



PM rörande vilseledande plagiat och annat vilseledande fusk genom användandet av textgenererande AI-verktyg inom utbildningen vid Lunds universitet.

Den 2 februari 2023 inrättade rektor en arbetsgrupp angående ChatGPT och undervisning, vars uppgift bland annat var att se över universitetets riktlinjer och föreskrifter rörande plagiering och vilseledande plagiering inom utbildning på grundnivå, avancerad nivå och forskarnivå (Dnr LS 2010/722) i ljuset av de nya omständigheter som AI-tekniken innebär. Sammankallanden för arbetsgruppen har bett mig se över ovannämnda riktlinjer med särskilt fokus på frågor rörande disciplinära åtgärder mot studenter enligt 10 kap. högskoleförordningen (1993:100).

Inledning

Termen 'fusk' saknar legaldefinition inom högskolerätten. Vad som utgör "fusk" inom ramen för utbildningen vid universitetet, eller hur "fusk" ska hanteras, är med andra ord inte tydligt definierat i någon lag. I *Svensk ordbok* (SO) definieras 'fusk' som "utnyttjande av otillåtna metoder eller hjälpmedel i syfte att vinna fördel; [bl.a.] vid prov...". Enligt högskoleförordningen (1993:100) får disciplinära åtgärder i form av varning eller avstängning vidtas mot studenter som "med otillåtna hjälpmedel eller på annat sätt försöker vilseleda vid prov eller när en studieprestation annars ska bedömas"¹. Enkelt uttryckt är alltså försök till *vilseledande* fusk påföljdssanktionerat genom bestämmelser i högskoleförordningen.²

Plagiat och vilseledande plagiat

Plagiering är exempel på en vanligt förekommande form av fusk. Enligt Lunds universitets riktlinjer för plagiering och vilseledande plagiering (Dnr LS 2010/722) är plagiering "en brist på självständighet vid formulering eller utformning av ett redovisat arbete jämfört med vad som krävs i relation till utbildningens mål och sammanhang", medan vilseledande plagiering definieras som "en brist på självständighet där studenten dessutom med avsikt ger sken av att andras arbete är studentens eget arbete". Även om användandet av AI-verktyg – så som ChatGPT –

¹ 10 kap. 1 § 1 st. 1 samt 2 § 1 st. högskoleförordningen.

² Det finns skäl som talar för att nyttjandet av otillåtna hjälpmedel i vissa fall skulle kunna anses utgöra grund för disciplinära åtgärder enligt 10 kap. högskoleförordningen även om nyttjandet i sig inte kan anses utgöra försök till vilseledande (här med stöd av 1 § 1 st. 2, som påföljdssanktionerar störande av prov och annan verksamhet inom ramen för utbildningen vid högskolorna). Så torde åtminstone vara fallet om det framstår som tydligt att nyttjandet varit otillåtet – det vill säga utgjort ett regelbrott – och att regelbrottet begåtts med uppsåt och stort verksamheten inom ramen för utbildningen på något konkret sätt.

vid formulering eller utformning av ett redovisat arbete ofta borde kunna falla in under definitionen av plagiat ovan, är det inte lika självklart att sådant användande faller in under definitionen av vilseledande plagiat, bland annat eftersom det inte är helt självklart att en text som genererats av ett AI-verktyg alltid kan sägas utgöra ”andras [eller annans, min anm.] arbete”.

Vilseledande fusk

När det gäller generativa AI-verktyg så som ChatGPT kan, utöver plagiat, även andra former av vilseledande fusk komma att bli aktuella. Att vara vilseledande är det samma som att ge någon (eller hålla kvar någon i) en felaktig uppfattning.³ Att använda ett otillåtet hjälpmedel i syfte att kunna genomföra ett prov på ett enklare sätt kan alltså utgöra vilseledande fusk om användandet medför att någon, exempelvis examinatorn, får en felaktig uppfattning om t.ex. provets genomförande eller fuskarens kunskaper. Det krävs dock inte att en student faktiskt lyckas vilseleda någon för att en disciplinär åtgärd ska kunna vidtas mot studenten, utan det räcker att studenten har *försökt* vilseleda vid prov eller när en studieprestation annars ska bedömas.⁴ Att studenten ska ha försökt vilseleda medför i sin tur ett krav på uppsåt;⁵ det krävs alltså att studenten faktiskt *försökt* ge någon (eller hålla kvar någon i) en felaktig uppfattning. Det är varken nödvändigt eller tillräckligt att en student försökt agera på ett sätt (vid prov eller när studieprestation annars ska bedömas) som faktiskt gett någon (o.s.v.) en felaktig uppfattning för att en disciplinär åtgärd ska kunna bli aktuell. Vad som är avgörande är istället huruvida studenten på något sätt handlat med uppsåt att ge (eller hålla kvar någon i) en sådan uppfattning.⁶ När det gäller verktyg som ChatGPT bör utgångspunkten för bedömningen inte vara huruvida studentens användande av verktyget faktiskt riskerat att vilseleda någon viss lärare, utan huruvida studenten använt verktyget i syfte att vilseleda eller åtminstone om studenten använt verktyget trots att hen måste ha insett risken att användandet skulle kunna komma att bli vilseledande – i båda fallen torde det krävas att verktyget faktiskt utgör ett otillåtet hjälpmedel, vilket i sin tur beror på utformningen av den aktuella kursen.

Tydlighet i information till studenter

Tydligheten i den information som tillhandahållits en student om vad som är tillåtet och inte, vad som ska bedömas och betygsättas och inte och så vidare kan ofta vara avgörande för bedömningen i disciplinärenden. Det är långt ifrån självklart om (och i så fall vilka) AI-verktyg som får användas vid prov och liknande, utan det beror på både vilken information om tillåtna hjälpmedel som har förmedlats till studenterna, och provet eller uppgiftens utformning och uttalade syfte. I vissa fall kan det tänkas ligga i sakens natur att det är vilseledande att använda exempelvis ChatGPT. Som exempel kan tänkas vissa uppgifter som är formulerade som uppmaningar till tentanderna (”Beskriv [med egna ord]...”, ”Redogör [kortfattat] för...” o.s.v.). Det torde inte vara särskilt kontroversiellt att hävda att en student som lämnar in en hemtentamen med sådana besvarade uppgifter som beskrivs ovan (utan att det framgår att svaren är AI-genererade) försöker ge någon (direkt eller indirekt examinatorn) den felaktiga uppfattningen att studenten följt anvisningarna

³ Nils Jareborg, ”Disciplinansvar för studenter som fuskar eller stör” (Högskoleverket, 2002), 5.

⁴ Se 10 kap. 1 § 1 st. 1 högskoleförordningen.

⁵ Jareborg, ”Disciplinansvar för studenter som fuskar eller stör”, 10.

⁶ Jmf. Jareborg, 9–10.

genom att själv beskriva, redogöra o.s.v.⁷ I andra fall kan det vara nödvändigt att vara extra tydlig med vad som förväntas av studenterna – tydligheten i kommunikationen med studenterna om vad som är tillåtet är alltså i egentlig mening inte preventiv, åtminstone inte i första hand. Istället avgör tydligheten vad som faktiskt utgör vilseledande fusk och inte. Information om vad som är tillåtet, och inte, ska med andra ord inte bara ses i relation till fuskprevention. Informationen avgör ofta vad som faktiskt utgör vilseledande fusk i det enskilda examinationsmomentet.

Vilseledande plagiat och annat vilseledande fusk som indirekt konsekvens av användandet av generativa AI-verktyg

Även om det inte nödvändigtvis är så att själva användandet av generativ AI i sig utgör försök till vilseledande i 10 kap. högskoleförordningens mening, kan användandet av AI-verktyg tänkas kunna innebära försök till vilseledande på andra sätt. För det första är det inte säkert att AI-genererad text är helt originell eller unik. AI-verktyg som ChatGPT bygger på stora mängder text som har producerats av människor, och kan därför komma att återanvända eller omarbeta delar av dessa texter.⁸ Det kan därför vara svårt för den enskilde studenten att upptäcka eller kontrollera om AI-genererad text innehåller stycken som vid inlämning skulle kunna uppfattas som vilseledande plagiering även om studenten är medveten om den risken (en sådan medvetenhet talar snarare till studentens nackdel, se nedan). Det är inte heller ovanligt att AI-verktyg fabricerar referenser till källor som inte existerar. Det finns goda skäl som talar för att det kan vara vilseledande i 10 kap. högskoleförordningens mening att lämna in en AI-genererad text som ger sken av att andra *personers* arbete är studentens eget, eller som innehåller fabricerade källhänvisningar, så länge studenten lämnat in texten trots att hen varit medveten om den risken.⁹ Sådant vilseledande skulle i så fall ofta kunna upptäckas med mer traditionella metoder så som *original* och stickprov av källhänvisningar.

Upptäckt och GDPR

När det gäller upptäckt av användande av AI-verktyg finns särskilda tjänster (AI-detektorer) som uppger ge en indikation på sannolikheten att en viss text är författad av till exempel ChatGPT. Det är dock svårt att bedöma sådana tjänsters tillförlitlighet, och dessutom är det troligtvis inte tillräckligt med ett resultat från en sådan tjänst för att det ska kunna anses vara styrkt att en student använt ett AI-verktyg för att generera en inlämnad text. Även om sådana tjänster förvisso skulle kunna användas som urvalsgrund för t.ex. kontroll av källhänvisningar med mera, så finns det många skäl som talar för att eventuellt nyttjande av tjänsterna bör ske

⁷ Om en student i stället använt någon annans text för att svara på frågorna, och därigenom framställt den som sin egen, rör det sig om tydligt plagiat. Att det i egentlig mening inte rör sig om någon annans text (åtminstone inte ”någon annans text” som i ”någon annans *persons* text”) vid användandet av ett AI-verktyg verkar dock inte innebära att det att försöka framställa sådan text på ett sätt som ger sken av att studenten följt anvisningarna i tentamensfrågorna är mindre missvisande.

⁸ ChatGPT genererade texter har i en studie visat sig kunna ge utslag i Ouriginal uppåt 13 %, se Randy Joy Magno Ventayen, ”OpenAI ChatGPT Generated Results: Similarity Index of Artificial Intelligence-Based Contents”, *Available at SSRN 4332664*, 2023.

⁹ En student som måste ha insett, men förhållit sig likgiltig inför, risken att en AI-genererad text som har lämnats in för bedömning innehåller vilseledande plagiat (det vill säga ger sken av att andras arbete är studentens eget arbete) eller fabricerade källhänvisningar torde i vissa fall kunna sägas ha så kallat ”likgiltighetsuppsåt”, jmf. NJA 2004 s. 176. Det är dock inte otänkbart att det att en student lämnar in en text för bedömning med källhänvisningar till material studenten inte har läst i sig skulle kunna anses vara vilseledande i 10 kap. högskoleförordningens mening.

med stor försiktighet. Att överföra studentarbeten till AI-detektionstjänster med servrar utanför EU skulle troligtvis strida mot EU:s dataskyddsförordning (GDPR) i många fall, för att bara nämna ett sådant skäl till försiktighet.

Angående Lunds universitets riktlinjer för plagiering och vilseledande plagiering (Dnr LS 2010/722).

Det är tveksamt om användandet av AI-verktyg som Chat GPT för att generera text i sig kan sägas innebära ”vilseledande plagiering” i den mening som avses i Lunds universitets riktlinjer för plagiering och vilseledande. Det hade förstås varit möjligt att ändra den nuvarande definitionen av vilseledande plagiering, ”en brist på självständighet där studenten dessutom med avsikt ger sken av att andras arbete är studentens eget arbete”, till en definition som även inkluderar AI-genererad text, men en sådan definition hade varit problematisk av flera anledningar – en AI-översatt text behöver inte nödvändigtvis utgöra just plagiat (även om, till exempel, användandet av en sådan text i en inlämningsuppgift i en språkkurs torde kunna anses vara vilseledande i många fall). En definition av vilseledande plagiat som innebär att man kan plagiera icke-personer verkar inte heller överensstämma särskilt bra med allmänt språkbruk. I *Svensk ordbok (SO)* definieras ’plagiat’ som en ”direkt upprepning eller efterbildning av eller lån ur (konstnärligt eller vetenskapligt) verk utan angivande av förlagan eller källan”, och *Svenska Akademiens ordlista (SAOL)* ger definitionen ”produkt av vetenskaplig [eller] konstnärlig stöld”, vilken stämmer överens med den äldre definitionen i *Svenska Akademiens ordbok (SAOB)*¹⁰. Att användande av AI-genererad text (utan att ange källa) per definition skulle utgöra någon form av stöld verkar inte alls särskilt självklart (även om det möjligen hade gått att hävda att sådant användande skulle kunna falla in under definitionen i *Svensk ordbok*).¹¹

Det är alltså tveksamt huruvida stora delar av universitetets riktlinjer är relevanta för vilseledande fusk genom användandet av AI-verktyg. Även om många av de tillämpningsregler som finns i riktlinjerna är högst relevanta även när det gäller vilseledande fusk genom användandet av AI-verktyg, exempelvis tillämpningsregel 2 om att man vid bedömning normalt bör ta hänsyn till ”kursens nivå, examinationsuppgiftens plats i utbildningen och ställda förkunskapskrav”, så är det inte lika självklart hur, för att bara nämna ett exempel, de för plagiat högst relevanta tillämpningsreglerna gällande principer för källhänvisningar skulle kunna tillämpas på vilseledande fusk genom AI-verktyg på ett rimligt sätt. Med anledning av vad som framförts ovan verkar det, i min mening, olämpligt att utöka definitionen av *vilseledande plagiat* så att den innefattar användande av AI-genererad text.

Slutsats

Sammantaget kan konstateras att det inte alltid måste ses som försök till vilseledande i 10 kap. högskoleförordningens mening om en student använder verktyg som ChatGPT i samband med examination, men det finns flera scenarier där användandet skulle kunna tänkas utgöra vilseledande plagiat eller annat vilseledande fusk. Det är ofta viktigt att beakta både vilken information som har förmedlats till studenterna om tillåtna hjälpmedel och provets eller uppgiftens

¹⁰ ”handlingen att plagiera, plagiering, litterär [eller] konstnärlig stöld; ofta konkretare, om det som stulits [eller] lånats (t. ex. ett avskrivet stycke).” (SAOB Spalt P 1005 band 20, 1953)

¹¹ Vad gäller definitionen av plagiat (plagiering), jmf. Mattias Alveteg ”PM rörande plagiering och vilseledande plagiering inom utbildning på grundnivå, avancerad nivå och forskarnivå vid Lunds universitet” (Lunds universitet, Dnr LS 2010/722), 2-3

utformning och syfte. Tydlig kommunikation med studenterna angående vad som är tillåtet och inte är avgörande för att kunna bedöma om användandet av AI-verktyg utgör vilseledande fusk eller inte.

Uppmärksamhet bör även riktas mot de indirekta konsekvenser av användandet av generativa AI-verktyg som kan innebära vilseledande plagiat eller annat vilseledande fusk, till exempel genererade texter med återanvända eller omarbetade delar av andras texter eller med fabricerade källhänvisningar. När det gäller upptäckt av användandet av AI-verktyg bör stor försiktighet iaktas vid användning av AI-detektionstjänster, inte minst på grund av de potentiella integritetsfrågor som kan uppstå i samband med överföring av studentarbeten till tjänster med servrar utanför EU.

För att säkerställa en rättvis och transparent examinationsprocess bör universitetet fortsätta att informera och utbilda både lärare och studenter om riskerna och konsekvenserna av användandet av AI-verktyg inom akademien. Detta kan innebära att förtydliga vilka typer av hjälpmedel som är tillåtna och vilka som inte är det samt att regelbundet uppdatera riktlinjer och föreskrifter för att anpassa dem till det snabbt föränderliga tekniska landskapet. Genom att göra detta kan universitetet bidra till att minska risken för vilseledande fusk och upprätthålla akademisk hederlighet i en tid av teknisk innovation.

Som ett led i detta anser jag att Lunds universitets riktlinjer och föreskrifter rörande plagiering och vilseledande plagiering inom utbildning på grundnivå, avancerad nivå och forskarnivå (Dnr LS 2010/722) bör utökas med separata riktlinjer om vilseledande fusk genom användandet av AI-verktyg med tillhörande föreskrifter motsvarande dem för plagiat och vilseledande plagiat.

Lund, dag som ovan,

Lehman Benson Jr.