

Hoe betrouwbaar is uw GenAI-toepassing?

03 september 2026, 12:00

Mihail Mihaylov

Ons Trustworthiness Playbook helpt u te bepalen wat u nodig hebt om uw GenAI-toepassing volledig te kunnen vertrouwen

Generatieve AI (GenAI) evolueert snel van een experimentele tool naar een essentiële technologie voor industriële en digitaal-technologische toepassingen. Bedrijven integreren GenAI in producten, werkstromen en besluitvormingsprocessen om efficiënter, flexibeler en toegankelijker te werken. Maar de echte doorbraak van GenAI hangt af van een vraag: “Kunt u uw toepassing vertrouwen?” Het antwoord verschilt per sector, toepassing en bedrijf. Ons Trustworthiness Playbook helpt u een heel eind op weg.

Verschillende vereisten per project en sector

Betrouwbaarheid zit niet alleen in het AI-model. De volledige werk- en datastroom en het procesontwerp zijn minstens even belangrijk. Bovendien verschilt het sterk naargelang de context. Als u elke mogelijke vereiste als even belangrijk behandelt, leidt dat tot onnodige complexiteit en overbodige engineeringinspanningen.

Neem een interne GenAI-assistent die een onderhoudstechnicus helpt informatie te vinden in technische documentatie. Een fout antwoord is ongewenst, maar de technicus behoudt de controle en kan het antwoord controleren.

Vergelijk dat met een AI-voorspellingssysteem waarvan de output automatisch wordt gebruikt om biedingen in te dienen op een energiemarkt. Fouten kunnen daar rechtstreekse financiële gevolgen hebben. Ook ruis of ontbrekende gegevens in de input kunnen een geautomatiseerd proces beïnvloeden. Beide zijn GenAI-toepassingen. Maar om ze te kunnen vertrouwen, moet u andere prioriteiten stellen.



Negen aspecten van betrouwbare GenAI

Het playbook onderscheidt negen aspecten van AI-betrouwbaarheid: **nauwkeurigheid**, **overeenstemming met de regelgeving**, **uitlegbaarheid**, **robuustheid**, **betrouwbaarheid**, **veiligheid**, **transparantie**, **eerlijkheid** en **latentie**.

Nauwkeurigheid gaat na of het systeem de juiste output produceert. **Betrouwbaarheid** beoordeelt of het systeem zich onder normale omstandigheden consistent gedraagt. **Robuustheid** geeft aan of het systeem blijft presteren wanneer de input ruis bevat, onvolledig is of afwijkt van de data waarvoor het werd ontworpen.

Andere aspecten worden belangrijk in andere omstandigheden. Een systeem dat toegankelijk is voor externe gebruikers kan extra nadruk op **veiligheid** vereisen. In gereguleerde omgevingen kunnen **naleving van regelgeving** en **transparantie** prioriteiten worden. **Uitlegbaarheid** is essentieel wanneer gebruikers aanbevelingen moeten begrijpen en valideren. **Eerlijkheid** wordt relevant wanneer de output verschillende gevolgen kan hebben voor personen of groepen. En in een realtime toepassing voor operatoren is zelfs een nauwkeurig antwoord nutteloos als de **latentie** te hoog is. U hoeft niet alle negen aspecten te maximaliseren. Focus op de aspecten die uw toepassing niet mag verwaarlozen.

Beslistabel als leidraad voor uw afwegingen

Hier speelt de **beslistabel** uit het playbook een belangrijke rol. In plaats van te vertrekken vanuit abstracte principes stelt de tabel praktische vragen over uw toepassing. Een antwoord met 'ja' wijst

op betrouwbaarheidsaspecten die u als absolute vereisten moet beschouwen.

Bijvoorbeeld:

- Ondersteunt de AI belangrijke operationele, financiële, gezondheids- of veiligheidsbeslissingen? Focus op **veiligheid** en **uitlegbaarheid**.
- Is de toepassing toegankelijk voor externe gebruikers of krijgt ze niet-vertrouwde input? Geef prioriteit aan **veiligheid**.
- Bevat de inputdata ruis, verandert ze in de loop van de tijd of is ze onvolledig? Maximaliseer de **robuustheid**.
- Gelden er regelgevende vereisten of auditvereisten? Focus op **naleving van de regelgeving** en **transparantie**.
- Moet de toepassing in realtime reageren? Investeer in een **lage latentie**.
- Moeten gebruikers of auditors het gedrag en de beperkingen van de toepassing begrijpen? Geef prioriteit aan **transparantie**.
- Voert de toepassing voorspellingen, prognoses, automatisering of controle uit? **Nauwkeurigheid en betrouwbaarheid** zijn dan uw belangrijkste aandachtspunten.

Deze beslistabel zet een brede discussie over AI-risicobeheer om in een praktisch instrument waarmee engineering-, product- en innovatieteams concreet aan de slag kunnen.

Voorbeeld: interactieve ondersteuning voor operatoren

Het playbook bevat ook concrete voorbeelden van hoe deze aanpak in de praktijk werkt. Neem bijvoorbeeld een productieomgeving waarin medewerkers een AI-assistent gebruiken om zelfstandig complexere taken uit te voeren. Het systeem combineert verschillende informatiebronnen, zoals technische documenten, beelden en gesproken vragen. Het communiceert ook met externe tools om operatoren contextuele informatie en aanbevelingen te geven.

Voor een toepassing die operatoren ondersteunt, zijn **nauwkeurigheid, betrouwbaarheid en lage latentie** logische uitgangspunten. Medewerkers hebben correcte en consistente ondersteuning nodig en verwachten binnen enkele seconden een antwoord. Maar de beslistabel brengt nog andere prioriteiten aan het licht.

De assistent genereert aanbevelingen en content. Daarbij wordt **uitlegbaarheid** een belangrijk aspect. De interactie met minder technisch onderlegde operatoren vraagt ook extra aandacht voor **veiligheid**. En omdat er in gesproken input ruis kan zitten en transcripties onvolmaakt kunnen zijn, wordt **robuustheid** essentieel. Het systeem moet nuttige ondersteuning blijven bieden, ook wanneer de input niet optimaal is.

De beoordeling resulteert zo in zes absolute vereisten: **nauwkeurigheid, betrouwbaarheid, uitlegbaarheid, veiligheid, robuustheid en lage latentie**. **Transparantie en naleving van de regelgeving** blijven waardevol, maar zijn voor deze specifieke toepassing niet noodzakelijk prioritair.

Vervolgens vertaalt u deze vereisten naar praktische ontwerpkeuzes. Zo verhoogt u de betrouwbaarheid door interacties met tools en validatiemechanismen te structureren. U versterkt de veiligheid door het systeem te beperken tot vooraf gedefinieerde acties en operatoren bij het proces betrokken te houden.

Voorwaarden voor GenAI-betrouwbaarheid

Zelfs een zeer intelligente GenAI-app is waardeloos als niet aan de juiste voorwaarden voor betrouwbaarheid is voldaan. Zelfs de beste inspanningen om GenAI betrouwbaar te maken leveren nog altijd een slecht resultaat op wanneer de inputdata van slechte kwaliteit zijn, outputs niet worden gevalideerd, gebruikers de beperkingen verkeerd begrijpen of de AI in de werkstroom te veel autonomie krijgt.

Er zijn daarentegen heel wat manieren om de betrouwbaarheid te verbeteren. Ons playbook somt er veel op, niet als abstracte theorie maar als praktische maatregelen. Baseer outputs op gezaghebbende gegevens, voeg validatiecontroles toe, standaardiseer prompts, monitor veranderingen in data, definieer terugvalscenario's, beperk gevoelige functionaliteiten of houd mensen betrokken bij het proces. Dat zijn slechts enkele voorbeelden van het praktische en onmiddellijk toepasbare advies in dit document.

Samengevat

Betrouwbare GenAI vraagt meer dan een goed AI-model. Wat 'betrouwbaar' betekent, hangt af van de taak, data, gebruikers, mate van autonomie, operationele omgeving, gevolgen van fouten en toepasselijke regelgeving.

Het 'Trustworthiness Playbook' van Sirris en Flanders Make helpt teams om die beoordeling systematisch uit te voeren. Het combineert negen betrouwbaarheidsaspecten met herkenbare clusters van gebruiksscenario's, een praktische beslistabel en concrete strategieën om de betrouwbaarheid te verbeteren.

Maak betrouwbaarheid concreet voor uw GenAI-toepassing

Welke aspecten van betrouwbaarheid zijn echt belangrijk voor uw GenAI-toepassing? Download het Trustworthiness Playbook om uw prioriteiten te bepalen en ontdek praktische maatregelen om ermee aan de slag te gaan.

[Download het playbook](#)

Vragen over betrouwbare GenAI?

Hebt u nog vragen over betrouwbaarheid of hoe u deze principes kunt toepassen op uw GenAI-toepassing? Mihail Mihaylov, Senior Engineer AI & GenAI bij Sirris, bespreekt ze graag met u.

[Contacteer onze expert](#)

Auteurs

Profile picture for user mihail.mihaylov@sirris.

Mihail Mihaylov