Auteur: drs. David Pavlovic, VP Global Digital Integrated Operations bij Kerry en voormalig strategisch adviseur bij Schuberg Philis

Samen beslissingen nemen met Explainable AI

Is AI een kans of een nachtmerrie voor u? De vraag daarachter is: hoe maken we AI veilig, transparant en betrouwbaar? Explainable AI (XAI) biedt inzicht in de besluitvormingsprocessen van AI, essentieel voor verantwoord gebruik in sterk gereguleerde en complexe omgevingen. In dit artikel bespreek ik de rol van XAI in het balanceren tussen efficiëntie van moderne bedrijfsprocessen en de informatie-beveiliging daarvan.

In mijn werk houd ik me dagelijks bezig met het ondersteunen van mensen in hun besluitvorming en het optimaliseren van bedrijfsprocessen door middel van digitale toepassingen en kunstmatige intelligentie (AI). Deze technologieën bieden ongekende mogelijkheden om mensen te ontlasten, hun beslissingen te verbeteren en processen efficiënter te maken. Tegelijkertijd stellen ze ons ook voor nieuwe uitdagingen op het gebied van informatiebeveiliging en transparantie, vooral wanneer het gaat om het begrijpen van de keuzes die AI-systemen maken.

XAI speelt een cruciale rol in hoe het ons kan helpen om AI-toepassingen niet alleen effectief, maar ook veilig en verantwoord in te zetten. Door AI begrijpelijker en transparanter te maken, kunnen we niet alleen vertrouwen op de resultaten, maar ook zorgen voor de veiligheid en beveiliging van deze systemen – een balans die essentieel is binnen een snel digitaliserende omgeving.

Transparantie en uitlegbaarheid

Veel AI-modellen, vooral de 'black box'-modellen zoals diepe neurale netwerken, zijn complex en bieden nauwelijks inzicht (transparantie) noch leggen zij de besluitvorming uit. Dit gebrek is vooral risicovol in sectoren waarin data- en informatiebeveiliging een grote rol spelen, zoals informatie informatie-interpretatie, formule- en receptgeneratie, productie-executie, supply-chain en andere sterk gereguleerde omgevingen.

Met XAI wordt de 'black box' opengebroken: het helpt gebruikers inzicht te krijgen in de factoren en variabelen die ten grondslag liggen aan AI-beslissingen. Dit is van onschatbare waarde om te begrijpen hoe algoritmen conclusies trekken, waardoor het makkelijker wordt om te vertrouwen op hun aanbevelingen in dagelijkse en strategische beslissingen. Transparantie is niet gebaseerd op efficiëntie of verantwoordelijkheid, maar wel op uitlegbaarheid en risicobeheersing.

Het belang van transparantie

Het bieden van inzicht in de logica (redenatie/uitleg = transparantie) achter een AI-systeem wekt vertrouwen in de technologie en zorgt voor een hogere mate van acceptatie. Wanneer de technologie transparant is, krijgen eindgebruikers een helder beeld van de factoren en patronen die door het systeem zijn geanalyseerd. Voor beslissers betekent dit dat zij de geldigheid en betrouwbaarheid van AI-aanbevelingen snel kunnen inschatten en kritisch kunnen beoordelen.

Momenteel wordt echter ook gedebatteerd over de mate waarin deze uitleg daadwerkelijk noodzakelijk is. Net zoals mensen die zelf beslissingen nemen, zonder precies te kunnen uitleggen hoe hun eigen zenuwstelsel werkt, vragen sommigen zich af of volledige uitlegbaarheid altijd nodig is. Dit blijft een open vraagstuk dat nog veel discussie zal oproepen.

Veiligheidsuitdagingen

AI-systemen die zelfstandig (geautomatiseerd) beslissingen nemen, kunnen ons werk enorm versnellen en verbeteren, maar ook nieuwe uitdagingen met zich meebrengen op het gebied van beveiliging. Wanneer AI-modellen gebruikmaken van vertrouwelijke bedrijfsgegevens om voorspellingen en adviezen te genereren, kunnen er potentiële risico's ontstaan, vooral als gebruikers niet begrijpen hoe beslissingen tot stand komen. Een gebrek aan inzicht in de besluitvorming van een AI-model kan leiden tot operationele risico's en onverwachte beveiligingsincidenten.

Beveiligingskwesties

Bij autonome AI-systemen is de beveiliging van zowel gegevens als algoritmen cruciaal, daarbij is het essentieel dat er een hoog niveau van bescherming bestaat tegen ongewenste toegang of manipulatie van gegevens. In dit kader speelt XAI een rol door te laten zien welke informatie wordt verwerkt en hoe deze informatie bijdraagt aan de beslissingen van het model.

'Bias'-risico's

Een ander veiligheids- en vertrouwensaspect is het risico op bias. In AI-besluiten kan bias ontstaan door vooringenomenheid in de data waarop een model is getraind. XAI stelt in staat om deze biases te identificeren en aan te pakken, wat de basis vormt voor eerlijke en betrouwbare besluitvorming. Zonder dit inzicht lopen bedrijven het risico op fouten die niet alleen inefficiënt, maar ook schadelijk voor hun reputatie kunnen zijn.

Daarom is informatiebeveiliging een dubbele prioriteit bij AI-systemen: we moeten niet alleen de gegevens beveiligen, maar ook de algoritmische processen die op deze data worden losgelaten. XAI biedt hierin een essentiële ondersteuning door inzicht te geven in de algoritmische logica, zodat mogelijke kwetsbaarheden en afwijkingen vroegtijdig gedetecteerd kunnen worden. XAI-deskundigen zijn dan ook altijd op zoek naar manieren om AI-besluiten te controleren en te begrijpen, zodat we in elke situatie in staat zijn tot weloverwogen en veilige keuzes.

Verantwoordelijkheid en veiligheid

XAI geeft ons de mogelijkheid om AI-besluiten te begrijpen en te controleren, waardoor verantwoorde en veilige besluitvorming mogelijk wordt. Wanneer AI binnen bedrijfsomgevingen autonoom beslissingen neemt, is het essentieel dat er een duidelijke verantwoording kan worden gegeven. Zonder transparantie zouden we beslissingen opvolgen waarvan we de basis niet

XAI speelt een cruciale rol in hoe het ons kan helpen om AI-toepassingen niet alleen effectief, maar ook veilig en verantwoord in te zetten

begrijpen, wat in kritieke situaties kan leiden tot onnodige risico's. Door XAI toe te passen, gaan we niet alleen na welke gegevens de AI heeft gebruikt, maar ook hoe deze gegevens bijdragen aan de uiteindelijke beslissing. Dit zorgt voor een robuuster proces en geeft de mensen die met AI werken de mogelijkheid om fouten of afwijkingen te herkennen en aan te pakken. Binnen supply-chain- en productieomgevingen, waar dagelijks planning, voorraadbeheer, inkoop, levering, prijsstelling, contractbeheer en kwaliteitscontrole een rol spelen, is de balans tussen context, robuustheid en transparantie onmisbaar in besluitvorming.

XAI en gereguleerde sectoren

In financiële sectoren, de gezondheidszorg en in productiesectoren gelden strenge regels voor de beveiliging en integriteit van gegevens. Hier speelt XAI een cruciale rol door bedrijven in staat te stellen om volledig te voldoen aan regelgeving omtrent gegevensbeheer en besluitvorming. Transparantie is hierbij niet alleen wenselijk, maar in daar ook wettelijk verplicht.

Praktische implementeringsuitgangspunten

Om XAI op een verantwoorde en transparante manier in te zetten, zijn er enkele belangrijke uitgangspunten waarmee we rekening houden:

- **Duidelijke en begrijpelijke beslissingslogica**
  Het is belangrijk dat AI-modellen begrijpelijke beslissingen opleveren. Dit kan door expliciet te kiezen voor transparante modellen of door extra uitlegmogelijkheden te bieden voor complexere algoritmen. Begrijpelijkheid verlaagt drempels voor gebruikers om AI-besluiten te valideren.
- **Beveiliging van data en algoritmen**
  Gevoelige data en algoritmen moeten op een veilige manier worden beschermd. XAI maakt inzichtelijk welke data is gebruikt en hoe, zodat gebruikers de logica achter AI-besluiten kunnen beoordelen en afwijkingen snel kunnen signaleren.
- **Menselijke controle en toezicht**
  XAI zorgt ervoor dat AI-besluiten altijd gecontroleerd kunnen worden door mensen, waardoor gebruikers kunnen valideren of een bepaalde beslissing consistent is met hun verwachtingen en ervaring. Dit verhoogt de betrouwbaarheid en maakt aanpassingen mogelijk waar nodig.
- **Verantwoording en foutopsporing**
  Transparante AI-besluiten maken het mogelijk om de oorzaak van fouten snel te achterhalen. Door de factoren te identificeren die aan een beslissing bijdragen, kunnen afwijkingen beter worden opgespoord en waar nodig aangepast.
- **Stapsgewijze implementatie van veilige XAI**
  Een zorgvuldige implementatie van XAI vergt samenwerking tussen data experts, IT-beveiligingsteams en operationele teams. Hierbij is het belangrijk dat elke stap in het beslissingsproces inzichtelijk wordt gemaakt voor alle betrokkenen, van IT-specialisten tot eindgebruikers.

Conclusie

Explainable AI vormt de kern van een verantwoorde en veilige inzet van kunstmatige intelligentie. Kijkende naar de uitdagingen in sterk gereguleerde sectoren en de behoefte aan transparantie, helpt XAI ons AI echt begrijpelijk en toegankelijk te maken. Vertrouwen in technologie is niet alleen belangrijk voor de directe gebruikers, maar ook voor iedereen die afhankelijk is van de door AI ondersteunde beslissingen. Dit vertrouwen kan alleen ontstaan als we begrijpen hoe de technologie werkt en hoe bepaalde keuzes tot stand komen.

Vaak worden gesprekken gevoerd over de balans tussen snelheid versus veiligheid en innovatie versus verantwoording. Juist in deze discussies is XAI onmisbaar voor ons allen. Het geeft niet alleen inzicht in de beslissingen van AI, maar ook de mogelijkheid om deze keuzes kritisch en ethisch te bekijken en aan te passen waar nodig. Zo kunnen we AI gebruiken als een waardevolle partner in besluitvorming zonder de controle of ons vertrouwen te verliezen.

Het blijft een fascinerend en soms lastig vraagstuk. Vooruitgang boeken betekent soms het onbekende te omarmen, maar met een sterk fundament van transparantie en veiligheid kunnen we AI inzetten om niet alleen efficiënter, maar ook zorgvuldiger te werken. XAI biedt precies dat evenwicht – een manier om het beste uit technologie te halen, terwijl we de mens centraal blijven stellen in het beslissingsproces.