

Verantwoord gebruik van artificiële intelligentie voor onderzoeksdoeleinden

Inhoudstabel

1. Introductie.....	2
1.1 Algemeen.....	2
1.2 Reikwijdte van richtlijnen	3
2. Algemene principes	3
3. Voor welke onderzoeksdoeleinden kunnen AI-tools ingezet worden?	5
3.1 Onderzoeksdoeleinden	5
3.2 Beschikbare AI-tools voor onderzoeksdoeleinden.....	13
3.3 Trainingen omtrent het inzetten van AI-tools voor deze doeleinden	13
4. Voordelen van het gebruik van AI in onderzoek.....	13
5. Beperkingen en potentiële problemen bij het gebruik van AI-tools in onderzoek.....	14
6. Refereren naar het gebruik van AI-tools in onderzoeksprocessen.....	18
4.1 Stand van zaken	18
4.2 Hoe te refereren naar AI?.....	18
7. Referentielijst.....	20

Note: Deze pagina zal regelmatig geüpdatet worden met nieuwe informatie naargelang het landschap verder evolueert. Er wordt aangeraden om altijd de meest recente versie van de richtlijnen te raadplegen.

Voor vragen rond deze richtlijnen en het gebruik van AI-tools in onderzoek kan u terecht bij AI4research@vub.be.

1. Introductie

1.1 Algemeen

Artificiële intelligentie, vaak afgekort als “AI”, is een verzamelnaam voor verscheidene technologieën die taken uitvoeren of zich gedragen op een “intelligente” wijze. De Organisatie voor Economische Samenwerking en Ontwikkeling (OESO) definieert een Artificieel Intelligentie Systeem als *“a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy”* (1,2).

De ontwikkelingen inzake AI zijn recent in een stroomversnelling terechtgekomen. De lancering van [OpenAI's ChatGPT](#) in november 2022 heeft hoge verwachten geschapen over de mogelijke inzetbaarheid van generatieve AI-tools (hierna “GenAI”) voor uiteenlopende doeleinden. Inmiddels zijn er tal van alternatieven voor ChatGPT ontwikkeld, zoals [Anthropic's Claude](#) en [Meta's LLaMa](#). Voor sommige van deze AI-tools zijn reeds verschillende plug-ins gecreëerd, waardoor het scala aan functionaliteiten en daarmee de grootschalige inzetbaarheid van deze tools nog vergroot. Deze tools kunnen niet enkel tekstinput verwerken, maar ook beelden (bv. DALL-E, Midjourney) en datasets (bv. ChatGPT Advanced Data Analysis) (3,4).

Het blijft nog onduidelijk hoe bovengenoemde AI-tools precies kunnen worden ingezet. Het is echter onmiskenbaar dat deze evolutie veel nieuwe mogelijkheden schept om processen binnen onderzoek (gedeeltelijk) te versterken en te automatiseren. Naar de toekomst toe kunnen onder onderzoekers verwachten dat de waaier aan beschikbare AI-tools (en daaraan verbonden plug-ins) enkel verder zal uitbreiden. Op termijn zullen deze AI-tools zullen ook in veelgebruikte softwareprogramma's en databanken geïntegreerd worden. De voordelen zijn duidelijk: De middelen en de tijd van onderzoekers zijn beperkt en het inzetten van deze tools zal er dus voor zorgen dat men meer kan bereiken met dezelfde middelen (m.a.w. de productiviteit verhoogt). Dit geldt met name voor activiteiten of processen die tijdrovend en repetitief zijn, en slechts in beperkte mate inzicht vereisen van onderzoekers. Het uitbesteden van deze taken aan AI-tools zal ervoor zorgen dat schaarse onderzoeksmiddelen beter benut kunnen worden, wat zowel onderzoekers als belastingbetalers ten goede komt.

Daarnaast kunnen deze tools ook de kwaliteit van wetenschappelijk werk verbeteren. Hier dienen AI-tools gezien te worden als *een hulpmiddel dat de eigen capaciteit van de onderzoekers om hoogstaand wetenschappelijk werk te creëren bevordert*. Zo kunnen AI-tools onderzoekers helpen om tot nieuwe ideeën te komen en zelfgeschreven teksten te verbeteren, vooralvoor niet-moedertaalsprekers van het Engels.

Desalniettemin werpen AI-tools ook vragen op over wat verantwoord gebruik hiervan precies inhoudt. De VUB wenst met deze beleidsrichtlijnen een strategische visie te ontwikkelen omtrent AI-tools en meer duidelijkheid te scheppen over wat onze universiteit gepast en ongepast gebruik van AI-tools binnen onderzoeksprocessen vindt. De VUB ondersteunt en moedigt het gebruik van AI-tools door onderzoekers aan. Eveneens wil onze universiteit onderzoekers wijzen op de beperkingen en risico's die het gebruik van AI-tools in onderzoeksprocessen met zich mee kan brengen.

1.2 Reikwijdte van richtlijnen

Deze richtlijnen zijn van toepassing op het **gebruik van generatieve AI-tools voor onderzoeksdoeleinden**. Wij richten ons hierbij voornamelijk op het gebruik van commercieel/publiek beschikbare AI-tools ongeacht hun scope (breed-inzetbaar of met specifieke doeleinde) en type input/output (bv. geschreven tekst, figuren, data). De richtlijnen zijn ook van toepassing op AI-tools die door middel van plug-ins beschikbaar zijn in softwareprogramma's. Het omvat zowel online- als offline-gebruik van commercieel/publiek beschikbare AI-tools (bv. binnen een afgesloten omgeving).

Deze richtlijnen zijn, strikt genomen, niet van toepassing op processen die gepaard gaan met het *ontwikkelen van AI-tools* als onderdeel van de academische activiteiten (alook het hergebruik van in-house AI-tools door VUB-collega's). Daarnaast ontzien de richtlijnen het *gebruik van AI-technieken voor data-analyse als data niet beschikbaar worden gemaakt doorheen een extern portaal*. AI-tools voor data-mining die formeel onderdeel uitmaken van de onderzoeksmethode- of design worden buiten beschouwing gelaten. Deze richtlijnen zijn ook niet van toepassing op onderwijsactiviteiten aan VUB, waaronder het organiseren van masterproeven. Voor meer informatie rond AI en onderwijs kunt u terecht op [Generatieve artificiële intelligentie en onderwijs](#). Sommige onderdelen van deze richtlijnen kunnen wel louter informatief blijven voor deze zaken.*

**De elementen die in acht genomen zijn in deze scope zullen verder bekeken worden met VUB-onderzoekers. De scope kan in de toekomst dus aangepast of gepreciseerd worden.*

2. Algemene principes

De VUB **staat open voor het inzetten van AI-tools** voor verschillende onderzoeksdoeleinden en legt hierin weinig specifieke beperkingen op.*

**Opgelet: Indien AI-tools ter beschikking gesteld worden, bijvoorbeeld via open source licenties, moet ook rekening gehouden worden met toepasbare wetgeving en in het bijzonder de AI Act van de Europese Unie. Deze verbiedt AI die een bedreiging vormt tegen mensen (bv. AI met onaanvaardbaar risico). Bij AI met hoog risico dienen wetgeving, en toepasselijke veiligheidsvoorschriften, gerespecteerd te worden in functie van veiligheid, gezondheid, welzijn en bescherming van de mensenrechten.*

Het inzetten van AI-tools voor onderzoeksdoeleinden moet echter **altijd in overeenstemming zijn met een reeks algemene principes**. Deze principes zijn:

- **Onderzoekers blijven volledig verantwoordelijk voor outputs van AI alsook de gepastheid van het onderzoeksproces waarbij AI ingezet wordt.** De verantwoordelijkheid kan dus niet afgeschoven worden op AI-tools. Dit betekent in de praktijk dat onderzoekers ter verantwoording kunnen geroepen worden over ongepast gebruik van AI-tools, zoals het niet naleven van de regels omtrent de wetenschappelijke integriteit (plagiaat, bronvermelding...). Deze regel is in overeenstemming met de positie van het *Committee of Publication Ethics (COPE)* dat *"authors are fully responsible for the content of their manuscript, even those parts produced by an AI-tool, and are thus liable for any breach of publication ethics"* (5). ALLEA's *2023 European Code of Conduct for Research Integrity* vereist dat onderzoekers het gebruik van AI-tools rapporteren conform de standaarden van het wetenschappelijke veld waartoe zij behoren (6).

- **De creatieve ideeën en eigen inbreng van onderzoekers blijven belangrijk.** De wisselwerking tussen onderzoeker en AI-tools kan nieuwe ideeën voortbrengen. Onderzoekers kunnen verder bouwen op eigen ideeën of kunnen nieuwe inzichten krijgen door AI-tools, zelfs indien het inzicht niet expliciet wordt voorgelegd door de AI-tool. Met betrekking tot ideevorming worden AI-tools best beschouwd als (veelzijdige) instrumenten die het verlengstuk kunnen zijn van de onderzoeker: Ze bevorderen de capaciteit van onderzoekers om ideeën te genereren.
- **Onderzoekers moeten een zo groot mogelijke mate van transparantie verschaffen** inzake het gebruik van AI-tools in hun wetenschappelijk activiteiten. Dit houdt in dat de manier waarop AI-tools gebruikt worden vaak in detail beschreven moet worden in de Materialen & Methoden-sectie en/of Supplementaire Data, inclusief prompts, output, naam en versie van de AI-tool. Onderzoekers moeten ook correct verwijzen naar AI-tools in wetenschappelijke publicaties (Zie sectie “Hoe te refereren naar AI?”)
- **Onderzoekers moeten de output van AI-tools controleren.** Zo moet men verschillende verificatiestappen inbouwen. Men moet steeds op zoek gaan naar de originele bronnen en waar van toepassing moet men licenties op het originele werk controleren. Originele ideeën moeten ook correct toegewezen worden aan de bedenker ervan. Training data die door AI-tools gebruikt worden kunnen ook auteursrechtelijk beschermd zijn of hergebruik kan beperkt zijn door licenties. In dat geval kan men de AI-output niet letterlijk overnemen. Daarnaast is het noodzakelijk om informatie (vnl. *facts and figures*) te *fact-checken*. Onderzoekers moeten nagaan of outputs voldoende diverse visies aan bod laten komen (bv. conflicterende resultaten van studies). Daarnaast moet men nagaan of wetenschappelijke stellingen voldoende gestaafd zijn met empirisch materiaal (indien relevant voor de onderzoeksdiscipline). Hoe “zelfstandiger” AI-tools werken, hoe groter de mate van controle die onderzoekers moeten uitvoeren. Hoe groter het risico op potentiële schade en de hoeveelheid schade die zou kunnen veroorzaakt worden, hoe groter de mate van controle die onderzoekers moeten uitvoeren.
- Onderzoekers moeten mogelijke **vakdiscipline-specifieke regels rond het verantwoord gebruik van AI-tools alsook richtlijnen van wetenschappelijke tijdschriften volgen.** Sommige tijdschriften verbieden het gebruik van AI-tools voor tekstproductie terwijl andere dit voorwaardelijk toelaten. Onderzoekers dienen zich te informeren over deze richtlijnen en deze te volgen als zij in een specifiek wetenschappelijk tijdschrift willen publiceren (Zie sectie “Refereren naar het gebruik van AI-tools in onderzoeksprocessen”).
- Onderzoekers moeten zich ervan bewust zijn dat **gebruik van AI-tools in sommige omstandigheden ook een nadeel kan zijn.** Zo kunnen AI-tools tekst met vrij algemene stellingen genereren. Het gebrek aan specificiteit in de tekst en/of het gebruik van AI-tools op zich (waarover onderzoekers transparant dienen te zijn) zouden ook kunnen meespelen als negatieve elementen in reviewprocessen.
- Onderzoekers dienen **de volgende gegevens niet in te geven in AI-tools:**
 - Persoonlijke data: data die (rechtstreeks of onrechtstreeks) kan leiden tot het identificeren van personen
 - Data die belangrijk zijn met oog op toekomstige valorisatie of data die beschermd (kunnen) worden door intellectueel eigendomsrecht. Zie ook “Tekstprompts en inputs van gebruikers bij AI-modellen” onder de sectie “Beperkingen en potentiële problemen”. Artikel 3 van het [Valorisatiereglement](#) stelt dat “Vóór openbaar-making moeten Valoriseerbare Onderzoeksresultaten aan de Knowledge & Technology Transfer Interface van de Instelling kenbaar worden gemaakt”. Contacteer techtransfer@vub.be bij vragen of voor ondersteuning.

- Data waarvan vrijgave ethisch problematisch kan zijn, zoals bij reële kans op ongepast gebruik of groepsschade (bv. stigmatisatie)
- Data waarbij bestaande contracten (bv. *non-disclosure agreements*) of regels vanuit fondsenverstrekkende organisaties beperkingen opleggen aan het delen ervan
- Data die onderhevig zijn aan contracten met derden (bv. bedrijven)
- Data die beschermd zijn door [auteursrecht of database rights](#) van een ander persoon, tenzij met diens toestemming.

Uitzonderingen: Voor bepaalde AI-tools en onder bepaalde voorwaarden, kan afgeweken worden van deze algemene richtlijn. Gelieve contact op te nemen met AI4research@vub.be indien u vragen heeft over mogelijke uitzonderingen.

Wij raden onderzoekers aan om de **privacy-instellingen van de AI-tools na te kijken** en deze aan te passen indien dit nodig wordt geacht. Zo kunt u *Chat History & Training bij ChatGPT* uitzetten.

- Het is **niet toegestaan om als onderzoeker AI-tools te gebruiken om de inhoudelijke aspecten van peer-reviews te schrijven**. Onderzoekers mogen wel:
 - De *lead author(s)* aanraden om zelf AI-tools te gebruiken om feedback te genereren.
 - AI-tools gebruiken om zelfgeschreven peer-reviews taalkundig beter te formuleren (bv. Grammarly) indien maatregelen getroffen worden om confidentialiteit te verzekeren.

3. Voor welke onderzoeksdoeleinden kunnen AI-tools ingezet worden?

3.1 Onderzoeksdoeleinden

Deze sectie dient om onderzoekers meer inzicht te geven in hoe AI-tools kunnen ingezet worden. In de volgende tabel worden mogelijke doeleinden beschreven. De opgelijste doeleinden zijn niet-exhaustief. Elk doeleinde wordt geïllustreerd met een korte casus waarbij ook de *Dos & Don'ts* bij het gebruik van AI-tools worden gegeven. Deze *Dos & Don'ts* zijn meer specifieke richtlijnen die in acht dienen genomen te worden bovenop de "Algemene principes". Als laatste wordt aan elk doeleinde een reeks AI-tools gelinkt. Deze lijst is eveneens niet exhaustief. Deze AI-tools kunnen toepasbaar zijn voor specifieke of verscheidene doeleinden. Er kunnen aanzienlijke verschillen in performantie van deze AI-tools, waarbij wij aanraden altijd de *state-of-the-art* AI-tools te gebruiken. In de praktijk betekent dit vaak het gebruik van de betalende versies (via individuele of groepslicenties). Als er verdere vragen zijn over deze casussen of gebruik van AI-tools voor andere onderzoeksdoeleinden, contacteer ons gerust op AI4research@vub.be.

Doeleinden	Type	Casus	Do's	Dont's	Voorbeelden van AI-tools
Het identificeren van nieuwe literatuur	Information retrieval	Bij het uitvoeren van een literatuurstudie wordt een AI-tool aanvullend ingezet om relevante literatuur in minder gekende vakgebieden te identificeren (bv. bij interdisciplinair onderzoek).	<ul style="list-style-type: none"> - Gepaste vermelding van AI-tool in "Methode"-sectie - Vrij gebruik indien niet binnen context van formele literatuurstudie 	<ul style="list-style-type: none"> - Gebruiken van AI-voorgestelde papers in literatuurstudie zonder inhoudelijke controle 	<ul style="list-style-type: none"> - Elicit (http://elicit.org) - Perplexity (https://www.perplexity.ai/) - Inciteful (http://inciteful.xyz) - Research Rabbit (http://researchrabbit.ai) - Connected Papers (http://connectedpapers.com) - Consensus (http://consensus.app) - Scite (http://scite.ai) - Iris (http://iris.ai)
Het samenvatten van artikelen	Tekstverwerking	De onderzoeker gebruikt een AI-tool die recente literatuur screent en samenvat. Vervolgens worden de samenvattingen in detail doorgenomen.	<ul style="list-style-type: none"> - Gebruik voor initiële screening alvorens de volledige teksten te lezen 	<ul style="list-style-type: none"> - Artikelen citeren/refereren op basis van samenvatting zonder deze zelf gelezen te hebben 	<ul style="list-style-type: none"> - Semantic Scholar (http://semanticscholar.org) - Humata (https://www.humata.ai/) - Iris (http://iris.ai) - ChatGPT (http://chat.openai.com) - Perplexity (https://www.perplexity.ai/)
Het selectief doorzoeken van inhoud van artikelen	Information retrieval	De onderzoeker gebruikt een AI-tool om "Recommendations" en "Perspectives" terug te vinden in wetenschappelijke publicaties om bij te blijven met de meest recente beleidsimplicaties van studies binnen zijn/haar domein.	<ul style="list-style-type: none"> - Screenen op specifieke onderdelen van artikelen of terminologie om vervolgens deze handmatig door te nemen 	<ul style="list-style-type: none"> - Specifieke onderdelen samenvatten overheen artikelen en deze individuele onderdelen niet grondig doorlezen - Gebrek aan referenties bij verwerken in eigen academisch werk 	<ul style="list-style-type: none"> - Iris (http://iris.ai) - Semantic Scholar (http://semanticscholar.org) - ChatGPT (http://chat.openai.com)
Het formuleren van wetenschappelijke stellingen	Tekstverwerking	De onderzoeker voert zelfbedachte wetenschappelijke	<ul style="list-style-type: none"> - Vermelden van gebruik AI-tool om stelling te herwerken, 		<ul style="list-style-type: none"> - ChatGPT (http://chat.openai.com) - Elicit (https://elicit.org/?via=topaitools)

		stellingen en hypothesen in in een AI-tool om feedback te krijgen. De AI-tool stelt mogelijke verbeteringen voor (bv. inhoudelijke verbeteringen of precisering)	inclusief prompts en outputs, indien testen van hypothese deel uitmaakt van wetenschappelijke publicatie		
Het bekritisieren van ingenomen stellingen of posities	Ideevorming	De onderzoeker gebruikt een AI-tool om een stelling te bekritisieren. Dit kan gebruikt worden om meer inzicht te krijgen in de tegenargumenten die opgeworpen kunnen worden.	<ul style="list-style-type: none"> - Zelftest of alle argumenten die de stelling ondermijnen kunnen weerlegd worden - Eventueel gebruik om studenten op te leiden door middel van debat 	<ul style="list-style-type: none"> - Het aanhalen van zeer persoonlijke verhalen uit het eigen leven om de eigen positie proberen te staven 	<ul style="list-style-type: none"> - DebateDevil (https://www.debate-devil.com/en) - Elicit (https://elicit.org/?via=topaitools) - Scite (https://scite.ai/?via=topaitools) - ChatGPT (http://chat.openai.com)
Een preliminaire screening uitvoeren van bewijslast voor en tegen wetenschappelijke stellingen	Information retrieval	De onderzoeker gebruikt een AI-tool om de resultaten van wetenschappelijke publicaties te screenen. Deze resultaten kunnen positieve of negatieve relaties zijn tussen bepaalde variabelen. De onderzoeker wil op deze manier inzicht bekomen in de bewijslast voor een bepaalde hypothese.	<ul style="list-style-type: none"> - Een initiële screening van bewijslast en tekortkomingen van studies, alvorens deze later grondig door te nemen - Aanvullende screening tijdens systematic review and meta-analysis van studies - Hou rekening met <i>publication bias</i> ten voordele van studies met positieve 	<ul style="list-style-type: none"> - Gebruik van output van AI-tool als "bewijslast" voor of tegen een stelling in maatschappelijk debat - Volledige afhankelijkheid van tool voor selecteren studies voor een <i>systematic review</i> 	<ul style="list-style-type: none"> - Consensus (https://consensus.app/search/) - Perplexity (https://www.perplexity.ai/) - System Beta (https://www.system.com/landing) - Scite (https://scite.ai/?via=topaitools) - ChatGPT (https://chat.openai.com/)

			resultaten tijdens deze oefening		
Brainstormen	Ideevorming	De onderzoeker legt eigen ideeën voor aan een AI-tool met de vraag of ze reeds behandeld worden in bestaande literatuur. De tool vindt enkele potentieel waardevolle publicaties die de onderzoeksvraag deels beantwoorden.	- Vermelden van interacties met AI-tool om ideeën te bekomen in publicaties, inclusief prompts en outputs	- Vermijd invoer van compleet nieuwe ideeën indien oog op latere valorisatie	- ChatGPT (http://chat.openai.com) - Elicit (https://elicit.org/?via=topaitools)
Genereren van nieuwe ideeën voor manuscripten	Ideevorming/producteren van nieuwe werken	De onderzoeker gebruikt een AI-tool en krijgt een output die een idee/argument bevat waarvan te vermoeden is dat ze nieuw is. De prompts die de onderzoeker heeft ingegeven hebben niet actief aangestuurd om dit idee te genereren.	- De oorsprong van de output wordt achterhaald via meerdere prompts. De onderzoeker zoekt tevens via andere bronnen. Indien nodig wordt de originele bron gerefereerd. De prompt van hoe het idee is bekomen wordt als Supplementaire Data toegevoegd in de Methode-sectie - Het idee wordt gebruikt als een aanzet/een startpunt	- Het kopiëren van ideeën en presenteren als de eigen intellectuele bijdrage - Geen controle of ideeën afkomstig zijn van andere bronnen en geen bronvermelding, terwijl dit noodzakelijk is	- ChatGPT (https://chat.openai.com/)
Het genereren van een initiële structuur	Tekstverwerking	De onderzoeker gebruikt een AI-tool om	- Het gebruik van AI-tools wordt vermeld in		- ChatGPT (https://chat.openai.com/)

voor het schrijven van subonderdelen van een wetenschappelijke publicatie		een initiële structuur te creëren voor de introductie-sectie. Nadien schrijft de onderzoeker de tekst zelf op basis van deze structuur.	de Methode-sectie van de publicatie		
Het genereren van een tekst door het ingeven van de initiële structuur van subonderdelen van een wetenschappelijke publicatie	Tekstverwerking/ Produceren van nieuwe werken	De onderzoeker geeft de titels van vijf paragrafen met de instructie om een zestal lijnen te schrijven per paragraaf. De AI-tool genereert een algemene tekst met bronvermelding.	<ul style="list-style-type: none"> - De onderzoeker dient alle geciteerde publicaties gelezen te hebben en te controleren of het citaat een correcte weergave is - De onderzoeker gaat na of de paragrafen voldoende diverse visies of bewijsmateriaal aanbrenen (bv. studies met contrasterende resultaten) - Het gebruik van de AI-tools wordt vermeld in de Methode-sectie van de publicatie en de prompt en output worden weergegeven in Supplementaire Data 	<ul style="list-style-type: none"> - Copy-pasten van tekst in eigen publicaties zonder controle op plagiaat, op mogelijke licenties die hergebruik beperken 	<ul style="list-style-type: none"> - ChatGPT (https://chat.openai.com/)
Het verbeteren van spelfouten, grammaticale fouten	Tekstverwerking	De onderzoeker gebruikt een AI-tool om de eigen tekst te	<ul style="list-style-type: none"> - Het gebruik van AI-tools wordt vermeld in 	<ul style="list-style-type: none"> - Het automatisch accepteren van alle mogelijke 	<ul style="list-style-type: none"> - Writefull (https://www.writefull.com/) - Quillbot (https://quillbot.com/) - ChatGPT (https://chat.openai.com/)

en structuur van paragrafen op basis van eigen tekstinput		verbeteren. De tool maakt voorstellen om de structuur van paragrafen te verbeteren en verbetert grammaticale en spelfouten. Daarnaast stelt de AI-tool voor om vage termen te vervangen.	de Methode-sectie van de publicatie	aanpassingen zonder na te kijken of de inhoud veranderd is	
Het analyseren van onderzoeksgegevens	Data-analyse en visualisatie	Diverse AI- methoden zijn reeds beschikbaar om onderzoeksgegevens te analyseren. Dit hoeft niet noodzakelijk te gaan over generatieve AI. Analyses kunnen eventueel gebeuren op basis van prompts binnen bestaande software.	<ul style="list-style-type: none"> - Uitvoeren van analyses met AI met voldoende inhoudelijke controle en in overeenstemming met ethische (bv. controle op bias in datasets) en wettelijke principes - Vermelden van AI-tool in Methode-sectie 	<ul style="list-style-type: none"> - Het laten uitvoeren van analyses enkel op basis van prompts zonder te controleren of de prompt correct geïnterpreteerd is - Het <i>post-hoc</i> formuleren van hypothesen op basis van een gegeven dataset zodat deze hypothesen altijd bevestigd worden 	<ul style="list-style-type: none"> - Code Interpreter in GPT-4 (https://chat.openai.com/?model=gpt-4-code-interpreter) - AutoML
Generatie van code	Produceren van nieuwe werken	De onderzoeker gebruikt een AI-tool om onderdelen van software te helpen schrijven.	<ul style="list-style-type: none"> - Vermelden van AI-tool in Methode-sectie 	<ul style="list-style-type: none"> - Volledige stukken code kopiëren zonder controle op uitvoerbaarheid van de code en of gewenst resultaat bereikt wordt 	<ul style="list-style-type: none"> - GitHub Copilot (https://github.com/features/copilot) - ChatWithGit plug-in in GPT-4 - AskTheCode plug-in in GPT-4 - Code Whisperer (https://aws.amazon.com/codewhisperer/) - Code LLama (https://ai.meta.com/llama/)

Generatie van data	Produceren van nieuwe werken	De onderzoeker gebruikt AI om synthetische data te genereren met behoud van statistische correlaties voor exploratief onderzoek.	<ul style="list-style-type: none"> - Publicaties vermelden dat synthetische data werden gebruikt en hoe deze geproduceerd werden in de Methode-sectie. - Gebruik van AI-tools is enkel aanvaardbaar als het duidelijk is wat met data-inputs gebeurt 	- Vervalsen van onderzoeksgegevens (fraude)	
Wetenschapscommunicatie	Schrijfhulp voor administratieve/aanvullende taken	De onderzoeker gebruikt een AI-tool om wetenschappelijke teksten samen te vatten voor het brede publiek. Deze berichten worden op sociale media en de pagina van de onderzoeksgroep gedeeld.	<ul style="list-style-type: none"> - Controleren van AI-samenvattingen voor ze uitgestuurd worden op inhoudelijke correctheid en stijl - Chatbot voor burgers die follow-up vragen kan beantwoorden 	- Overmatig verspreiden van wetenschapscommunicatie ("spammen")	- ChatGPT (https://chat.openai.com/)
Schrijven van Data Management Plan (DMP)	Schrijfhulp voor administratieve/aanvullende taken	De onderzoeker gebruikt een AI-tool om gerichte vragen te stellen over het invullen van het DMP (bv. bepaalde secties). De AI-tool geeft gerichte antwoorden over mogelijkheden die de onderzoeker heeft.	<ul style="list-style-type: none"> - Gebruik van AI om specifieke vragen beter te kunnen beantwoorden (bv. welke databanken zijn geschikt voor datatype X?) - Schrijven van DMPs door AI indien inhoudelijke controle en begrip van implicaties DMP 	- Vermijd het genereren van zeer generieke DMPs met weinig detail	- ChatGPT (https://chat.openai.com/)

			- Hou rekening met VUB-specifieke richtlijnen inzake Research Data Management		
Schrijven projectaanvragen	Schrijfhulp voor administratieve/aanvullende taken	De onderzoeker gebruikt een AI-tool om bepaalde onderdelen van een projectaanvraag te schrijven.	<ul style="list-style-type: none"> - De onderzoeker die de aanvraag indient heeft de eindverantwoordelijkheid over de wetenschappelijke inhoud en moet zeker zijn dat het over een origineel voorstel gaat - Bekend maken van het gebruik van AI-tool voor het genereren van de tekst - Controleer of AI-gegenereerde tekst voldoende precies is - Controleer of de projectaanvraag voldoet aan de voorwaarden van de <i>project call</i> 	- Volledig uitbesteden van het schrijven van projectaanvraag aan AI-tools	<ul style="list-style-type: none"> - Grantable (https://grantable.co/) - Granted AI (https://grantedai.com/) - ChatGPT (https://chat.openai.com/)

3.2 Beschikbare AI-tools voor onderzoeksdoeleinden

U kunt op de onze [“Praktische informatie voor het gebruiken van AI-tools”](#) pagina meer gedetailleerde informatie vinden over hoe u aan de slag kunt met AI-tools.

3.3 Trainingen omtrent het inzetten van AI-tools voor deze doeleinden

Als u interesse heeft in het mee helpen organiseren van lessen rond het gebruik van AI-tools voor specifieke doeleinden, gelieve ons te contacteren op AI4research@vub.be. Als u reeds lessen organiseert over deze thematiek, gelieve ons ook te contacteren daarover. Op deze manier krijgen wij inzicht in het reeds bestaande trainingsmateriaal en zo kunnen wij verder bouwen op uw werk en u ook het werk van anderen.

4. Voordelen van het gebruik van AI in onderzoek

Het inzetten van AI-tools voor verschillende onderzoeksdoeleinden meerdere voordelen hebben. We kunnen de voordelen (en gevolgen) voor onderzoekers opdelen in verschillende niveaus:

- **De productiviteit van onderzoekers verhogen.** We beschouwen productiviteit hier als de hoeveelheid (wenselijke) onderzoeksuitkomsten die men voortbrengt gegeven een vaste hoeveelheid middelen. Wetenschappelijke studies geven aan dat het gebruik van GenAI-tools de productiviteit van werknemers aanzienlijk kan doen toenemen (7–12). Met betrekking tot onderzoek voorzien wij dat tijdsintensieve (en/of repetitieve) taken die geen substantiële intellectuele bijdrage vereisen van onderzoekers aan AI-tools deels uitbesteed kunnen worden. Daarnaast kunnen AI-tools ook de eigen competenties van onderzoekers versterken (bv. om ideeën te genereren). Tevens kunnen AI-tools gebruikt worden om complexe patronen te identificeren in grote datasets, wat zinvol kan zijn in de context van exploratief onderzoek. Tezamen kunnen deze de productiviteit van onderzoekers verhogen.
- **De kwaliteit van het werk van onderzoekers verhogen.** Observaties uit andere sectoren wijzen uit dat het gebruik van GenAI-tools kan helpen om bepaalde taken op meer kwalitatieve wijze uit te voeren (7–9). Evenzeer verwachten wij dat onderzoekers hun vooropgestelde doelstellingen in het kader van onderzoek op een meer kwalitatieve wijze zullen kunnen volbrengen. Zo zou een literatuurstudie meer allesomvattend kunnen gemaakt worden. AI-tools kunnen ook de inzichten, creativiteit en zelfreflectie van onderzoekers bevorderen, wat tot meer kwalitatief werk kan leiden.
- **Reflectie op het eigen werk bevorderen.** Net zoals uw collega-wetenschappers kunnen AI-tools een nieuwe blik werpen op de eigen overtuigingen, argumenten, ideeën en werken. Allerlei factoren kunnen echter het proces van feedback geven en krijgen bemoeilijken. AI-tools kunnen dus het voordeel hebben snel een “externe en onpersoonlijke” blik te kunnen werpen op het eigen werk.
- **Tot een hogere wendbaarheid leiden bij onderzoekers.** Binnen de hedendaagse onderzoeksomgeving worden onderzoekers vaak geconfronteerd met uitdagingen die nieuwe vaardigheden vereisen. Zo moeten onderzoekers projectaanvragen leren schrijven, DMPs opstellen en omgaan met dataproductiewetgeving en ethische richtlijnen. Een aanzienlijke barrière hierbij is dat het leerproces heel wat tijd kan vergen. Wij verwachten dat de beschikbaarheid van AI-tools de tijd die het leerproces vergt bij

“nieuwe” taken kan verminderen. Ze kunnen dus de wendbaarheid van de onderzoeker verhogen om met deze uitdagingen om te gaan.

- **Voor een verschuiving zorgen in dagdagelijks werk van onderzoekers.** Eén studie toont aan dat AI, net zoals andere technologieën, verschuivingen kan veroorzaken in het takenpakket (9). Voor onderzoekers bestaat dit takenpakket uit coördinatie, kernonderzoekstaken, onderwijstaken en administratieve taken. AI-tools zijn inzetbaar voor al deze taken. Wij voorzien echter dat ze, op dit moment, in grotere mate inzetbaar zullen zijn voor repetitieve taken van niet-intellectuele aard (bv. administratief werk). Voor complexere taken kunnen ze ook ingezet worden, voornamelijk in samenwerking met mensen. Als we ervan uitgaan dat repetitieve taken minder tijd zullen vragen (en ze niet toenemen in aantal), dan zullen de complexere taken in de toekomst een groter aandeel van het dagdagelijks werk uitmaken (bv. kernonderzoekstaken).
- **Voor bepaalde vaardigheden zou de kloof tussen onderzoekers kunnen verminderen.** Sommige studies wijzen er op dat minder vaardige werknemers meer voordeel halen uit het gebruik van AI-tools dan vaardige werknemers (7–11). Men kan zich bijvoorbeeld voorstellen dat een bekwaam onderzoeker die niet vaardig is in het neerschrijven van zijn/haar bevindingen, een AI-tool kan gebruiken om dit gebrek te compenseren. Met andere woorden, AI-tools zouden een nivellerend effect kunnen hebben op de kloof tussen onderzoekers voor bepaalde vaardigheden.
- **De kloof tussen burger en wetenschapper verkleinen.** AI-tools kunnen ingezet worden om wetenschappelijke teksten samen te vatten en uit te leggen op niveau van de modale burger. Dit helpt wetenschappers hun onderzoek op voldoende begrijpelijke wijze uit te leggen. In feite laten AI-tools op deze manier toe een belangrijke cognitieve beperking van mensen te omzeilen (de zogenaamde “*curse of knowledge*”) (13).
- **Interdisciplinaire dialoog en onderzoek bevorderen.** AI-tools faciliteren het exploreren en begrijpen van overvloedige informatie, zoals het diverse academisch werk overheen vakgebieden. Ze kunnen ook verbanden leggen overheen deze diverse literatuur die niet meteen zichtbaar zijn voor onderzoekers (In feite is dit een vorm van creativiteit) (14). Hier kunnen AI-tools ook de rol innemen van *match-makers* om onderzoekers met elkaar in contact te brengen, wat interdisciplinariteit kan bevorderen (15).

Op systeemniveau kunnen we anticiperen dat deze voordelen voor individuele onderzoekers zich vertalen **in een grotere snelheid van wetenschappelijke vooruitgang en totale maatschappelijke impact van onderzoek** (gegeven eenzelfde hoeveelheid middelen). Dit betekent dat de voorziene middelen die aangewend worden in wetenschappelijke onderzoek in het algemeen beter besteed zouden worden.

5. Beperkingen en potentiële problemen bij het gebruik van AI-tools in onderzoek

Zoals hierboven beschreven kan het gebruik van AI-tools duidelijke voordelen hebben. Onderzoekers dienen zich er echter bewust van te zijn dat AI-tools ook een aantal ernstige tekortkomingen en beperkingen kan hebben. Hieronder worden deze per categorie toegelicht.

Trainingsdata en leerproces voor AI-modellen

- **Data bias kan zich voordoen in trainingsdata**
 - a. **Trainingsdata kunnen reeds beïnvloed zijn door bestaande structurele ongelijkheden tussen groepen.** Verschillen tussen groepen, zoals tussen

verschillende genders, etniciteiten of sociaaleconomische klassen kunnen resulteren in een over- of ondervertegenwoordiging in *real-world data* (bv. personeelsbestanden, gezondheidszorggegevens). Structurele ongelijkheden inzake groeps participatie of -uitkomsten kunnen sterk afhankelijk zijn van de databron en context. Ze zijn niet beperkt tot de bovengenoemde categorieën.

- b. **Trainingsdata kunnen discriminerend taalgebruik of racistische ondertoon, foutieve informatie en argumentatie (m.a.w. *logical fallacies*) bevatten.** Dit geldt voornamelijk wanneer geschreven tekst als trainingsdata voor Large Language Models (LLMs) wordt gebruikt.
 - c. **Trainingsdata kunnen andere vormen van bias vertonen** afhankelijk van de soort bron. Zo kan *publication bias* ten voordele van positieve resultaten nog steeds een impact hebben op AI-tools waarmee een initiële screening wordt uitgevoerd voor bewijslast voor wetenschappelijke stellingen.
- **Trainingsdata voor AI-tools kunnen gedateerd zijn.** Het is mogelijk dat AI-tools, waaronder LLMs, geen toegang hebben tot de meest recente gebeurtenissen of publicaties. Zo is de *knowledge cut-off* van GPT-3.5 september 2021 omdat bronnen van na deze datum niet geïncorporeerd zijn in de training data. Sinds 27 september hebben GPT4 (Plus en Enterprise) deze beperking niet meer en GPT-3.5 zal in de toekomst deze ook niet meer hebben.
 - **Trainingsdata voor AI-tools kunnen reeds (deels) beïnvloed zijn door LLMs.** Als het gebruik van LLMs wijdverbreid is zal dit een invloed hebben op de inhoud van nieuwe teksten. Deze nieuwe teksten kunnen in theorie opnieuw gebruikt worden als “trainingsdata” voor LLMs. Dit is een vorm van “circulariteit” waarbij het risico bestaat dat, als de mate van menselijke interventie te gering is, machines de fouten van andere machines gaan imiteren (16).
 - **Verschillende onderdelen van de trainingsdata zijn niet even informatief.** De trainingsdata voor LLMs kan reeksen informatie bevatten die, onder normale omstandigheden, verschillende niveaus van geloofwaardigheid zouden hebben. Zo zouden resultaten uit één enkele empirische studie minder geloofwaardig geacht worden dan een meta-analyse van alle beschikbare empirische studies. Op dezelfde manier worden blogs, krantenartikelen en wetenschappelijke publicaties niet noodzakelijk dezelfde mate van geloofwaardigheid toegeschreven. LLMs zoals ChatGPT kunnen, in principe, rekening houden met deze verschillen, bijvoorbeeld doorheen *fine-tuning* of *reinforcement learning* from human feedback (RLHF) stappen tijdens ontwikkeling. Op dit moment kan men echter nog steeds links naar minder “betrouwbare bronnen”, zoals blogs, krijgen bij sommige LLMs als men geen gebruik maakt plug-ins of gespecialiseerde AI-tools.
 - **Trainingsdata kunnen persoonlijke gegevens of auteursrechtelijk beschermd materiaal bevatten die onrechtmatig vrijgegeven zijn.** Persoonlijke gegevens die onrechtmatig vrijgegeven zijn kunnen vervat zitten in trainingsdata. Het gebruik van deze data kan zo een (verdere) inbreuk vormen op de onderzoeksintegriteit en/of de AVG/GDPR.
 - **Trainingsdata die gebruikt worden door AI-tools zouden deels gereconstrueerd kunnen worden onder bepaalde omstandigheden.** Verschillende vormen van *adversarial privacy attacks*, waarbij doelgerichte aanvallen gemaakt worden om de werking van AI-modellen te verstoren of informatie te ontvreemden, zouden een risico kunnen vormen. De theoretische mogelijkheid hiertoe is gedocumenteerd bij AI-tools die persoonlijke data gebruiken voor data-analyse (en dus niet noodzakelijk voor GenAI). In theorie zouden externen, onder specifieke voorwaarden de aanwezigheid van de gegevens van een individu in trainingsdata (“Membership Inference Attacks”) na kunnen gaan of

zouden onderdelen van trainingsdata gereconstrueerd kunnen worden (“Model Inversion Attacks”) (17–20).

Tekstprompts en inputs van gebruikers bij AI-modellen

- **Het invoeren van persoonlijke gegevens in AI-tools kan in strijd zijn met de Algemene Verordening Gegevensbescherming (AVG/GDPR) van de Europese Unie.** Persoonlijke gegevens die ingegeven worden in AI-tools zouden opgeslagen kunnen worden buiten de Europese Unie en derden zouden toegang kunnen krijgen tot deze gegevens. Gelieve contact op te nemen met AI4research@vub.be alvorens persoonsgegevens in te voeren in een AI-tool.
- **Het invoeren van persoonlijke gegevens in AI-tools kan in strijd zijn met de geïnformeerde toestemming die bekomen is van de onderzoeksdeelnemer.** Ethische richtlijnen rond deelname aan onderzoek kunnen vereisen dat toestemming slechts geïnformeerd als de risico’s op voorhand bekend gemaakt zijn aan onderzoeksdeelnemers. Als het gebruik van AI-tools risico’s voortbrengt die niet bekend gemaakt zijn, kan dit eventueel betekenen dat de toestemming niet voldoende “geïnformeerd” was. Daarnaast kan het in de geïnformeerde toestemming opgenomen zijn dat gegevens niet zullen vrijgegeven worden aan commerciële bedrijven. Gelieve contact op te nemen met AI4research@vub.be alvorens persoonsgegevens in te voeren in een AI-tool.
- **Het gebruik van AI-tools om feedback te geven op volledig nieuwe onderzoeksideeën zou kunnen gelijkgesteld worden met de “openbaring” van deze ideeën voor specifieke doeleinden.** Zo bekijken werknemers van OpenAI gesprekken van gebruikers met hun AI-tool om het algoritme te verbeteren (21). In dat geval kan het zijn dat onderzoekers de mogelijkheid verliezen om uitvindingen te patenteren en valoriseren. Wees dus voorzichtig met wat u ingeeft als u (op lange termijn) een patent beoogt.
- **Uw inputs in AI-tools kunnen een invloed hebben op de outputs die andere krijgen.** In tegenstelling tot de betalende GPT-4 stelt OpenAI dat de gratis versie van GPT-3.5 “[...] may use content such as prompts, responses, uploaded images, and generated images to improve our services. (...) ChatGPT, for instance, improves by further training on the conversations people have with it, unless you choose to disable training” (22,23). Dat betekent dat uw interacties met ChatGPT, waaronder uw prompts, kunnen gebruikt worden om de performantie van het model te verbeteren. Het is onduidelijk of het hier gaat over basistraining, *fine-tuning* of *reinforcement learning* from human feedback (RLHF). Als het hier gaat over basistraining gaat, zouden uw prompts/inputs terug tevoorschijn kunnen komen als outputs voor andere gebruikers. **Er is hiervoor echter geen sluitend bewijs op dit moment.** Als het enkel gaat over verdere RLHF, dan zullen uw prompts niet terugkomen bij anderen.
- **Uw inputs zelf (bv. kwantitatieve data) kunnen ook van lage kwaliteit zijn.** Er kunnen bijvoorbeeld tekortkomingen geweest zijn in het uitvoeren van de steekproef waardoor de bekomen data minder informatief zijn om uitspraken te doen over de populatie (m.a.w. methodologische tekortkomingen). Tevens kunnen uw data onvoldoende “metadata” (m.a.w. informatie over experimentele omstandigheden) bevatten om op gepaste wijze door AI-modellen verwerkt te worden. Als laatste kunnen AI-modellen op dit moment ook niet altijd gepast omgaan met metadata. Wij raden onderzoekers aan om het idee “Garbage in, garbage out” in het achterhoofd te houden.

Output van AI-modellen

- **Outputs van AI-tools kunnen bias bevatten.** De vormen van bias zoals beschreven onder “Trainingsdata voor AI-modellen” kunnen invloed hebben op de outputs van AI-tools. Op deze manier kunnen Large Language Models (LLMs) seksistische, racistische of andere discriminerende ideeën reproduceren (24–26). Daarnaast kan het fine-tunen van LLMs voor specifieke taken, waarbij mensen outputs van AI-tools moeten beoordelen, bias introduceren (27).
- **Gebrek aan inzicht in hoe AI-tools precies werken.** Het is vaak niet te achterhalen hoe algoritmes precies tot een bepaalde output komen. Zelfs als het duidelijk is dat een bepaald onderdeel van trainingsdata een grotere invloed heeft gehad, dan ontbreekt nog steeds een verklaring die aannemelijk is voor mensen. Als men bijvoorbeeld observeert dat een bepaalde regio op beelden een grotere invloed heeft, dan ontbreekt dus een causale verklaring voor *waarom* dit zo is. Een gebrek aan inzicht in de exacte werking van AI-tools kan ernstige gevolgen hebben. Dit is voornamelijk het geval als onbedoelde correlaties met betrekking tot beschermde karakteristieken (bv. gender, geaardheid, religie...) of *proxies* hiervoor worden vastgesteld (28).
- **Waarheidsgehalte.** De outputs van LLMs kunnen gepresenteerd worden als overtuigend maar kan inaccuraat, volledig fout of gewoon irrelevant zijn (29). Eén enkele studie toonde al aan dat, onder sommige omstandigheden, gebruik van hoogkwalitatieve AI de kwaliteit van het werk kan schaden als gebruikers te veel afhankelijk worden (vnl. voor taken waar AI-tools nog niet geschikt voor zijn). Ze vallen als het ware in slaap achter het stuurwiel (30). De zelfzekerheid waarmee AI-tools stellingen poneren kan dus misleidend zijn. Onderzoekers moeten dus zeer kritisch omgaan met tekstoutputs van AI-tools.
- **Beperkte mate van reproduceerbaarheid.** Generatieve AI-tools creëren outputs die niet altijd exact gerepliceerd kunnen worden. Zo kan eenzelfde prompt leiden tot verschillende tekstfragmenten. Dit betekent ook dat volledige reproduceerbaarheid van onderzoeksprocessen niet altijd mogelijk is, zoals bij het ingeven van dezelfde prompt bij een tool die literatuur aanraadt. Daarnaast kunnen algoritmes, voornamelijk dynamischlerende AI-tools, op zichzelf ook onstabiel zijn in de zin dat inputs niet op dezelfde manier verwerken doorheen de tijd, wat het moeilijker maakt om (reproduceerbare) workflows uit te werken.
- **Outputs kunnen geplagieerde elementen bevatten.** Vormelementen of ideeën uit bestaande teksten of beelden kunnen integraal overgenomen worden en zo een inbreuk zijn op onderzoeksintegriteit. Niet alle AI-tools zijn in staat om kwalitatieve referenties te genereren naar de gebruikte bronnen. Correct citeren kan in de praktijk gepaard gaan met enkele complexiteiten, zoals de aanwezigheid van bronnen binnen andere bronnen (bv. citaties binnen reviewpaper) en het toewijzen van een verkeerd geïnterpreteerde tekst aan een bepaalde bron.
- **AI-tools kunnen nieuwe data produceren die conflicteren met bestaande privacyregels, zoals de AVG/GDPR.** Er kunnen ook (algemene) problemen zijn omtrent de problematiek van *data linkage* en/of afgeleide informatie waarbij individueel geanonimiseerde data tezamen aanleiding kunnen geven tot nieuwe privacygevoelige data. Zo kunnen AI-tools, in principe, nieuwe informatie creëren die privacygevoelig is of die als proxy kunnen gebruikt worden voor beschermde karakteristieken (bv. combinatie inkomensklasse en postcode). Dit wil ook zeggen dat de originele datasets niet werkelijk “anoniem” (m.a.w. irreversibel gedeïdentificeerd) waren, wat ook de fundamentele vraag doet rijzen of “anonimiteit” wel degelijk kan bestaan.

6. Refereren naar het gebruik van AI-tools in onderzoeksprocessen

4.1 Stand van zaken

Het wijdverbreid gebruik van AI-tools in onderzoeksprocessen creëert nieuwe uitdagingen voor de attributie van wetenschappelijke bijdragen. Op dit moment is er nog geen consensus of AI-tools toegestaan zijn en op welke manier het gebruik ervan transparant moet gemaakt worden. Wetenschappelijke tijdschriften en uitgevers hebben hun eigen posities omtrent gebruik van en refereren naar AI-tools. Merk op dat sommige uitgevers restricties opleggen in het gebruik van AI-tools. **Wij raden onderzoekers aan om zich te informeren hierover.** Dit kan gebeuren alvorens het onderzoek uit te voeren indien het gebruik van AI-tools formeel behoort tot de methode of onderzoeksopzet. Dit kan ook gebeuren vooraleer het manuscript te schrijven indien het gebruik van AI-tools ingezet wordt voor het genereren van tekst of figuren. De manier waarop gerefereerd dient te worden zal wellicht ook nog verder evolueren. Een regelmatige check is dan ook aan te raden.

4.2 Hoe te refereren naar AI?

De VUB raadt onderzoekers aan om **allereerst de richtlijnen te bekijken van het tijdschrift of de uitgever** waar ze willen publiceren! Volg deze richtlijnen indien ze beschikbaar zijn.

Indien het tijdschrift of de uitgever geen expliciete richtlijnen heeft rond het gebruik van AI-tools, kan volgende algemene richtlijn gevolgd worden:

Onderzoekers moeten het gebruik van AI-tools beschrijven in de **Methode-sectie**, inclusief de technische specificaties ervan. Deze specificaties omvatten maar beperking zich niet tot de volledige naam, versie, link naar een repository, gebruikte parameters en configuraties. Als de AI-tool in detail beschreven werd in wetenschappelijke publicaties, dan kunnen deze papers geciteerd worden. De prompts die gebruikt werden en outputs die verkregen werden, worden in de Methode-sectie beschreven of als Supplementaire Data toegevoegd. Als een tekstoutput van AI-tools tekstfragmenten of ideeën bevat die van anderen afkomstig zijn, dient men in eerste instantie de originele bronnen te **citeren**.

In **twee specifieke gevallen** dient er echter anders gerefereerd te worden:

- a. *AI-tools zijn het object van de studie (bv. bestuderen van bias in ChatGPT outputs) en/of de output van AI-tools wordt letterlijk geciteerd*

De onderzoekers dienen alle prompts accuraat weer te geven en outputs van AI *verbatim* te citeren. Dit kan bijvoorbeeld door de naam van de AI-tool op het einde van de quote tussen haakjes zetten. Dit behoeft dus, in principe, geen referentie naar AI-tools als "auteur" van de quotes in de Referenties-sectie. Indien het tijdschrift een referentiestijl aanhoudt die dit wel toelaat, dienen onderzoekers deze richtlijnen te volgen.

- b. *AI-tools creëren ogenschijnlijk nieuwe inzichten of argumenten zonder aanzienlijke intellectuele bijdrage van onderzoekers*

De onderzoekers dienen eerst en vooral de oorsprong van de output na te gaan om plagiaat te vermijden. Indien er sprake is van een bestaand idee moet de bron geciteerd worden. Indien de oorsprong van het idee niet kan teruggevonden worden na extensief zoekwerk, mogen onderzoekers het idee quoteren **maar moet het duidelijk zijn dat het niet van de onderzoeker**

zelf komt. De acties die de onderzoekers ondernomen hebben om de oorsprong te achterhalen moeten beschreven worden in de Methode-sectie. De prompts zelf moeten toegevoegd worden als Supplementaire Data.

7. Referentielijst

1. OECD. Recommendation of the Council on OECD Legal Instruments Artificial Intelligence. 2022. Available from: <https://legalinstruments.oecd.org/public/doc/648/1df51f15-53fc-43ef-9f13-ee9f957076bc.htm>
2. European Parliament. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. 2021. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>
3. Midjourney [Internet]. [cited 28 June 2023]. Midjourney. Available from: <https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F>
4. DALL·E 2 [Internet]. [cited 28 June 2023]. Available from: <https://openai.com/dall-e-2>
5. COPE: Committee on Publication Ethics [Internet]. [cited 29 June 2023]. Authorship and AI tools. Available from: <https://publicationethics.org/cope-position-statements/ai-author>
6. ALLEA. The European Code of Conduct for Research Integrity REVISED EDITION 2023. Available from: <https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf>
7. Dell'Acqua F, McFowland E, Mollick ER, Lifshitz-Assaf H, Kellogg K, Rajendran S, et al. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. SSRN Electron J. 2023 [cited 22 September 2023]. Available from: <https://www.ssrn.com/abstract=4573321>
8. Brynjolfsson E, Li D, Raymond L. Generative AI at Work. Cambridge, MA: National Bureau of Economic Research; 2023 Apr [cited 26 September 2023]. Report No.: w31161. Available from: <http://www.nber.org/papers/w31161.pdf>
9. Noy S, Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence. Science. 2023 Jul 14;381(6654):187-192. DOI: 10.1126/science.adh2586
10. Peng S, Kalliamvakou E, Cihon P, Demirer M. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. 2023 [cited 26 September 2023]. Available from: <https://arxiv.org/abs/2302.06590>
11. Kanazawa K, Kawaguchi D, Shigeoka H, Watanabe Y. AI, Skill, and Productivity: The Case of Taxi Drivers [Internet]. Cambridge, MA: National Bureau of Economic Research; 2022 Oct [cited 6 October 2023]. Report No.: w30612. Available from: <http://www.nber.org/papers/w30612.pdf>
12. Campero A, Vaccaro M, Song J, Wen H, Almaatouq A, Malone TW. A Test for Evaluating Performance in Human-Computer Systems. arXiv; 2022 [cited 27 September 2023]. Available from: <http://arxiv.org/abs/2206.12390>
13. The Curse of Knowledge: A Difficulty in Understanding Less-Informed Perspectives – Effectiviology [Internet]. [cited 27 September 2023]. Available from: <https://effectiviology.com/curse-of-knowledge/>
14. Mollick E. Automating creativity [Internet]. 2023 [cited 29 September 2023]. Available from: <https://www.oneusefulthing.org/p/automating-creativity>
15. Chubb J, Cowling P, Reed D. Speeding up to keep up: exploring the use of AI in the research process. AI & Soc. 2022;37:1439–1457. DOI: 10.1007/s00146-021-01259-0
16. Kenton Z, Everitt T, Weidinger L, Gabriel I, Mikulik V, Irving G. Alignment of Language Agents. 2021 [cited 6 October 2023]. Available from: <https://arxiv.org/abs/2103.14659>
17. Rigaki M, Garcia S. A Survey of Privacy Attacks in Machine Learning. 2020 [cited 6 October 2023]. Available from: <https://arxiv.org/abs/2007.07646>
18. Zhang G, Liu B, Zhu T, Zhou A, Zhou W. Visual privacy attacks and defenses in deep learning: a survey. Artif Intell Rev. August 2022;55(6):4347-401. DOI: 10.1007/s10462-021-10123-y
19. Balle B, Cherubin G, Hayes J. Reconstructing Training Data with Informed Adversaries [Internet]. arXiv; 2022 [cited 26 June 2023]. Available from: <http://arxiv.org/abs/2201.04845>
20. Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, et al. Extracting Training Data from Large Language Models. 2020 [cited 6 October 2023]. Available from: <https://arxiv.org/abs/2012.07805>

21. What is ChatGPT? | OpenAI Help Center [Internet]. [cited 10 July 2023]. Available from: <https://help.openai.com/en/articles/6783457-what-is-chatgpt>
22. Data usage for consumer services FAQ | OpenAI Help Center [Internet]. [cited 29 September 2023]. Available from: <https://help.openai.com/en/articles/7039943-data-usage-for-consumer-services-faq>
23. How your data is used to improve model performance | OpenAI Help Center [Internet]. [cited 29 September 2023]. Available from: <https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance>
24. ChatGPT and Large Language Model bias. Available from: <https://www.cbsnews.com/news/chatgpt-large-language-model-bias-60-minutes-2023-03-05/>
25. Omiye JA, Lester J, Spichak S, Rotemberg V, Daneshjou R. Beyond the hype: large language models propagate race-based medicine [Internet]. Health Informatics; 2023 Jul [cited 6 October 2023]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2023.07>
26. Deshpande A, Murahari V, Rajpurohit T, Kalyan A, Narasimhan K. Toxicity in ChatGPT: Analyzing Persona-assigned Language Models. 2023 [cited 6 October 2023]. Available from: <https://arxiv.org/abs/2304.05335>
27. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, Mishkin P, et al. Training language models to follow instructions with human feedback. arXiv; 2022 [cited 27 September 2023]. Available from: <https://doi.org/10.48550/arXiv.2203.02155>
28. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 25 October 2019;366(6464):447-453. DOI: 10.1126/science.aax2342
29. Hosseini M, Rasmussen LM, Resnik DB. Using AI to write scholarly publications. Account Res. 25 January 2023;1-9. <https://doi.org/10.1080/08989621.2023.2168535>
30. Fabrizio Dell'Acqua. Falling Asleep at the Wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters. Available from: <https://static1.squarespace.com/static/604b23e38c22a96e9c78879e/t/62d5d9448d061f7327e8a7e7/1658181956291/Falling+Asleep+at+the+Wheel+-+Fabrizio+DellAcqua.pdf>