

[Transcript] Forklart / Har KI-kjempene stjålet?

Det koker blant verdens kjente forfattere om dagen.

400 page novel.

Hvis i 15 minutter det kan stå og redistribute.

For i sommer ble det avslørt at tekstelskaper som OpenAI og META

i det stille har foret kunstintelligens med romanene deres.

Målet er å lage den aller beste språkmodellen.

Og nå starter en avgjørende kamp i amerikanske rettsaler

der forfattere kjemper for sin eksistens og kooselskapene for teknologiske fremskritt.

Er menneskelig kreativitet ferdig?

Du hører på forklart fraftenposten en podcast der vi forklager deg en nyhet i hver episode.

I dag om hvordan kunstintelligens truer forfattere i hele verden.

Jeg heter Sunne Søhohl. I dag er det fredag 29. september.

Det som står på spel er at flere store forfattere har gått til flere ulike søksmål mot store kooselskap.

Og helt enkelt, så mener jeg at kooselskapene har stjåle verken eller herans.

Andreas Bratt og Kristøva jobber som kultursnolist i aftenposten

og skriver blant annet om denne saken mellom koosjempene og forfattere.

Og det hele startet med en sak i magasinet Atlantikk.

Det er funnet ut at det finnes et datasett som heter Bux3.

Og i datasettet er det over 180.000 bøker. Det er flest der jeg er ute i de siste 20 år med.

Da de får nytt videre var det store kooselskap som Meta, Bloomberg og sånn.

De har brukt dette datasettet til å trene språkmodellene sine.

Hvordan bøker det?

Det er alle mulige slags bøker. Det er for eksempel bøker fra George R. R. Martin.

Det er her i Norge Knøvskor, Nesbø.

Da som er litt centralt er at disse bøkene er piratkopierte.

Det er gjort utdann til atelser for forfattere med.

Bøkene blir altså brukt til å gjøre språkmodellene bedre.

For jo mer tekstifor, jo mer lærer de å etterligne språkmønstre.

For eksempel å gette hvilket ord som mest ansynlig kommer etter et annet.

Og derfor matestechnologien bak tjenester som ChatGPT, Bard og Bing med bestsellende romaner.

Hvis du bruker for eksempel e-bøker som er gjort i dette tilfellet,

så er det jo på en måte høkt nivå på språket.

Det er masse tekst, og det er bra.

For jo høyere kvalitet på tekst når jo mer tekst,

jo bedre treningskunderlag har språkmodellen.

Og da får du et enda bedre språk som framstår som mer menneskelig og organisk.

Det er jo ikke sikkert at du bruker ChatGPT hverdagen,

men kunstintelligens er det ganske vanskelig å umgå.

Kundeservice sier på mobilen, hei Google og ansiktsjennkjenning er bare noen eksempler.

Og takket være skjønnliteratur fra hele verden, blir språkmodellene bare bedre og bedre.

Forfattere, jeg mener jo at hovedproblemet her er at det er piratkopiert innhold.

Det er ikke jævlig tillatelse, så de ser oppe at det er en tjuveri av deire onsverk.

Nå kan vi kanskje si at de har spesielle penne, du har frase, du har stil som er din egen,

som de plutselig kan show i en av disse pratebottene.

[Transcript] Forklart / Har KI-kjempene stjålet?

Men det er ikke sånn at selskapene publicerer bøkene nå sted?
De bruker ikke bare til å trene språkmodellerne?
Det er helt sant, og forfattere, jeg mener jo i flere av disse søkstmålene, så mener jeg at bæreda å bruke verkene derens til trening er en ulovlig reproduksjon isengsjøl. Og det er jo ikke kår i selskapene energi.
Nei, og de har også en fordel.
For arbeidet med kunstindelig jens går så fort at juristene klarer ikke å holde følge.
Mils Kavjar har vært en del av hverdagen til nurmen i over 70 år.
Men hvor lang tid tar det egentlig å lage?
Fra januar til april leverer over 600 fisker av skreirongt fra Nordnorge.
Mils håndplukker rongna, og den beste legges på tønna for langtidsmodning i hele 9-11 måned.
Til slutt røykes rongna på ekte bøkeflis.
Så bak Mils Kavjar likk faktisk et helt år med håndtverk.
Andreas, hva sier egentlig kår i selskapene?
De mener at måten de bruker verkenet er forfatterende på, fell inn under noe så heiter Fair Use i USA eller rimelig bruk her i Norge, som åpner egentlig for å avgrense gjennjeving av en tekst.
Altså det vil si at det er lov å gjenni en vis mengde tekst fra et vilket som helst verk?
Ja, så kår i selskapene argumenterer i alle fall for det.
Men likevel, nå foregår det mange rettsaker og søksmål fra forfattere i USA.
Og spørsmålet er om det kår i selskapene nå gjør, bryter med opphavsretten?
Da er vel det en prov å forklare heitt i noe i flere disse sakerne som går i USA.
Men det er ikke bare amerikanske forfattere som føler at kunstig intelligens steler arbeidet deres.
Det kan godt hende at norske forfattere velger å sagsøke a meter.
Samtidig ser det flere regler i EU som kan gjøre det vanskelig å vina fram med en sånn sak.
Hvordan da?
For eksempel i noe så heitt digitalmarkedsdirektive som er en sånn lov.
Der blir det åpnet for at kår i selskap har rett til tekst og datautvinning eller så kalla dataskraping.
Det er så det er rett å hente ut data fra internet.
Forfattere kan si nei til sånn dataskraping, men de må ge helt konkret beskjed om dette i forkant.
Så er det sånn at utviklingen av kår i den har gått så fort at ljusen slitt med å halve fylget.
Og lov som ble trodde kraftbære for noe å se an, de har ikke noe om kår i seg.
Kunstig intelligens utvikler seg altså raskere enn det jurister klarer å lage nye lover.
Men nå jobber det med å lage en nye lov i Europa.
Det er først og fremst en lov om produktansvar som stiller krav til ulike kår i tjeneste baseret på en sånn risikovodering.
Og i den loven er det også noe å teke inn forslag om at de som lager sånne modeller må være åpne om hva materiale de har brukt til å trene språkmodellerne sine.
Og som jeg har sett nå er jo mange av disse tekstene opphavsrett beskyttet.
Men denne kår i loven, den vil ikke tre kraft før tidlig ekst i 2026.
Meta sier at de er sluttet å fore språkmodellen med bøker.
I tillegg sier de at innholdet de producerer, altså svarende fra språkmodellen, ikke er vesendelig likt innholdet i forfatternes bøker.
Altså at tekstene ikke skal kunne igjenkjennes.

[Transcript] Forklart / Har KI-kjempene stjålet?

Men Andreas, finnes det noen måter hvor forfattere kan bruke kår i til sin fordel?

Ja, da mener jeg i alle fall kår i selskapene at det er.

Det er jeg mener jo at kår igjen ikke må show oss på som en trussel, men som et nyttig verktøy i kreative prosessen.

Så jeg mener jo at forfattere må beleare seg å bruke dette her på en måte i arbeid sitt.

Hvordan da?

Nei, altså da kan jo tenkjast at kår igjen kan bruke sånn en slags sparingpartner, en slags lav terskeld redaktør.

En kan diskutere karakterer, ordval og den slags.

Så det er nok da kår i selskapene mener.

Uansett hvordan kampen mellom kår i kjempen og forfattere ender, så har kunstintelligens allerede fått en plass i kunstnerisk uttrykk.

På Spotify kan du finne kår i musikk, og Amazon selger faktisk egne bøker som er skrevet av kunstintelligens, bøker som ikke har en forfatter.

Og for et år siden ble et maleri laget av kunstintelligens solgt til nesten en halv million dollar.

Og så mange andre saker, så er ikke dette enten eller.

Da TV kom for eksempel, så var mange helt sikre på radio ville forsvinne.

Og debatten vi har nå med klima og naturvern, så er det ikke sånn at vi bare satser på vindmøller, eller at vi bare tenker på å verne naturen.

Men hvor grensen for hvor langt køy kan gå nå, avgjøres i amerikanske rett saller.

Det er jo både forfattere og kår i selskap som innskylder i saletta her med dialog.

Men så er det jo sånn at det er flere søksmål som allerede er i prosess.

Og hvis jeg er sikker at forfattere vinner fram med noen av disse søksmålene, som går oppe og bryd på opphavesretten,

så vil jeg en turlig forklare heit til hvor grensene går og hva som er greit å ikke å gjøre.

Og da har vi jo selv søkt igjen før og til innstraminga å gjøre det vanskeligere og dyrere for kår i selskapene og skaffa seg treningsdata.

Med andre ord. Hvis forfattere vinner denne kampen, så vil bøkene deres og opphavesretten bli beskyttet fremover.

Men på den andre siden, hvis kår i selskapene ikke kan trene modelene sine med andres tekster, så vil utviklingen for kunstentlig ganske gå treigere og bli litt for enkelt dummere.

Og akkurat det har vi faktisk lagt en hel episode om tidligere som du kan søke opp om du vil.

Men Andreas, hva betyr det da?

Og da har vi trulig bremse utviklingen ganske kraftig.

Og om forfattere ikke får mehaleretten, så betyr det egentlig at den ganske uregulerte og grenseløvs måten kår selskapene skrapet data på i dag innanfor lovensrammet.

Du har hørt en podcast fra Aftenposten.

Det var Andreas Brattokke Støva som tok deg gjennom konflikten mellom forfattere og køyselskapene.

Denne episoden er laget og produsert gjennom i følelen og meg sinnes øyehol.

Og resten av forklart er David Vekone, Ola Veggesvik og Anders Weberg.

Lydende hørte er fra ABC News, CNN, Good Morning America og nyhjotsberået AP.

Hei, jeg heter Haral Eija og hva heter du for nå?

Jeg heter Nils Brann.

[Transcript] Forklart / Har KI-kjempene stjålet?

Og vi driver en podcast sammen hva gjør vi der, Nils?

Du, der la vi folk ta en personelsest og så analyserer vi dem etter det.

Det er så kjennemorsomt, det er så kjennunderholdende og det er så kjennelærigt for deg som hører på for å skjønne hvem du er også.

Ja, også så nært som jeg kommer.

Nils Brann, Haral Eija, Big Five og Spodmi, jeg er bare glad i seg.

Det er jeg.