

## Implementatie van de AI-verordening

Definitie van een AI-systeem

Februari 2025

## 1. Introductie

De AI-verordening stelt eisen aan de inzet van artificiële intelligentie (AI) in de Europese Unie (EU). De productveiligheidswetgeving ziet toe op verantwoorde ontwikkeling en gebruik van AI door publieke en private organisaties. Hiermee worden de veiligheid, gezondheid en grondrechten van burgers van de EU beschermd. Implementatie van de AI-verordening brengt echter lastige vraagstukken met zich mee, bijvoorbeeld welke algoritmische toepassingen binnen de reikwijdte van de verordening vallen.

Noch in het maatschappelijke debat rondom deze technologie, noch binnen de academische en technische wereld is in de afgelopen 50 jaar een vaste definitie gebruikt voor het begrip AI. Men maakte gebruik van intuïtieve, ongeschreven definities. Wat als AI wordt gezien groeit mee met wat geldt als de technologische *cutting edge*: zodra algemeen toegankelijke software complexe taken kan uitvoeren die voorheen waren voorbehouden aan 'AI', wordt het al snel niet meer als AI gezien.<sup>1</sup>

Met de komst van de AI-verordening wordt AI echter gevat in een juridisch bindende definitie. Het doel van de EU is om met deze definitie AI-systemen te onderscheiden van eenvoudigere traditionele software-systemen of programmeringsbenaderingen, en daarmee rechtszekerheid te bieden, brede acceptatie van de technologie te creëren, en toekomstbestendigheid

te faciliteren.<sup>2</sup> De definitie zoals gebruikt door de Europese wetgever is niet nieuw: het volgt op de definitie van AI zoals ontwikkeld door de Organisatie voor Economische Samenwerking en Ontwikkeling (OESO).

Aan de hand van deze juridische definitie moeten organisaties aan de slag met implementatie van de AI-verordening. Dit blijkt ingewikkeld. Zo hebben juristen vaak weinig praktijkervaring met de technologieën onderliggend aan AI en zijn technici juist onervaren met juridische definities. Daarnaast bevat de definitie van een AI-systeem termen die niet allemaal evenveel gewicht hebben. Voor implementatie van de verordening is het daarom nodig om bruggen te slaan tussen deze verschillende werelden, waarbij niet alleen volledigheid maar ook pragmatiek in ogenschouw moet worden genomen. Deze white paper doet hiertoe een voorzet. We analyseren de kernelementen uit de definitie van een AI-systeem zowel vanuit juridisch als statistisch perspectief, waarmee de reikwijdte van de AI-verordening wordt verkend. De richtlijnen van de Europese Commissie (EC) over de definitie van een AI-systeem zijn in deze analyse meegenomen.<sup>3</sup>

Daarnaast wordt de samenhang tussen AI-systemen en impactvolle algoritmen besproken. Dit begrip wordt al langere tijd door de Nederlandse overheid gebruikt om algoritmen<sup>4</sup> met een verhoogd risico aan te duiden.<sup>5</sup> Deze algoritmen zijn geen AI-systeem volgens de AI-verordening, maar kunnen wel grote invloed hebben op betrokkenen.<sup>6</sup>

<sup>1</sup> Gezichtsherkenning en schaakcomputers werden lang als het ultieme voorbeeld van AI gezien, terwijl de toepassingen nu zijn geïntegreerd in het alledaagse leven en er niet meer als zodanig naar wordt verwezen. Dit fenomeen is door Pamela McCorduck beschreven als "het AI effect".

<sup>2</sup> Zie overweging 12 AI-verordening

<sup>3</sup> 'Guidelines on the definition of an artificial intelligence system established by AI Act', Europese Commissie (2025).

<sup>4</sup> 'Algoritme' zoals gedefinieerd door de Algemene Rekenkamer (2021): "Een set van regels en instructies die een computer geautomatiseerd volgt bij het maken van berekeningen om een probleem op te lossen of een vraag te beantwoorden".

<sup>5</sup> Figuur 3 [Handreiking Algoritmeregister](#) van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Dat impactvolle algoritmen een belangrijke categorie algoritmen is blijkt onder meer uit [Kamerstukken II 2024-25 2025D00512](#).

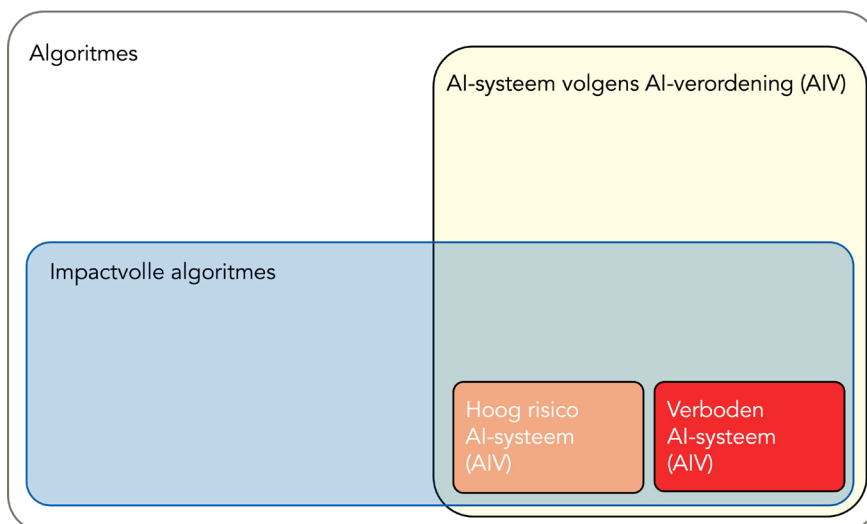
<sup>6</sup> Zo kwam bijvoorbeeld in 2023 na onderzoek van onder mee Investico, NOS op 3 en De Groene Amsterdammer aan het licht dat DUO een algoritme gebruikte voor risicogerichte selectie tijdens controle op de uitwendendenbeurs. Na onderzoek van PwC en Stichting Algorithm Audit bleek dat er sprake was van (indirecte) discriminatie. De Minister van Onderwijs, Cultuur en Wetenschap heeft een hersteloperatie [aangekondigd](#), waarbij ruim tienduizend studenten gecompenseerd worden.

Een schematisch overzicht van de verschillende soorten algoritmen en de relatie tussen deze begrippen kan worden gevonden in [Figuur 1](#).

We introduceren drie vragen aan de hand waarvan gebruikers kunnen bepalen of een algoritmische toepassing een AI-systeem betreft. Aan de hand van maximaal vier extra vragen kan worden bepaald of er sprake is van een ‘impactvol algoritme’. Deze vragenlijsten zijn online beschikbaar onder de naam ‘AI-verordening implementatie tool’.<sup>7</sup> Deze tool is gepubliceerd onder de EUPL-1.2 licentie.<sup>8</sup>

Identificatie van hoog risico en verboden AI-systemen worden in een aparte white paper besproken.<sup>9</sup>

In deze white paper volgt een analyse van de zeven kenmerken van de definitie van een AI-systeem ([sec. 2](#)). Waarbij speciale aandacht wordt besteed aan het begrip inferentievermogen ([sec. 3](#)) en autonomie ([sec. 4](#)). Daarnaast wordt de samenhang tussen AI-systemen en het begrip impactvolle algoritmen besproken ([sec. 5](#)). De paper sluit af met vragenlijsten aan de hand waarvan AI-systemen en impactvolle algoritme kunnen worden geïdentificeerd ([sec. 6](#)).



**Figuur 1** - Overzicht van besproken type algoritmen in deze white paper. Hoog-risico en verboden AI-systemen worden in een aparte white paper besproken.<sup>4</sup>

<sup>7</sup> <https://algorithmaudit.eu/technical-tools/implementation-tool/#tool>

<sup>8</sup> <https://github.com/NGO-Algorithm-Audit/AI-Act-Implementation-Tool?tab=EUPL-1.2-1-ov-file>

<sup>9</sup> <https://algorithmaudit.eu/technical-tools/implementation-tool/#documentation-high-risk>

## Box 1

### Voorbehouden naleving AI-verordening

Dit document is een interpretatie van de wettekst van de AI-verordening en aanvullende richtlijnen zoals gepubliceerd door de Europese Commissie (EC) door stichting Algorithm Audit. Aan deze analyse kunnen geen rechten worden ontleend. Zoals opgemerkt in paragraaf (7) van de door de EC uitgebrachte richtlijnen oordeelt uiteindelijk het Hof van Justitie van de Europese Unie (HvJ-EU) wat de correcte interpretatie van een AI-systeem is.

## 2. AI-systeem definitie

De definitie van een AI-systeem wordt geïntroduceerd in artikel 3(1) van de AI-verordening. Alleen systemen die aan deze definitie voldoen vallen onder de wet.

### Artikel 3(1) van de AI-verordening definieert een AI-systeem als volgt:

“een op een machine gebaseerd systeem dat is ontworpen om met **verschillende niveaus van autonomie** te werken en dat na het inzetten ervan **aanpassingsvermogen kan vertonen**, en dat, voor **expliciete of impliciete doelstellingen**, **uit de ontvangen input afleidt hoe output te genereren** zoals **voorspellingen, content, aanbevelingen of beslissingen** die van **invloed kunnen zijn op fysieke of virtuele omgevingen**.”

We analyseren en interpreteren de bovenstaande zeven gekleurde begrippen. De belangrijkste bronnen voor deze interpretatie zijn overweging 12 uit de preambule van de verordening en de richtlijnen zoals gepubliceerd door de EC (hierna: ‘richtlijnen’).<sup>10</sup>

<sup>10</sup> Supra note 3.

Overweging 12 bestaat uit 13 zinsdelen die in [Appendix A](#) worden vermeld. De overwegingen (Engels: recitals) geven inzicht in de intenties van de uniewetgever bij het opstellen van de wetteksten en geven daarmee duiding hoe de begrippen geïnterpreteerd moeten worden. Voor de interpretatie zijn zowel de Engelse als Nederlandse wetteksten geraadpleegd.

De richtlijnen zijn gepubliceerd om aanvullende uitleg te geven over de definitie van een AI-systeem. In de analyse van de definitie van een AI-systeem, volgend in sectie 2-4, wordt doorlopend verwezen naar specifieke passages uit deze richtlijnen. De richtlijnen bevatten naast een sec interpretatie van de elementen uit de definitie ook een aantal controversiële uitzonderingen. Deze uitzonderingen vertroebelen de interpretatie van de definitie in plaats van hierover helderheid te verschaffen. Een inhoudelijke toelichting op deze kritiek kan worden gevonden in [Box 3](#).

### Box 2

## Inconsistente vertalingen: Beslissingen of besluit? Content of inhoud?

Merk op: Verschillende begrippen in de AI-verordening zijn niet consequent vertaald van het Engels naar het Nederlands.

In artikel 3 van de Nederlandse wetteksten wordt het woord “*afleidt*” gebruikt, waar de Engelse wetteksten spreken over “*infers*”. In overweging 12 wordt gesproken over “*inferentievermogen*” en in het Engels over “*capability to infer*”. Op een vergelijkbare manier wordt in artikel 3 gesproken over “*beslissingen*”, terwijl in overweging 12 wordt gesproken over “*besluiten*”. In de Engelse wetteksten wordt consequent “*decisions*” gebruikt. Ook “*content*” wordt niet consequent vertaald: in de Nederlandse wetteksten wordt in artikel 3 gesproken over “*inhoud*”, en in overweging 12 over “*content*”.

Deze verschillen in de vertaling benadrukken dat het van waarde is om bij interpretatie van internationale wetgeving ook de Engelse wetteksten in ogenschouw te nemen.

In deze white paper wordt ook verwezen naar het memorandum van de OESO over de definitie van een AI-systeem (hierna: 'OESO-memorandum').<sup>11</sup> Tijdens onderhandelingen over de AI-verordening is dit memorandum, inclusief eerdere conceptversies, gebruikt om tot een definitie te komen van een AI-systeem in de wettekst. In dit licht benoemt overweging 12 expliciet dat de EU de wens heeft *"nauw aan te sluiten op het werk van internationale organisaties die zich bezighouden met AI, om rechtszekerheid te waarborgen, internationale convergentie en brede acceptatie te faciliteren"*.

Iedere analyse van bovenstaande zeven gekleurde begrippen sluiten we af met een oordeel in hoeverre het als criterium kan dienen om AI-systemen te onderscheiden van algoritmen. Met oog op het belang van de begrippen inferentie en autonomie in de definitie van een AI-systeem worden deze zelfstandig geanalyseerd in 3. [Inferentie](#) en 4. [Autonomie](#).

Voortbouwend op deze analyse wordt een dynamische vragenlijst geïntroduceerd aan de hand waarvan binnen drie vragen AI-systemen geïdentificeerd kunnen worden. Zie sectie 6.1-6.3 in 6. [Dynamische vragenlijst](#).

<sup>11</sup> [Explanatory Memorandum on the Updated OECD definition of an AI system](#) (2024).

### Box 3

## Uitzonderingen in de richtlijnen vertroebelen de interpretatie van de AI-systeem definitie

De Europese Commissie (EC) heeft richtlijnen gepubliceerd hoe de definitie van een AI-systeem in de AI-verordening geïnterpreteerd kan worden.<sup>12</sup> In de richtlijnen worden uitzonderingen geïntroduceerd van algoritmische systemen die niet als AI-systeem kwalificeren op basis van argumenten die niet aansluiten bij de wettekst. Dit is merkwaardig aangezien richtlijnen zijn bedoeld om de interpretatie van de wettekst te verduidelijken in plaats van aanvullende bepalingen te introduceren. De richtlijnen vertroebelen daarmee de interpretatie van de wettekst in plaats van hierover helderheid te verschaffen.

Specifiek leiden de uitzonderingen voor systemen die mathematische optimalisatie verbeteren (paragrafen (42)-(45)) en simpele voorspellingssystemen (paragrafen (49)-(51)) voor problemen.

<sup>12</sup> *Supra* note 3.

In paragraaf (48) wordt bijvoorbeeld gesteld dat simpele machine learning systemen die voorspellingen maken geen AI-systemen zijn. In de definitie van een AI-systeem in artikel 3 van de wettekst en in de toelichting in overweging 12 wordt de complexiteit van een systeem niet genoemd als bepalende factor om systemen als AI-systeem te kwalificeren.<sup>13</sup> De uitzondering zorgt voor verwarring: wanneer is een systeem ‘simpel’ genoeg om onder deze uitzondering te vallen?

In paragraaf (42) wordt gesteld dat systemen die worden ingezet voor mathematische optimalisatie niet kwalificeren als AI-systeem. De toepassing doet er volgens de definitie uit de wettekst echter niet toe of een systeem wel of geen AI-systeem betreft. In deze paragraaf wordt toegelicht dat gevestigde methoden, zoals lineaire en logistische regressie, geen AI-systeem zijn omdat “ondanks dat deze methoden inferentievermogen bezitten, de elementaire verwerking van data niet wordt overstegen”. Deze passage spreekt overweging 12 uit de AI-verordening direct tegen. Uit deze overweging volgt dat “dat inferentievermogen zich laat kenmerken doordat de elementaire verwerking van data wordt overstegen”. Zowel het punt dat mathematische optimalisatie buiten de reikwijdte van de definitie valt als de toelichting hierop staan op gespannen voet met de wettekst.

Een andere inconsistentie doet zich voor bij interpretatie van het begrip ‘aanpassingsvermogen’. In paragrafen (22)-(23) van de richtlijnen wordt toegelicht dat aanpassingsvermogen geen harde eis is om aan de definitie van een AI-systeem te voldoen. Later, in paragraaf (48) van de richtlijnen, wordt gesteld dat heuristieken niet kwalificeren als AI-systeem omdat er sprake is van een gebrek aan aanpassingsvermogen. Met de uitzonderingen in deze richtlijnen lijkt de EC de met het Parlement en Raad politiek uitonderhandelde definitie te versmallen en haar eigen zienswijze op de verordening op te leggen. Vanuit democratisch oogpunt lijkt de EC hiermee buiten haar mandaat te treden.

Met oog op de spanningen tussen de richtlijnen en de AI-verordening is het relevant om op te merken dat de richtlijnen ten aanzien van de wettekst een ondergeschikte juridische status hebben in de hiërarchie der reguleringsinstrumenten. Totdat jurisprudentie van het HvJ-EU beschikbaar is over deze kwestie adviseert Algorithm Audit organisaties, in lijn met de opvatting van de Autoriteit Persoonsgegevens (AP), om op veilig te spelen als zij bepalen of AI-systemen wel of niet onder de reikwijdte van de AI-verordening vallen.<sup>14</sup>

Tot slot merken we op dat bovenstaande ontwikkelingen bijdragen aan een poging de reikwijdte van de AI-verordening in te perken. Ditmaal niet via discussie over de reikwijdte van risicoclassificatie van AI-systemen, enkel hoog-risico AI-systemen moeten immers voldoen aan verplichte beheersmaatregelen, maar via de vraag of algoritmische systemen überhaupt vallen onder de definitie van een AI-systeem. Hiermee wordt zowel via de route van AI-systeem identificatie als risicoclassificatie de reikwijdte van de AI-verordening gepoogd te versmallen.

<sup>13</sup> In overweging 12 wordt wel benoemd dat de definitie zelf de kenmerken moet bevatten die AI-systemen onderscheiden van eenvoudigere traditionele softwaresystemen. Daarom moet dus verondersteld worden dat een systeem dat voldoet aan de kenmerken in de definitie, daarmee geen eenvoudig traditioneel software systeem is.

<sup>14</sup> [Rapportage AI- & Algoritmerisico's Nederland](#), Winter 2024/2025 (Editie 4, februari 2025), Directie Coördinatie Algoritmes (DCA) – Autoriteit Persoonsgegevens (AP).

## 2.1 Interpretatie van de definitie van een AI-systeem aan de hand van overweging 12

Overweging 12 bevat een aantal zinsneden die helpen om interpretatie van de definitie van een AI-systeem te kaderen:

- i) *“de [AI-systeem] definitie moet gebaseerd zijn op de belangrijkste kenmerken van AI-systemen die het onderscheiden van eenvoudigere traditionele softwaresystemen of programmeringsbenaderingen”;*
- ii) *“[de AI-systeem definitie] mag geen betrekking hebben op systemen die gebaseerd zijn op regels die uitsluitend door natuurlijke personen zijn vastgesteld om automatisch handelingen uit te voeren.” – Zie overweging 12 zin 2.*

Uit zinsnede i) volgt de lens waar doorheen we de definitie van een AI-systeem interpreteren: de kenmerken in de definitie moeten het onderscheid tussen AI-systemen en andere softwaresystemen mogelijk maken. De zin is ook een ondergrens waarmee de wetgever aangeeft dat de reikwijdte van de definitie van een AI-systeem niet alle programmeringsbenaderingen betreft. Onder ‘eenvoudige traditionele softwaresystemen’ zou simpele dataverwerking in Excel of SQL kunnen worden verstaan. AI kan in deze programmeringsbenaderingen ook meer geavanceerde dataverwerking worden uitgevoerd wat mogelijk wél een AI-systeem betreft. Zinsnede i) is daarom niet relevant om te bepalen of een toepassing wel of niet een AI-systeem betreft. Dit beeld wordt bevestigd door paragraaf (26) uit de richtlijnen.

Zinsnede ii) refereert naar regelgebaseerde algoritmen waarbij de regels door natuurlijke personen zijn opgesteld. Een voorbeeld van een regel is ‘als leeftijd <65 jaar, dan geen recht op

seniorenkorting’. Als de variabele leeftijd en de drempelwaarde 65 jaar uitsluitend door natuurlijke personen zijn vastgesteld om de automatische handelingen van het bepalen van een korting uit te voeren, is het regelgebaseerde algoritme niet een AI-systeem. Dit is ook het geval wanneer dit algoritme wordt ingezet voor impactvolle doeleinden, zoals risicoprofilering. Zinsnede ii) heeft een sterk vermogen om AI-systemen te onderscheiden van algoritmen. Dit kenmerk is daarom opgenomen als antwoordoptie in de derde vraag van de dynamische vragenlijst. Zie [6.3 Q3 – Is de toepassing automatisering van door mensen opgestelde regels?](#) De richtlijnen geven geen specifieke uitleg over de verhouding tussen regelgebaseerde algoritmen en AI-systemen.

## 2.2 Machine gebaseerd systeem

Overweging 12 vermeldt dat *“op een machine gebaseerd systeem”* uit de AI-systeem definitie de volgende betekenis kent:

*“De term ‘op een machine gebaseerd’ verwijst naar het feit dat AI-systemen op machines draaien.” – zie overweging 12 zin 7.*

Aangezien vrijwel alle moderne softwaresystemen of programmeringsbenaderingen gebruikmaken van een machine, zijnde een computer, server of virtual machine (VM)<sup>15</sup>, voldoen vrijwel alle softwaresystemen en algoritmen aan deze vereiste. We stellen daarom vast dat de ‘machine-gebaseerd systeem’-vereiste geen onderscheidend vermogen heeft om AI-systemen van andere algoritmen te scheiden, omdat alle moderne softwaresystemen of programmeringsbenaderingen machine-gebaseerd zijn. Deze lezing wordt bevestigd door paragraaf (11)-(13) uit de richtlijnen.

<sup>15</sup> Met een VM wordt verwezen naar een microprocessor die op een PC, laptop, of in een cloud-omgeving, algoritmen uitvoert. Zie ook 3.32 uit ISO/IEC 13522-6:1998 Information technology — Coding of multimedia and hypermedia information

## 2.3 Verschillende niveaus van autonomie

Overweging 12 vermeldt dat “*verschillende niveaus van autonomie*” uit de AI-systeem definitie de volgende betekenis kent:

*“AI-systemen worden zodanig ontworpen dat zij in verschillende mate autonoom kunnen functioneren, wat betekent dat zij een zekere mate van onafhankelijkheid van menselijke betrokkenheid bezitten en zonder menselijke tussenkomst kunnen functioneren.”* – zie [overweging 12](#) zin 12.

In de Engelse tekst wordt gesproken van “*some degree of independence*”.

Er moet dus sprake zijn van enige mate van autonomie. Daarom zien we autonomie als een factor waarmee AI-systemen onderscheiden kunnen worden van algoritmen. In [4. Autonomie](#) wordt uitgebreider ingegaan op de betekenis en interpretatie van autonomie.

## 2.4 Aanpassingsvermogen kan vertonen

Overweging 12 vermeldt dat “*na het inzetten ervan aanpassingsvermogen kan vertonen*” uit de AI-systeem definitie de volgende betekenis kent:

*“Het aanpassingsvermogen dat een AI-systeem na het inzetten ervan kan vertonen, heeft betrekking op zelflerende capaciteiten, waardoor het systeem tijdens het gebruik kan veranderen.”*

Door gebruik van het werkwoord *kan* is het op basis van de Nederlandse wettekst onduidelijk of aanpassingsvermogen een vereiste eigenschap is van een AI-systeem. De Engelse versie van de overweging luidt:

*“The adaptiveness that an AI system could exhibit after deployment, refers to self-learning capabilities, allowing the system to change while in use.”* – zie [overweging 12](#) zin 12.

Door gebruik van het werkwoord *could* (zou kunnen) wordt opgemaakt dat aanpassingsvermogen van een AI-systeem geen vereiste is. Ook OESO ziet aanpassingsvermogen na het inzetten ervan als optioneel, in het memorandum benoemt het expliciet ook een systeem dat eenmalig is geleerd uit data als een AI-systeem.<sup>16</sup> Veel AI-systemen die momenteel in gebruik zijn vertonen geen aanpassingsvermogen na het inzetten ervan. Gezichtherkenningssoftware, waar de AI-verordening op verschillende plekken naar verwijst, zijn een voorbeeld waarbij modelparameters over het algemeen niet tijdens gebruik maar enkel voorafgaand aan een softwarerelease worden geüpdatet. Kortom, ook AI-systemen die tijdens gebruik geen aanpassingsvermogen vertonen kunnen nog steeds een AI-systeem zijn, als aan de andere voorwaarden wordt voldaan. Deze lezing wordt bevestigd door paragraaf (22)-(23) uit de richtlijnen.

We concluderen dat ‘aanpassingsvermogen’ geen vereiste is voor de AI-systeem definitie. Daarmee is het geen onderscheidende factor om AI-systemen te scheiden van andere algoritmen.

## 2.5 Expliciete of impliciete doelstellingen

Overweging 12 vermeldt dat “voor *expliciete of impliciete doelstellingen*” uit de AI-systeem definitie de volgende betekenis kent:

*“De verwijzing naar expliciete of impliciete doelstellingen onderstreept dat AI-systemen kunnen functioneren volgens expliciete, gedefinieerde doelstellingen, of volgens impliciete doelstellingen. De doelstellingen van een AI-systeem kunnen verschillen van het beoogde doel van het AI-systeem in een specifieke context.”* – zie [overweging 12](#) zin 8.

<sup>16</sup> Supra noot 10



Een toepassing streeft altijd een doel na, dat ofwel expliciet ofwel impliciet gedefinieerd kan zijn. De reden dat dit element is opgenomen in de definitie is om uit te drukken dat een expliciete doelstelling geen vereiste is voor een AI-systeem.<sup>17</sup> Bijvoorbeeld met reinforcement learning kunnen AI-systemen zelf doelstellingen afleiden, die niet expliciet zijn geformuleerd maar wel impliciet vervat zijn in het AI-systeem. Dit is ook het geval bij Large Language Models (LLMs) zoals ChatGPT en andere toepassingen van generatieve AI. Dit beeld wordt bevestigd door paragraaf (24) uit de richtlijnen. Uit paragraaf (25) volgt dat het 'beoogde doel' niet alleen betrekking heeft op interne operaties die door een systeem worden uitgevoerd, maar ook op de externe context waarbinnen het systeem wordt toegepast.

De 'doelstelling'-vereiste heeft geen onderscheidend vermogen om AI-systemen van andere algoritmen te scheiden.

## 2.6 Uit de ontvangen input afleidt hoe output te genereren

Overweging 12 vermeldt dat *"uit de ontvangen input afleidt hoe output te genereren"* uit de AI-systeem definitie de volgende betekenis kent:

*"Een belangrijk kenmerk van AI-systemen is hun inferentievermogen. Dit inferentievermogen slaat op het proces waarbij output, zoals voorspellingen, content, aanbevelingen of besluiten, wordt verkregen waarmee fysieke en virtuele omgevingen kunnen worden beïnvloed, en op het vermogen van AI-systemen om modellen of algoritmen, of beide, af te leiden uit input of data."* – zie [overweging 12](#) zin 3-4.

Het begrip 'afleiden' uit de definitie wordt aan de hand van het begrip 'inferentie' in overweging 12 toegelicht. We concluderen dat inferentievermogen het belangrijkste element is van de definitie om AI-

systemen te onderscheiden van andere algoritmen. Dit wordt bevestigd in paragraaf 26 van de richtlijnen, die inferentie een *"onmisbare voorwaarde die AI-systemen onderscheidt van andere soorten systemen"* noemen.

In [3. Inferentie](#) wordt ingegaan op de betekenis en interpretaties van inferentie.

## 2.7 Voorspellingen, inhoud, aanbevelingen of beslissingen

Overweging 12 vermeldt dat *"voorspellingen, content, aanbevelingen of besluiten"* uit de AI-systeem definitie de volgende betekenis kent:

*"... de output die door het AI-systeem wordt gegenereerd is een uiting van de verschillende functies van AI-systemen en kan de vorm aannemen van voorspellingen, content, aanbevelingen of besluiten."* – zie [overweging 12](#) zin 10.

Deze passage houdt verband met inferentie, het afleiden van output uit input, waarvan een analyse volgt in [3. Inferentie](#). Met betrekking tot *"voorspellingen, inhoud, aanbevelingen of beslissingen"* gaat het hier om verschillende vormen van output die worden afgeleid:

- 1. Voorspellingen:** Hieronder vallen ingeschatte scores, rangschikkingen, kansen, labels en classificaties. Dit hoeft niet per se een voorspelling over de toekomst te zijn, aangezien een voorspelling ook betrekking kan hebben op een niet eerder geobserveerd datapunt. Het statistische begrip 'schatting' wordt in dit geval ook een voorspelling genoemd (Engels: *estimation, prediction*).
- 2. Inhoud/content:** Hieronder valt gegenereerde tekst, beeld en spraak, bijvoorbeeld gecreëerd middels generatieve AI.
- 3. Aanbevelingen:** Hieronder vallen aanbevelings-systemen, zoals gepersonaliseerde tijdlijnen op social media platforms, zoekmachineresultaten

<sup>17</sup> Supra noot 10

en online advertenties. Tot deze categorie behoren ook aanbevolen handelingen, zoals een aanbeveling voor extra controle die volgt op een toegekende risicoscore voor onrechtmatig gebruik van een sociale voorziening, of een auto die aanbeveelt om naar een andere versnelling te schakelen.<sup>18</sup> Scores of classificaties waaraan beleidsmatig een vaste actie of handeling verbonden is, kunnen ook worden gezien als aanbevelingen. Denk aan: een toegekende risicoscore in transactiemonitoring binnen banken, aan de hand waarvan een werkinstructie voorschrijft dat aanvullend onderzoek moet worden uitgevoerd. Aanbevelingen volgen vaak ook een voorspelling.

**4. Beslissingen/besluiten:** Hieronder lijken beslissingen (“*decisions*”) in de breedste zin des woords te vallen, zoals de beslissing om een actie of handeling uit te voeren, bijvoorbeeld een auto die geautomatiseerd remt voor een voetganger<sup>19</sup>, de keuze om een controle uit te voeren, het vaststellen van iemands identiteit (verificatie) of een formeel besluit zoals gedefinieerd in de Algemene wet bestuursrecht (Awb art.1:3).<sup>20</sup> Voor de publieke sector is het belangrijk op te merken dat algoritmische output die wordt gebruikt in de voorbereidende fase van een besluit ook als onderdeel van het gehele besluitvormingsproces beschouwd dient te worden en daarmee ook dient te voldoen aan de algemene beginselen van behoorlijk bestuur (abbb), zoals het motiveringsbeginsel, het zorgvuldigheidsbeginsel en het beginsel van *fair play*.<sup>21</sup> Wanneer de output een aanbeveling of besluit is, is het begrip ‘geautomatiseerde

besluitvorming’ uit de Algemene Verordening Gegevensbescherming (AVG) relevant.<sup>22</sup>

Bovenstaande beschouwing wordt ondersteund door paragrafen (52)-(59) uit de richtlijnen. Over de rol van menselijke betrokkenheid bij de totstandkoming van gegenereerde output door (impactvolle) algoritmes wordt nader ingegaan in [5.3. Heeft het algoritme een significant effect op de uitkomst van het proces?](#)

De voorbeelden (voorspellingen, inhoud/content, aanbevelingen of besluiten/beslissingen) zijn een belangrijk signaal wat de wetgever ziet als output van AI-systemen tijdens gebruik. Zie ook paragraaf (28) uit de richtlijnen. Aan de hand van deze opsomming kunnen een aantal type algoritmen uitgesloten worden die niet kwalificeren als AI-systemen. Zo stellen we vast dat algoritmen die beschrijvende (populatie)statistieken berekenen, zoals gemiddelden en standaardafwijkingen, geen AI-systeem zijn. Bij het berekenen van het gemiddelde inkomen van een groep natuurlijke personen is de output geen “*voorspelling, content, aanbeveling of beslissing/besluit*”. Deze lezing wordt bevestigd door paragraaf (46) uit de richtlijnen. Wanneer een statistisch model gebruikt wordt om een score te schatten voor een nieuw datapunt, dan is er wel sprake van een voorspelling. Volgens deze redeneerwijze gelden eenvoudige dataverwerking- en visualisatiesystemen, zoals dashboards die populatiestatistieken weergeven, niet als AI-systeem. Deze lezing wordt bevestigd door paragraaf (47) uit de richtlijnen.

<sup>18</sup> Supra noot 10

<sup>19</sup> Supra noot 10

<sup>20</sup> Zie ook Advies geautomatiseerde besluitvorming, Autoriteit Persoonsgegevens <https://www.autoriteitpersoonsgegevens.nl/documenten/advies-geautomatiseerde-besluitvorming>

<sup>21</sup> Hoe ‘algotrudentie’ kan bijdragen aan een verantwoorde inzet van machine learning-algoritmen, A. Meuwese, J. Parie, A. Voogt, 2024, Nederlands Juristenblad (NJB) [https://algorithmaudit.eu/nl/knowledge-platform/knowledge-base/white\\_paper\\_algotrudentie/](https://algorithmaudit.eu/nl/knowledge-platform/knowledge-base/white_paper_algotrudentie/)

<sup>22</sup> Art. 22 AVG. Zie ook supra noot 16.

We zien kenmerken van de output van een AI-systeem daarom als een belangrijke factor om AI-systemen te onderscheiden van andere algoritmen, zeker in combinatie met en in relatie tot de begrippen autonomie en inferentie. Het type output dat door een algoritmische toepassing wordt gegenereerd wordt daarom als eerste uitgevraagd de dynamische vragenlijst om AI-systemen te identificeren. Zie [6.1 Q1 – Welke van de volgende categorieën beschrijft de uitkomst van de toepassing het beste?](#)

Voor de vraag of een algoritme met een “voorspelling, content, aanbeveling of beslissing/besluit” als output ook daadwerkelijk een AI-systeem betreft, is het belangrijk om na te gaan hoe de output tot stand komt. Manieren om het proces van de verkregen output, vanuit het licht van de AI-systeem definitie, nader te onderzoeken wordt toegelicht in [3. Inferentie](#).

## 2.8 Fysieke en virtuele omgeving

Overweging 12 vermeldt dat “*invloed kunnen zijn op fysieke of virtuele omgevingen*” uit de AI-systeem definitie de volgende betekenis kent:

*“Voor de toepassing van deze verordening moeten onder omgevingen de contexten worden verstaan waarin de AI-systemen werken, terwijl de output die door het AI-systeem wordt gegenereerd een uiting is van de verschillende functies van AI-systemen en de vorm kan aannemen van voorspellingen, content, aanbevelingen of besluiten.”* – zie [overweging 12](#) zin 10.

De fysieke en virtuele omgeving zijn complementair. De combinatie van de twee omgevingen is uitputtend. Het gaat hier dus om systemen die überhaupt invloed uitoefenen, op welke omgeving dan ook. Dit sluit systemen uit die helemaal geen invloed uitoefenen, bijvoorbeeld omdat ze nog niet in gebruik genomen zijn. Dit beeld wordt bevestigd door paragraaf (60) uit de richtlijnen. Verder bieden geen van de gebruikte bronnen behulpzame

duiding voor het begrip ‘invloed’. Er lijkt bijna geen systeem denkbaar dat niet invloed uitoefent op een omgeving.

In ieder geval is de vereiste van ‘invloed op de fysieke of virtuele omgeving’ geen criterium waarmee AI-systemen van algoritmen onderscheiden kunnen worden. Het begrip invloed wordt indirect ook besproken in de begrippen in [3. Inferentie](#) en [4. Autonomie](#).

## 3. Inferentie

Inferentievermogen is het belangrijkste element van de definitie om AI-systemen te onderscheiden van reguliere algoritmen. In deze sectie worden verschillende passages uit overweging 12 geanalyseerd en gerelateerd aan de AI-systeem definitie.

Overweging 12 vermeldt dat inferentievermogen de volgende betekenis kent:

*“Een belangrijk kenmerk van AI-systemen is hun inferentievermogen. Dit inferentievermogen slaat op het proces waarbij output, zoals voorspellingen, content, aanbevelingen of besluiten, wordt verkregen waarmee fysieke en virtuele omgevingen kunnen worden beïnvloed, en op het vermogen van AI-systemen om modellen of algoritmen, of beide, af te leiden uit input of data.”* – Zie [overweging 12](#) zin 3-4.

*“De technieken die inferentie mogelijk maken bij de opbouw van een AI-systeem, omvatten benaderingen op basis van machinaal leren waarbij aan de hand van data wordt geleerd hoe bepaalde doelstellingen kunnen worden bereikt, alsook op logica en kennis gebaseerde benaderingen waarbij iets wordt geïnfereerd uit gecodeerde kennis of uit een symbolische weergave van de op te lossen taak.”* – Zie [overweging 12](#) zin 5.

*“Het inferentievermogen van een AI-systeem overstijgt de elementaire verwerking van data door leren, redeneren of modelleren mogelijk te maken.”*  
– Zie [overweging 12](#) zin 6.

De eerste en laatste zin kaderen de interpretatie: het inferentievermogen is een belangrijk kenmerk waaraan AI-systemen geïdentificeerd kunnen worden en het is specifiek dit kenmerk dat AI-systemen onderscheidt van andere dataverwerking door *“leren, redeneren of modelleren”*. Merk op dat enkel sprake hoeft te zijn van een van deze drie kenmerken: leren, redeneren óf modelleren.

Aan de hand van deze drie kernbegrippen worden bovenstaande zinnen uit [overweging 12](#) geanalyseerd.

### 3.1 Leren en modelleren

[Overweging 12](#) benoemt dat inferentievermogen betrekking heeft op:

*“het vermogen van AI-systemen om modellen of algoritmen, of beiden, af te leiden uit input of data.”*  
– zie [overweging 12](#) zin 4.

Wanneer modellen of algoritmen zijn afgeleid uit data, is er sprake van modelleren of leren. Voorbeelden hiervan zijn het leren van de gewichten van een neurale netwerk gebruikt voor spraakherkenning of een algoritme dat kenmerken selecteert voor profilering. Verschillende experts gebruiken hiervoor verschillende termen zoals leren, modeleren, trainen of fitten. Zie ook de genoemde voorbeelden over (un)supervised-, reinforcement- en deep learning in paragrafen (33)-(38) uit de richtlijnen. Ongeacht de gebruikte terminologie, volgt uit deze passage van [overweging 12](#) dat er sprake is van inferentie wanneer een model of algoritme wordt afgeleid uit input of data. Uit deze passage blijkt dat AI-systemen het vermogen moeten hebben om af te leiden. Hieronder verstaan we dat er een mate van automatisering moet zijn

bij het afleiden van modellen of algoritmen uit data, specifiek in de ontwikkelingsfase van het AI-systeem. Zie ook paragraaf (28) uit de richtlijnen. Wanneer eerst een data-analyse wordt uitgevoerd, bijvoorbeeld om de gemiddelde leeftijd van een populatie vast te stellen, wat als input dient voor domeinexperts die handmatig een algoritme opstellen, dan is er geen sprake van een situatie waarin een AI-systeem een algoritme afleidt uit data.

[Overweging 12](#) benoemt verder:

*“De technieken die inferentie mogelijk maken bij de opbouw van een AI-systeem, omvatten benaderingen op basis van machinaal leren ...”* – zie [overweging 12](#) zin 6.

Bij machinaal leren wordt een model ‘geleerd’ uit een dataset, vaak training data genoemd. In veel gevallen wordt statistiek gebruikt om modelparameters te berekenen die het beste passen bij de beschikbare dataset. Voor datawetenschappers laat het berekenen van parameters op basis van inputdata zich het beste uitdrukken als de `.fit()`-functie, zoals gebruikt in scikit-learn en statsmodels Python-software. Het berekenen van een gemiddelde, aan de hand van een simpele formule, is een voorbeeld van een parameter. Zo ook het berekenen van lineaire regressie-coëfficiënten, aan de hand van een meer uitgebreide formule, of de gewichten van een neurale netwerk aan de hand van een zeer complexe formule.

Machinaal leren omvat ook het leren van de variabelen en drempelwaarden van een beslisboom voor regressie en classificatie. Dit betreft het leren van een simpele beslisboom, maar ook het leren van groepen beslisbomen, zoals ensemble-based tree learning. Denk aan: random forest, xgboost, explainable boosting etc. Dit zijn allen voorbeelden van machinaal leren. Deze lezing wordt bevestigd door paragrafen (30)-(33) uit de richtlijnen.

Of een datagedreven toepassing machinaal leren (Engels: 'machine learning') genoemd wordt verschilt per domeinexpertise. Een econometrist of statisticus zal het ontwikkelen van een lineair model zoals een regressievergelijking of general linear model (GLM) waarschijnlijk geen machine learning noemen. In deze gevallen wordt een model afgeleid van een beschikbare dataset. We zien op basis van de tekst van overweging 12 geen richtlijnen welke techniek wel of geen machinaal leren betrekken. De definitie van machinaal leren is niet relevant voor de definitie van een AI-systeem. We concluderen dat alle gevallen wanneer een model gefit, getraind of geleerd wordt uit data vallen onder inferentievermogen.

Het enkel afleiden van modelparameters of regels uit input data, bijvoorbeeld het leren van regressiecoëfficiënten, maakt een model of algoritme nog geen AI-systeem. Overweging 12 vermeldt dat inferentievermogen slaat op:

- a) *"het proces waarbij output voorspellingen, content, aanbevelingen of besluiten wordt verkregen [...] waarmee omgevingen kunnen worden beïnvloed";*
- b) *"het vermogen van AI-systemen om modellen of algoritmen, of beide, af te leiden uit input of data" – zie overweging 12 zin 4.*

Bij het leren van regressiecoëfficiënten wordt voldaan aan b) – namelijk: `.fit()` – maar niet aan a). Bij het leren van regressiecoëfficiënten worden immers geen voorspellingen gemaakt voor nieuwe datapunten. Bij a) gaat het over het toepassen van het geleerde model of algoritme op nieuwe data. Naar dit proces wordt door datawetenschappers verwezen als `.predict()`, zoals gebruikt in scikit-learn en statsmodels Python-software. Dit relateert ook aan de door de wetgever gespecificeerde output van een AI-systeem, namelijk: *"voorspellingen, content/inhoud, aanbevelingen of beslissingen/besluiten"*. Enkel na het toepassen van deze

`.predict()`-functie wordt dus output gegenereerd die volgens de definitie van een AI-systeem vereist is. In het geval van aanbevelingen en beslissingen wordt vaak eerst een score voorspeld, middels een geleerd model, waarna aan de hand van deze score een aanbeveling wordt gedaan of beslissing wordt genomen. Een model – gebaseerd op statistiek of machine learning – is een AI-systeem als modelparameters of regels worden berekend én daarna een voorspelling of soortgelijk volgt. Zie ook [2.7 Voorspellingen, inhoud, aanbevelingen of beslissingen](#).

Het 'genereren van output' is een belangrijke factor om AI-systemen van algoritmen te onderscheiden en is daarom opgenomen in de dynamische vragenlijst. De eerste vraag uit de vragenlijst houdt verband met de gegenereerde output. De tweede vraag gaat over het afleiden van modellen of algoritmen uit data. Middels een vervolgvraag, waarin door de gebruiker een toelichting moet worden gegeven over de gebruikte methoden, kan per geval worden beoordeeld of er sprake is van een AI-systeem. Zie [6. Dynamische vragenlijst](#).

### 3.1.1 Tegenstrijdigheden in de richtlijnen van de Europese Commissie

Zoals in [2. AI-systeem definitie](#) toegelicht ondersteunen paragrafen (1)-(41) uit de richtlijnen onze zienswijze op de definitie van een AI-systeem. Voor paragrafen (42)-(45) en (49)-(51) is dit niet het geval. In deze paragrafen worden uitzonderingen beschreven die lastig verenigbaar zijn onze zienswijze op leren en modelleren. De uitzonderingen uit de richtlijnen worden onderbouwd met argumenten die tegenstrijdig zijn met de tekst van de AI-verordening. In deze subsectie volgt een toelichting welke specifieke passages ambiguïteit introduceren. In paragrafen (42)-(45) uit sectie 5 van de richtlijnen wordt beargumenteerd dat: *"systemen gebruikt om mathematische optimalisatie te verbeteren of om traditionele optimalisatiemethoden te versnellen of*

benaderen, zoals lineaire en logistische regressie, buiten de reikwijdte van de definitie van een AI-systeem vallen. Dit is het geval omdat de elementaire verwerking van data niet wordt overstegen, ondanks dat er sprake is van inferentievermogen". De eerste zin hangt deze uitzondering op aan de manier waarop een systeem gebruikt wordt – namelijk voor het verbeteren van optimalisatie – in plaats van aan eigenschappen het systeem zelf. De definitie gaat echter uitsluitend over kenmerken van het systeem. De manier waarop een systeem gebruikt wordt is geen deel van de definitie en kan dus niet gebruikt worden als argument voor een uitzondering op basis van de wettekst. De tweede zin staat nog directer op gespannen voet met de wettekst. Uit overweging 12 van de AI-verordening volgt dat met *"inferentievermogen een AI-systeem elementaire verwerking van data ... overstijgt"*. Paragraaf 42 uit de richtlijnen spreekt de beschrijving van inferentievermogen zoals toegelicht in overweging 12 van de AI-verordening dus tegen.

Verwarring volgt ook door het in deze paragrafen aangehaalde argument dat: *"een indicatie dat systeem de elementaire verwerking van data niet overstijgt afgeleid kan worden uit het feit dat de methode al voor vele jaren wordt gebruikt"*. Statistisch gezien doet het er niet toe of de methode waar een AI-systeem gebruik van maakt al lange of korte tijd wordt gebruikt. Zoals eerder toegelicht zijn lineaire en logistische regressies elementaire vormen van machine learning die vervolgens gebruikt kunnen worden om voorspellingen te genereren. Ook neurale netwerken zijn inmiddels decennia oude technologie. Het lijkt onhoudbaar om op grond van het 'gevestigde methoden'-argument deze systemen buiten de reikwijdte van de AI-verordening te plaatsen.

In paragraaf (48) wordt geopperd dat een gebrek aan aanpassingsvermogen een reden is om systemen uit te zonderen van de definitie. Dit is tegenstrijdig

met de strekking van paragrafen (42)-(43) uit dezelfde richtlijnen, waar expliciet wordt gesteld dat aanpassingsvermogen geen voorwaarde is voor de definitie van een AI-systeem. Het ontbreken van een optioneel kenmerk kan niet gebruikt worden om te beargumenteren dat een systeem geen AI-systeem is.

In paragrafen (49)-(51) wordt gesteld dat alle machine-gebaseerde systemen die functioneren aan de hand van simpele regels – die mogelijk middels statistisch leren tot stand zijn gekomen en daarmee machine learning betreffen – niet kwalificeren als AI-systeem, *"vanwege hun prestaties"*. Ook deze passage staat op gespannen voet met de wettekst. In de definitie van AI-systeem en overweging 12 wordt complexiteit van het systeem of de kwaliteit van prestaties niet toegeschreven als kenmerk van inferentie. Wel wordt in zin 5 uit overweging 12 toegelicht dat *"de technieken die inferentie mogelijk maken bij de opbouw van een AI-systeem, omvatten benaderingen op basis van machinaal leren waarbij aan de hand van data wordt geleerd hoe bepaalde doelstellingen kunnen worden bereikt"*. Ook met machine learning methodes die resulteren in eenvoudige regels, zoals lineaire regressie of beslisboom leren, wordt aan de hand van data geleerd hoe een doelstelling bereikt kan worden. Het opbrengen van argumenten rond complexiteit en prestaties zijn niet alleen nieuw ten aanzien van de wettekst om systemen uit te zonderen van de AI-verordening, ze brengen ook extra onduidelijkheid: wanneer een systeem 'simpel' genoeg is om onder deze uitzondering te vallen?

Als voorbeeld van simpele systemen die niet kwalificeren als AI-systeem worden in paragrafen (50)-(51) een 'gemiddelde baseline predictie' en een 'statisch schattingsysteem' aangehaald. Deze methode zijn echter geen vorm van machine learning – er worden geen modelparameters of beslisregels afgeleid. De voorbeelden kwalificeren volgens de



definitie uit de AI-verordening niet als AI-systeem omdat er geen sprake is van leren of redeneren. Met de verwijzing naar machine learning in paragraaf (49) wordt onnodige verwarring gecreëerd over de betekenis van machine learning en hoe dit begrip zich verhoudt tot de definitie van een AI-systeem.

Met oog op paragraaf (7) uit de richtlijnen – waaruit volgt dat de richtlijnen een lagere juridische status hebben dan de wettekst – wordt in deze white paper de wettekst uit de AI-verordening gevolgd om de reikwijdte van een AI-systeem te bepalen. De bovenstaande uitzonderingen, zijn vanwege hun inconsistentie niet meegenomen in onze interpretatie van de definitie.

### 3.2 Redeneren: op logica en kennis gebaseerde benaderingen

Inferentievermogen kan ook betrekking hebben op het vermogen van een AI-systeem om te redeneren – zie [overweging 12](#) zin 6. Hieruit blijkt dat er een type systemen is waarbij geen sprake is van leren of modelleren, maar wel van inferentie.

Dit roept de vraag op: bij welk type algoritmen is er sprake van redeneren? Overweging 12 noemt een aantal voorbeelden van systemen die hier niet onder vallen: *“regels die uitsluitend door natuurlijke personen zijn vastgesteld om automatisch handelingen uit te voeren”* en *“elementaire verwerking van data”* – zie [overweging 12](#) zin 2 en 6.

Verder biedt overweging 12 weinig aanvullende duiding voor het begrip “redeneren”. In overweging 12 wordt wel het volgende benoemd:

*“De technieken die inferentie mogelijk maken bij de opbouw van een AI-systeem, omvatten ... op logica en kennis gebaseerde benaderingen waarbij iets wordt geïnfereerd uit gecodeerde kennis of uit een*

*symbolische weergave van de op te lossen taak.”* – zie [overweging 12](#) zin 5.

Bij op logische en kennis gebaseerde benaderingen van AI is geen sprake van machinaal leren, er is hier sprake van inferentievermogen omdat er sprake is van redeneren.

Op logica en kennis gebaseerde benaderingen van AI worden in de academische wereld ook wel symbolische AI genoemd, zo ook in het OESO-memorandum.<sup>23</sup> Symbolische AI is sinds de jaren ‘80 en ‘90 gebruikt in bijvoorbeeld schaakcomputers of medische beslissingsondersteuningssystemen. Met de grote vooruitgang op het gebied van machine learning, deep learning en generatieve AI, is er echter steeds minder aandacht uitgegaan naar deze vorm van AI.

Overweging 12 bevat geen aanvullende informatie over de definitie en interpretatie van op logica en kennis gebaseerde benaderingen van AI-systemen. In het originele voorstel van de AI-verordening is wel aanvullende duiding opgenomen: *“Op logica en op kennis gebaseerde benaderingen, waaronder kennisrepresentatie, inductief (logisch) programmeren, kennisbanken, inferentie- en deductiemachines, (symbolisch) redeneren en expertsystemen”*.<sup>24</sup> Deze voorbeelden zijn in lijn met interpretaties van symbolische AI in de academische wereld. Deze voorbeelden zijn ook opgenomen in de richtlijnen in paragraaf (39).

Om op logica en kennis gebaseerde AI-systemen te onderscheiden van algoritmen moeten we onderscheiden wat deze technieken anders maakt dan *“regels die uitsluitend door natuurlijke personen zijn vastgesteld om automatisch handelingen uit te voeren”* en *“de elementaire verwerking van*

<sup>23</sup> Supra noot 10

<sup>24</sup> Zie Annex I van Voorstel voor een verordening van het Europees Parlement en de Raad tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie (Wet op de artificiële intelligentie) en tot wijziging van bepaalde wetgevingshandelingen van de Unie. <https://eur-lex.europa.eu/legal-content/NL/TXT/HTML/?uri=CELEX:52021PC0206>

data". We duiden op logica en kennis gebaseerde benaderingen aan de hand van twee academische standaardwerken in AI: Artificial Intelligence van Russel and Norvig en [Artificial Intelligence](#) van Poole and Mackworth.<sup>25</sup> Samengevat bestaan op logica en kennis gebaseerde benaderingen van AI uit een knowledge-base en redeneercomponent. Paragraaf (30) uit de richtlijnen bevestigen deze lezing.

- i) **Knowledge-base:** Een expliciete representatie van (domein)kennis. Deze kennis bestaat uit bijvoorbeeld uit regels, feiten en relaties. Zie paragraaf (39) uit de richtlijnen. Hiervoor wordt vaak logica gebruikt, waarbij kennis wordt uitgedrukt in *proposities* en *connectieven*, zoals  $\neg A$ ,  $A \wedge B$ ,  $A \vee B$ , waarbij een propositie (bijv. A) enkel waar of onwaar kan zijn. Andere bekende vormen van knowledge bases zijn knowledge graphs.
- ii) **Redeneer component:** Deze component definieert hoe het systeem kan redeneren over de kennis in de knowledge-base en input data, bijvoorbeeld door middel van formele logica. Deze component wordt ook wel een *inference engine* genoemd, bijvoorbeeld een deductive of inductive engine. Redeneren kan ook plaatsvinden door operaties als sorteren, zoeken of matchen. Zie ook paragraaf (39) uit de richtlijnen. Door middel van de redeneercomponent kunnen nieuwe kennis én nieuwe regels worden afgeleid.

De richtlijnen benoemen dit onderscheid tussen een knowledge-base en redeneer component niet expliciet. Uit paragraaf (39) volgt dat: *"AI systems learn from knowledge including rules, facts and relationships encoded by human experts. Based on the human experts encoded knowledge, these systems can 'reason' via deductive or inductive engines or using operations such as sorting, searching, matching, chaining."*

Beide componenten worden zorgvuldig opgebouwd en vereisen veel domeinkennis. Vaak worden deze benaderingen gebruikt wanneer er sprake is van een grote hoeveelheid vaste kennis en regels in een domein, waarover vervolgens geredeneerd kan worden. Denk aan medisch beslissingsondersteuningsysteem waarbij de kennisbasis medische feiten bevat over symptomen, diagnoses en mogelijke behandelingen, het redeneersysteem kan dan op basis van input data van symptomen een mogelijke behandeling voorstellen.

Op logica en kennis gebaseerde benaderingen van AI vormen tegenwoordig een minderheid. Meestal worden deze technieken vandaag de dag gebruikt in combinatie met vormen van machine learning. In dat geval zou het systeem vanwege het gebruik van machine learning een AI-systeem zijn, zie [3.1 Leren en modelleren](#). Ontwikkelaars die dit type technologie gebruiken zijn hier waarschijnlijk op de hoogte dat zij dit type AI-systeem gebruiken. We zien het 'op logica en kennis gebaseerde benaderingen'-aspect alleen in die zeldzame gevallen dat geen ML gebruikt wordt, als een belangrijk vereiste om AI-systemen van algoritmen te onderscheiden.

Op logica en kennis gebaseerde benaderingen zijn een factor waarmee AI-systemen onderscheiden kunnen worden van algoritmen. Er is daarom een vraag over ingevoegd in de dynamische vragenlijst. Zie [6.3 Q3 – Is de toepassing automatisering van door mensen opgestelde regels?](#)

### 3.2.1 Redeneren, gecodeerde kennis en rule-based systemen

Er zijn, anders dan op logica en op kennis gebaseerde benaderingen, geen andere benaderingen waar met betrekking tot redeneren naar wordt verwezen in de AI-verordening. In de context van de AI-

<sup>25</sup> Artificial Intelligence: foundations of computational agents. Poole, D.L. and Mackworth, A.K., 2010. Cambridge University Press. Artificial intelligence: a modern approach. Russell, Stuart J., and Peter Norvig. Pearson, 2016. Zie voor een begrijpelijke uitleg ook: [https://en.wikipedia.org/wiki/Knowledge-based\\_systems](https://en.wikipedia.org/wiki/Knowledge-based_systems)



verordening houdt redeneren dus alleen verband met op logica en kennis gebaseerde benaderingen. Er kan worden beargumenteerd dat in het geval van eenvoudige handmatig opgestelde regelgebaseerd algoritme sprake is van redeneren. Dit is echter onverenigbaar met de toelichting die op de definitie van een AI-systeem wordt gegeven: “de definitie [moet] gebaseerd zijn op de belangrijkste kenmerken van AI-systemen die het onderscheiden van eenvoudigere traditionele softwaresystemen of programmeringsbenaderingen...”. Als regelgebaseerde algoritmen redeneren, redeneren alle soorten softwaresystemen en dat gaat tegen de strekking van voorgaande zin in. Ongeacht of er sprake is van redeneren geldt dat “regels die uitsluitend door natuurlijke personen zijn vastgesteld om automatisch handelingen uit te voeren” geen AI-systeem zijn – [overweging 12](#) zin 2.

De passage over “gecodeerde kennis” – [overweging 12](#) zin 6 – moet ook in licht van op logica en kennis gebaseerde benaderingen worden gezien. Gecodeerde kennis relateert in deze context aan de vorm waarin kennis wordt gecodeerd in een knowledge-base, zoals hierboven beschreven. Regelgebaseerde algoritmen, waarin menselijke kennis is gecodeerd, worden in de praktijk niet toegepast middels een knowledge-base (ook wel: een ‘op kennis gebaseerde benadering’). De passage “gecodeerde kennis” heeft dus geen betrekking op regelgebaseerde algoritmen die we kennen uit de praktijk.

## 4. Autonomie

Overweging 12 vermeldt dat “*verschillende niveaus van autonomie*” uit de AI-systeem definitie de volgende betekenis kent:

*“AI-systemen worden zodanig ontworpen dat zij in verschillende mate autonoom kunnen functioneren, wat betekent dat zij een zekere mate*

*van onafhankelijkheid van menselijke betrokkenheid bezitten en zonder menselijke tussenkomst kunnen functioneren.”* – zie [overweging 12](#) zin 11.

Om te voldoen aan de ‘autonomie’-vereiste moet er sprake zijn van *enige mate* van autonomie, zoals ook besproken in [2.3 Verschillende niveaus van autonomie](#) en paragraaf (14) uit de richtlijnen.

‘Enige mate’ is een zwakke vereiste: een systeem hoeft niet geheel autonoom te zijn om aan deze vereiste te voldoen.

Het OESO-memorandum vermeldt dat: “de autonomie van een AI-systeem betrekking heeft op de mate waarin een systeem kan *leren* of *handelen* zonder menselijke betrokkenheid”. Dit impliceert dat ieder lerend algoritme in een zekere mate autonoom is. Oftewel, als aan de inferentie-vereiste wordt voldaan, wordt ook aan de autonomie-vereiste voldaan. Verder wordt in het OESO-memorandum autonomie gekoppeld aan de verschillende types gegenereerde output, waarbij beslissingen het meest autonoom en voorspellingen het minst autonoom zijn. Uit deze formulering maken we op dat OESO ook voorspellingen als in ‘enige mate’ van autonoom beschouwd.

De richtlijnen benoemen in paragraaf (18) dat een systeem dat handmatig ingevoerde input gebruikt om zelf een output te genereren, een systeem is met “een zekere mate van onafhankelijkheid van handelen”, oftewel een zekere mate van autonomie. Deze ondergrens geldt voor alle systemen die uit input data de genoemde output types genereren. Met een beschouwing van het type output van een algoritme ([2.7 Voorspellingen, inhoud, aanbevelingen of beslissingen](#)) en het inferentievermogen ([3. Inferentie](#)) kan daarmee dus ook de autonomie-vereiste worden ingevuld.

Al met al concluderen we dat de ‘autonomie’-vereiste geen aanvullend onderscheidend vermogen heeft ten opzichte van de andere vereisten om AI-systemen van algoritmen te scheiden. Wel speelt menselijke tussenkomst een belangrijke rol in het onderscheiden van impactvolle algoritmen van reguleren algoritmen. Er is daarom een vraag over menselijke tussenkomst opgenomen in als vraag in de dynamische vragenlijst. Zie [6.7 Q7 – Welke van de volgende opties beschrijft het effect van de toepassing op de uitkomst het beste?](#)

Paragraaf (21) uit de richtlijnen benadrukt dat een gebrek aan menselijke tussenkomst aanvullende risico’s met zich mee brengt en daarom extra maatregelen vergt. Dit sluit aan bij het concept ‘geautomatiseerde besluitvorming’ zoals gebruikt in de Algemene Verordening Gegevens verwerking (AVG) en recente adviezen die over dit onderwerp door de Autoriteit Persoonsgegevens (AP) en de landsadvocaat zijn gepubliceerd.<sup>26</sup>

## 5. Impactvolle algoritmen

Ook andere algoritmen dan AI-systemen kunnen grote invloed hebben op betrokkenen. Een voorbeeld zijn profileringsalgoritmen waarvan de regels zijn opgesteld door experts. Een voorbeeld is het algoritme van de Dienst Uitvoering Onderwijs (DUO) dat in de periode 2012-2023 is gebruikt om studenten een risicoscore toe te kennen voor misbreuk van de uitwonendenbeurs.<sup>27</sup> Zie ook [2.1 Interpretatie van de definitie van een AI-systeem aan de hand van overweging 12](#).

Uit de ‘Voortgang vulling algoritmeregister november 2024’ van het Ministerie van Financiën volgt dat van de Belastingdienst, Douane, Dienst Toeslagen en het beleidsdepartement geen hoog-risico AI-systemen in gebruik hebben.<sup>28</sup> Daarentegen gebruiken deze organisaties respectievelijk 77, 55, 22 en 8 impactvolle algoritmen. Deze voortgang maakt duidelijk dat het belangrijk is om niet alleen AI-systemen, maar ook impactvolle algoritmen te identificeren zodat hier gepaste beheersmaatregelen op toegepast kunnen worden. Deze white paper en bijbehorende dynamische vragenlijst richten zich daarom niet enkel op het identificeren van AI-systemen, zoals gedefinieerd in de AI-verordening, maar ook op het identificeren van impactvolle algoritmen.

Waar in de vorige secties de vraag centraal staat hoe AI-systemen te identificeren, staat in deze sectie de vraag centraal hoe ‘impactvolle algoritmen’ van andere algoritmen onderscheiden kunnen worden. Er volgt een toelichting aan de hand van welke kenmerken impactvolle algoritmen herkend kunnen worden. Daartoe wordt de Handreiking Algoritmeregister<sup>29</sup> (hierna: ‘Handreiking’) van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties gevolgd. Impactvolle algoritmen moeten gepubliceerd worden in het Algoritmeregister, tenzij er sprake is van een uitzonderingsgrond.<sup>30</sup>

Aan de hand van deze analyse wordt een dynamische vragenlijst voor identificatie van AI-systemen aangevuld met vier vragen aan de hand waarvan impactvolle algoritmen geïdentificeerd kunnen worden. Zie sectie 6.4-6.7 in [6. Dynamische vragenlijst](#).

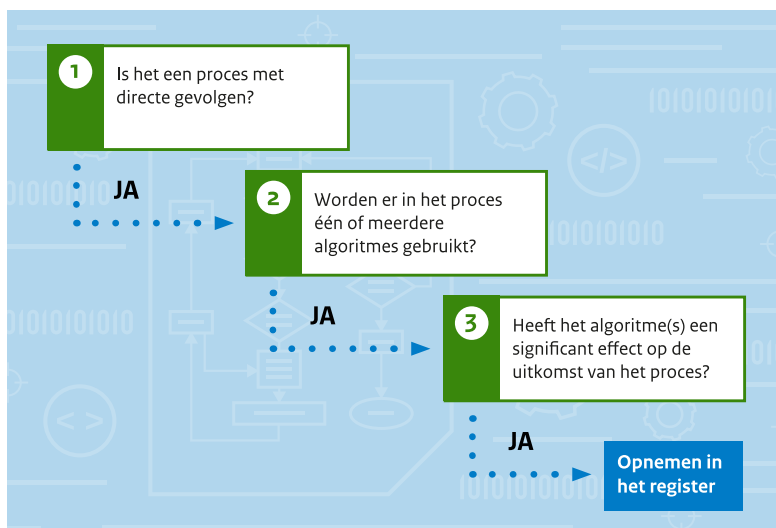
<sup>26</sup> Art. 22 AVG; [Advies geautomatiseerde besluitvorming](#), Autoriteit Persoonsgegevens (2024); [Advies landsadvocaat over geautomatiseerde selectietechnieken](#), Pels Rijcken.

<sup>27</sup> [Vooringenomenheid voorkomen](#), Algorithm Audit (2024)

<sup>28</sup> [Kamerstukken II 2024-25 2025D00512](#)

<sup>29</sup> [Handreiking Algoritmeregister](#) van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2023).

<sup>30</sup> Uitzonderingsgronden zoals vermelden in de Handreiking zijn: “wettelijke uitzonderingsgronden zoals gespecificeerd in de Wet open overheid (Woo) en Wet publieke gezondheid (Wpg), of ‘gaming the system’.”



Figuur 2 - Vragen uit de Handreiking Algoritmeregister aan de hand waarvan bepaald kan worden of een algoritme een 'impactvol algoritme' betreft.

De definitie van een 'impactvol algoritme' uit de Handreiking luidt als volgt:

- > **Directe gevolgen:** algoritme heeft directe gevolgen voor betrokkenen (burger, organisatie), denk aan: opleggen van boete of weigeren subsidie;
- > **Classificatie:** Algoritme beïnvloedt hoe de overheid een betrokkene of groep categoriseert of benadert, denk aan: profilering of risico-indicatie voor controle.

Bovenstaande categorieën worden in de Handreiking aan de hand van drie vragen toegelicht. Zie Figuur 2.

De drie vragen luiden als volgt:

1. Is het een proces met directe gevolgen?
2. Worden er in het proces één of meerdere algoritmen gebruikt?
3. Heeft het algoritme een significant effect op de uitkomst van het proces?

Merk op dat de Handreiking eerst nagaat wat de impact is van het proces op betrokkenen (burgers, organisaties) voordat het effect van het algoritme op dat proces wordt uitgevraagd.

## 5.1. Is het een proces met directe gevolgen?

Een impactvol algoritme wordt ingezet in een proces dat directe gevolgen heeft voor betrokkenen.

1. *"Het betreft hier processen met impact, dat zullen over het algemeen besluitvormingsprocessen zijn. Of het proces draagt bij aan hoe de overheid een betrokkene of groep categoriseert of benadert, bijvoorbeeld door het gebruiken van weegfactoren of voorspellingen. Dit kan gevolgen hebben voor de benadering of behandeling. Voorbeelden van dit laatste zijn risicoscans en algoritmen voor fraudesignalering."*
2. *"Onder de gevolgen vallen in ieder geval rechtsgevolgen. Een rechtsgevolg houdt in dat het besluit volgens de Algemene wet bestuursrecht (Awb) van invloed is op de wettelijke rechten van een betrokkene, iemands juridische status of zijn rechten uit hoofde van een overeenkomst. Ook betreft het feitelijke gevolgen die raken aan belangen van een persoon zoals financiële gevolgen (wel/geen toeslag), gevolgen voor grondrechten (wel/geen rechtsbescherming) en juridische gevolgen (wel/niet in NL blijven, een woning toegewezen*

krijgen). Ook de selectie voor een inspectie of controle wordt gezien als een gevolg.”

3. *“Betrokkenen betreft iedereen die te maken krijgt met de Nederlandse overheid. We vatten dit samen als burgers en organisaties.”*

Directe gevolgen zijn breed gedefinieerd, waardoor veel algoritmen potentieel impactvolle algoritmen zijn. Om na te gaan of er sprake is van directe gevolgen in het algoritme-gedreven besluitvormingsproces wordt allereerst nagegaan of er sprake is van een beslissing. Een besluit moet breed geïnterpreteerd worden. Niet alleen een formeel besluit, zoals gedefinieerd in de Algemene wet bestuursrecht (Awb art.1:3), heeft impact op burgers en organisaties, ook andere beslissingen kunnen aanzienlijke gevolgen hebben voor betrokkenen en kunnen daarmee behoren tot de categorie impactvolle algoritmen. Daarom wordt in het vervolg van deze paper en in de dynamische vragenlijst gesproken over beslissingen.

Het kan hier gaan om ofwel beslissingen voor individuele betrokkenen, ofwel om het benaderen of categoriseren van groepen betrokkenen. Wanneer is vastgesteld dat sprake is van een beslissingsproces voor individuele burgers, moet worden nagegaan wat voor een soort beslissing wordt genomen in het algoritme-gedreven beslissingsproces (prioritering, formele klacht of bezwaar, aanvraag met financiële gevolgen etc.). Op deze manier wordt bepaald of deze beslissing directe impact heeft. Voor deze aspecten zijn 3 vragen opgenomen in de vragenlijst. Zie [6.4 Q4 – Wordt in het proces een beslissing genomen voor individuele burger, organisatie of medewerker?](#), [6.5 Q5 – Wat voor soort beslissing wordt er genomen in dit proces?](#) en [6.6 Q6 – Draagt het proces bij aan hoe de overheid \(groepen\) burgers, organisaties of ambtenaren categoriseert of benadert?](#)

## 5.2. Worden er in het proces één of meerdere algoritmen gebruikt?

Een impactvol algoritme heeft betrekking op een proces waarin één of meerdere algoritmen gebruikt worden.

1. *“In het proces worden één of meerdere algoritmen gebruikt.”*
2. *“Een proces bestaat vaak uit meerdere stappen, waarbij sommige door algoritmen en sommige door mensen worden uitgevoerd.”*
3. *“Een organisatie weet vaak zelf het best welke set aan stappen gezamenlijk het proces met impact op betrokkenen vormen.”*

Uit bovenstaande beschrijving volgt niet waarom het relevant is dat één of meerdere algoritmen betrokken zijn in een besluitvormingsproces. Het is geen onderscheidende factor om impactvolle algoritmen te scheiden van andere algoritmen en is daarom niet opgenomen als vraag in de vragenlijst.

## 5.3. Heeft het algoritme een significant effect op de uitkomst van het proces?

Een impactvol algoritme heeft een significant effect op de uitkomst van het proces.

1. *“Hierbij gaat het niet over processen waarin het algoritme een handmatige werkinstructie automatiseert/ digitaliseert. Zoals algoritmen waarbij alle parameters wettelijk vaststaan en het algoritme een (complexe) beslisboom afloopt enkel op basis van deze parameters”.*
2. *“Hierbij gaat het wel over processen waarbij het algoritme een besluit beïnvloedt. Zoals algoritmen waarbij een weegfactor wordt meegegeven wat de vervolgstap in het proces (mede) bepaalt. De weegfactoren worden daarbij ingevuld door de ruimte of vrijheid die een bestuursorgaan toekomt bij het uitvoeren van diens taken.”*

Een significant effect op de uitkomst van het proces kan worden gelezen als: hoe direct is de uitkomst van het proces verbonden met de output van het algoritme? Hierin zit ook impliciet een ander element: zou het proces zonder het algoritme anders verlopen? Er is hier een parallel met de definitie van AI-systeem in de AI-verordening. Daarin wordt autonomie gekoppeld aan de verschillende types uitkomsten en de mate van menselijke tussenkomst.

De Handreiking benoemt expliciet dat een *“algoritme [dat] een handmatige werkinstructie automatiseert/ digitaliseert”* geen impactvol algoritme is. Deze handmatige instructies zijn te onderscheiden in a) wet- en regelgeving en formeel beleid en b) meer informele werkinstructies inclusief andere door mensen opgestelde regels. Categorie a) is formeel vastgelegd in wet- en regelgeving en daarmee onderhevig aan democratische en institutionele controle. De automatisering van deze regels verandert de uitkomst van het proces niet. Daarmee is dit type algoritmen minder impactvol dan algoritmen waarvoor de regels niet zo duidelijk zijn gespecificeerd.

Wet- en regelgeving of formeel beleid bevatten echter niet altijd bepalingen die direct in een beslisregel geïmplementeerd kunnen worden. Soms is er sprake van ‘0.8-op-één automatisering’, waarbij ruimte is voor de organisatie om een bepaling zelf te interpreteren en te formuleren als beslisregel.<sup>31</sup> Er is dan geen sprake van een AI-systeem aangezien *“het regels [betreft] die uitsluitend door natuurlijke personen zijn vastgesteld om automatisch handelingen uit te voeren”* (overweging 12 AI-verordening) en daarmee buiten de reikwijdte van de AI-verordening valt. Dit type algoritmen behoort tot de hierboven beschreven categorie b).

Een voorbeeld van dit type algoritme is het risicoprofiel dat gebruikt werd door DUO in het CUB-proces. Dit algoritme bestond uit simpele, door mensen opgestelde regels. Deze regels hadden ook handmatig kunnen worden uitgevoerd als werkinstructie. Toch leidde dit algoritme tot indirecte discriminatie en uiteindelijk compensatie van meer dan 10.000 studenten.<sup>32</sup> Hoewel deze categorie volgens de Handreiking niet als impactvol algoritme geldt, blijkt dit algoritme in praktische zin zeer impactvol. Daarom wijkt deze paper op dit punt af van de Handreiking. Wanneer een algoritmische toepassing een automatisering is van niet in formeel beleid vastgestelde regels, zien wij deze toepassing wél als mogelijk impactvol algoritme.

Het effect van een algoritme op de uitkomst van een besluitvormingsproces is een belangrijke factor om impactvolle algoritmen te onderscheiden van andere algoritmen en is daarom opgenomen als vraag in de dynamische vragenlijst. Hierbij wordt uitgevraagd in welke vorm er sprake is van (betekenisvolle) menselijke tussenkomst. Zie [6.7 Q7 – Welke van de volgende opties beschrijft het effect van de toepassing op de uitkomst het beste?](#) Daarnaast bepaalt het type uitkomst ook het effect van het algoritme op de uitkomst van het proces. Zie [6.1 Q1 – Welke van de volgende categorieën beschrijft de uitkomst van de toepassing het beste?](#) De onder a) beschreven type algoritme dat een directe automatisering is van een handmatige werkinstructie is verwerkt in [6.3 Q3 – Is de toepassing automatisering van door mensen opgestelde regels?](#)

<sup>31</sup> Een voorbeeld uit de Werkloosheidswet (WW) kan worden gevonden in ‘De A staat voor algoritme: hoe de Awb op dit vlak te versterken’ O.A. al Khatib, M.H.A.F. Lokin, R.J.H. Bruggeman & A.C.M. Meuwese (2024).

<sup>32</sup> Supra noot 6

## 6. Dynamische vragenlijst

Bovenstaande analyses van de definitie van een AI-systeem en van een impactvol algoritme hebben geresulteerd in een dynamische vragenlijst om deze typen algoritmen te identificeren. Aan de hand van drie vragen kan worden bepaald of een toepassing een AI-systeem betreft. Deze vragen worden geïntroduceerd en toegelicht in 6.1-6.3. Aan de hand van vier vervolgvragen kan worden bepaald of een toepassing een impactvol algoritme betreft. Deze vragen worden geïntroduceerd en toegelicht in 6.4-6.7.

De vragenlijst kent vier uitkomsten:

- > **Niet in scope:** algoritme is geen AI-systeem en ook geen impactvol algoritme;
- > **Impactvol algoritme:** algoritme is een impactvol algoritme, maar geen AI-systeem;
- > **AI-systeem:** algoritme is een AI-systeem, maar geen impactvol algoritme;
- > **Impactvol algoritme en AI-systeem:** algoritme is zowel een AI-systeem als een impactvol algoritme.

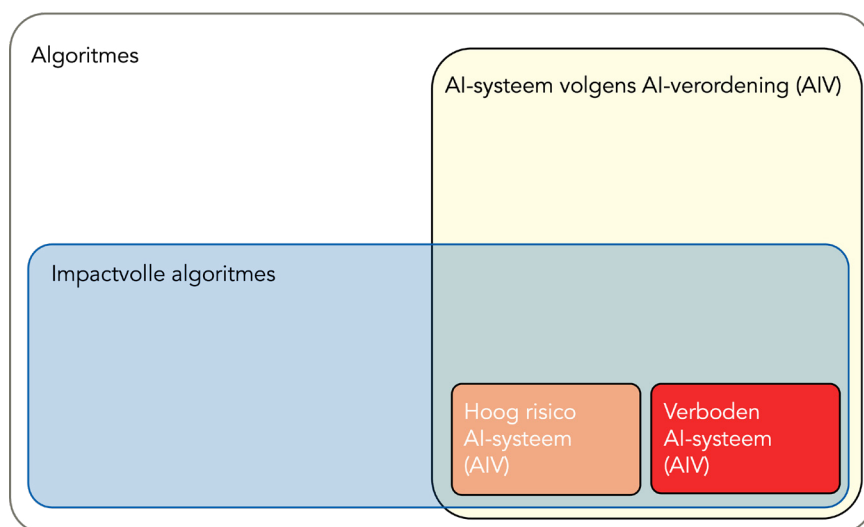
Een overzicht van alle uitkomsten kan worden gevonden in [Figuur 3](#). Een flowchart van alle vragen

en uitkomsten kan worden gevonden in [Appendix B](#).

### 6.1 Q1 – Welke van de volgende categorieën beschrijft de uitkomst van de toepassing het beste?

Zoals toegelicht in [2.7 Voorspellingen, inhoud, aanbevelingen of beslissingen](#) bevat de door een algoritme gegenereerde output belangrijk informatie of de algoritmische toepassing mogelijk een AI-systeem betreft. Omdat algoritme-ontwikkelaars, producteigenaren, lijnmanagers en andere uitvoerende gebruikers doorgaans bekend zijn met de output van algoritmen, en zij de doelgroep zijn die de tool gebruikt, begint de dynamische vragenlijst met een uitvraag naar de output die algoritmische toepassing produceert. Zie [Figuur 4](#).

De AI-verordening benoemt ‘voorspellingen’ als mogelijk vorm van output. Voorspelling is een breed begrip dat niet door iedereen op dezelfde wijze wordt geïnterpreteerd. In de datawetenschappen hoeft een voorspelling niet over de toekomst te gaan. Een voorspelling kan ook betrekking hebben op een niet eerder geobserveerd datapunt. In de antwoordopties zijn toelichtende



Figuur 3 - Venndiagram van alle uitkomsten van de beslisboom

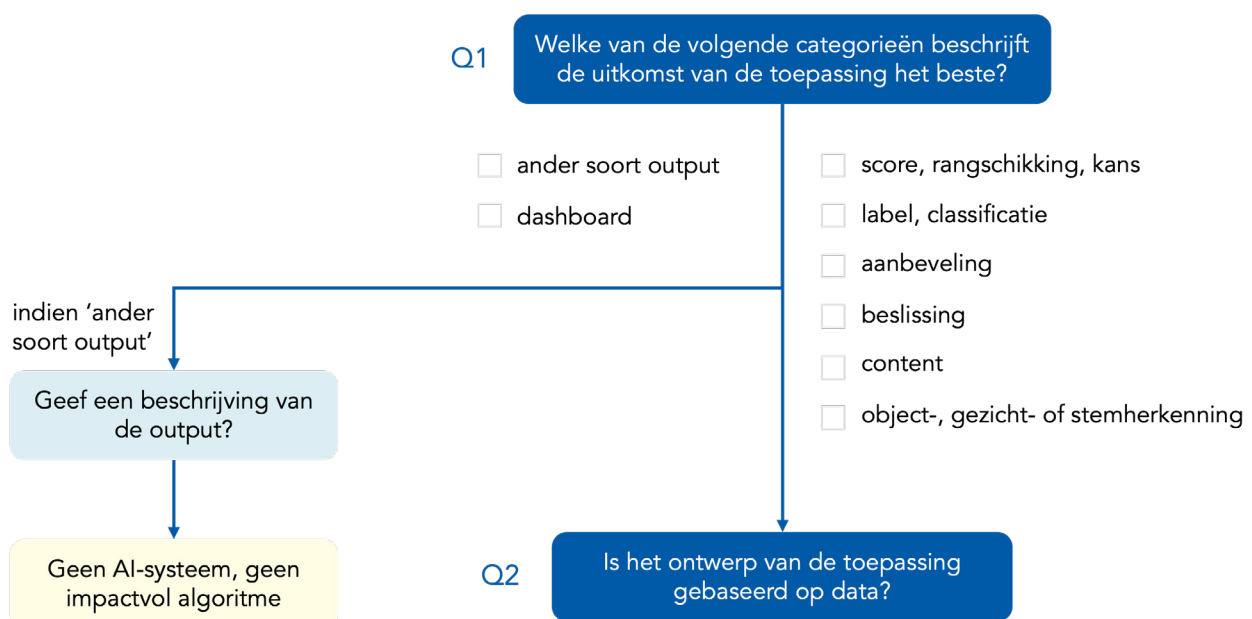
termen voor types voorspellingen opgenomen die herkenbaar kunnen zijn voor gebruikers van de vragenlijst (score, rangschikking, label, object-, gezicht- of stemherkenning). Ondanks dat iedere score, rangschikking, label, classificatie, en beeldherkenning in de kern een voorspelling is, wordt hiermee toegankelijke taal geprefereerd boven het voorkomen van duplicatie.

Omdat dashboards een veelvoorkomende datagedreven-toepassing zijn die vragen oproept over de reikwijdte van de AI-verordening is het als aparte antwoordoptie opgenomen. Doorgaans worden dashboards enkel gebruikt voor datavisualisatie en voldoen daarmee niet aan de definitie van een AI-systeem. Deze dashboards overstijgen namelijk niet *“de elementaire verwerking van data door leren, redeneren of modelleren mogelijk te maken.”* – Zie [overweging 12](#) zin 6 en [3. Inferentie](#). Onderliggende algoritmen, waarvan de uitkomst op een dashboard wordt getoond kunnen een AI-systeem zijn, maar dit geldt niet voor het dashboard zelf. Indien een dashboard scores, rangschikkingen of soortgelijk visualiseert, dient bij Q1 een ander antwoord te worden ingevuld waarmee de toepassingen mogelijk wel binnen de

reikwijdte van de AI-verordening valt. Aan de hand van antwoordopties ‘dashboard’ wordt daarom geconcludeerd dat er geen sprake is van een AI-systeem.

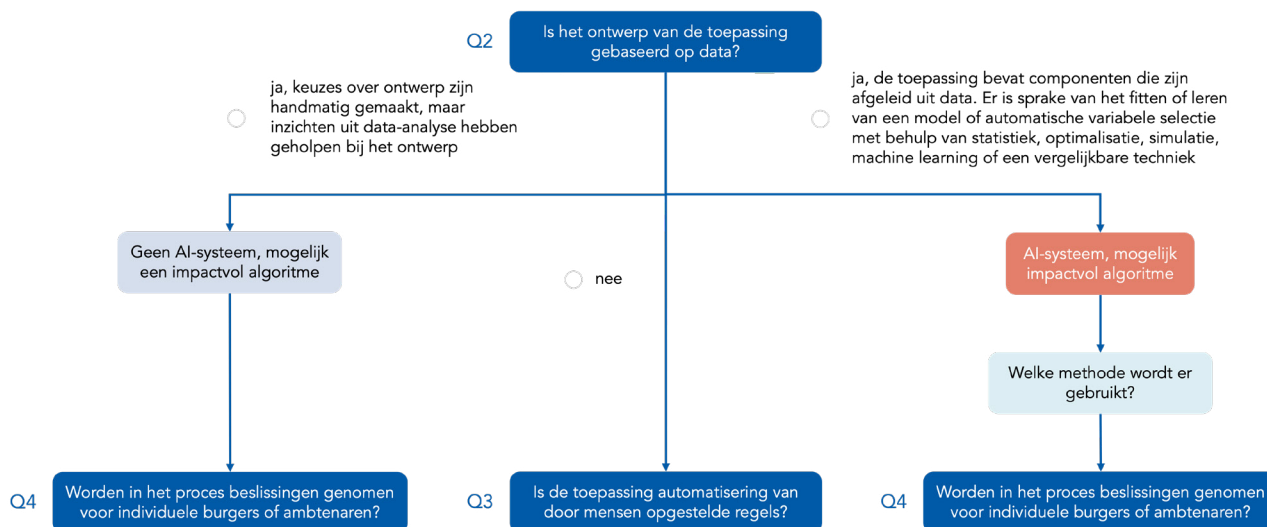
De antwoordopties – score, rangschikking, label, aanbeveling, beslissing, content, object-, gezicht- of stemherkenning – zijn ook behulpzaam om algoritmes te onderscheiden van impactvolle algoritmen. Algoritmen met dit type output hebben een effect op de uitkomst van het proces en zijn daarom een impactvol algoritme. Deze vormen van output sluiten aan bij de beschrijving die hiervan wordt gegeven in de AI-verordening: *“voorspellingen, content, aanbevelingen of besluiten”*.

Er is ook geen sprake van een AI-systeem of impactvol algoritme wanneer wordt gekozen voor de antwoordoptie ‘ander soort output’. In dit geval wordt de gebruiker gevraagd een beschrijving te geven van de output. Bij alle andere antwoordopties is er mogelijk sprake van een AI-systeem en een impactvol algoritme. In dit geval wordt de gebruiker doorgestuurd naar Q2.



Figuur 4 - Q1 vraagt uit wat voor type output de algoritmische toepassing genereert.





Figuur 5 - Q2 gaat over hoe het model of algoritme tot stand is gekomen dat de output genereert.

## 6.2 Q2 – Bevat de toepassing een model of algoritme dat uit data is afgeleid?

Waar Q1 de output van de algoritmische toepassing uitvraagt, richt Q2 zich op de manier waarop de output wordt gegenereerd. Zoals in 3. Inferentie toegelicht is "het vermogen van AI-systemen om modellen of algoritmen, of beiden, af te leiden uit input of data" een belangrijke factor om een algoritme te onderscheiden van een AI-systeem. Nadat de gebruiker in Q1 een van de antwoordopties heeft geselecteerd die leiden tot Q2, wordt uitgevraagd of het ontwerp van de algoritmische toepassing is gebaseerd op data.

Wanneer de toepassing componenten bevat die zijn afgeleid uit data, dan is de toepassing een AI-systeem. Dit is bijvoorbeeld het geval wanneer een model of algoritme wordt geleerd of gefit met behulp van statistiek, optimalisatie, simulatie of machine learning of een vergelijkbare techniek. Zie 3.1 Leren en modelleren.

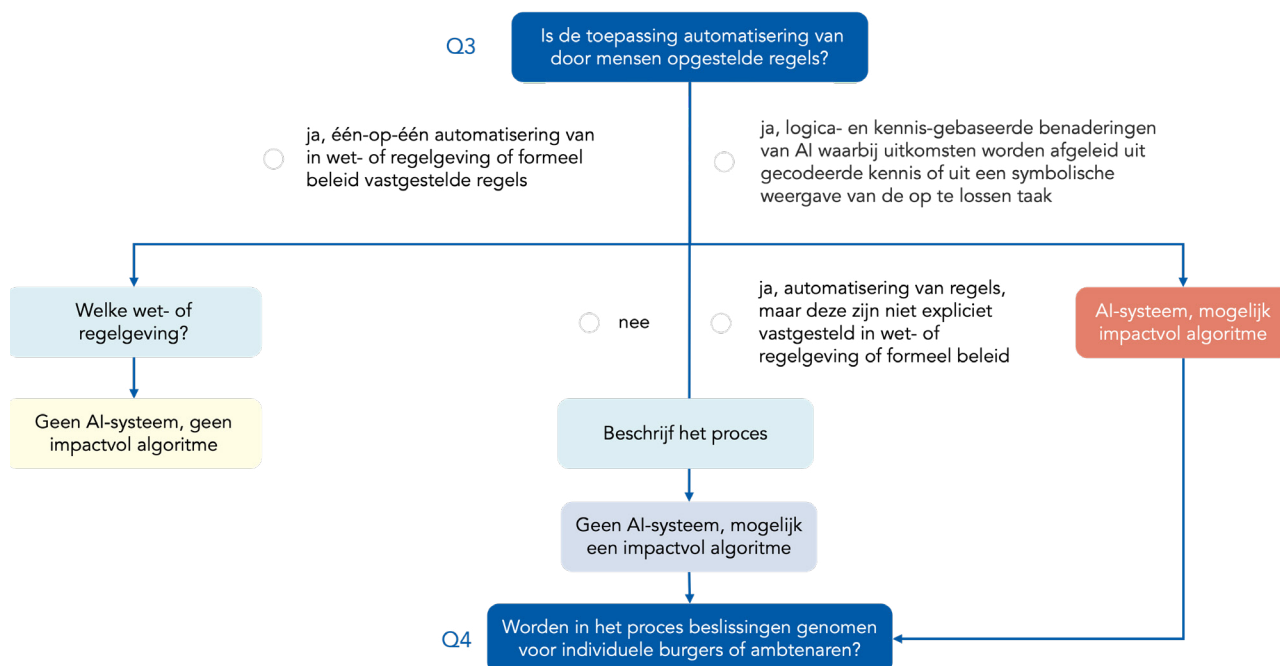
Op basis van de richtlijnen volgt dat niet alle toepassingen waarvan componenten afgeleid zijn

uit data een AI-systeem zijn. Uitzonderingen op deze regel worden besproken in 3. Inferentie. Om deze afweging te maken wordt bij de antwoordoptie 'ja, de toepassing bevat componenten die afgeleid zijn uit data' een toelichting gevraagd van de gebruiker. Deze informatie helpt om per geval te bepalen of een algoritme wel of geen AI-systeem betreft.

Wanneer keuzes in het ontwerp van de toepassing handmatig zijn gemaakt is de toepassing, met oog op overweging 12 uit de AI-verordening, waarschijnlijk geen AI-systeem maar mogelijk wel een impactvol algoritme. Zie 2.1 Interpretatie van de definitie van een AI-systeem aan de hand van overweging 12. Om na te gaan of de toepassing een impactvol algoritme betreft, wordt de gebruiker doorverwezen naar Q4.

Ook als het ontwerp van de toepassing niet gebaseerd is op data, dan kan de toepassing alsnog een AI-systeem zijn. Zie 3.2 Redeneren: op logica en kennis gebaseerde benaderingen. Om na te gaan of dit het geval is, worden gebruikers doorgestuurd naar Q3.





Figuur 6 - Q3 gaat na in hoeverre er sprake is van menselijke betrokkenheid bij het opstellen van regels die worden gebruikt in de algoritmische toepassing.

### 6.3 Q3 – Is de toepassing automatisering van door mensen opgestelde regels?

Indien regels worden gebruikt in algoritmische toepassingen is er mogelijk geen sprake van een AI-systeem. Zie 2.1 [Interpretatie van de definitie van een AI-systeem aan de hand van overweging 12](#). Dit wordt uitgevraagd in Q3.

Wanneer regels tot stand zijn gekomen met behulp van op logica- en kennis-gebaseerde benaderingen kwalificeert het algoritme wel als AI-systeem. Op logica- en kennis-gebaseerde benaderingen zijn echter zeldzame vormen van AI. Zie 3.2 Redeneren: op logica en kennis gebaseerde benaderingen. In andere gevallen waar de toepassing uitsluitend gebaseerd is op "regels die uitsluitend door natuurlijke personen zijn vastgesteld", dan is de toepassing geen AI-systeem. Zie 2.1 [Interpretatie van de definitie van een AI-systeem aan de hand van overweging 12](#). Indien er sprake zou zijn geweest van data-analyse die een rol heeft gespeeld in

de totstandkoming van de beslisregel, zou in een eerdere vraag een ander antwoord moeten zijn gegeven, waardoor de gebruiker bij Q4 in plaats van Q3 uit zou zijn gekomen.

Q3 is verder met name belangrijk voor de identificatie van impactvolle algoritmen. De manier waarop door mensen opgestelde regels tot stand zijn gekomen bepaalt mede of een algoritmische toepassing een impactvol algoritme is. We maken hier onderscheid tussen a) wet- en regelgeving en formeel vastgesteld beleid en b) meer informele werkinstructies en andere door mensen opgestelde regels. Zie ook 5.3. [Heeft het algoritme een significant effect op de uitkomst van het proces?](#)

Wet- en regelgeving of formeel beleid bevatten echter niet altijd bepalingen die direct in een beslisregel geïmplementeerd kunnen worden. Soms is er sprake van '0.8-op-één automatisering', waarbij ruimte is voor de organisatie om een bepaling zelf te interpreteren en te formuleren als beslisregel.<sup>33</sup>

<sup>33</sup> Een voorbeeld uit de Werkloosheidswet (WW) kan worden gevonden in 'De A staat voor algoritme: hoe de Awb op dit vlak te versterken' O.A. al Khatib, M.H.A.F. Lokin, R.J.H. Bruggeman & A.C.M. Meuwese (2024).

Deze gevallen zijn geen één-op-één automatisering van wet- of regelgeving of formeel beleid. Deze systemen zijn mogelijk wél een impactvol algoritme. Wanneer Q3 met 'nee' wordt beantwoord, wordt de gebruiker gevraagd een beschrijving te geven van de toepassing om daarna naar Q4 door te worden verwezen.

#### 6.4 Q4 – Wordt in het proces een beslissing genomen voor individuele burger, organisatie of medewerker?

Aan de hand van Q1-Q3 is bepaald of de algoritmische toepassing een AI-systeem betreft. Aan de hand van maximaal vier aanvullende vragen kan bepaald worden of de toepassing een impactvol algoritme is. In navolging van de Handreiking wordt eerst gevraagd naar het proces, ongeacht het gebruik van een algoritme (Q4-Q6) en daarna pas naar de rol van het algoritme in dit proces.

De eerste vraag gaat na of er in het proces een beslissing wordt genomen dat betrekking heeft op een individuele burger, organisatie of ambtenaar. Dit aspect is van belang om impactvolle algoritmen te onderscheiden van overige algoritmen. Zie 5.1. [Is het een proces met directe gevolgen?](#) De nadruk in deze vraag ligt op individuen. In Q6 ligt de nadruk op groepen.

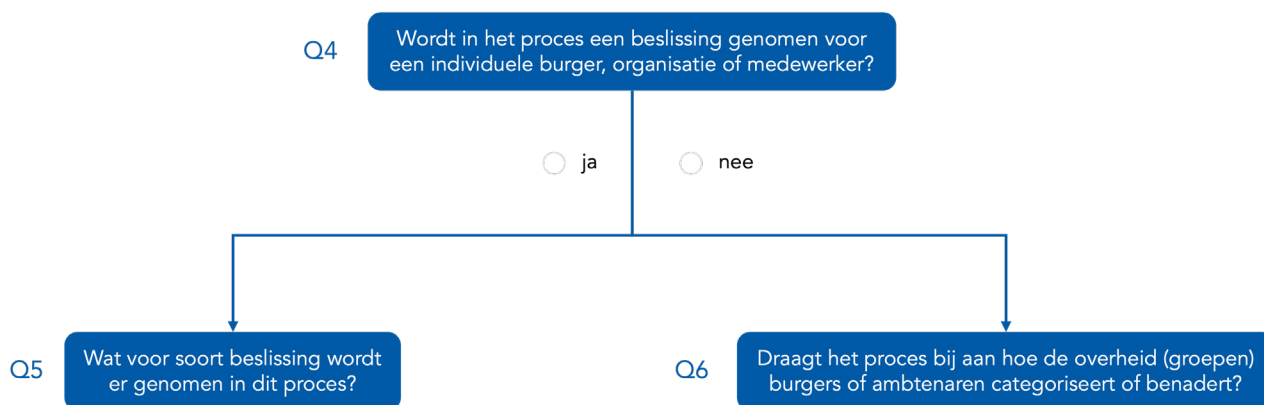
Indien er wel een beslissing wordt genomen in het proces, is het van belang na te gaan wat voor een soort beslissing in het proces wordt genomen. De gebruiker wordt doorverwezen naar Q5.

Indien er geen besluit wordt genomen in het proces, wordt uitgevraagd in hoeverre het proces bijdraagt aan de manier waarop de overheid groepen burgers of ambtenaren categoriseert of benadert. De gebruiker wordt doorverwezen naar Q6.

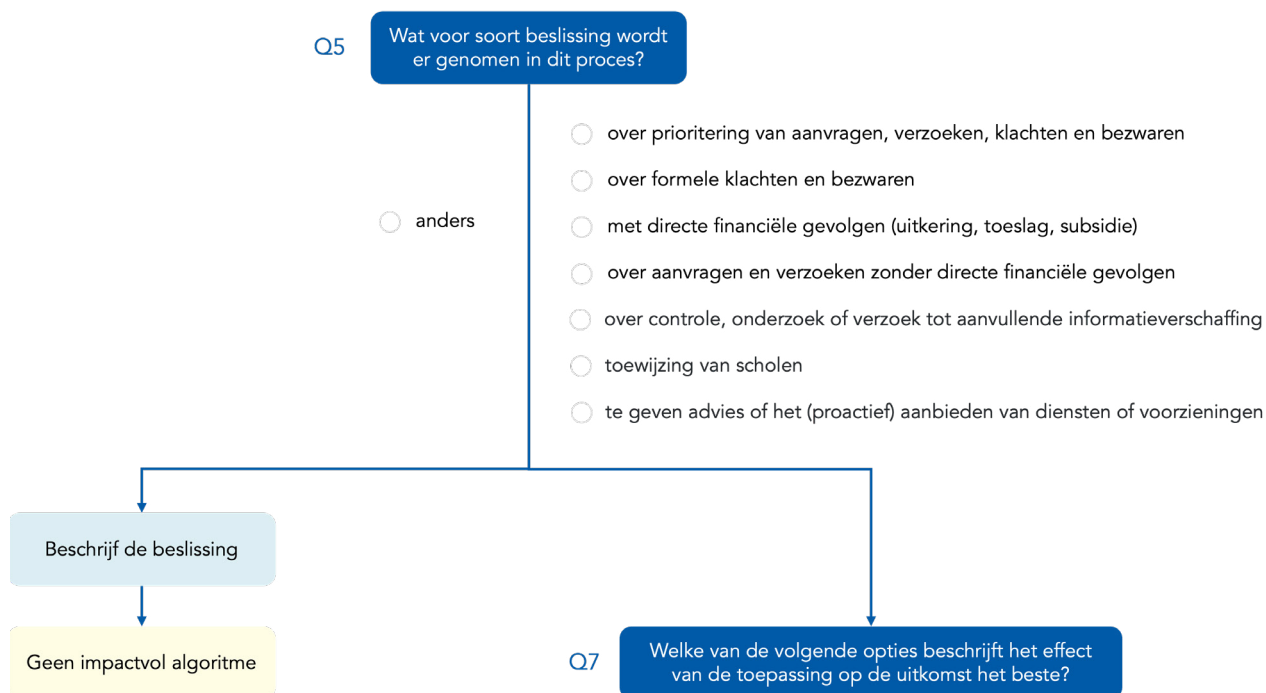
#### 6.5 Q5 – Wat voor soort beslissing wordt er genomen in dit proces?

Gegeven dat er een beslissing in het proces wordt genomen voor een individuele burger, organisatie of medewerker, wordt er in Q5 uitgevraagd wat voor soort beslissingen in het proces wordt genomen. Het type beslissing is een belangrijke factor om impactvolle algoritmen te onderscheiden van reguliere algoritmen. Zie 5.1. [Is het een proces met directe gevolgen?](#)

Een beslissing is breed gedefinieerd. Onder meer kan gedacht worden aan beslissingen over prioritering van aanvragen, verzoeken, klachten of bezwaren; beslissingen over formele klachten en bezwaren; beslissingen met directe financiële gevolgen, zoals beslissingen over uitkering, toeslag, subsidie, boete, terugbetaling of mogelijkheid tot



Figuur 7 - Q4 gaat na of er sprake is van een beslissing in het proces waarbij gebruik wordt gemaakt van een algoritmische toepassing.



Figuur 8 - Q5 vraagt uit wat voor een soort beslissing er in het proces wordt genomen.

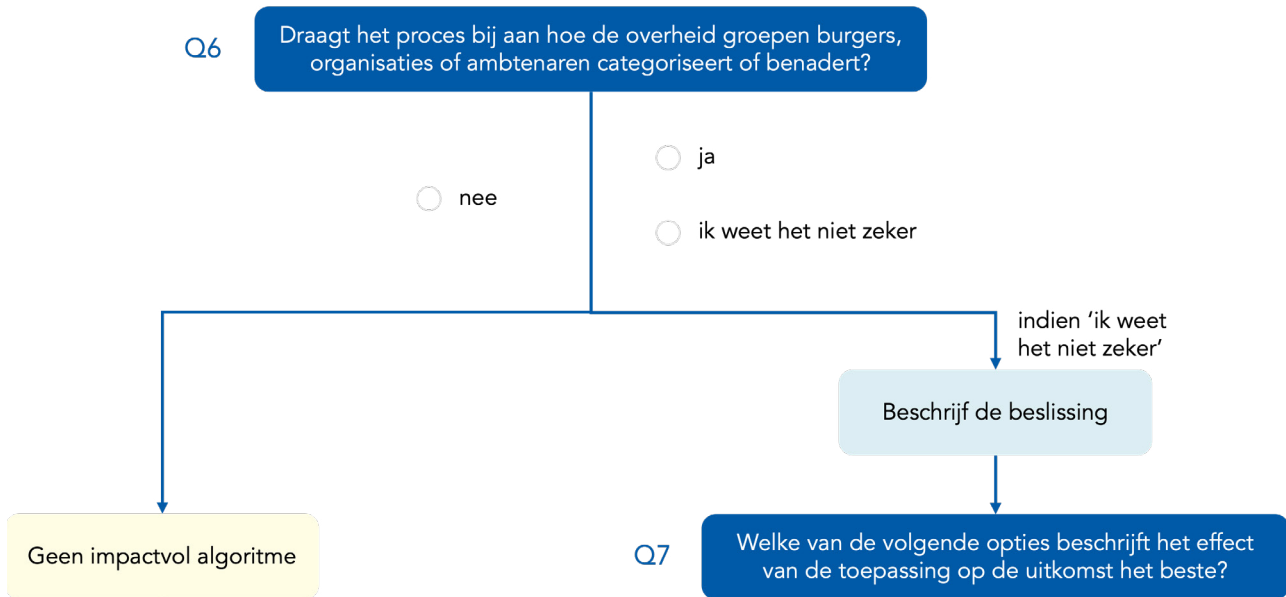
betalingsregeling; beslissingen over aanvragen en verzoeken zonder directe financiële gevolgen, zoals toekenning van een aanvraag van diensten of vergunningtoekenning; beslissingen over controle, onderzoek of verzoek tot aanvullende informatieverschaffing; beslissing over toewijzing van scholen; beslissing over te geven advies of het (proactief) aanbieden van diensten of voorzieningen. In al deze gevallen worden gebruikers doorgestuurd naar Q7 om het effect van de toepassing op de uitkomst te bepalen.

Wanneer een soort beslissing niet tot een van de voorgaande categorieën behoort kan worden geconcludeerd dat de impact van het proces beperkt is en dat het algoritme in kwestie geen impactvol algoritme is. In dit geval wordt gevraagd om een beschrijving van het type beslissing.

## 6.6 Q6 – Draagt het proces bij aan hoe de overheid (groepen) burgers, organisaties of ambtenaren categoriseert of benadert?

Gegeven dat er geen beslissing in het proces wordt genomen voor een individuele burger, organisatie of medewerker, gaat Q6 na of het proces bijdraagt hoe de overheid groepen burgers, organisaties of ambtenaren categoriseert of benadert. Het belang hiervan wordt toegelicht in 5.1. [Is het een proces met directe gevolgen?](#)

Indien het proces niet bijdraagt aan hoe de overheid groepen burgers, organisaties of ambtenaren categoriseert of benadert, wordt geconcludeerd dat het onderhavige algoritme geen impactvol algoritme betreft. Wanneer dit niet met zekerheid



Figuur 9 - Q6 gaat na of het proces bijdraagt hoe groepen burgers, organisaties of ambtenaren door de overheid worden gecategoriseerd of benaderd

gezegd kan worden, wordt gevraagd om een toelichting, waarna de gebruiker alsnog door wordt gestuurd naar Q7. De gebruiker wordt ook naar Q7 doorgestuurd wanneer Q6 met 'ja' wordt beantwoord.

## 6.7 Q7 – Welke van de volgende opties beschrijft het effect van de toepassing op de uitkomst het beste?

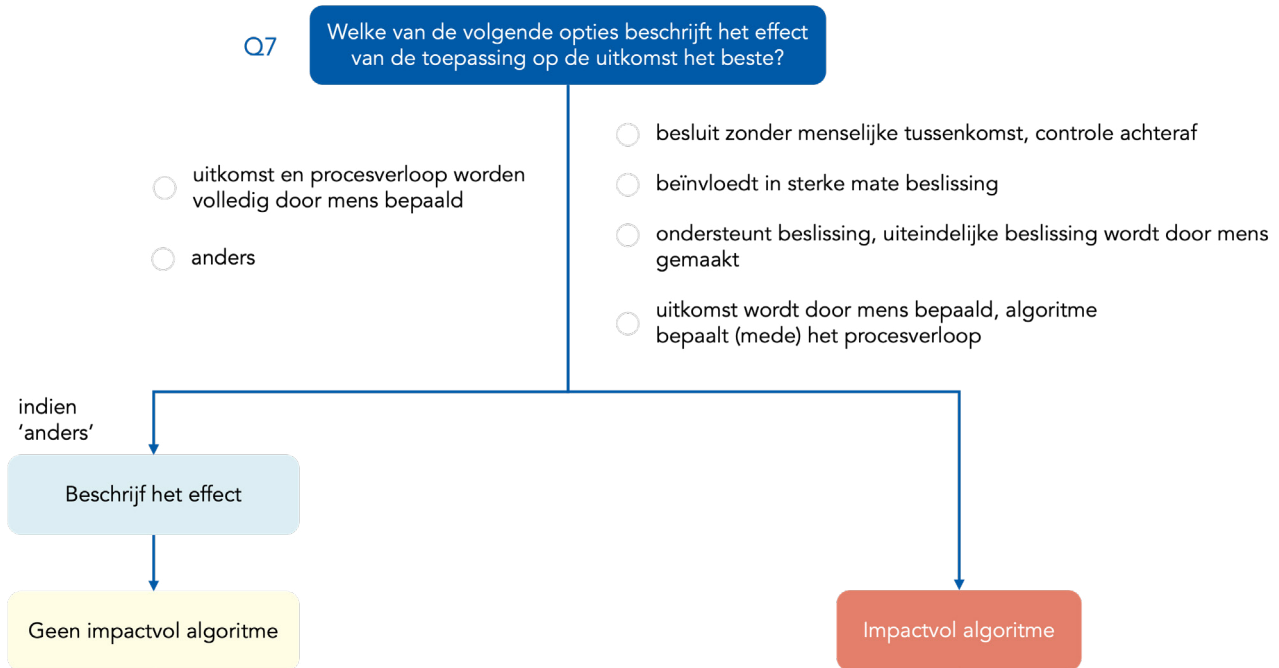
In alle gevallen dat een beslissing wordt genomen voor burgers, organisaties of ambtenaren (individuele of groepen) is het relevant te achterhalen wat het effect is van de algoritmische toepassing op het besluitvormingsproces. Dit effect bepaalt of het algoritme wel of geen impactvol algoritme is. Zie [5.3. Heeft het algoritme een significant effect op de uitkomst van het proces?](#)

Indien de uitkomst van het proces direct wordt bepaald door de algoritmische toepassing, zonder dat daarbij sprake is van menselijke tussenkomst, dan is er sprake van een impactvol algoritme. Dit

is ook het geval wanneer na voltooiing van het proces menselijke analisten de resultaten kunnen controleren

Er is ook sprake van een impactvol algoritme wanneer het proces sterk wordt beïnvloed door de algoritmische toepassing. Bijvoorbeeld doordat werkvoorschriften bepalen wat het gevolg is van een bepaalde uitkomst van de toepassing. In dit geval kan een medewerker in sommige gevallen andere keuzes maken, maar meestal bepaalt het resultaat van het systeem wat het eindresultaat van het proces zal zijn.

Ook wanneer de uitkomst van het proces deels wordt beïnvloed door de toepassing, is er sprake van een impactvol algoritme. Hierbij is het resultaat van de toepassing belangrijk voor het eindresultaat, maar de uiteindelijke beslissing wordt genomen door een medewerker. Deze medewerker heeft de juiste informatie, ervaring/kunde, mandaat en beschikbare tijd om de beslissing te maken.



Figuur 10 - Q7 vraagt uit wat het effect is van de algoritmische toepassing op de uitkomst van het proces.

Wanneer de algoritmische toepassing (mede) het procesverloop bepaalt, maar de uitkomst van het proces volledig door een medewerker wordt bepaald is er ook sprake van een impactvol algoritme. Dit is bijvoorbeeld het geval wanneer de uitkomst van de toepassing een risicoscore is aan de hand waarvan een controleproces wordt geïnitieerd of een meer intensieve dossierevaluatie plaatsvindt, maar de controle of evaluatie daarna volledig door een medewerker wordt uitgevoerd.

Indien de uitkomst van het proces en het procesverloop volledig door een mens bepaald worden, is er geen sprake van een impactvol algoritme. Dit is ook het geval wanneer er sprake is van een ander effect van de toepassing op de uitkomst. In dit laatste geval wordt een beschrijving van dit effect gevraagd.

## Appendix A – Overweging 12

Overweging 12 uit de preambule van de AI-verordening.

### Zin 1 – geanalyseerd in 1. Introductie

Het begrip “AI-systeem” in deze verordening moet duidelijk worden gedefinieerd en moet nauw aansluiten op het werk van internationale organisaties die zich bezighouden met AI, om rechtszekerheid te waarborgen, internationale convergentie en brede acceptatie te faciliteren, en tegelijkertijd de nodige flexibiliteit te bieden om op de snelle technologische ontwikkelingen op dit gebied te kunnen inspelen.

### Zin 2 – geanalyseerd in 2.1 Interpretatie van de definitie van een AI-systeem aan de hand van overweging 12

Bovendien moet de definitie gebaseerd zijn op de belangrijkste kenmerken van AI-systemen die het onderscheiden van eenvoudiger traditionele softwaresystemen of programmeringsbenaderingen, en mag het geen betrekking hebben op systemen die gebaseerd zijn op regels die uitsluitend door natuurlijke personen zijn vastgesteld om automatisch handelingen uit te voeren.

### Zin 3-4 – geanalyseerd in 3.1 Leren en modelleren

Een belangrijk kenmerk van AI-systemen is hun inferentievermogen. Dit inferentievermogen slaat op het proces waarbij output, zoals voorspellingen, content, aanbevelingen of besluiten, wordt verkregen waarmee fysieke en virtuele omgevingen kunnen worden beïnvloed, en op het vermogen van AI-systemen om modellen of algoritmen, of beide, af te leiden uit input of data.

### Zin 5-6 – geanalyseerd in 3.2 Redeneren: op logica en kennis gebaseerde benaderingen

De technieken die inferentie mogelijk maken bij de opbouw van een AI-systeem, omvatten benaderingen op basis van machinaal leren waarbij aan de hand van data wordt geleerd hoe bepaalde doelstellingen kunnen worden bereikt, alsook op logica en kennis gebaseerde benaderingen waarbij iets wordt geïnferreerd uit gecodeerde kennis of uit een symbolische weergave van de op te lossen taak. Het inferentievermogen van een AI-systeem overstijgt de elementaire verwerking van data door leren, redeneren of modelleren mogelijk te maken.

### Zin 7 – geanalyseerd in 2.2 Machine gebaseerd systeem

De term “op een machine gebaseerd” verwijst naar het feit dat AI-systemen op machines draaien.

### Zin 8-9 – geanalyseerd in 2.5 Expliciete of impliciete doelstellingen

De verwijzing naar expliciete of impliciete doelstellingen onderstreept dat AI-systemen kunnen functioneren volgens expliciete, gedefinieerde doelstellingen, of volgens impliciete doelstellingen. De doelstellingen van een AI-systeem kunnen verschillen van het beoogde doel van het AI-systeem in een specifieke context.

### Zin 10 – geanalyseerd in 2.7 Voorspellingen, inhoud, aanbevelingen of beslissingen en 2.8 Fysieke en virtuele omgeving

Voor de toepassing van deze verordening moeten onder omgevingen de contexten worden verstaan waarin de AI-systemen werken, terwijl de output die door het AI-systeem wordt gegenereerd een uiting is van de verschillende functies van AI-systemen en de vorm kan aannemen van voorspellingen, content, aanbevelingen of besluiten.

**Zin 11 – geanalyseerd in 2.3 Verschillende niveaus van autonomie en 4. Autonomie**

AI-systemen worden zodanig ontworpen dat zij in verschillende mate autonoom kunnen functioneren, wat betekent dat zij een zekere mate van onafhankelijkheid van menselijke betrokkenheid bezitten en zonder menselijke tussenkomst kunnen functioneren.

**Zin 12 – geanalyseerd in 2.4 Aanpassingsvermogen kan vertonen**

Het aanpassingsvermogen dat een AI-systeem na het inzetten ervan kan vertonen, heeft betrekking op zelflerende capaciteiten, waardoor het systeem tijdens het gebruik kan veranderen.

**Zin 13 – niet geanalyseerd, want niet van specifieke toegevoegde waarde**

AI-systemen kunnen op standalonebasis of als component van een product worden gebruikt, ongeacht of het systeem fysiek in het product is geïntegreerd (ingebed) dan wel ten dienste staat van de functionaliteit van het product zonder daarin te zijn geïntegreerd (niet-ingebed).





## Over Algorithm Audit

Algorithm Audit is een Europees kennisplatform voor AI bias testing en normatieve AI-standaarden. De doelen van de stichting zijn drieledig:



### Kennisplatform

Samenbrengen van kennis en experts voor collectief leerproces over verantwoorde inzet van algoritmes, bijvoorbeeld ons [AI Policy Observatory](#) en [position papers](#)



### Normatieve adviescommissies

Adviseren over ethische kwesties in concrete algoritmische toepassingen door het samenbrengen van deliberatieve, diverse adviescommissies, met [algotrudentie](#) als resultaat



### Technische hulpmiddelen

Implementeren en testen van technische methoden voor bias-detectie en -mitigatie, zoals onze [bias detection tool](#)



### Projectwerk

Ondersteuning bij specifieke vragen vanuit de publieke en private sector over de verantwoorde inzet van algoritmes.

## Structurele partners van Algorithm Audit

### SIDNfonds

#### SIDN Fonds

Het SIDN Fonds staat voor een sterk internet voor iedereen. Het Fonds investeert in projecten met lef en maatschappelijke meerwaarde, met als doel het borgen van publieke waarden online en in de digitale democratie.

European  
Artificial Intelligence  
& Society Fund

#### European AI&Society Fund

Het European AI&Society Fund ondersteunt organisaties uit heel Europa die AI beleid vormgeven waarin mens en maatschappij centraal staan. Het fonds is een samenwerkingsverband van 14 Europese en Amerikaanse filantropische organisaties.



Ministerie van Binnenlandse Zaken en  
Koninkrijksrelaties

#### Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

Het ministerie van BZK maakt zich sterk voor een democratische rechtsstaat, met een slagvaardig bestuur. Ze borgt de kernwaarden van de democratie. BZK staat voor een goed en digitaalvaardig openbaar bestuur en een overheid waar burgers op kunnen vertrouwen.

Opbouwen van **publieke kennis**  
over verantwoorde AI **zonder winstoogmerk**



[www.algorithmaudit.eu](http://www.algorithmaudit.eu)



[www.github.com/NGO-Algorithm-Audit](https://www.github.com/NGO-Algorithm-Audit)



[info@algorithmaudit.eu](mailto:info@algorithmaudit.eu)



Stichting Algorithm Audit is geregistreerd bij de  
Kamer van Koophandel onder nummer 83979212