



Richtlijn voor het verantwoord gebruik van generatieve kunstmatige intelligentie (GenAI) bij Buildwise

Laatste update: 19 februari 2025

Gezien de snelle ontwikkelingen van de mogelijkheden en risico's van GenAI, wordt deze richtlijn regelmatig geüpdatet en is hiermee een dynamisch document.

Deze richtlijn is mede tot stand gekomen met gebruik van GPT4 (versie 24 mei 2023), een taalmodel van OpenAI. De door GPT4 gegenereerde tekstonderdelen zijn door Buildwise gecontroleerd op feitelijke juistheid, waar nodig aangepast, en vervolgens goedgekeurd. De eindverantwoordelijkheid voor de uiteindelijke tekst van deze richtlijn ligt bij Buildwise. De blauwe kaders in dit document bevatten de essentiële punten voor het correcte en verantwoordelijke gebruik van GenAI voor tekstgeneratie, tekstredactie, en andere vormen van generatieve output zoals afbeeldingen, video's, audio en code. De richtlijn is opgesteld met gebruik van de referentiedocumenten zoals vermeld in sectie 9, in het bijzonder de richtlijn van KU Leuven omtrent het verantwoord gebruik van GenAI in onderzoek, hierna de richtlijn van KU Leuven genoemd.

Inhoud

1. Inleiding
2. Mogelijkheden en toepassingen
3. Inherente beperkingen en risico's van GenAI
4. Risico's omtrent data- en privacy beheer bij het gebruik van GenAI
5. Transparantie en labeling
6. Gebruik van GenAI voor subsidieaanvragen en wetenschappelijke publicaties
7. Gebruik van GenAI voor programmeercode
8. Bekendstelling, herziening en bekrachtiging van de richtlijn
9. Referenties

Disclaimer:

Deze AI policy is bedoeld om algemene richtlijnen en inspiratie te bieden voor het verantwoord gebruik van generatieve kunstmatige intelligentie (GenAI). Buildwise heeft alle mogelijke inspanningen geleverd om de juistheid en volledigheid van de hierin opgenomen informatie te waarborgen. Buildwise geeft echter geen enkele vorm van verklaring of garantie, expliciet of impliciet, met betrekking tot de volledigheid van dit beleid. Dit beleid is geen juridisch advies en mag niet als zodanig worden beschouwd. Gebruikers van dit beleid zijn zelf verantwoordelijk voor het waarborgen dat hun gebruik van GenAI voldoet aan alle toepasselijke wetten en regelgeving.

1. Inleiding

GenAI, ook bekend als Generative Artificial Intelligence, is een subveld van kunstmatige intelligentie dat zich richt op het creëren van nieuwe inhoud of gegevens. GenAI-systemen zijn ontworpen om op een autonome manier output te genereren die niet noodzakelijkerwijs uit bestaande gegevens hoeft te worden gekopieerd. Ze kunnen tekst, afbeeldingen, muziek, spraak, of zelfs complexe datastructuren genereren.

In de context van AI modellen zoals GPT (Generative Pretrained Transformer) wordt GenAI voornamelijk gebruikt om zinvolle en coherente tekst te genereren. Deze modellen worden getraind op enorme datasets van menselijke taal. De output van GenAI kan onvoorspelbaar zijn omdat het AI model, uitzonderingen daargelaten, nieuwe inhoud genereert in plaats van bestaande inhoud te hergebruiken. Het is belangrijk om op te merken dat hoewel GenAI in staat is om indrukwekkend realistische uitvoer te genereren, het geen bewust begrip heeft van de inhoud die het genereert. Het baseert zijn generatie puur op statistische patronen die het heeft geleerd uit de data waarop het is getraind.

Als innovatiecentrum in de bouwsector onderkent Buildwise het potentieel van GenAI, inclusief modellen zoals ChatGPT, Google Gemini, Claude, en andere, waaronder ook de talrijke modellen voortgebracht door de open source gemeenschap. GenAI is niet beperkt tot tekstgeneratie en is afhankelijk van het gebruikte model multimodaal, met een voorspelbaar toenemende tendens naar meer multimodaliteit in de toekomst. Dit stelt GenAI in staat om zelfstandig te leren over de patronen aanwezig in audio, foto's, video's, sensordata en andere modaliteiten, en deze te gebruiken als input en om te zetten naar andere gewenste modaliteiten. Deze richtlijn is van toepassing op alle vormen van generatieve AI (GenAI), inclusief maar niet beperkt tot tekst, afbeeldingen, video, audio en code, die binnen Buildwise worden gebruikt. Naarmate de technologie en onze toepassingen ervan evolueren, zal deze richtlijn worden aangepast om nieuwe ontwikkelingen en risico's te adresseren. Dit punt wordt nader toegelicht in sectie 8 van deze richtlijn. Buildwise streeft ernaar om GenAI op een verantwoorde en ethische manier in te zetten voor het verbeteren van de eigen werking en haar dienstverlening voor de Belgische bouwsector.

2. Mogelijkheden en toepassingen

Tekstgeneratie en tekstredactie biedt een breed scala aan mogelijkheden, die naar verwachting snel verder zullen toenemen en diversifiëren. Onder de huidige mogelijkheden en toepassingen vallen het collaboratief produceren van nieuwe teksten zoals artikels, subsidieaanvragen, beleidsstrategieën, social media berichten, programmeercode, het samenvatten en herschrijven van teksten, het creëren van afbeeldingen, video's, het beantwoorden van vragen, en, met aankoppeling van extra tools of plug-ins, het autonoom uitvoeren van taken zoals het opzoeken van informatie, en het updaten van bestaande teksten door raadpleging van internetinformatie, om maar enkele voorbeelden te noemen.

3. Inherente beperkingen en risico's van GenAI

De huidige state-of-the-art GenAI taalmodellen hebben meerdere beperkingen en risico's, die ervoor zorgen dat hun output verschillende soorten van problematiek kan bevatten:

- Taalmodellen beschikken niet over een menselijk begrip van tekst. Hierdoor kan de output van taalmodellen vanuit menselijk perspectief overduidelijke fouten en incorrecte redeneringen bevatten.
- Taalmodellen kunnen geleerd hebben van teksten en datasets met inherente bias, vooroordelen en/of feitelijk onjuiste informatie, waardoor de taalmodellen deze ook zelf kunnen voortbrengen.
- Taalmodellen kunnen hallucineren, wat wil zeggen dat het taalmodel creatief is, maar niet op de gewenste manier, waardoor het bijvoorbeeld feitelijk onjuiste beweringen maakt, complete citaten verzint of verkeerde interpretaties geeft.
- De output van taalmodellen kan een 1-op-1 kopie bevatten van bestaand werk of daar een grote overeenkomst mee hebben, zonder dit te vermelden. Wordt de tekst ongecontroleerd overgenomen, dan bestaat daarmee de kans op plagiaat.
- Taalmodellen kunnen meegaan met de mening zoals ingegeven door de menselijke gebruiker, ook als deze niet is gestoeld op de werkelijkheid.
- Taalmodellen zijn niet expliciet geprogrammeerd, maar gegroeid. Dit, samen met de hoge mate van complexiteit maakt dat taalmodellen daardoor een zogenoemde 'black box' kunnen zijn, waarvan het lastig is om te bepalen hoe het model precies tot een bepaalde output is gekomen. Om de output van het model te sturen, kan er worden gewerkt met ingebouwde zogeheten guardrails (veiligheidsmaatregelen) en een eigen instructieset (system message), echter dit blijft bij een sturing, en biedt geen complete controle over de output. Het is daarnaast mogelijk om met ingave van een specifieke instructie (prompt injection) een zogeheten jailbreak te forceren, waarmee de guardrails van het model teniet worden gedaan en het mogelijk maakt om het model potentieel schadelijke informatie te laten genereren.
- Bij het overstijgen van een bepaald capaciteitsniveau ontstaat bij taalmodellen het fenomeen van emergent properties, nieuwe capaciteiten die spontaan, en onverwacht kunnen ontstaan. Een voorbeeld daarvan is een hogere capaciteit voor redenerie. Voor huidige state-of-the-art modellen zoals GPT-4o is gesteld dat er geen sprake is van problematische emergent properties. Bij toekomstige modellen kan die kans niet worden uitgesloten. Het is daarom dat er voorstellen zijn van AI-experten om het gebruik van GenAI boven een bepaald capaciteitsniveau aan banden te leggen, en een toezichhoudend wereldwijd agentschap in te stellen, gelijkend aan het Internationaal Atoomenergie Agentschap (IAEA).

Wees ervan bewust dat tekst geproduceerd door GenAI onjuiste of verzonnen informatie en bias kan bevatten, en/of een mogelijke inbreuk op het auteursrecht of plagiaat. Daarnaast kan GenAI meegaan met de mening van de gebruiker, zonder de feitelijke juistheid daarvan in twijfel te trekken. Het is de verantwoordelijkheid van de medewerker om ethisch en verantwoordelijk te handelen, en de GenAI tekst grondig op deze potentiële problemen te controleren en waar nodig aan te passen. De verantwoordelijkheid voor de tekst ligt altijd geheel bij de medewerker zelf.

4. Risico's omtrent data- en privacy beheer bij het gebruik van GenAI

- **Verhoogde kans op overtreding van de GDPR** Anders dan bij het gebruik van traditionele zoekmachines is de mate en aard van de informatie die wordt ingevoerd in taalmodellen zoals chat interfaces anders. Het kan bijvoorbeeld gaan om het ingeven van complete Word documenten, e-mails, gespreksnotities en andere documentatie, waarvan de kans groter is dat deze persoonsgegevens of vertrouwelijke informatie bevatten. Bij het gebruik van taalmodellen aangeboden door externe diensten, waardoor de gegevens de beveiligde Buildwise omgeving verlaten, is het daarmee eens te meer essentieel dat de dienst waarvan gebruik wordt gemaakt een data- en privacy beheer heeft dat overeenkomstig is met de GDPR-wetgeving. Daarnaast brengt het ingeven van informatie in een taalmodel van een externe dienst een uniek aanvullend risico met zich mee: Zonder de nodige maatregelen bestaat er een kans dat de ingegeven informatie wordt gebruikt in een toekomstige versie van het gebruikte taalmodel, en potentieel door derden kan worden onttrokken.
- **Risico voor IP en octrooien** Zonder de juiste maatregelen kan ingegeven informatie in een taalmodel als zijnde publiekelijk beschikbaar worden beschouwd, met een grote kans dat er geen bescherming meer kan worden verkregen voor ingegeven IP of een octrooi kan worden aangevraagd op die gegevens.

Voorkom GDPR-overtredingen door alleen gegevens in te voeren in taalmodellen van bedrijven en instellingen die voldoen aan de GDPR-wetgeving.

Buitenom beveiligde omgevingen toegestaan door Buildwise dienen er geen persoons- of vertrouwelijke gegevens te worden ingevoerd in taalmodellen. Dit is eens te meer belangrijk, omdat er bij oneigenlijk gebruik van taalmodellen kans is op onbedoelde openbaarmaking van de ingegeven informatie. Een bijzonder voorbeeld is IP en octrooierbare informatie. Door deze in te geven in een onbeveiligde omgeving is de kans groot dat deze niet langer kan worden beschermd/octrooierbaar is.

De opvolgende secties 5 en 6 van de Buildwise richtlijn leunen in het bijzonder zwaar op de richtlijn van KU Leuven

5. Transparantie en labeling

Gebruik van GenAI moet transparant en duidelijk worden vermeld (gelabeld als gemaakt met behulp van GenAI), zodat het niet over het hoofd gezien kan worden. De situatie bepaalt of hiertoe een verplichting bestaat. In deze sectie gaat het over transparantie en labeling in het algemeen. Latere secties behandelen de specifieke gevallen van het schrijven van subsidieaanvragen, wetenschappelijke publicaties en programmeercode. Om de mate van AI-bijdrage verder te verduidelijken, maken we gebruik van de labels “Aided by AI” en “Made with AI”, hieronder grafisch weergegeven. Het label “Aided by AI” wordt gebruikt voor materialen die gedeeltelijk zijn gemaakt met behulp van AI, waarbij significante menselijke aanpassingen zijn gedaan. Voorbeelden hiervan zijn een AI-gegenereerde foto die nadien manueel is aangepast, een bewerkte AI-gegenereerde tekst, of een video met menselijke bewerkingen. Het label “Made with AI” wordt gebruikt voor materialen die volledig door AI zijn gegenereerd zonder significante menselijke bijdragen, zoals een volledig door AI gegenereerde afbeelding, tekst of video. Deze labels helpen om transparantie te waarborgen en duidelijk te maken in welke mate AI heeft bijgedragen aan de creatie van visuele materialen.



AIDED BY AI



MADE WITH AI

Vermelding verplicht

Teksten die tot stand zijn gekomen met innovatieve ideeën bijgedragen door GenAI, bedoeld voor publicatie, zowel schriftelijk als online, moet duidelijk als zodanig worden gelabeld om verwarring met menselijk geschreven tekst te voorkomen en transparantie te waarborgen. Dit geldt ook voor intern gebruik.

Vermelding niet verplicht

Als GenAI louter werd gebruikt als taal-assistent om een tekst te herschrijven, om te vormen of te vertalen zonder toevoeging van innovatieve ideeën.

Als GenAI feedback heeft gegeven ter verbetering van menselijk geschreven tekst, maar die feedback louter door de medewerker wordt gebruikt om de tekst zelf te verbeteren.

Ongeacht of vermelding verplicht is of niet: de medewerker is altijd volledig verantwoordelijk voor de tekst.



6. Gebruik van GenAI voor subsidieaanvragen en wetenschappelijke publicaties

Het is toegestaan om GenAI in te zetten voor het schrijven van subsidieaanvragen en wetenschappelijke publicaties, onder de volgende voorwaarden en doeleinden:

Subsidieaanvragen en wetenschappelijke publicaties

Controleer en respecteer een mogelijke GenAI richtlijn van de subsidieverstrekker of redactiecommissie. Maak vooraf melding bij strijdige richtlijnen met die van Buildwise.

GenAI mag worden gebruikt voor het genereren van tekst over algemeen bekende informatie. Voor niet-algemene informatie mag GenAI onverminderd worden gebruikt voor andere ondersteunende doeleinden zoals ideevorming, onderzoek en feedback. Controleer GenAI tekst grondig, wees transparant en vermeld de aard van het gebruik. Vermelding is niet nodig als GenAI alleen wordt ingezet als taal-assistent.

Claim GenAI nooit als (mede-)auteur; de volledige verantwoordelijkheid voor de tekst ligt bij de opsteller(s).

Subsidievertrekkers of redactiecommissies kunnen het gebruik van GenAI traceren. Documenteer het gebruik van GenAI buitenom de inzet als taal-assistent, opdat je voorbereid bent op mogelijke vragen. Een export van de chatdata helpt hierbij.

7. Gebruik van GenAI voor programmeercode

GenAI kan zeer tijdbesparend werken voor het genereren of verbeteren van code. Zoals bij gewone GenAI teksten bestaat de kans op plagiaat doordat GenAI bestaande code kan hergebruiken of daar een grote gelijkenis mee kan hebben zonder dit te vermelden. Daarbij kan die code onder een licentie vallen die het eigen voorziene gebruik niet toestaat. Tools zoals Amazon CodeWhisperer kunnen helpen om dergelijke problematiek te signaleren. De eindverantwoordelijkheid voor de code ligt onverminderd bij de programmeur zelf. Gebruik van GenAI voor programmeercode dient transparant te worden vermeld.

Wees je bewust van mogelijke plagiaatrisico's en gelijkenissen met bestaande code die door GenAI kan worden gegenereerd, en controleer of de gegenereerde code mogelijk onder een licentie valt die het beoogde gebruik niet toestaat.

Gebruik hulpmiddelen, zoals Amazon CodeWhisperer, om eventuele problemen te identificeren. De eindverantwoordelijkheid voor de code ligt bij de Buildwise medewerker.

Wees transparant en vermeld het gebruik van GenAI in het programmeringsproces.



8. Bekendstelling, herziening en bekrachtiging van de richtlijn

Deze richtlijn is bekend gesteld aan de medewerkers door integrale plaatsing op het intranet van Buildwise. Als praktisch instrument voor een correct en verantwoordelijk gebruik van GenAI werd eveneens een 'cheat sheet' lijst met Do's en Don'ts verstrekt waarin essentiële punten van de richtlijn zijn vervat.

Deze richtlijn wordt regelmatig herzien en geüpdatet naar aanleiding van nieuwe toepassingen, veranderingen en technologische ontwikkelingen. Hiervan worden de medewerkers op de hoogte gebracht, en de lijst met Do's en Don'ts zal hierop eveneens worden aangepast. De directeur-generaal van Buildwise zal de richtlijn en haar toekomstige versies ondertekenen, teneinde te bekrachtigen dat deze de goedkeuring heeft van de organisatie en van belang is voor de operationele bedrijfsvoering.

9. Referenties

Voor de opstelling van deze richtlijn werden, in volgorde van belangrijkheid, de volgende referenties gebruikt:

1. KU Leuven. (2023, 13 juni). *Verantwoord gebruik van generatieve artificiële intelligentie (GenAI) in onderzoek*. Geraadpleegd op 19 juni 2023, van https://research.kuleuven.be/en/integrity-ethics/integrity/practices/genAI_NL
2. OpenAI. (2022, 14 november). *Sharing & publication policy*. <https://openai.com/policies/sharing-publication-policy>
3. Center for AI Safety. (n.d.). *Statement on AI risk*. <https://www.safe.ai/statement-on-ai-risk>
4. Altman, S., Brockman, G., & Sutskever, I. (2023, 22 mei). *Governance of superintelligence*. OpenAI. <https://openai.com/blog/governance-of-superintelligence>
5. (2023). *Tools such as ChatGPT threaten transparent science; here are our ground rules for their use [Editorial]*. *Nature*, 613, 612. <https://doi.org/10.1038/d41586-023-00191-1>
6. Soice, E. H., Rocha, R., Cordova, K., Specter, M., & Esvelt, K. M. (2023). Can large language models democratize access to dual-use biotechnology? arXiv preprint arXiv: 2306.03809. <https://doi.org/10.48550/arXiv.2306.03809>
7. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. arXiv preprint arXiv: 2303.12712. <https://doi.org/10.48550/arXiv.2303.12712>
8. Roberts, S. (2023, 13 april). Amazon CodeWhisperer, free for individual use, is now generally available. Amazon. <https://aws.amazon.com/blogs/aws/amazon-codewhisperer-free-for-individual-use-is-now-generally-available/>