

Mensgerichte AI

Jos van der Wijst

BG.legal / legalAIR

Maandag 4 april 2022

Wie ben ik

- Advocaat bij BG.legal (1991)
- IE/IT/privacy praktijk
- Kantoor in Eindhoven (en Den Bosch); 20 advocaten/ juristen
- Actief bij de NL AIC
- Oprichter kennisplatform LegalAIR



Waar gaat het vandaag over?

- MIT: R&D-samenwerkingsprojecten Artificiële Intelligentie (AI)
- Staatscourant, 14 maart 2022
 - Criterium: Oog voor betrouwbare en mensgerichte AI
- Wat betekent betrouwbare en mensgerichte AI?
- Hoe voldoe je aan dit criterium?

STAATSCOURANT

Officiële uitgave van het Koninkrijk der Nederlanden sinds 1814.

Nr. 6889

14 maart

2022

Voorwaarden MIT R&D AI

- generieke voorwaarden MIT R&D samenwerkingsprojecten.
- bijlage 3.4.1 van de Regeling nationale EZK- en LNV-subsidies:
- Toegevoegd: par. 5.a. Digital Technologies: Artificiële Intelligentie (AI)
- om MKB-ondernemers te ondersteunen bij de ontwikkeling en toepassing van AI.
- *“AI verwijst naar systemen die intelligent gedrag vertonen door hun omgeving te analyseren en – met een zekere mate van zelfstandigheid – actie te ondernemen om specifieke doelen te bereiken”* = definitie Eur. Commissie.

Welke projecten

projecten die bijdragen aan de generieke ontwikkeling van AI, doordat:

1. zij gericht zijn op de inzet van Machine Learning;
2. zij gericht zijn op een aantoonbaar nieuwe oplossing voor één of meerdere toepassingsgebieden van de Nederlandse AI Coalitie;
3. zij maken gebruik van een aanpak die oog heeft voor betrouwbare en mensgerichte AI;

2. toepassingsgebieden

Alle toepassingsgebieden:

- ◆ Cultuur en Media
- ◆ Defensie
- ◆ Energie en Duurzaamheid
- ◆ Financiële Dienstverlening
- ◆ Gebouwde Omgeving
- ◆ Gezondheid en Zorg
- ◆ Haven en Maritiem
- ◆ Landbouw en Voeding
- ◆ Mobiliteit, Transport en Logistiek
- ◆ Onderwijs
- ◆ Publieke Diensten
- ◆ Technische Industrie
- ◆ Veiligheid, Vrede en Recht



Bron: <https://nlaic.com/toepassingsgebied/>

3. Betrouwbare en mensgerichte AI

Staatscourant:

“Ten derde is het van belang dat AI door mensen en organisaties op een verantwoorde manier wordt toegepast. De inzet op mensgerichte en ethisch verantwoorde AI is een belangrijke doelstelling van het **Strategisch Actieplan voor Artificiële Intelligentie.**”

<https://www.rijksoverheid.nl/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie>

3. Betrouwbare en mensgerichte AI



Strategisch Actieplan voor Artificiële Intelligentie



Inhoud

Samenvatting	7
Artificiële intelligentie is dé sleuteltechnologie voor maatschappelijke en economische kansen	9
Spoor 1: Maatschappelijke en economische kansen benutten	13
AI biedt oplossingen voor maatschappelijke uitdagingen	15
De overheid maakt optimaal gebruik van AI bij publieke taakuitvoering	19
AI-ondernemerschap wordt gestimuleerd	22
Spoor 2: Scheppen van de juiste voorwaarden	25
AI-onderzoek en innovatie in Nederland zijn van hoge kwaliteit en toonaangevend in Europa	27
Nederland heeft excellente scholingsmogelijkheden om met AI te kunnen leven en meer talent om met AI te werken	30
Nederland heeft meer bruikbare data voor AI-toepassingen om betere AI-ontwikkelingen te realiseren	34
Nederland heeft een voorhoedepositie in Europa in kwalitatief hoogwaardige digitale en intelligente connectiviteit en rekenkracht voor effectieve AI-toepassingen	37
Spoor 3: Versterken van de fundamenteën	39
Publieke waarden en mensenrechten blijven beschermd	41
AI wordt zo ingezet dat iedereen het kan vertrouwen	43
Markten zijn open, competitief en bieden consumenten goede bescherming	47
De veiligheid van burgers, bedrijven en overheden in Nederland blijft beschermd	50
Bijlagen	53
Bijlage 1 Overzicht van de acties	54
Bijlage 2 Financiële middelen voor AI	60

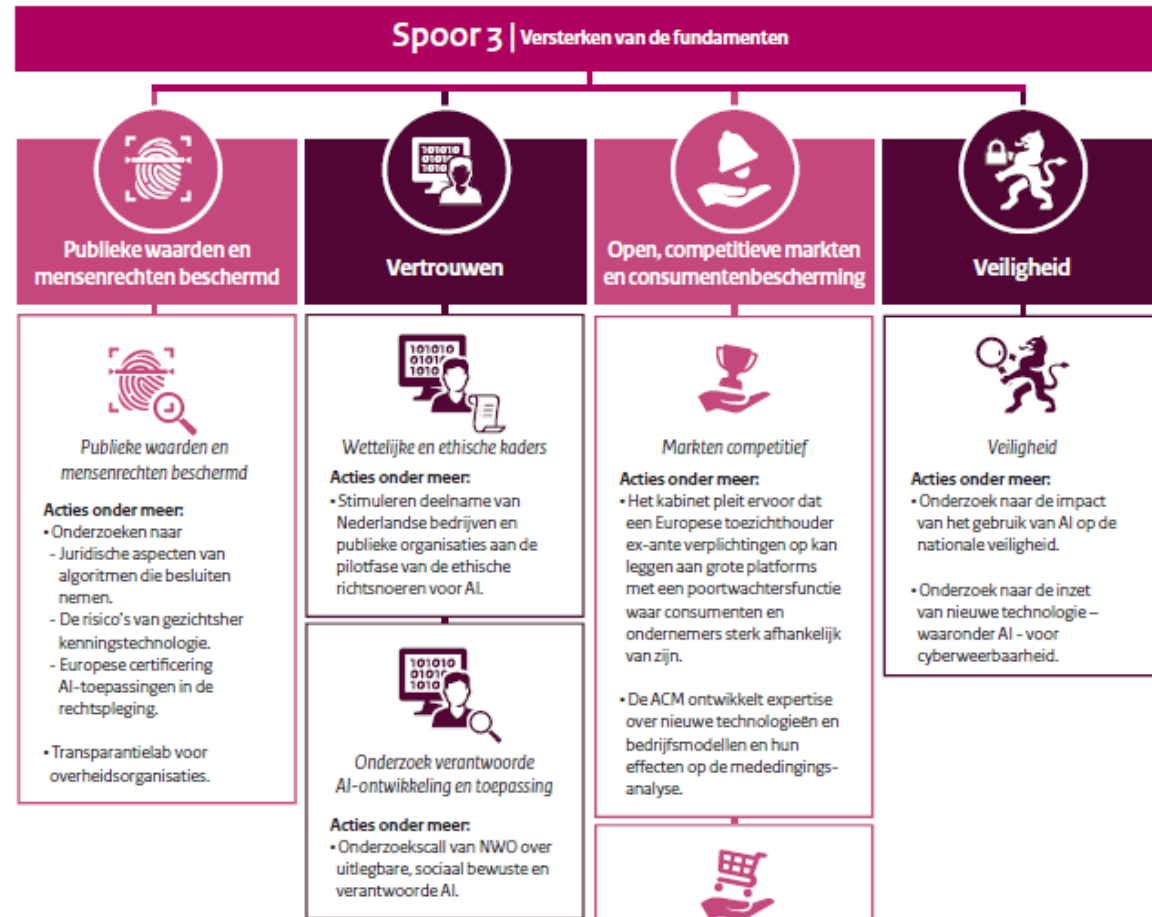
3. Betrouwbare en mensgerichte AI

De volgende uitgangspunten zijn leidend voor de Nederlandse aanpak:

- We werken publiek-privaat samen, onder meer in de Nederlandse AI-Coalitie om de maatschappelijke en economische kansen van AI te benutten.
- Nederland gaat voor internationale samenwerking, vooral met Europese partners.
- We profileren ons als een land dat voorop loopt met AI-toepassingen die de belangen dienen van mens en maatschappij.
- We kiezen voor een inclusieve aanpak die mensen centraal stelt, waarbij we streven naar betrouwbare AI.

Bron

<https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie/Rapport+SAPAI.pdf>



Ethische richtsnoeren voor betrouwbare AI

- Een op Europese waarden gebaseerde benadering van AI
- EC = > High-level expert group voor AI
- Rapport 2019



Betrouwbare en Mensgerichte AI

- Componenten
- Grondslagen
- Verwezenlijking
- controle



Bron: <https://legalair.nl/kennisbank/ai-wet-en-regelgeving/>

betrouwbare AI

3 componenten:

- a. De AI toepassing moet **wettig** zijn, door te voldoen aan alle toepasselijke wet- en regelgeving;
- b. De AI toepassing moet **ethisch** zijn, door naleving van ethische beginselen en waarden te waarborgen;
- c. De AI moet **robuust** zijn uit zowel technisch als sociaal oogpunt, aangezien AI-systemen ongewild schade kunnen aanrichten, zelfs al zijn de bedoelingen goed.

Componenten betrouwbare AI

Wettig:

O.a.

- Europees Handvest grondrechten
- AVG
- Machinerichtlijn
- Productaansprakelijkheid
- Consumentenrecht
- Medische hulpmiddelen verordening

Componenten betrouwbare AI

Ethisch:

- Wetgeving loopt vaak achter bij technologische ontwikkeling
- 'ethische normen'
- Soms ook vastgelegd in wet- en regelgeving
- Soms in gedragscodes

Componenten betrouwbare AI

Robuust:

- AI moet niet ongewild schade aanrichten
- Systeem moet veilig en betrouwbaar zijn
- Cyberveiligheid
- Waarborgen tijdens gehele gebruik AI

Grondslagen betrouwbare AI

4 ethische beginselen (ethische geboden):

- Respect voor menselijke autonomie
- Preventie van schade
- Rechtvaardigheid
- verantwoordiging

Ethische beginselen

AI EN PUBLIEKE WAARDEN EN MENSENRECHTEN

Verbod op discriminatie

Het verbod op discriminatie kan worden geschonden door bias in de onderliggende data, bias in het algoritme, of bij fouten in classificatie. Dat zou er bijvoorbeeld toe kunnen leiden dat ongerechtvaardigd onderscheid wordt gemaakt tussen mannen en vrouwen.

Bescherming van de privacy

De privacy kan worden geschonden als de verwerking van persoonsgegevens niet voldoet aan de eisen van behoorlijkheid en transparantie uit de AVG. Privacyvraagstukken spelen rondom onder meer gezichtsherkenningstechnologie, big data en het afleiden van persoonsgegevens uit andere gegevens.

Vrijheid van meningsuiting

Bij de toepassing van AI kan de vrijheid van meningsuiting onder druk komen te staan. Enerzijds gaat het dan om toegang tot informatie (bijv. personalisatie en ordening van zoekresultaten), anderzijds om de werking van algoritmes die geautomatiseerd content verwijderen.

Menselijke waardigheid en autonomie

Dit betreft onder meer het gevaar van dehumanisering en de invloed van AI op het maken van keuzen.

Het recht op een eerlijk proces

Dit recht speelt onder meer bij geautomatiseerde besluitvorming. Als ondoorzichtig is of en zo ja welke algoritmen zijn gebruikt om een besluit te nemen of voor te bereiden of welke aannames en gegevens daaraan ten grondslag liggen, ontstaat druk op de aanvechtbaarheid en motivering van besluiten en uitspraken en op het evenwicht tussen procespartijen (equality of arms).

Bron

<https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie/Rapport+SAPAI.pdf>

Europees Handvest

Bron Ethische geboden

Het **Handvest** bevat een preambule en is onderverdeeld in zeven titels, te weten:

1. **waardigheid** (bevat bepalingen over o.a. het recht op leven en het verbod van folteringen en van onmenselijke of vernederende behandelingen of straffingen),
2. **vrijheden** (bevat bepalingen over o.a. de eerbiediging van het privé-leven en van het familie- en gezinsleven en de bescherming van persoonsgegevens),
3. **gelijkheid** (bevat bepalingen over o.a. gelijkheid voor de wet, non-discriminatie, de gelijkheid van vrouwen en mannen en de rechten van het kind),
4. **solidariteit** (bevat bepalingen over o.a. het verbod van kinderarbeid en de bescherming van jongeren op het werk),
5. **burgerschap** (bevat bepalingen over o.a. het recht op behoorlijk bestuur en het recht van inzage in documenten),
6. **rechtspleging** (bevat bepalingen over o.a. het recht op een doeltreffende voorziening in rechte en op een onpartijdig gerecht en het vermoeden van onschuld en rechten van de verdediging), en
7. algemene bepalingen betreffende de uitlegging en de toepassing van het Handvest (bevat bepalingen over o.a. het toepassingsgebied, de reikwijdte en uitlegging van de gewaarborgde rechten en beginselen).

Verwezenlijking van betrouwbare AI

vereiste	toelichting
Menselijke controle en menselijk toezicht	grondrechten, menselijke controle en menselijk toezicht
Technische robuustheid en veiligheid	Weerbaarheid tegen aanvallen en beveiliging, uitwijkplan, nauwkeurigheid, betrouwbaarheid en reproduceerbaarheid
Privacy en datagovernance	Respect voor privacy, kwaliteit en integriteit gegevens, toegang tot gegevens
transparantie	Traceerbaarheid, verklaarbaarheid, communicatie

Verwezenlijking van betrouwbare AI

vereiste	Toelichting
Diversiteit, non-discriminatie, rechtvaardigheid	Voorkomen van onrechtvaardige vertekening, toegankelijkheid en participatie van belanghebbenden
Maatschappelijk en milieubewustzijn	Duurzaamheid, milieuvriendelijkheid, sociale gevolgen, samenleving en democratie
verantwoording	Controleerbaarheid, minimalisering en verslaglegging van negatieve gevolgen, afwegingen en beroep

Controleren van betrouwbare AI

CONTROLELIJST VOOR BETROUWBARE KI (TESTVERSIE)

1. Menselijke controle en menselijk toezicht

Grondrechten:

- ✓ Hebt u, in de gebruikssituaties waarin er mogelijk sprake kon zijn van negatieve gevolgen voor de grondrechten, een effectbeoordeling op het gebied van grondrechten uitgevoerd? Hebt u mogelijke afwegingen tussen de verschillende beginselen en rechten vastgesteld en gedocumenteerd?
- ✓ Is er sprake van een wisselwerking tussen het KI-systeem en de beslissingen van menselijke eindgebruikers (bijv. aanbevolen handelingen of beslissingen, aanbod van opties)?
 - Is er in die gevallen een risico dat het KI-systeem de menselijke autonomie aantast door het beslissingsproces van de eindgebruiker op een onbedoelde manier te beïnvloeden?
 - Hebt u overwogen of het KI-systeem aan de gebruikers moet aangeven dat inhoud of een beslissing, advies of uitkomst het resultaat is van een algoritmische beslissing?
 - Indien het KI-systeem een chatbot of gesprekssysteem bevat: zijn de menselijke eindgebruikers in kennis gesteld van het feit dat ze te maken hebben met een niet-menselijke actor?

Controleren van betrouwbare AI



Impactanalyse Richtlijnen voor
het toepassen van algoritmen
door gemeenten (tweede fase)

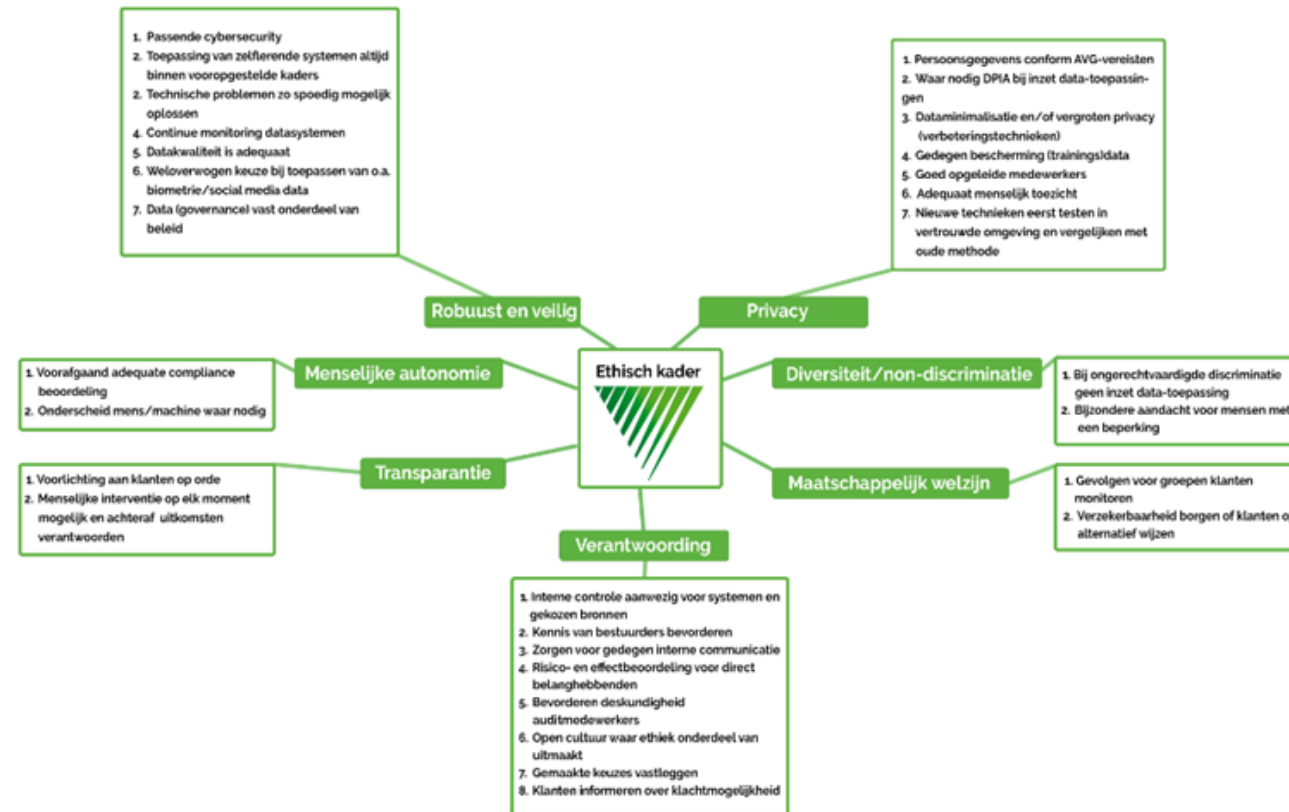
Eindrapportage

Oktober 2020

Toolbox voor Ethisch
Verantwoorde Innovatie



Voorbeeld self assessment



Bron: <https://www.verzekeraars.nl/publicaties/actueel/verzekeraars-aan-de-slag-met-ethisch-kader>

Geen betrouwbare en mensgerichte AI

- het sociaal labelen van burgers door overheden
- Het op afstand biometrisch identificeren in de publieke ruimte.
- AI-systemen die gebruikmaken van de kwetsbaarheden van kinderen

"MY DOLL SAYS IT'S OK"

A STUDY OF CHILDREN'S CONFORMITY TO A TALKING DOLL

INTRODUCTION

Many children are growing up with smart devices such as toys, personal assistants and interconnected toys. Children in all ages will believe that these devices are trustworthy and friendly, and the number of the experience and return of smart toys may make children susceptible to their direct influence. On one hand, we know from prior work that interacting with robots may cause children's behavior to change. On the other hand, previous technology studies with adults show that their robots and avatars are not nearly as persuasive as expected.

To address this pressing concern, we investigated the ability of a talking doll to directly influence children on a conformity task and a disobedience task.

EXPERIMENT SETUP

Can children be directly persuaded by a talking doll to change their moral judgements or disobey an instruction?

We suspected that a talking doll would have more influence over the child, but not as much as a human. Therefore we used two tasks: a conformity task and a disobedience task for that children in two conditions: no external influence (control), an adult human influence (human), and a

PROCEDURE

Children were brought to a private room with a live camera.

Conformity Pretest: Children completed the conformity questionnaire by themselves.

Building Rapport: Children played for the first against the human or talking doll.

Disobedience Test: The researcher presented the child with a box, then said they had to leave the room. The researcher asked the child to not look at the box.

Conformity Post-test: Children completed the conformity questionnaire again, this time listening to an influencer's second voice.

The disobedience test always happened second since it was an exception of the conformity test.

FINDINGS

Changed Answers in Conformity Test by

DISCUSSION

This work first assessed that as many children changed their answers on moral questions, mostly children were more likely to change their answers on non-conventional questions. This is because the questions are subjective. Our results suggest that conformity may not be effective with smart toys.

One explanation for children changing their answers on moral questions is that children were just being polite. Our analysis observed that answers are more likely to change when names were interacting with robots.

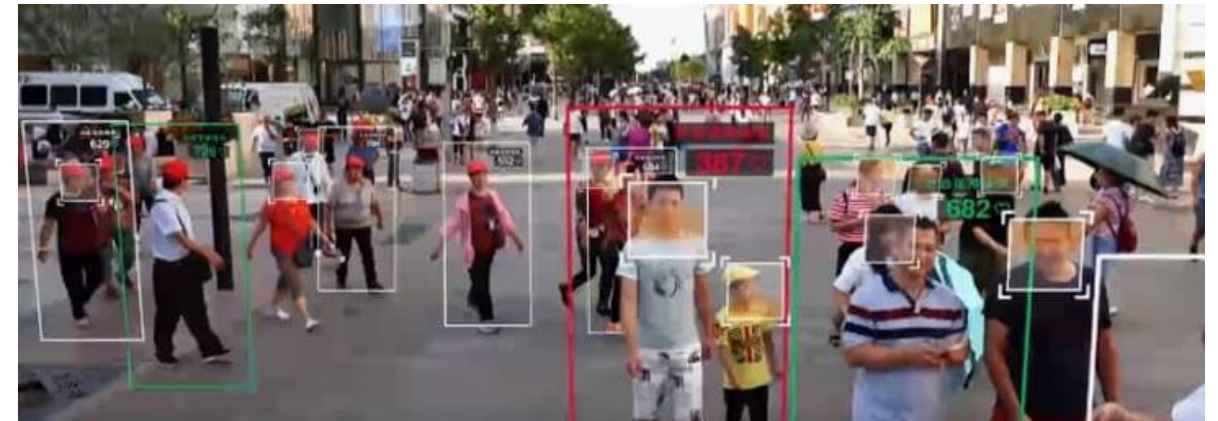
"It is OK to have another child" was the label. "It is OK to be" was the label. Some children changed their answers to a second, then answer "not OK" on the next question. Some children were "not OK to be" some kids at their ages. One answer "OK" for this question and the next one as well.

It is possible that children's behavior on non-conventional questions and the disobedience task is that they believed that they did not have smart toys.

"I don't look in the box" would be OK. "What do you think?" Easy was getting published with dolls. "The dolls, you're being very naughty." No about the box. Further away from dolls. "The Disobedience Test was OK to be OK."

Children's answers are optimized their behavior.

They were conforming to one of their questions and often said "no". "The dolls, that's wrong" because in the experiment scenario we said that dolls were always told the truth. "Was a very nice doll." The introduction addresses the discrepancy between what children see on a screen and how they really feel and behave.



- <https://www.youtube.com/watch?v=NXyzpMDtpSE>

Bron: www. t

Beoordelen aanvraag MIT

- Hoe wordt beoordeeld “plan heeft oog voor betrouwbare en mensgerichte AI”?
- Staatscourant enig document + verwijzing naar SAPAI
- RVO: uit aanvraag moet blijken in hoeverre draag je bij aan dit onderwerp
- Nog onduidelijk hoe dit bij de ranking van de aanvragen een rol speelt.

Beoordelen aanvraag MIT

Advies

- Laat in aanvraag blijken dat dit aspect aandacht krijgt in het project
- Bijvoorbeeld door een apart werkpakket
- Bijvoorbeeld door aandacht aan grondslagen, verwezenlijking en controle van betrouwbare AI

LegalAIR

- kennisplatform met informatie over juridische en ethische aspecten van AI/data
- Deel publiek/deel voor abonnees:
 - Meer informatie/documenten
 - Helpdesk/vraagbaak
 - Loket naar netwerk met experts (ethiek, cyberveiligheid, data science, juridisch)
- Delen opgedane kennis via platform



contact



Jos van der Wijst

E : wijst@bg.legal

M : 0650695916

Jos van der Wijst Tech lawyer

company owner at BG.legal, lawyer for Tech companies, new technologies and law (artificial intelligence and blockchain)

Eindhoven, North Brabant, Netherlands · [Contact info](#)