

An illustration of a business meeting in a modern office. Several people in business attire are seated around a large white table, looking at documents and laptops. A man in a suit stands on the left, gesturing towards a large screen displaying charts and graphs. A woman in a white blazer stands near the center. The room has large windows in the background and a modern pendant light hangs from the ceiling.

Artificial Intelligence, kansen en risico's voor de RvC

Hoe kunnen raad van commissarissen voorkomen dat Artificial Intelligence technologie verkeerde of onjuiste besluitvormingsinformatie geeft in belangrijke processen?

Inhoudsopgave

Voorwoord.....	3
Samenvatting.....	4
1. Inleiding.....	5
1.1 Achtergrond van het onderwerp	5
1.2 Belang van betrouwbare AI-besluitvormingsinformatie	6
1.3 Definitie van verkeerde of onjuiste besluitvormingsinformatie	6
1.4 Doelstelling en onderzoeksvragen	7
2. De Rol van de Raad van Commissarissen (RvC) in AI-besluitvorming	7
2.1 Taken en verantwoordelijkheden van de RvC	7
2.2 Specifieke uitdagingen bij AI-toepassingen	8
2.3 Toezicht op technologische innovatie en leiderschap	8
3. Risico's van verkeerde of onjuiste AI-besluitvormingsinformatie	9
3.1 Fouten in data en algoritmen	9
3.2 Bias en discriminatie in AI-systemen	10
3.3 Gebrek aan transparantie (black box-probleem).....	10
3.4 Overmatige afhankelijkheid van AI-systemen.....	10
4. Governance en toezichtmechanismen voor betrouwbare AI	10
4.1 AI-ethiek en -compliance	10
4.2 Belang van transparantie en uitlegbaarheid (Explainable AI)	10
4.3 Kwaliteitsborging van data en algoritmen.....	11
4.4 Rol van externe audits en onafhankelijke toetsing.....	11
5. Praktische handvatten voor de RvC.....	11
5.1 Vragen die de RvC zou moeten stellen	11
5.2 Vaardigheden en kennisvereisten binnen de RvC.....	13
5.3 Best practices uit de praktijk.....	14
6. Case studies	14
6.1 Voorbeeld van een AI-besluitvormingsfout en de lessen hieruit.....	14
6.2 Succesvol toezicht op AI in een organisatie	14
7. Conclusie en aanbevelingen.....	14
7.1 Samenvatting en aanbevelingen	14
7.2 Praktische implementatie.....	16
7.3 Toekomstige ontwikkelingen en aandachtspunten	16
8. Bronnenlijst.....	17

Voorwoord

Technologische ontwikkelingen voltrekken zich in een razendsnel tempo. Dit biedt niet alleen talloze kansen, maar brengt ook significante risico's met zich mee voor onze maatschappij. Mijn overtuiging om deel te nemen aan de Nyenrode Commissarissen Cyclus is geworteld in de overtuiging dat de rol van de Chief Information Officer (CIO) een onmisbare schakel vormt binnen de Raad van Toezicht (RvT) en de Raad van Commissarissen (RvC). Deze rol is cruciaal om technologische kansen en risico's beter te kunnen beoordelen en te integreren in strategische besluitvorming.

Al voorafgaand aan de cyclus besepte ik dat het programma mij zou inspireren om een paper te schrijven over de impact van technologie. Dit vermoeden werd direct bevestigd tijdens de eerste sessies. Van de ruim 45 deelnemers waren er slechts twee met een CIO-achtergrond; de overige deelnemers hadden voornamelijk een bedrijfskundig of financieel profiel. Dit contrast prikkelde mij, mede door de inzichten van de sprekers over de toenemende complexiteit van de toezichthoudende rol. Een rol die wordt gekenmerkt door beperkte tijd, complexe dilemma's, afstand tot de organisatie en het werken met een beperkte set informatie, terwijl de toezichthouder tegelijkertijd aansprakelijk blijft voor de genomen besluiten.

De eerste dagen hebben mijn onderwerp verder vormgegeven. De ontwikkeling van technologieën, zoals Artificial Intelligence, biedt onmiskenbaar kansen, maar ook aanzienlijke risico's. In mijn carrière heb ik veel samengewerkt met toezichthouders en ervaren hoe het ontbreken van diepgaande kennis en ervaring op het gebied van digitale transformatie en technologische ontwikkelingen kan leiden tot aannames en een gebrek aan kritische diepgang. Er wordt vaak onvoldoende doorgevraagd en gespard op strategisch-technologisch vlak.

Mijn overtuiging is dat mijn betoog andere toezichthouders en bestuurders zal inspireren. Bewustwording van de invloed van besluitvormingsprocessen in combinatie met de inzet van Artificial Intelligence kan al bijdragen aan verandering. Dit bewustzijn kan niet alleen leiden tot betere besluitvorming, maar ook tot meer veerkrachtige en toekomstbestendige organisaties.

Samenvatting

a. Motivatie voor het onderzoek en de centrale onderzoeksvraag

De opkomst van kunstmatige intelligentie (AI) in besluitvormingsprocessen biedt aanzienlijke kansen voor organisaties, maar brengt ook complexe risico's met zich mee. Foutieve of bevooroordeelde data (bias), gebrekkige algoritmische modellen en gebrek aan transparantie kunnen leiden tot verkeerde of onjuiste besluitvormingsinformatie. Dit vormt een uitdaging voor de Raad van Commissarissen (RvC), die verantwoordelijk is voor effectief toezicht en risicobeheersing. De centrale onderzoeksvraag luidt: *“Hoe kunnen Raden van Commissarissen voorkomen dat AI verkeerde of onjuiste besluitvormingsinformatie geeft in belangrijke processen?”*

b. Methodiek/benaderingswijze

Het onderzoek is uitgevoerd door middel van een literatuurstudie, waarin bestaande theorieën en praktijkvoorbeelden zijn geanalyseerd. Daarnaast zijn inzichten van experts op het gebied van AI, ICT recht (legal), corporate governance en ethiek geïntegreerd, waaronder het werk van Werner Zuurbier over leiderschap in het AI-tijdperk. Deze benadering biedt een breed perspectief op zowel technologische als organisatorische aspecten van AI-toezicht. Er zijn veel YouTube video's en artikelen beschikbaar die een verdieping en verrijking geven voor dit onderzoek.

c. Bevindingen van het onderzoek

AI introduceert unieke risico's, waaronder fouten in data en algoritmen, bias en discriminatie, gebrek aan transparantie (het “black box”-probleem) en overmatige afhankelijkheid van AI-systemen. De RvC kan deze risico's mitigeren door:

- Investeren in technologische kennis binnen de RvC.
- Implementatie van governance-structuren gericht op ethiek en transparantie.
- Actieve betrokkenheid bij risicobeheer en data-kwaliteitsborging.
- Toezicht op de maatschappelijke impact van AI-besluitvorming.

d. Beperkingen van het onderzoek

Dit onderzoek is gebaseerd op secundaire bronnen en theoretische modellen, zonder empirische casestudy's. Hierdoor kunnen specifieke praktijkinzichten ontbreken.

Bovendien is de technologische ontwikkeling van AI dynamisch, wat kan leiden tot snelle veranderingen in relevante risico's en toezichtbehoeften. Ik heb wel een aantal experts en RvC-leden geïnterviewd.

e. Praktische en/of theoretische implicaties

Praktisch biedt het onderzoek concrete handvatten voor RvC's om AI-risico's te identificeren en te beheersen. Theoretisch draagt het bij aan de discussie over de rol van toezichthouders in een wereld die steeds digitaler wordt, waarbij ethiek, governance en technologische innovatie samenkomen.

f. Bijdrage aan maatschappelijke of organisatorische issues

Dit onderzoek benadrukt het belang van verantwoord AI-gebruik in organisaties, met aandacht voor ethische besluitvorming en maatschappelijke impact. Door de rol van de RvC te versterken in het toezicht op AI, kan bijgedragen worden aan transparantere, eerlijkere en betrouwbaardere besluitvormingsprocessen in zowel profit- als non-profitorganisaties.

1. Inleiding

1.1 Achtergrond van het onderwerp

Kunstmatige intelligentie (AI) wordt steeds vaker ingezet om besluitvormingsprocessen te ondersteunen. Van risicobeoordeling en strategische planning tot operationele optimalisatie, AI biedt geavanceerde analysemogelijkheden en voorspellingen die de efficiëntie en effectiviteit van organisaties kunnen vergroten. AI maakt het mogelijk om complexe datasets te analyseren, patronen te herkennen en beslissingsondersteuning te bieden die verder gaat dan menselijke capaciteiten.

Deze technologische vooruitgang brengt echter ook nieuwe uitdagingen met zich mee. Naarmate organisaties steeds meer vertrouwen op AI, groeit het risico dat foutieve of bevooroordeelde informatie de basis vormt voor belangrijke beslissingen. Dit kan leiden tot operationele fouten, strategische misstappen en zelfs juridische of reputatieschade. De integratie van AI in besluitvorming vereist daarom een kritische benadering, waarbij niet alleen de technische werking van AI, maar ook de ethische en juridische implicaties zorgvuldig worden overwogen.

1.2 Belang van betrouwbare AI-besluitvormingsinformatie

Voor de Raad van Commissarissen (RvC), die verantwoordelijk is voor het toezicht op het bestuur en het waarborgen van de continuïteit van de organisatie, is het essentieel om zich bewust te zijn van de potentiële risico's van AI. Foutieve of bevooroordeelde data (bias), gebrekkige algoritmische modellen of een gebrek aan transparantie in AI-processen kunnen de kwaliteit van de besluitvorming negatief beïnvloeden. De betrouwbaarheid van AI-besluitvormingsinformatie is een cruciale factor in strategische besluitvorming, aangezien het directe implicaties heeft voor risicobeheer, compliance en de algehele bedrijfsvoering. Een onjuist geïnterpreteerd AI-advies kan leiden tot verkeerde investeringsbeslissingen, inefficiënte processen of zelfs ethische dilemma's, zoals discriminatie of privacy-schendingen. Dit onderstreept de noodzaak voor de RvC om niet alleen te vertrouwen op AI-outputs, maar ook actief te toetsen hoe deze outputs tot stand komen.

1.3 Definitie van verkeerde of onjuiste besluitvormingsinformatie

In de context van dit paper verwijst ik met "verkeerde of onjuiste besluitvormingsinformatie" naar informatie die leidt tot beslissingen die niet in lijn zijn met de strategische doelen van de organisatie of die een verhoogd risico vormen voor operationele, financiële of reputatieschade. Dit kan voortkomen uit:

- **Foutieve data:** Onnauwkeurige, onvolledige of verouderde gegevens die als input dienen voor AI-modellen. Dit kan bijvoorbeeld het gevolg zijn van slecht databeheer, gebrekkige dataverzamelmethodeën of menselijke fouten.
- **Bevooroordeelde data of algoritmen:** Informatie of beslisregels die systematische vooroordelen bevatten, wat kan leiden tot discriminerende of oneerlijke uitkomsten. Bias kan (bewust of onbewust) zowel in de data als in de modelontwikkeling sluipen.
- **Gebrekkige modellen:** Algoritmen die onvoldoende zijn gevalideerd of niet robuust genoeg zijn voor de context waarin ze worden toegepast. Dit kan resulteren in onbetrouwbare voorspellingen of suboptimale beslissingen.
- **Gebrek aan transparantie:** Situaties waarin het moeilijk is om te begrijpen hoe een AI-systeem tot een bepaald besluit is gekomen (het zogenaamde "black

box"-probleem). Dit belemmert het vermogen van toezichthouders om de validiteit van beslissingen te beoordelen.

1.4 Doelstelling en onderzoeksvragen

Dit paper onderzoekt hoe de RvC kan voorkomen dat AI verkeerde of onjuiste besluitvormingsinformatie genereert. De belangrijkste onderzoeksvragen en hun behandelde hoofdstukken zijn:

- Welke risico's zijn verbonden aan het gebruik van AI in besluitvormingsprocessen? (Hoofdstuk 3)
- Hoe kan de RvC effectief toezicht houden op de kwaliteit en betrouwbaarheid van AI-systemen? (Hoofdstuk 4 en 5)
- Welke governance- en controlemechanismen zijn noodzakelijk om de integriteit van AI-besluitvorming te waarborgen? (Hoofdstuk 6 en 7)

Door deze vragen te beantwoorden, wil ik de RvC praktische handvatten bieden om de juiste vragen te stellen, de juiste competenties te ontwikkelen en een effectief toezichtkader te implementeren dat AI-gerelateerde risico's minimaliseert. Dit paper beoogt niet alleen een theoretisch kader te bieden, maar ook concrete aanbevelingen voor de praktijk.

2. De Rol van de Raad van Commissarissen (RvC) in AI-besluitvorming

2.1 Taken en verantwoordelijkheden van de RvC

De Raad van Commissarissen heeft als kerntaak toezicht te houden op het beleid van het bestuur en de algemene gang van zaken binnen de organisatie. Dit omvat strategisch toezicht, risicobeheersing en het waarborgen van de naleving van wet- en regelgeving. In de context van AI betekent dit dat de RvC niet alleen moet toezien op de implementatie van AI-technologieën, maar ook op de ethische, juridische en maatschappelijke implicaties ervan. Belangrijke verantwoordelijkheden van de RvC met betrekking tot AI zijn onder andere:

- **Toezicht op strategische inzet van AI:** Waarborgen dat AI-toepassingen aansluiten bij de langetermijnstrategie van de organisatie en bijdragen aan duurzame waardecreatie.
- **Risicobeheer:** Identificeren en beoordelen van de risico's die gepaard gaan met AI-gebruik, waaronder tactische, operationele, strategie, juridische en ethische risico's.
- **Beoordeling van governance-structuren:** Zorgen dat er een adequaat intern controlesysteem aanwezig is om de betrouwbaarheid en integriteit van AI-besluitvorming te waarborgen.
- **Ethisch toezicht:** Bewaken van de maatschappelijke verantwoordelijkheid van de organisatie bij het gebruik van AI.

2.2 Specifieke uitdagingen bij AI-toepassingen

AI introduceert unieke uitdagingen voor toezichthouders, waaronder:

- **Complexiteit van technologie:** De technische aard van AI kan het moeilijk maken voor toezichthouders zonder technische achtergrond om de implicaties volledig te begrijpen. Dit vereist voortdurende bijscholing en de inzet van externe experts.
- **Dynamiek en snelheid van technologische veranderingen:** AI-technologieën ontwikkelen zich snel, wat betekent dat toezichthouders flexibel en adaptief moeten zijn in hun toezichtpraktijken.
- **Ethiek en verantwoordelijkheid:** AI-besluiten kunnen ethische dilemma's opleveren, bijvoorbeeld rond privacy, bias en autonomie, waarbij onduidelijk kan zijn wie eindverantwoordelijk is voor de uitkomsten.
- **Vertrouwen en transparantie:** De "black box"-natuur van sommige AI-systemen maakt het moeilijk om inzicht te krijgen in de besluitvormingsprocessen, wat het vertrouwen in de technologie kan ondermijnen.

2.3 Toezicht op technologische innovatie en leiderschap

Effectief toezicht op AI vereist een proactieve benadering van technologische innovatie. De RvC moet niet alleen reageren op bestaande risico's, maar ook anticiperen op toekomstige ontwikkelingen. Dit kan door:

- **Vergroten van technologische kennis:** Investeren in opleidingen en trainingen voor RvC-leden om hun begrip van AI te verdiepen en hen in staat te stellen kritische vragen te stellen.
- **Diversiteit in de raad:** Zorgen voor een diverse samenstelling van de RvC, inclusief leden met technologische expertise en verschillende perspectieven.
- **Stakeholderdialoog:** Actief in gesprek gaan met interne en externe belanghebbenden om een breed perspectief op AI-gerelateerde vraagstukken te verkrijgen.

Werner Zuurbier benadrukt in zijn boek *"Leiderschap in het AI-tijdperk"* het belang van transformationeel leiderschap in deze context. Hij onderscheidt vier cruciale leiderschapsrollen binnen de RvC:

- **Visionaire leiders:** Zij zien de kansen van technologie en durven toekomstgericht te denken, ook zonder altijd concrete antwoorden te hebben.
- **Risicoleiders:** Zij benoemen de onzekerheden van technologische transformatie en nemen verantwoordelijkheid voor het bespreekbaar maken van complexe dilemma's.
- **Morele leiders:** Zij waarborgen de maatschappelijke en ethische impact van AI-besluitvorming binnen de organisatie.
- **Culturele leiders:** Zij inspireren en motiveren teams om met enthousiasme de digitale transformatie te omarmen.

Idealiter combineert een effectieve toezichthouder eigenschappen van al deze leiderschapsrollen, om zo een krachtig, adaptief en ethisch verantwoord toezicht te bieden in het AI-tijdperk.

3. Risico's van verkeerde of onjuiste AI-besluitvormingsinformatie

3.1 Fouten in data en algoritmen

AI-systemen zijn afhankelijk van de kwaliteit van de data die ze verwerken. Onjuiste, verouderde of onvolledige data kunnen leiden tot foutieve analyses en verkeerde beslissingen. Dit kan bijvoorbeeld ontstaan door gebrekkige dataverzamelmethode, menselijke fouten of verouderde datasets. Daarnaast

kunnen fouten in algoritmische code of modelconfiguraties onverwachte uitkomsten veroorzaken.

3.2 Bias en discriminatie in AI-systemen

AI-modellen kunnen bestaande vooroordelen in data versterken, wat kan leiden tot discriminerende beslissingen. Dit vormt niet alleen een ethisch probleem, maar kan ook juridische risico's met zich meebrengen. Bias kan optreden in zowel de trainingsdata als in de algoritmische logica zelf.

3.3 Gebrek aan transparantie (black box-probleem)

Veel AI-systemen zijn complex en moeilijk te doorgronden, waardoor het lastig is te begrijpen hoe beslissingen tot stand komen. Dit gebrek aan transparantie kan het vertrouwen in de besluitvorming ondermijnen en het moeilijk maken om verantwoording af te leggen bij fouten of incidenten.

3.4 Overmatige afhankelijkheid van AI-systemen

Organisaties kunnen te veel vertrouwen op AI-uitkomsten zonder voldoende kritische evaluatie. Dit kan leiden tot onvoldoende menselijke controle en beoordelingsvermogen, wat risico's vergroot in complexe situaties. Menselijke tussenkomst en kritisch denken blijven essentieel om AI-beslissingen te valideren.

4. Governance en toezichtmechanismen voor betrouwbare AI

4.1 AI-ethiek en -compliance

Een effectief governance-framework voor AI begint bij ethiek en compliance. AI-ethiek richt zich op het waarborgen van rechtvaardigheid, transparantie en verantwoordelijkheid in het gebruik van AI. Voor de RvC betekent dit het bewaken van ethische richtlijnen binnen de organisatie en ervoor zorgen dat deze zijn geïntegreerd in de AI-strategieën. AI Act stelt strikte eisen aan ethische AI-systemen, met een focus op risicobeoordeling en compliance.

4.2 Belang van transparantie en uitlegbaarheid (Explainable AI)

Een van de grootste uitdagingen bij AI is het gebrek aan transparantie, ook wel het "black box"-probleem genoemd. Explainable AI (XAI) streeft ernaar AI-systemen

begrijpelijker te maken voor zowel technische als niet-technische stakeholders. Transparantie is een wettelijke verplichting volgens de AI Act, vooral bij hoog-risico systemen.

4.3 Kwaliteitsborging van data en algoritmen

De kwaliteit van AI-besluitvorming is direct afhankelijk van de kwaliteit van de gebruikte data en algoritmen. Data moeten accuraat, compleet en representatief zijn. Voor algoritmen geldt dat ze grondig getest en gevalideerd moeten worden.

4.4 Rol van externe audits en onafhankelijke toetsing

Onafhankelijke toetsing door externe auditors kan bijdragen aan objectieve controle van AI-systemen. Dit vergroot de transparantie en het vertrouwen in de technologie. AI Act vereist dat AI-systemen traceerbaarheid en verantwoording waarborgen.

5. Praktische handvatten voor de RvC

5.1 Vragen die de RvC zou moeten stellen

Om effectief toezicht te houden op AI-besluitvormingsprocessen, moeten RvC-leden kritische vragen kunnen stellen over de implementatie, risico's en governance van AI-systemen. De EU AI Act introduceert verplichte risicoclassificaties en beheersmaatregelen, die als leidraad kunnen dienen. Hieronder volgt een uitgebreide set vragen die de RvC zou moeten stellen:

A. Kwaliteit van gebruikte data en bias in AI-modellen

- Hoe wordt de kwaliteit van de gebruikte data gegarandeerd?
- Zijn de AI-modellen getest op bias en discriminatie?
- Welke maatregelen zijn genomen om data schoon, accuraat en representatief te houden?
- Hoe wordt gecontroleerd of trainingsdata niet onbedoeld bevooroordeeld zijn?
- Worden er externe audits uitgevoerd op de gebruikte datasets?

B. Risicobeoordeling en classificatie conform de AI Act

- In welke risicocategorie valt ons AI-systeem volgens de EU AI Act (hoog, beperkt, minimaal)?

- Worden er periodieke risicobeoordelingen uitgevoerd op onze AI-toepassingen?
- Hoe wordt het risico van fouten en onvoorziene gevolgen van AI-beslissingen beheerd?
- Is er een plan voor crisismanagement als een AI-systeem onbedoelde schade veroorzaakt?

C. Menselijk toezicht en controle

- Op welke manier is menselijk toezicht ingebouwd in de besluitvorming van AI-systemen?
- Welke teams of functionarissen zijn verantwoordelijk voor het valideren van AI-uitkomsten?
- Hoe wordt voorkomen dat AI-besluiten volledig geautomatiseerd worden zonder menselijke tussenkomst?
- Zijn er duidelijke protocollen om AI-beslissingen te herzien en bij te sturen?

D. Transparantie en uitlegbaarheid (Explainable AI)

- Kunnen we als organisatie duidelijk uitleggen hoe onze AI-beslissingen tot stand komen?
- Hoe wordt de uitlegbaarheid van AI-beslissingen gewaarborgd?
- Welke mechanismen zijn ingebouwd om gebruikers en stakeholders inzicht te geven in hoe AI werkt?
- Zijn AI-beslissingen traceerbaar en goed te herleiden naar specifieke datapunten en algoritmische processen?

E. Cybersecurity en robuustheid

- Welke maatregelen zijn genomen om AI-systemen te beschermen tegen cyberaanvallen of manipulatie?
- Worden er penetratietests en stresstests uitgevoerd op kritische AI-systemen?
- Hoe wordt de betrouwbaarheid van AI-algoritmen bewaakt en verbeterd na implementatie?
- Zijn er incidentresponsplannen voor AI-gerelateerde cyberdreigingen?

F. Wettelijke en ethische compliance

- Zijn onze AI-toepassingen in lijn met de AI Act en andere relevante regelgeving zoals de AVG?
- Hoe wordt ethisch toezicht op AI binnen de organisatie georganiseerd?
- Wat zijn de procedures voor het melden en corrigeren van AI-gerelateerde fouten of klachten?
- Hoe zorgen we ervoor dat AI-gebruikers en stakeholders weten wanneer ze met een AI-systeem te maken hebben?
- Zijn er interne of externe audits die toezien op naleving van wetgeving en ethische richtlijnen?

5.2 Vaardigheden en kennisvereisten binnen de RvC

Om effectief toezicht te houden op AI-besluitvormingsprocessen is het essentieel dat RvC-leden beschikken over relevante technologische en analytische kennis. De complexiteit van AI vereist dat toezichthouders niet alleen vertrouwd zijn met traditionele governance- en risicobeheersingsmodellen, maar ook inzicht hebben in de technische, ethische en juridische aspecten van AI-systemen.

Opleiding en permanente educatie:

Een belangrijke aanbeveling is het structureel opnemen van AI-gerelateerde opleidingen in de ontwikkeling van RvC-leden. Dit kan in de vorm van:

- **Masterclasses en trainingen** over kunstmatige intelligentie, data-ethiek, en algoritmisch toezicht, aangeboden door gerenommeerde onderwijsinstellingen zoals Nyenrode of internationale AI-expertisecentra.
- **Workshops over risicomanagement** met betrekking tot AI, waarin scenario-gebaseerde analyses en casestudy's worden besproken om RvC-leden praktische inzichten te geven.
- **Bijscholing over technologische trends**, zoals machine learning, explainable AI (XAI) en de impact van AI-wetgeving (zoals de EU AI Act).

Door permanente educatie op te nemen als vast onderdeel van de RvC-agenda, kan de raad proactief inspelen op technologische ontwikkelingen en effectiever toezicht houden op AI-risico's.

5.3 Best practices uit de praktijk

Organisaties die succesvol AI integreren, kenmerken zich door een cultuur van transparantie, continue evaluatie en ethisch bewustzijn. Tijdens het Nationaal Congres AI Governance & Compliance hebben toonaangevende organisaties zoals NS, ING, Randstad en het Ministerie van Financiën hun strategieën en best practices gedeeld met betrekking tot AI-governance en compliance. Deze organisaties hebben inzichten gegeven in hoe zij zich voorbereiden op de naleving van de AI Act en welke stappen zij ondernemen om veilige en verantwoorde AI-toepassingen te waarborgen.

6. Case studies

6.1 Voorbeeld van een AI-besluitvormingsfout en de lessen hieruit

Een internationale bank implementeerde een AI-algoritme voor kredietbeoordelingen. Na verloop van tijd bleek het model discriminatoire beslissingen te nemen, gebaseerd op onbewuste bias in de trainingsdata. Dit leidde tot reputatieschade en juridische procedures.

6.2 Succesvol toezicht op AI in een organisatie

Een technologiebedrijf voerde een robuust AI-governance framework in, inclusief een ethische AI-commissie. Dit comité evalueert regelmatig de ethische impact van AI-projecten.

7. Conclusie en aanbevelingen

7.1 Samenvatting en aanbevelingen

AI biedt aanzienlijke kansen voor organisaties, maar brengt ook complexe risico's met zich mee, vooral op het gebied van besluitvorming. Betrouwbare data, transparante algoritmen en ethisch bewustzijn zijn cruciaal. De RvC speelt hierin een essentiële rol en kan door middel van effectieve governance bijdragen aan verantwoord AI-gebruik. Op basis van dit onderzoek worden de volgende aanbevelingen gedaan:

- **Investeer in technologische kennis en opleiding:** RvC-leden dienen structureel opleidingen te volgen over AI, data-ethiek en digitale transformatie. Dit kan via masterclasses, workshops en bijscholingsprogramma's, zodat zij beter in staat zijn kritische vragen te stellen en toezicht te houden op AI-implementaties.
- **Implementeer duidelijke governance-structuren:** Ontwikkel beleid gericht op ethiek, transparantie en risicobeheersing binnen AI-toepassingen, conform de EU AI Act en ISO/IEC 42001:2023.
- **Stimuleer een cultuur van ethisch bewustzijn:** Integreer ethiek in alle lagen van de organisatie, met speciale aandacht voor de impact van AI-besluitvorming op de samenleving.
- **Diversiteit in expertise:** Zorg voor een diverse samenstelling van de RvC, met leden die verschillende perspectieven en technologische expertise inbrengen.
- **Implementeer ISO/IEC 42001:2023 als AI-managementsysteem:** Dit biedt organisaties een gestructureerd kader voor verantwoord AI-beheer, inclusief risicomanagement, ethische richtlijnen en compliance met wet- en regelgeving.
- **Zorg voor menselijk toezicht op AI-besluitvorming:** De EU AI Act vereist dat AI-systemen niet volledig autonoom beslissen in hoog-risicosituaties. Organisaties moeten duidelijke protocollen implementeren om menselijke controle te waarborgen.
- **Transparantie en verantwoording borgen:** AI-gebruikers en stakeholders moeten inzicht krijgen in de werking en impact van AI-besluitvormingssystemen. Het documenteren en rapporteren van AI-besluiten wordt steeds belangrijker.

7.2 Praktische implementatie

Naast de strategische aanbevelingen en toekomstgerichte inzichten, is een effectieve implementatie van AI-governance cruciaal voor organisaties. De volgende concrete stappen kunnen helpen bij de integratie in de praktijk:

- **Operationele verankering van AI-governance:** Organisaties moeten AI-governance koppelen aan bestaande structuren, zoals risicobeheer, compliance en interne audits, zodat AI niet als een losstaand domein wordt behandeld.
- **AI-comité of ethiekcommissie oprichten:** Het aanstellen van een multidisciplinair team dat AI-beslissingen monitort, ethische kwesties bespreekt en zorgt voor naleving van de AI Act en ISO/IEC 42001:2023.
- **Continue monitoring en rapportage:** AI-systemen moeten regelmatig worden geëvalueerd op prestaties, bias en ethische compliance, met periodieke rapportage aan de RvC.
- **Externe samenwerking:** Organisaties kunnen waardevolle inzichten verkrijgen door samen te werken met toezichthouders, AI-experts en sectororganisaties om best practices te implementeren.
- **Stakeholdercommunicatie:** Transparantie over AI-gebruik binnen en buiten de organisatie vergroot het vertrouwen en minimaliseert juridische en ethische risico's.

7.3 Toekomstige ontwikkelingen en aandachtspunten

De rol van de RvC zal zich blijven ontwikkelen naarmate AI een grotere impact krijgt op organisaties. Proactief toezicht, adaptieve governance en voortdurende dialoog zijn cruciaal. Regelgeving zoals de AI Act en normen zoals ISO/IEC 42001:2023 helpen bedrijven AI op een verantwoorde en strategische manier te integreren. De belangrijkste vraag voor elke RvC is dan ook: hoe ziet onze organisatie er over vijf jaar uit? Welke afdelingen, rollen en processen veranderen door AI? Door vooruit te kijken en AI niet alleen als risico, maar ook als kans te benaderen, kan de RvC sturen op een toekomstbestendige en wendbare organisatie.

8. Bronnenlijst

- Autoriteit Persoonsgegevens (2021). *Richtlijnen voor AI en data governance*.
- Autoriteit Persoonsgegevens (2024). *Manipulatieve, misleidende en uitbuitende AI-systemen. Verbodsbepalingen in EU-verordening 2024/1689 (AI-verordening)*.
- Brynjolfsson, E. & McAfee, A. (2017). *Machine, Platform, Crowd: Harnessing Our Digital Future*. Norton & Company.
- European Commission (2020). *Ethics Guidelines for Trustworthy AI*.
- Nyenrode Business Universiteit (2024). *Handboek Toezicht en Technologie*.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.
- Zuurbier, W. (2024). *Leiderschap in het AI-tijdperk*. Brave New Books.
- World Economic Forum (2025). *Artificial Intelligence and Cybersecurity: Balancing Risks and Rewards*.
- Stanford University Human Centered Artificial Intelligence (2023). *Artificial Intelligence Index Report*.
- MIT Technology Review (2021). *Bias isn't the only problem with credit scores and no, AI can't help*.
- De Nederlandse Corporate Governance Code (2022). [https://www.mccg.nl/publicaties/codes/2022/12/20/corporate-governance-code- 2](https://www.mccg.nl/publicaties/codes/2022/12/20/corporate-governance-code-2).
- Harvard Business Review (2012). *Managing Risks: A New Framework*.
- The Institute of Internal Auditors (2020). *Het three-lines-model van het IIA*.

Elon Musk: "I have a concern really for all the major AI programs. I mean, the two biggest ones obviously are Google Gemini and Microsoft OpenAI. The concern I have, and this seem like a small concern, but I think it's actually a very big issue, is that they are not maximally truth-seeking. They are pandering to political correctness."