



Vägledning för etisk användning av generativ AI i forskning

Inledning

Denna vägledning är avsedd att stödja alla forskande medarbetare i att fatta välgrundade beslut kring användningen av generativ AI i forskningsprocessen, inklusive om, när och hur de väljer att använda generativ AI. Vägledningen fokuserar på principer för etisk och ansvarsfull användning av AI, samtidigt som forskare måste ha en grundläggande kunskap om vad generativ AI är, hur den fungerar och varför den kan vara problematisk ur ett forskningsetiskt perspektiv. Då generativ AI förändras väldigt snabbt och tillämpningen av den ser olika ut inom olika forskningsområden, behöver dessa principer vara allmänt formulerade.

EU-kommissionen har tagit fram riktlinjer om ansvarsfull användning av generativ AI inom forskning ("Living guidelines on the responsible use of generative AI in research"¹). Dessa bygger bland annat på ALLEAs "The European Code of Conduct for Research Integrity"² och tolkar utifrån denna hur AI används på ett etiskt ansvarsfullt sätt inom forskning. Riktlinjerna är en viktig källa till denna vägledning och har utformats med utgångspunkt i nuvarande tekniska utvecklingsnivå och omgivande policyramverk. Dessa är båda dynamiska i sin natur, vilket innebär att riktlinjerna, och som en följd även vägledningen, kommer behöva anpassas, utvecklas och uppdateras regelbundet. Etikskottet, i samråd med forskningsnämnden, ansvarar för årliga översyner av vägledningen.

Ansvar

Den enskilde forskaren ansvarar alltid för forskningens kvalitet, korrekthet och integritet, även när generativa AI-verktyg används. Det förutsätter ett kritiskt förhållningssätt till AI-genererat material och en medvetenhet om verktygens begränsningar, såsom systematiska snedvridningar (bias), påhittat innehåll (hallucinationer), otillförlitlighet i form av bristande reproducerbarhet (varierande output vid identisk input), låg specificitet i genererad output och sakfel. AI får aldrig fungera som den drivande aktören i forskningsprocessen, utan måste begränsas till en stödjande funktion. Generativ AI kan vara ett bra hjälpmedel i olika delar av processen, men AI-verktyg saknar förmåga att ersätta forskarens mänskliga omdöme, ansvarstagande och kontextuella förståelse i komplexa sammanhang.

Universitetet har ett ansvar för att alla forskande medarbetare erbjuds utbildning inom etisk AI-användning. När det gäller doktorander ansvarar fakulteterna för att de får nödvändig utbildning om etisk AI-användning i forskning.Handledare verkar för att doktorander utvecklar förmågan att etiskt reflektera över AI-användning, både inom forskning generellt och inom sitt specifika forskningsområde.

Öppenhet och transparens

När generativa AI-verktyg har använts i forskningsprocessen ska detta redovisas öppet, på ett sätt som är vedertaget inom det aktuella forskningsområdet. Mer omfattande användning, så som exempelvis vid dataanalys, litteraturgenomgångar, skrivande, identifiering av kunskapsluckor, målformulering, eller hypotesutveckling ska beskrivas. I relevanta fall bör även prompter och genererat material göras tillgängliga, i linje med principerna för öppen vetenskap. Vilka delar av forskningsprocessen där stöd av AI anses lämpligt eller inte kan dock skilja sig åt mellan olika forskningsområden.

Forskare ska beakta att generativa AI-verktyg kan producera olika output från samma input och sträva efter att möjliggöra granskning och reproducerbarhet av sina resultat och slutsatser. De ska även redovisa verktygens begränsningar, inklusive potentiella bias samt eventuella åtgärder för att minska dessa.

Känslig information och känsliga sammanhang

Vid användning av generativ AI är det viktigt att vara medveten om att det kan finnas en risk för att inskickat eller genererat material kan komma att användas vidare, exempelvis vid utveckling och träning av AI-modeller. Opublicerat, konfidentiellt eller på annat sätt skyddsvärt material ska därför inte delas med externa AI-tjänster utan tydliga juridiska garantier för att informationen inte lagras eller återanvänds. Vissa AI-verktyg som tillhandahålls av Göteborgs universitet använder inte information som laddas upp för att träna AI-modeller.³ Detta innebär att vissa typer av opublicerat material (som inte innehåller sådan information som beskrivs nedan) kan användas i dessa verktyg. Samtidigt rekommenderar universitetet att information som innehåller känsliga och extra skyddsvärda personuppgifter, sekretessreglerade uppgifter eller annat känsligt material (exempelvis information som enligt avtalsvillkor ska hanteras särskilt) inte ska laddas upp⁴, eftersom den förs över till de bolag som tillhandahåller verktygen och detta skulle kunna innebära ett otillåtet röjande eller strida mot avtalsvillkor.

Output som produceras av generativ AI kan vara särskilt känslig i relation till skyddet av immateriella rättigheter och personuppgifter. AI-genererat material kan innehålla delar som härrör från andras arbete, vilket ställer krav på korrekt källhänvisning och respekt för upphovsmannaskap. Risken för plagiat, oavsett om det gäller text, kod eller visuellt material, måste därför uppmärksammas. Om personuppgifter förekommer i AI-genererat innehåll är forskaren ansvarig för att dessa behandlas på ett lagenligt och etiskt korrekt sätt.

I processer där användning av generativ AI kan få konsekvenser för andra forskare eller organisationer, såsom vid sakkunniggranskning, bedömning av ansökningar eller andra utvärderingsprocesser, kan AI användas som ett verktyg i vissa delar av processen, men inte i forskarens egen sammanställning, bedömning eller ställningstagande. Detta för att minska risken för orättvisa eller snedvridna bedömningar till följd av verktygens begränsningar, samt för att undvika oklar ansvarsfördelning vid användning av generativ AI i sådana processer. Exempelvis anger Vetenskapsrådet i sina instruktioner att generativa AI-verktyg inte får användas i den vetenskapliga bedömningen av ansökningar, utan endast för att förbättra språket i de skriftliga omdömena.⁵

Kompetensutveckling och etisk reflektion

Eftersom generativa AI-verktyg utvecklas snabbt behöver forskare kontinuerligt uppdatera sin kunskap om hur de kan användas på ett ansvarsfullt och effektivt sätt. Detta kan ske genom utbildning, erfarenhetsutbyte och delning av goda exempel.

Forskning kan påverkas av AI på olika sätt beroende på ämnesområde och fokus för den etiska reflektionen kring AI-användning kan därför se olika ut inom olika discipliner. Vid gemensamt författarskap krävs gemensam reflektion kring projektets AI-användning, så att alla författare kan enas om ramarna och ta ansvar för användningen. När det gäller doktorander är en öppen dialog och reflektion tillsammans med handledare avgörande för att säkerställa att generativ AI används på ett genomtänkt och etiskt sätt. Ogenomtänkt AI-användning riskerar dessutom att undergräva färdighets- och kunskapsutveckling vad gäller den vetenskapliga forskningsprocessen.

Utvecklingen och användningen av generativ AI kan ha en negativ påverkan på människor och miljö, exempelvis genom resursförbrukning, arbetsvillkor i samband med databehandling och hantering av data. Möjligheten att minska denna bör beaktas genom att välja ändamålsenliga verktyg och använda dem på ett genomtänkt, hållbart och resurseffektivt sätt (exempelvis genom att minimera onödiga körningar, begränsa omfattningen av genererad output och välja modeller som är proportionerliga i förhållande till uppgiften).

Referenser

1. European Commission. (2024). *Living Guidelines on the responsible use of generative AI in research*. https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf
2. All European Academies (ALLEA). (2023). *The European Code of Conduct for Research Integrity*. <https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf>
3. Göteborgs universitet. *Vägledning vid användning av generativa AI-verktyg*. <https://gunet.sharepoint.com/sites/mp-stod-och-service/SitePages/V%C3%A4gledning-vid-anv%C3%A4ndning-av-AI-verktyg.aspx>
4. Göteborgs universitet. *AI för medarbetare*. <https://gunet.sharepoint.com/sites/mp-stod-och-service/SitePages/AI-verktyg-f%C3%B6r-medarbetare.aspx>
5. Vetenskapsrådet. *Riktlinjer kring användning av AI-verktyg*. <https://www.vr.se/soka-finansiering/att-soka-bidrag/riktlinjer-kring-anvandning-av-ai-verktyg.html>