WAAR MOET JE AANDACHT VOOR HEBBEN ALS JE GENAI GEBRUIKT?

Generatieve AI (GenAI) is een technologie die zelf onder andere teksten, beelden of muziek kan maken. Dat biedt **veel mogelijkheden** voor creativiteit, efficiëntie en productiviteit, zowel voor bedrijven als voor burgers. Tegelijk roept GenAI ook **belangrijke vragen** op: Wat betekent het voor onze jobs? Hoe zit het met privacy? Wat is de impact op het milieu? En wie is eigenaar van de inhoud die GenAI maakt?

Om organisaties te helpen **bewust om te gaan met deze risico's**, ontwikkelde het Kenniscentrum Data & Maatschappij de GenAI Blind Spots. Deze kaartenset helpt professionals **ethische valkuilen te herkennen** en er verstandig mee om te gaan.

In deze brAInfood vind je meer informatie over de GenAI Blind Spots en hoe je ze kan gebruiken in je organisatie.

ETHISCHE VALKUILEN VAN GENAI

Ethische valkuilen bij GenAI gebruik in een organisatie situeren zich in **vier categorieën**:

■ TOEPASSING

Bij zowel **het ontwerp, de training en de ontwikkeling** van GenAI-toepassingen zijn er belangrijke ethische aandachtspunten. Enkele voorbeelden:

- **Datakwaliteit:** GenAI-toepassingen worden getraind met erg veel gegevens - vaak verkregen van het internet - waarbij **de juistheid van deze gegevens** niet altijd wordt geverifieerd.
- **Auteursrecht:** Of bepaalde gegevens, zoals boeken of teksten, ook **mogen gebruikt worden** voor de training van GenAI-toepassingen is niet altijd gegarandeerd.

Als organisatie heb je vaak geen controle over de specificaties van een GenAI-toepassing of over de gegevens waarop het model is getraind. Je kunt deze aspecten wel mee in overweging nemen bij **het kiezen van een toepassing en bij hoe je die inzet**.

● GEBRUIK

Wanneer je GenAI-toepassingen gaat **inzetten binnen een organisatie** is het belangrijk rekening te houden met verschillende overwegingen in verband met het eigenlijke gebruik. Enkele voorbeelden:

- **Onvoorspelbaarheid:** Een GenAI-toepassing kan op allerlei **voorziene en onvoorziene manieren** gebruikt worden. Een toepassing die ontwikkeld werd om sociale media-berichten te genereren kan (oneigenlijk/ongewenst) gebruikt worden voor doeleinden waarvoor het algoritme niet getraind is, bijvoorbeeld om rapporten op te stellen.
- **Vaardigheden en competenties:** GenAI-toepassingen kunnen zorgen voor tijdwinst, inspiratie of een betere dienstverlening. Maar kan iedereen gemakkelijk met de nieuwe toepassing overweg of hebben collega's **ondersteuning nodig** bij het verwerven van deze nieuwe competenties?

In een **pilootproject** kan je duidelijke afspraken maken rond het gebruik van GenAI voor specifieke taken. Daarnaast helpt het je ook belangrijke zaken **in kaart te brengen** zoals de mogelijke impact, de gebruiksvriendelijkheid van de toepassing en de nodige bijscholing.

▶ UITKOMSTEN

Ook aan de uitkomsten, zoals teksten, beelden, geluiden, inzichten, etc., **die een GenAI-toepassing genereert**, kunnen risico's verbonden zijn. Enkele voorbeelden:

- **Hallucinaties:** GenAI-toepassingen zijn ontworpen om menselijke taal en vaardigheden te imiteren. De toepassing heeft **geen besef van wat feitelijk correct** is aangezien het louter voorspellingen maakt van wat logisch volgt. Zo kan het dat de toepassing dingen gaat "verzinnen", ook wel hallucinaties genoemd.
- **(Indirecte) vooroordelen:** De gegevens waar modellen van GenAI-toepassingen op zijn getraind, kunnen **onvoldoende diversiteit** bevatten, bijvoorbeeld door onder- of overrepresentatie van bepaalde groepen of historische gebeurtenissen waar minder informatie over beschikbaar is. Dat maakt de uitkomsten van een GenAI-toepassing vaak **stereotiep**.

Zorg dat je de uitkomsten van een GenAI-toepassing altijd **kritisch bekijkt en controleert** op mogelijke onjuistheden, vooroordelen, en of het resultaat representatief, verklaarbaar en relevant is.

⬣ ORGANISATIE & SAMENLEVING

Het bestaan en gebruik van GenAI heeft ook **bredere gevolgen** voor een organisatie of zelfs de samenleving. Enkele voorbeelden:

- **Werkgelegenheid en -tevredenheid:** GenAI-toepassingen kunnen bepaalde **jobs volledig overnemen**, terwijl bij andere jobs de inhoud mogelijk **ingrijpend verandert**. Organisaties die GenAI inzetten voor het creëren van afbeeldingen zullen mogelijk minder designers in dienst nemen. De rol van een designer kan hierdoor verschuiven van 'maker' naar 'corrigeerder', wat als minder creatief ervaren kan worden.
- **Doelbewust misbruik:** Een persoon kan met de hulp van GenAI **deepfakes** creëren die personen in een compromitterende situatie afbeelden, op grote schaal (academische) teksten genereren om een **persoonlijke ideologische agenda te legitimeren**, een AI-developer chatbot gebruiken om een **virus** te schrijven, etc. De impact ervan reikt ver buiten de grenzen van een organisatie, en heeft **invloed op de samenleving**.

Voor je een GenAI-toepassing uitrolt, denk je best na over de bredere maatschappelijke impact. Probeer **bewuste keuzes** te maken over hoe je hier als organisatie mee wil omgaan.

NEXT STEPS

Je kan de kaartenset **op verschillende manieren** gebruiken:

- om meer **bewustzijn** te creëren over aandachtspunten bij GenAI-gebruik,
- als **leidraad om afspraken te maken** over het gebruik van GenAI (zoals een gedragscode),
- om **risico's te ontdekken** tijdens het ontwikkelen of invoeren van een GenAI-toepassing,
- als **startpunt voor gesprekken** over mogelijke gevolgen van GenAI binnen de organisatie.

We bevelen aan een **diverse groep** van personen te betrekken bij de bespreking van de kaarten. Zeker als je een gedragscode wil ontwikkelen of nadenkt over de inzet van GenAI-toepassingen in de organisatie, is het van belang om zo veel mogelijk **verschillende perspectieven** mee te nemen.

Download hier de GenAI Blind Spots die je helpen op een **verantwoorde** manier met GenAI om te gaan in je organisatie. Veel succes!

Kenniscentrum Data & Maatschappij

Artificiële Intelligentie Vlaanderen

Kenniscentrum Data & Maatschappij (juni 2025). Waar moet je aandacht voor hebben als je GenAI gebruikt? Brussel: Kenniscentrum Data & Maatschappij.

Deze brAInfood is beschikbaar onder een CC BY 4.0 licentie.