabele risico's.

Principe 7: Risico gebaseerde inzet
Een Pensioenfonds voert voorafgaan aan het inzetten van AI-toepassingen, een risicoanalyse uit, waarbij het Pensioenfonds een bewuste keuze maakt met betrekking tot de geconstateerde risico's en maatregelen in vergelijking tot meer traditionele technieken en processen.

Principe 8: Voorkomen van vooroordelen in AI-toepassing
Een Pensioenfonds treft maatregelen ter voorkoming van vooroordelen (onder meer 'confirmation bias' - voorkeur voor bevestiging) en met het oog op behoud van menselijke autonomie.

Principe 9: Menselijk toezicht
Het gebruik van AI-toepassingen in de praktijk vindt altijd plaats onder menselijk toezicht en menselijke verantwoordelijkheid, bijvoorbeeld door waar nodig AI te hertrainen/corrigeren.

Technische robuustheid en veiligheid

Betrouwbare AI vereist dat algoritmen veilig, betrouwbaar en robuust genoeg zijn om fouten of inconsistenties tijdens alle fasen van de levenscyclus van het AI-systeem aan te pakken en adequaat om te gaan met foutieve uitkomsten. AI-systemen moeten betrouwbaar en veilig genoeg zijn om weerbaar te zijn tegen zowel openlijke aanvallen als subtielere pogingen om gegevens of algoritmen zelf te manipuleren, en ze moeten zorgen voor een terugvalplan in geval van problemen. Hun beslissingen moeten nauwkeurig zijn, of op zijn minst hun nauwkeurigheidsniveau correct weerspiegelen, en hun resultaten moeten reproduceerbaar zijn. Bovendien moeten AI-systemen ingebouwde veiligheids- en beveiligingsmechanismen integreren om ervoor te zorgen dat zij bij elke stap aantoonbaar veilig zijn, waarbij de fysieke en mentale veiligheid van alle betrokkenen ter harte wordt genomen. Dit omvat het minimaliseren en waar

mogelijk omkeren van onbedoelde gevolgen of fouten in de werking van het systeem. Er moeten processen worden ingevoerd om de potentiële risico's in verband met het gebruik van AI-systemen op verschillende toepassingsgebieden te verduidelijken en te beoordelen.

Een Pensioenfonds hanteert met betrekking tot de technische robuustheid en veiligheid de volgende principes:

Principe 10: Informatiebeveiliging incl. cybersecurity
Een Pensioenfonds treft ten aanzien van AI-toepassingen (inclusief databeheer) passende technische en organisatorische beveiligingsmaatregelen overeenkomst de geldende wet- en regelgeving (onder meer DORA-verordening vanaf 17-01-2025) alsook de Good Practices van de toezichthouders.

Principe 11: Betrouwbaarheid/reproduceerbaarheid
Een Pensioenfonds monitort periodiek of gebruikte AI-toepassingen in overeenstemming met vooraf gestelde doelen, doelstellingen en beoogde toepassingen werken.

Principe 12: Kwaliteit en integriteit van data
Een Pensioenfonds borgt de kwaliteit van (trainings)data die gebruikt wordt voor AI-toepassingen.

Privacy en data governance

Privacy en gegevensbescherming moeten in alle stadia van de levenscyclus van het AI-systeem worden gewaarborgd. Digitale registraties van menselijk gedrag kunnen AI-systemen in staat stellen niet alleen de voorkeuren, leeftijd en geslacht van individuen af te leiden, maar ook hun seksuele geaardheid, religieuze of politieke opvattingen. Om personen in staat te stellen vertrouwen te hebben in de gegevensverwerking, moet ervoor worden gezorgd dat zij volledige controle hebben over hun eigen gegevens en dat de gegevens die op hen betrekking hebben, niet worden gebruikt om hen te schaden of te discrimineren. Naast het waarborgen van privacy en persoonsgegevens moet aan eisen worden voldaan om AI-systemen van hoge kwaliteit te waarborgen. De kwaliteit van de gebruikte datasets is van het grootste belang voor de prestaties van AI-systemen. Wanneer gegevens worden verzameld, kunnen deze sociaal geconstrueerde vooroordelen weerspiegelen of onnauwkeurigheden, fouten en vergissingen bevatten. Dit moet worden aangepakt voordat een AI-systeem met een bepaalde dataset wordt getraind. Daarnaast moet de integriteit van de gegevens gewaarborgd zijn. De

gebruikte processen en datasets moeten bij elke stap, zoals planning, training, testen en implementatie, worden getest en gedocumenteerd. Dit zou ook moeten gelden voor AI-systemen die niet in eigen huis zijn ontwikkeld, maar elders zijn aangeschaft. Ten slotte moet de toegang tot gegevens adequaat worden geregeld en gecontroleerd.

Een Pensioenfonds hanteert met betrekking tot privacy en data governance de volgende principes:

Principe 13: Beheer van gegevens
Een Pensioenfonds zorgt als onderdeel van integere en beheerste bedrijfsvoering voor verantwoord beheer van data en borging van de data governance. Het Pensioenfonds neemt AI-gerelateerde data-aspecten expliciet op in zijn beleidskaders.

Principe 14: Persoonsgegevens
Bij gebruik van persoonsgegevens voor AI-toepassingen werkt het Pensioenfonds in overeenstemming met de Algemene Verordening Gegevensbescherming (AVG) en de Uitvoeringswet AVG (UAVG) en de Gedragslijn Verwerking Persoonsgegevens van de Pensioenfederatie.

Transparantie

De traceerbaarheid van AI-systemen moet worden gewaarborgd; Het is belangrijk om zowel de beslissingen die door de systemen zijn genomen als het hele proces (inclusief een beschrijving van het verzamelen en labelen van gegevens en een beschrijving van het gebruikte algoritme) dat tot de beslissingen heeft geleid, te loggen en te documenteren. Hieraan gekoppeld moet zoveel mogelijk worden gezorgd voor verklaarbaarheid van het algoritmische besluitvormingsproces, aangepast aan de betrokken personen. Lopend onderzoek naar de ontwikkeling van verklaarbaarheidsmechanismen moet worden voortgezet. Daarnaast moet er uitleg beschikbaar zijn over de mate waarin een AI-systeem het besluitvormingsproces van de organisatie beïnvloedt en vormgeeft, de ontwerpkeuzes van het systeem en de reden voor de inzet ervan (waardoor niet alleen de transparantie van gegevens en het systeem wordt gewaarborgd, maar ook de transparantie van het bedrijfsmodel). Ten slotte is het belangrijk om de mogelijkheden en beperkingen van het AI-systeem adequaat te communiceren aan de verschillende betrokken belanghebbenden op een manier die past bij de use case in kwestie. Bovendien moeten AI-systemen als zodanig herkenbaar zijn, zodat gebruikers weten dat zij interactie hebben met een AI-systeem en welke personen daarvoor verantwoordelijk zijn.

Een Pensioenfonds hanteert met betrekking tot transparantie de volgende principes:

Principe 15: Uitlegbaarheid
Een Pensioenfonds zet enkel AI-toepassingen in waarvan zij een adequate toelichting kunnen geven over de werking en de uitkomsten van het AI-systeem.

Principe 16: Herstelbaar
Bij de inzet van AI-toepassingen zal altijd een beroep gedaan kunnen worden op menselijke tussenkomst en uitleg verkregen kunnen worden door belanghebbenden over de uitkomsten bij een toepassing.

Principe 17: Communicatie over inzet
Bij gebruik van AI-toepassingen zoals chatbots zullen wij waar nodig vermelden dat de deelnemer met een AI-systeem van doen heeft en niet met een mens, om verwarring of onduidelijkheid hierover te voorkomen. Ook richting interne medewerkers is het Pensioenfonds transparant en helder over AI-ondersteuning.

Diversiteit, non-discriminatie en rechtvaardigheid

Datasets die door AI-systemen worden gebruikt (zowel voor training als voor exploitatie) kunnen te lijden hebben onder de opname van onbedoelde historische vooringenomenheid, onvolledigheid en slechte bestuursmodellen. Het voortduren van dergelijke vooroordelen kan leiden tot (in)directe discriminatie. Schade kan ook het gevolg zijn van het opzettelijk uitbuiten van (consumenten)vooroordelen of door het aangaan van oneerlijke concurrentie. Bovendien kan ook de manier waarop AI-systemen worden ontwikkeld (bijvoorbeeld de manier waarop de programmeercode van een algoritme is geschreven) last hebben van vooroordelen. Dergelijke problemen moeten vanaf het begin van de ontwikkeling van het systeem worden aangepakt. Het opzetten van diverse ontwerpteams en het opzetten van mechanismen die de participatie van met name burgers in de ontwikkeling van AI waarborgen, kunnen ook helpen om deze problemen aan te pakken. Het is raadzaam om belanghebbenden te raadplegen die gedurende de hele levenscyclus direct of indirect door het systeem kunnen worden beïnvloed. AI-systemen moeten rekening houden met het hele scala aan menselijke capaciteiten, vaardigheden en vereisten, en toegankelijkheid waarborgen door middel van een universele ontwerpbenadering om te streven naar gelijke toegang voor personen met een handicap.

Een Pensioenfonds hanteert met betrekking tot diversiteit, non-discriminatie en rechtvaardigheid de volgende principes:

Principe 18: Open Cultuur
Een Pensioenfonds zorgt voor een open cultuur waarin medewerkers worden aangemoedigd om ethische afwegingen te maken en een gedegen systeem waarbij (potentiële) negatieve gevolgen van het gebruik van een AI-toepassing kunnen worden gemeld en adequaat worden afgehandeld.

Principe 19: Voorkomen onrechtvaardige vooroordelen
Wanneer inbreuk op grondrechten, waaronder ongerechtvaardigde discriminatoire bias in AI- toepassingen niet vermeden of uitgesloten kan worden, zet een Pensioenfonds de AI-toepassing niet in.

Maatschappelijk welzijn

Om AI betrouwbaar te laten zijn, moet rekening worden gehouden met de impact ervan op het milieu en andere levende wezens. Idealiter zouden alle mensen, inclusief toekomstige generaties, moeten profiteren van biodiversiteit en een leefbare omgeving. Duurzaamheid en ecologische verantwoordelijkheid van AI-systemen moeten daarom worden aangemoedigd. Hetzelfde geldt voor AI-oplossingen die gericht zijn op gebieden van mondiaal belang, zoals bijvoorbeeld de Duurzame Ontwikkelingsdoelstellingen van de VN. Bovendien moet de impact van AI-systemen niet alleen vanuit een individueel perspectief worden bekeken, maar ook vanuit het perspectief van de samenleving als geheel. Het gebruik van AI-systemen moet zorgvuldig worden overwogen, met name in situaties die verband houden met het democratische proces, met inbegrip van meningsvorming, politieke besluitvorming of electorale contexten. Bovendien moet rekening worden gehouden met de maatschappelijke impact van AI. Hoewel AI-systemen kunnen worden gebruikt om sociale vaardigheden te verbeteren, kunnen ze evenzeer bijdragen aan de verslechtering ervan.

Een Pensioenfonds hanteert met betrekking tot Maatschappelijk welzijn de volgende principes:

Principe 20: Sociale gevolgen
In de risicoanalyse bij de inzet van AI-toepassingen wordt aandacht besteed aan de sociale gevolgen van de inzet voor belanghebbenden.

Principe 21: Samenleving en democratie
Het Pensioenfonds borgt dat het gebruik van AI-toepassingen in lijn is met het algemene beleid van het Pensioenfonds.

Naleving gedragslijn

Het Pensioenfonds is zich bewust van de risico's die samenhangen met de inzet van AI-toepassingen en de uitstraling naar de pensioensector als een ongewenst risico zich voordoet. Daarom zal het bestuur van het Pensioenfonds in de meeste gevallen een lage risicobereidheid vaststellen met betrekking tot de inzet van AI binnen zijn processen. Als een Pensioenfonds een andere risicobereidheid voor de inzet van AI-toepassingen vaststelt, zal zij dit expliciet maken richting de belanghebbenden van deze AI-toepassingen.

De wijze waarop invulling gegeven wordt aan de naleving van deze gedragslijn is aan het bestuur van het Pensioenfonds. Hierbij kan gedacht worden aan een AI-verklaring op de website van het Pensioenfonds alsook vermelding van de naleving van de gedragslijn in het jaarverslag van het Pensioenfonds. Op dit moment is er geen assurance-verplichting op de naleving van deze gedragslijn.

De Pensioenfederatie ontwikkelt een vragenlijst/self assessment die het Pensioenfonds kan hanteren om de naleving van deze gedragslijn inzichtelijk te maken en de mate van voldoen aan deze gedragslijn te onderbouwen.

Bijlage: Definitielijst

Begrip	Betekenis
AI	een machinaal systeem dat is ontworpen om met verschillende niveaus van autonomie te werken en dat na de uitrol aanpassingsvermogen kan vertonen en dat, voor expliciete of impliciete doelstellingen, uit de ontvangen input afleidt hoe output te genereren zoals voorspellingen, inhoud, aanbevelingen of beslissingen die van invloed kunnen zijn op fysieke of virtuele omgevingen.
AI-betrouwbaarheid	Een AI-systeem zou betrouwbaar zijn als het zich gedraagt zoals verwacht, zelfs voor nieuwe input waarop het niet eerder is getraind of getest. Controleerbaarheid: controleerbaarheid verwijst naar het vermogen van een AI-systeem om de beoordeling van de algoritmen, gegevens en ontwerpprocessen van het systeem te ondergaan. Dit betekent niet noodzakelijkerwijs dat informatie over bedrijfsmodellen en Intellectueel Eigendom met betrekking tot het AI-systeem altijd openlijk beschikbaar moet zijn. Zorgen voor traceerbaarheid en loggingmechanismen vanaf de vroege ontwerpfase van het AI-systeem kan helpen de controleerbaarheid van het systeem mogelijk te maken.
AI-vooroordeel	AI (of algoritmische) vooroordeel beschrijft systematische en herhaalbare fouten in een computersysteem die oneerlijke resultaten opleveren, zoals het bevoordelen van een willekeurige groep gebruikers boven andere. Vertekening kan ontstaan door vele factoren, inclusief maar niet beperkt tot het ontwerp van het algoritme of het onbedoelde of onverwachte gebruik of beslissingen met betrekking tot de manier waarop gegevens worden gecodeerd, verzameld, geselecteerd of gebruikt om het algoritme te trainen. Bias kan algoritmische systemen binnendringen als gevolg van reeds bestaande culturele, sociale of institutionele verwachtingen; vanwege technische beperkingen van hun ontwerp; of door te worden

	gebruikt in onverwachte contexten of door doelgroepen waarmee geen rekening is gehouden in het oorspronkelijke ontwerp van de software. AI-vooroordelen komen voor op alle platforms, inclusief maar niet beperkt tot resultaten van zoekmachines en sociale mediaplatforms, en kunnen gevolgen hebben die variëren van onbedoelde privacy schendingen tot het versterken van sociale vooroordelen op het gebied van ras, geslacht, seksualiteit en etniciteit.
Betrouwbare AI	Betrouwbare AI heeft drie componenten: (1) het moet wettig zijn en ervoor zorgen dat alle toepasselijke wet- en regelgeving wordt nageleefd (2) het moet ethisch zijn en blijk geven van respect voor en naleving van ethische principes en waarden en (3) het moet zowel technisch als maatschappelijk robuust zijn, aangezien AI-systemen ook met goede bedoelingen onbedoeld schade kunnen aanrichten.37 Betrouwbare AI betreft niet alleen de betrouwbaarheid van het AI-systeem zelf, maar omvat ook de betrouwbaarheid van alle processen en actoren die deel uitmaken van de levenscyclus van het AI-systeem.
Vooroordeel	Vooringenomenheid of bias in data. Vooroordeel kan op verschillende manieren ontstaan en heeft vaak een negatieve impact op een ML-model.
Chatbot	Een computerprogramma dat is ontworpen om een gesprek met een menselijke gebruiker te simuleren, meestal via internet; vooral een die wordt gebruikt om informatie of assistentie te bieden aan de gebruiker als onderdeel van een geautomatiseerde service.
Classificatie	ML-taak waarin een ML-model voorbeelden in categorieën of klassen onderverdeelt. Classificatie is een vorm van supervised learning en vereist gelabelde data.
Data governance	Data governance is een term die zowel op macro- als op microniveau wordt gebruikt. Op macroniveau verwijst data governance naar het besturen van grensoverschrijdende datastromen door landen, en wordt daarom preciezer internationaal data governance genoemd. Op microniveau is datagovernance een datamanagementconcept dat betrekking heeft op het vermogen dat een organisatie in staat stelt om te zorgen voor een hoge datakwaliteit gedurende de

	volledige levenscyclus van de data, en om datacontroles te implementeren die de bedrijfsdoelstellingen ondersteunen. De belangrijkste aandachtsgebieden van gegevensbeheer zijn onder meer de beschikbaarheid, bruikbaarheid, consistentie, integriteit en delen van gegevens. Het gaat ook om het opzetten van processen om te zorgen voor effectief gegevensbeheer in de hele onderneming, zoals verantwoording voor de nadelige effecten van slechte gegevenskwaliteit en ervoor zorgen dat de gegevens waarover een onderneming beschikt door de hele organisatie kunnen worden gebruikt.
(Eind)gebruiker	Een eindgebruiker is de persoon die het AI-systeem uiteindelijk gebruikt of uiteindelijk gaat gebruiken. Dit kan een consument zijn of een professional binnen een publieke of private organisatie. De eindgebruiker staat in contrast met gebruikers die het product ondersteunen of onderhouden, zoals systeembeheerders, databasebeheerders, IT-experts, softwareprofessionals en computertechnici.
Ethiek en AI	Bij ethiek in AI gaat het principes en normen die worden toegepast bij het ontwikkelen, implementeren en gebruiken van kunstmatige intelligentie (AI) systemen. Het omvat overwegingen met betrekking tot de morele en sociale implicaties van AI, evenals de verantwoordelijkheid van AI-ontwikkelaars, -beheerders en -gebruikers. Onderwerpen: • • Transparantie en verantwoording; • • Privacy en gegevensbescherming; • • Vooringenomenheid en discriminatie; • • Veiligheid; • • Werkgelegenheid en maatschappelijke impact.
Human-On-The-Loop	Een vorm van interactie tussen mens en machine waarbij het AI-systeem zelfstandig beslissingen kan nemen zonder dat er een mens aan te pas komt. Wel heeft een mens inzicht in het proces en is die in staat om in te grijpen en wijzigingen aan te brengen.

Levenscyclus	De levenscyclus van een AI-systeem omvat verschillende onderling afhankelijke fasen, gaande van het ontwerp en de ontwikkeling (inclusief subfasen zoals behoefteanalyse, gegevensverzameling, training, testen, integratie), installatie, implementatie, bediening, onderhoud en verwijdering. Gezien de complexiteit van AI-systemen (en in het algemeen informatiesystemen), zijn er verschillende modellen en methodologieën gedefinieerd om deze complexiteit te beheersen, vooral tijdens de ontwerp- en ontwikkelingsfasen, zoals waterval, spiraal, agile softwareontwikkeling, rapid prototyping en incrementeel.
Nauwkeurigheid	Het doel van een AI-model is om patronen te leren die goed generaliseren voor ongeziene gegevens. Het is belangrijk om te controleren of een getraind AI-model goed presteert op ongeziene voorbeelden die niet zijn gebruikt voor het trainen van het model. Om dit te doen, wordt het model gebruikt om het antwoord op de testdataset te voorspellen en vervolgens wordt het voorspelde doel vergeleken met het daadwerkelijke antwoord. Het concept van nauwkeurigheid wordt gebruikt om het voorspellende vermogen van het AI-model te evalueren. Informeel is nauwkeurigheid de fractie van de voorspellingen die het model goed deed. Bij machine learning (ML) wordt een aantal statistieken gebruikt om de voorspellende nauwkeurigheid van een model te meten. De keuze van de te gebruiken nauwkeurigheidsmetriek is afhankelijk van de ML-taak.
Reproduceerbaarheid	Reproduceerbaarheid verwijst naar de nabijheid tussen de resultaten van twee acties, zoals twee wetenschappelijke experimenten, die dezelfde input krijgen en de methodologie gebruiken, zoals beschreven in een overeenkomstig wetenschappelijk bewijs (zoals een wetenschappelijke publicatie). Een verwant concept is replicatie, wat het vermogen is om onafhankelijk niet-identieke conclusies te trekken die op zijn minst vergelijkbaar zijn, wanneer er verschillen zijn in bemonstering, onderzoeksprocedures en data-analysemethoden.

	Reproduceerbaarheid en repliceerbaarheid behoren samen tot de belangrijkste instrumenten van de wetenschappelijke methode.
Robuustheid AI	Robuustheid van een AI-systeem omvat zowel de technische robuustheid (passend in een bepaalde context, zoals het toepassingsdomein of de levenscyclusfase) als de robuustheid vanuit maatschappelijk perspectief (ervoor zorgen dat het AI-systeem terdege rekening houdt de context en omgeving waarin het systeem opereert). Dit is cruciaal om ervoor te zorgen dat er, zelfs met goede bedoelingen, geen onbedoelde schade kan ontstaan. Robuustheid is de derde van de drie componenten die nodig zijn om betrouwbare AI te bereiken.
Trainingsdata	Gedeelte van een dataset dat wordt gebruikt om een ML-model te trainen. Bevat vaak 60% tot 80% van de beschikbare data.
Uitlegbaarheid	Kenmerk van een AI-systeem dat begrijpelijk is voor niet-experts. Een AI-systeem is begrijpelijk als de functionaliteit en werking ervan niet-technisch kan worden uitgelegd aan een onervaren persoon.