

# AI och disciplinärenden

17 december 2024 kl. 14:00-16:00

Seminarium nr 5 i serien **Högskolan och AI – en seminariserie i samverkan** som SUHF genomför under 2024.

## Workshop Fall 4

Felaktiga AI-genererade citat

## Fall 4 – Felaktiga AI-genererade citat

När seminariet planerades valde arbetsgruppen att söka exempel där användning av AI i utbildningssammanhang hade lett till disciplinära åtgärder. I detta rättsfall användes ett AI-verktyg (ChatGPT) av en student för att söka fram citat till en hemtentamen. Studenten trodde sig hämta riktiga citat ur litteraturen, men AI-verktyget genererade i själva verket påhittade citat (s.k. hallucinationer). Universitetets disciplinnämnd gjorde gällande att studenten försökt vilseleda vid den aktuella examinationen och stängde av studenten i sex veckor. Studenten menade att hen använt verktyget i tron att det fungerade som en sökmotor och att inga tydligt uppsatta regler för användning av AI-verktyg hade brutits. Förvaltningsrätten upphävde disciplinnämndens och ansåg inte att det var tillräckligt klarlagt att studenten medvetet försökt vilseleda. Kammarrätten gjorde samma bedömning. Istället menade domstolarna att händelsen berodde på studentens okunskap om hur AI-verktyget fungerade och att det snarare var en fråga om studieprestationens kvalitet än en disciplinär förseelse.

### Exempel på diskussionsfrågor:

- a) Hur bedömer ni domstolarnas resonemang kring studentens avsikt? Är det rimligt att okunskap om ett AI-verktygs funktion ursäktar användandet av fabricerade citat?
- b) Hade ni gjort en annan bedömning om universitetet tydligt hade förbjudit användning av AI-verktyg eller krävt källkontroll av citat i instruktionerna?
- c) Hur ändras bedömningen om andrahandscitatet hade varit hämtade från tidskriftsartikel med fabricerade citat?
- d) Vilka krav bör ställas på studenter vad gäller kontroll av citat hämtade från andrahandskällor, och hur tydligt måste ett lärosäte kommunicera detta ansvar?

## **KAMMARRÄTTENS DOM**

[...]

Kammarrätten avslår överklagandet.

### **YRKANDEN M.M.**

Disciplinnämnden vid X universitet yrkar att förvaltningsrättens dom ska upphävas och disciplinnämndens beslut fastställas. Nämnden anför i huvudsak följande.

N.N. har med hjälp av AI-verktyget ChatGPT försökt vilseleda examinatorn gällande sin arbetsinsats vid författandet av examinationsuppgiften. N.N. har i texten byggt sina resonemang på citat som framställs som att de är hämtade i litteraturen. Citaten är dock felaktiga och skapade med hjälp av ChatGPT för att passa för ändamålet. Hen har genom detta agerande försökt vilseleda examinatorn. Genom att använda sig av ChatGPT har hen haft ett ansvar för att ta reda på hur verktyget fungerar och att kontrollera att de aktuella textstyckena utgjorde korrekta citat. Hen ska genom sitt agerande därmed anses ha varit i vart fall likgiltig inför om examinationen har genomförts på ett korrekt sätt.

### **SKÄLEN FÖR KAMMARRÄTTENS AVGÖRANDE**

Kammarrätten gör samma bedömning som förvaltningsrätten.

## **FÖRVALTNINGSRÄTTENS DOM**

[...]

Förvaltningsrätten upphäver det överklagade beslutet.

### **DET ÖVERKLAGADE BESLUTET**

Disciplinnämnden vid X universitet beslutade den 7 juni 2023 att stänga av N.N. från utbildningen vid X universitet i sex veckor från och med den 28 augusti till och med den 8 oktober 2023. Som grund för beslutet anförde nämnden bl.a. följande. N.N. har vidgått att hen har använt AI-verktyget ChatGPT för att skapa text som framställts som citat i uppsats. Utredningen ger inte tillräckligt stöd för slutsatsen att texten i övrigt har genererats av ChatGPT, eller något annat AI-verktyg. De av ChatGPT genererade textstyckena utgjorde en inte oväsentlig del av N.N:s resonemang, och framställdes i examinationsuppgiften som att det rörde sig om citat som hen själv hittat, satt i sina sammanhang, och referat till. N.N. har alltså överlåtit skapandet av delar av sin inlämnade text till ett AI-verktyg och måste ha varit medveten om att detta tillvägagångssätt inte var tillåtet. Hen har därmed försökt vilseleda vid examinationen och har gjort sig skyldig till en disciplinär förseelse enligt högskoleförordningen. Ingen annan disciplinär åtgärd än avstängning från utbildningen kan komma i fråga och sex veckor är en väl avvägd avstängningstid.

## YRKANDEN M.M.

N.N. överklagar beslutet och yrkar att avstängningsbeslutet upphävs, att det görs tydligt att hen utifrån universitetets eget regelverk för hemtentamen inte kan anses ha begått något regelbrott, att förvaltningsrätten påpekar för disciplinnämnden att användning av AI-verktyg för att söka och generera citat är tillåtet vid hemtentamina enligt universitetets egna regler och inte kan anses vara vilseledande vid examination genom otillåtet hjälpmedel samt att X universitet uppmanas att genomgå och tydliggöra sina regler och riktlinjer gällande användningen av AI-verktyg i undervisningen och hemtentamina för att förhindra liknande missförstånd i framtiden. N.N. begär även att disciplinnämndens avstängningsbeslut inhämmas tills förvaltningsrätten har fattat beslut om överklagandet. Hen anför därvid bl.a. följande. Vid en hemtentamen i engelska för blivande mellanstadielärare gavs en hemtentamen med tre essäuppgifter där man (1) skulle utveckla ett eget resonemang kring hur läst litteratur påverkar läsaren och förhåller sig till skolans värdegrund, (2) diskutera nyttan med att använda dikter samt (3) diskutera hur huvudkaraktären utvecklas och hur detta kan relateras till miljön i boken. I samtliga uppgifter skulle studenten stödja sitt resonemang med citat från litteraturen. Hens tentamen lämnades in den 13 februari 2023. För uppgifterna 1 och 3 använde hen sig av verktyget ChatGPT för att, som hen trodde, få fram giltiga citat som stöd för sitt resonemang i uppgifterna. Det är vanligt bland lärarstudenter att man använder ett verktyg (vanligen Google) för att få fram citat till inlämningsuppgifter eller hemtentamen då man inte har tillgång till litteraturen hemma. I uppgift 2 hade hen emellertid tillgång till litteraturen hemma och använde i den uppgiften då den litteraturen som källa för sina citat. Då ChatGPT inte är ett AI-baserat sökverktyg (som t.ex. Google, Bing, etc.) utan ett generativt AI-verktyg blev resultatet av att använda det verktyget emellertid att citaten för uppgifterna 1 och 3 blev felaktiga. Att använda ChatGPT för att söka fram citat som sedan utgör en utgångspunkt för ett resonemang i en hemtenta kan inte vara ett försök att vilseleda vid examination. Enligt X universitets regler för examination på grundnivå och avancerad nivå är alla hjälpmedel tillåtna om annat inte anges i kursplan eller i tentamensuppgiften. Några sådana begränsningar fanns inte i den aktuella kursplanen. Disciplinnämndens beslut måste därför vara ogiltigt då det strider mot det egna regelverket och då nämnden därmed tycks tillämpa det egna regelverket godtyckligt. Hen och nämnden har samsyn kring att det enda som är skapat av ChatGPT är citaten. Det råder därmed också en samsyn kring att resonemangen (som är det som examineras) är hens egna och är skrivna av hen själv. Disciplinnämndens påstående att de av ChatGPT genererade textstyckena utgjorde en inte oväsentlig del av hens resonemang och framställdes i examinationsuppgiften som att det rörde sig om citat som hen själv hittat, satt i sina sammanhang, och referat till, är emellertid inte korrekt. En korrekt beskrivning av situationen är i stället att de genererade ”textstyckena” inte utgör någon del av resonemangen utan framställdes som citat (hen har alltså inte påstått sig ha skrivit dessa själv) samt att hen sökte fram citaten med hjälp av ett generativt verktyg, vilket är ett tillåtet tillvägagångssätt och sedan använde citaten som stöd för sitt självständiga resonemang. ChatGPT-verktyget visade sig emellertid vara olämpligt för uppgiften och citaten blev därför felaktiga, vilket bör leda till att hen får poängavdrag för uppgiften på samma sätt som när man gör andra fel. Det måste dock understrykas att X universitets egna regler säger att alla hjälpmedel är tillåtna vid hemtentamen. Disciplinnämndens slutsatser är alltså felaktiga eftersom det måste anses tillåtet att använda ChatGPT för att söka fram citat samt då resonemangen helt och hållet är hens egna

och inga hjälpmedel har använts för att framställa de resonemangen. Det finns inte heller någon grund för disciplinnämndens bedömning av att hens skulle ha varit medveten om att tillvägagångssättet inte var tillåtet. Disciplinnämnden pekar inte heller på några regler som har brutits och utifrån universitetets eget regelverk så måste det vara ytterst rimligt för en student att anta att tillvägagångssättet har varit korrekt då alla hjälpmedel är tillåtna. Hen har agerat i enlighet med universitetets regelverk och i likhet med många andra studenter inte varit medveten om vilka riktlinjer och policys som finns kring AI-verktyg. Det finns inte heller något som visar att hen har haft för avsikt att försöka vilseleda vid examinationen. Hen har följt gällande regelverk och använt tillåtna verktyg för att få fram citat och sedan självständigt resonerat kring citaten samt har dessutom i uppgift två inte använt något verktyg då hen där har haft en egen bok att hämta citat ur som stöd för sina resonemang. Valet av disciplinär åtgärd saknar också proportion och är inte i linje med vad andra lärosäten har beslutat i liknande fall. I Uppsala och Lund varnades t.ex. studenter efter att ha använt ChatGPT vid inlämningsuppgifter och i Växjö stängdes en student av i två veckor efter att en robot hade svarat på inlämningsuppgifter för studenten. I samtliga av de fallen hade studenterna inte heller själva skrivit sina inlämningsuppgifter utan hade låtit ChatGPT göra det. För hens del skulle det därmed ha motsvarats av att hen hade låtit ChatGPT skriva resonemanget kring citaten, vilket hen inte har gjort. Det går därmed inte att komma fram till att annan disciplinär åtgärd än avstängning inte kan komma i fråga och att sex veckor är väl avvägd tid.

Disciplinnämnden har inte funnit skäl att ändra beslutet.

## **SKÅLEN FÖR FÖRVALTNINGSRÄTTENS AVGÖRANDE**

Förvaltningsrätten tar upp målet till omedelbart avgörande. Yrkandet om inhibition föranleder därmed ingen ytterligare åtgärd.

Förvaltningsrätten kan vid sin prövning av det överklagade beslutet bara pröva om disciplinnämnden har haft fog för sitt avstängningsbeslut eller inte och kan därmed inte pröva N.N:s yrkanden om tydliggöranden av universitetets regelverk samt om uttalanden kring den generella tillåtligheten av användandet av AI-verktyg.

Enligt 10 kap. 1 § högskoleförordningen (1993:100) får disciplinära åtgärder bl.a. vidtas mot studenter som med otillåtna hjälpmedel eller på annat sätt försöker vilseleda vid prov eller när en studieprestation annars ska bedömas.

En disciplinär åtgärd i form av avstängning utgör ett allvarligt ingrepp mot den enskilde studenten. För att ett beslut om avstängning ska få meddelas krävs därför att det av utredningen klart framgår att studenten har gjort sig skyldig till den påtalade förseelsen (jfr RÅ 1996 ref. 15). I praxis har även beaktats om studenten har varit medveten om att agerandet har varit otillåtet. Studenter har dock samtidigt ett ansvar för att ta reda på vilka regler som gäller inför en examination (RÅ 2010 ref. 6).

Förvaltningsrätten delar disciplinnämndens bedömning av att utredningen inte ger tillräckligt stöd för slutsatsen att N.N:s inlämnade text ska ha genererats av ChatGPT eller något annat AI-verktyg i en vidare omfattning än vad avser sammanlagt tre utpekade citat ur två av hens inlämnade uppsatser.

Utredningen visar alltså inte att N.N. skulle ha försökt vilseleda i sin studieprestation genom att ha gjort gällande att ChatGPT-generad textmassa avseende resonemang och analyser skulle ha skapats av hen själv. De aktuella citaten uttrycker i stället tydligt att de har sitt ursprung från någon annan källa än från N.N.

Det framgår av utredningen att X universitets allmänna regler för examination på grundnivå och avancerad nivå föreskriver att alla hjälpmedel ofta är tillåtna vid hemtentamen och inlämningsuppgift samt att om det finns begränsningar, till exempel avseende samarbete, ska det anges i kursplanen. Utredningen utvisar inte att den nu aktuella kursplanen skulle ha ställt upp några andra mer begränsande förhållningsregler i den delen. Förvaltningsrätten bedömer därmed att användande av AI-verktyg i sig inte kan anses ha utgjort ett otillåtet hjälpmedel för studenterna i detta fall. Utredningen i målet utvisar emellertid att de tre citaten har varit påhittade. De aktuella citaten har också utgjort en mer integrerad del i den av N.N. även i övrigt inlämnade textmassan för hemtentamensuppgiften. Då citaten har varit felaktiga och påhittade understödjer de t.ex. inte heller de textmassor som citaten hänvisar till. I den delen har N.N. anfört att hen numera har förstått att det är en väsentlig skillnad mellan t.ex. Google och ChatGPT genom att Google söker fram text i originaldokument medan ChatGPT genererar text och därmed kan skapa citat som inte finns, vilket också har skett i hens fall. Uppgifterna framstår i sig inte som orimliga och utredningen visar inte heller i övrigt att N.N. genom hänvisningen till de aktuella citaten medvetet skulle ha försökt att vilseleda kring citatens äkthet eller källa.

Annan slutsats kan därmed inte dras än att N.N. efter sin ChatGPT-sökning har försummat en vidare granskning av om citaten var korrekta och därigenom på grund av slarv kom att lämna in en studieprestation som innehöll tre påhittade citat. I avsaknad av tydliga riktlinjer och information från universitet till studenterna om tillåtligheten eller otillåtligheten av olika förhållningssätt till ChatGPT eller annan form av AI bedömer förvaltningsrätten därmed att det inte står klart att N.N. har försökt att vilseleda i sin studieprestation med otillåtna hjälpmedel eller på annat sätt på det sätt som föreskrivs i 10 kap. 1 § högskoleförordningen, eller att hen har varit medveten om att hens agerande skulle ha varit otillåtet (jfr t.ex. Förvaltningsrätten i Växjö's dom den 20 juni 2023 i mål nr 1998-23). N.N:s felaktiga citeringar i hemtentamensuppgifterna har enligt förvaltningsrätten under sådana förhållanden i stället varit en fråga för den aktuella fakultetens materiella bedömning avseende studieprestationens kvalitet.

Med ledning av det anförda bedömer förvaltningsrätten att utredningen i målet inte klart utvisar att N.N. har gjort sig skyldig till en disciplinär förseelse enligt högskoleförordningen. Det överklagade beslutet ska därför upphävas.