

Forfatteren_GPT-2: Relevansen av forfatterens intensjon i estetisk tolkning

I filosofisk estetikk herjer en debatt om hvorvidt en forfatteres intensjon med et verk er viktig for den estetiske tolkningen og for å «finne ut» hva meningsinnholdet i et verk er for noe.

Med utgangspunkt i en kort presentasjon av tre ulike syn innen denne debatten, ønsker vi å se på hvordan postulatet om en algoritmisk forfatter fungerer innenfor disse teoretiske rammene. Med «algoritmisk forfatter» mener vi at forfatteren er selve algoritmen som genererer teksten, og ikke verken de som har gitt opphav til datasettet eller de som har utviklet algoritmen.

Vi vil først presentere E. D. Hirsches syn på saken. Han står bak det såkalte “identity theisis”¹, hvor meningen med et litterært verk i stor grad determineres utfra forfatterens intensjon. Som han selv skriver i kapittelet «In Defense of the Author» i boken *Validity in Interpretation*; “if the meaning of a text is not the author’s, then no interpretation can possibly correspond to *the* meaning of the text, since the text can have no determinate or determinable meaning.”²

Hovedargumentet hans for et slikt syn kommer til uttrykk gjennom “the determinacy argument”. Sett at en rekke ord satt sammen, men i sin samlethet sett for seg selv, har mange ulike betydninger. Med andre ord er ordrekkens mening veldig vanskelig å skulle angi, sett kun i lys av seg selv. Her mener Hirsch at vi kun kan avdekke ordrekkens mening ved å se på produsenten bak disse ordene sin intensjon; “The array of possibilities only begins to become a more selective system of *probabilities* when, instead of confronting merely a word sequence, we also posit a speaker who very likely means something”³.

Hvis vi ikke gjør dette, later Hirsch til å tenke at vi ender opp i en situasjon hvor vi ikke kan si noen ting om poenget med et litterært verk. At vi ikke kan si noen ting om et litterært verks poeng virker lite plausibelt for ham. Han mener derfor at det må være et identisk forhold mellom det som er et litterært verks mening og den meningen som er intensjonsbærende hos forfatteren bak verket.

Vi tenker at en algoritmisk forfatters intensjon er vanskelig å determinere. Vi kan ikke postulere at den har noen intensjon om en mening annet enn gjennom observasjonene av dens prosess. Prosessen til en tekstgenererende algoritme som GPT-2 er unsupervised, altså er ikke definisjonene bak ordene hardkodet, men et resultat av hvordan det nevralt nettverket er trent til å se relasjoner mellom ulike ord. Sett fra algoritmens ståsted vil “kitten” og “cat” være

¹ Beardsley, *The Possibility of Criticism*, 210

² Hirsch, *Validity in Interpretation*, 196

³ Ibid. 199

nærmere hverandre enn “kitten” og “dog”, men “kitten” og “dog” vil være nærmere hverandre enn “kitten” og “typewriter”.

I den grad det finnes en intensjon hos algoritmen, er denne altså orientert mot å generere en form for matematisk og statistisk mening. Dette leder oss kort oppsummert til et sted hvor algoritmegenerert kunst kommer veldig flatt ut gitt Hirsches tanker om at forfatterens intensjon bak kunsten er viktig for å determinere det estetiske outputet, dersom vi ser selve algoritmen som forfatteren, siden denne intensjonen kun dreier seg om å skape en mening i seg selv, og ikke å nødvendigvis kommunisere noe utover ordenes form.

Rammene Beardslys syn, som skiller seg radikalt fra Hirsches, kan derimot åpne opp for at en algoritmisk forfatter kan generere et estetisk output på lik linje med et menneske. I *The Possibility of Criticism* forsvarer han et anti-intensjonalistisk syn, som også kan kalles konvensjonalistisk. For å introdusere synets betydning, kan man kanskje si at det anti-intensjonalistiske synet ikke tillegger eventuelle intensjoner hos forfatteren av verket mye vekt når man skal søke meningen med litterære verk.

Beardsley mener at en tekst kan ha to ulike typer mening. Den første defineres av ham som “textual meaning”, og den andre som “authorial meaning”.⁴ Her sikter Beardsley til tekstuell mening som i noe alé Hirsches intensjonelle mening. I “authorial meaning” legger han ords konvensjonelle og normative meningsinnholdet til grunn. Meningen med et litterært verk er i denne forstand den meningen som er gitt utfra de konvensjonelle betydninger av ord og setninger verket består i, sett i lys av generelle og offentlige regler i et gitt språk. Det kan også innebære mening i lys av kulturelle og historiske konvensjoner, eller kontekst.

Det virker som at Beardsley finner denne siste betydning av mening mest meningsfull i at “the text acquire determinate meaning through the interactions of their words without the intervention of an authorial will.”⁵ Meningen finnes utenfor forfatteren og dennes vilje. Beardsley tenker gjennom sin argumentasjon at vi ikke har direkte tilgang på en forfatters intensjon i utgangspunktet, og at vi følgelig ikke kan determinere den intensjonelle meningen av et verk. Det forblir noe mystisk. Til sammenligning vil den konvensjonelle meningen ved et litterært verk uten problemer være tilgjengelig for oss, siden det eneste som kreves av informasjon i denne sammenheng er informasjon om verket frittstående av forfatteren, noe vi faktisk har tilgang på.

Dette perspektivet på forfatterens relevans for et det estetiske ved et verk gjør det kanskje enklere å gå med på at en algoritme kan være en forfatter i seg selv og produsere et

⁴ Beardsley, *The Possibility of Criticism*, 210

⁵ Ibid. 217

estetisk output. Her fungerer outputets estetiske uttrykk og meningsinnhold autonomt fra hva det nå enn var som brakte det inn i verden.

Til sist vil vi presentere Levinsons syn som kalles hypotetisk intensjonalisme. Dette synet legger til grunn flere ting, men først og fremst den at et litterært verks meningsinnhold samsvarer med følgende: Den beste hypotesen om et verks intenderte meningsinnhold. Den beste hypotesen og hva som er det intenderte meningsinnholdet må justeres tatt i betraktning de konvensjonelle betydninger av ord og setninger, kulturell og historisk kontekst, samt relevant informasjon om forfatteren som kan tjene oss i å indikere hvilken intensjon hen måtte ha hatt eller har med sitt litterære verk.⁶

Jamfør Levinson virker det som om at det er én type intensjon som alltid vil være relevant for en fortolkning av et litterært verks meningsinnhold. Han spesifiserer det for oss. Intensjonen han sikter til er forfatterens “categorical intention”⁷. Denne intensjonen er rotet i spørsmålet om “something is regarded as literature (or as art) at all” og refererer til de ulike kategorier kunst som sådan.⁸

Man kan kanskje si at ontologien bak algoritmen som forfatter blir viktig for å avgjøre hva slags kategori den gir sitt output inn i. Den blir en del av dens kontekst. Ontologien bak en algoritme som GPT-2 fungerer positivistisk; heller enn å bestemme *a priori* (forut for erfaring) hva et ord betyr, er ideen bak algoritmene denne: med grundig analyse av den kontekstuelle betydningen vil det bli mulig å «gjette» en tilnærmet plassering av et ord innen det virtuelle rommet hvor teksten blir konstruert. Likevel: kan man si at det er «gjetting»? Hvis vi legger en animistisk tanke til grunn, så ja, kanskje. Samtidig: selv om definisjonene av ordene ikke er gitt *a priori*, vil fremgangsmåten til algoritmer basere seg på aksiomatiske beregninger, - noe vi vanligvis ikke knytter opp mot kunstnerisk intuisjon eller skaperkraft.

Kort oppsummert kan dette bety at meningen bak algoritmegenerert kunst i lys av den kategoriske intensjonen og ontologiske konteksten hos Levinsons hypotetiske intensjonalisme, samt i Hirsches faktiske intensjonelisme, faktisk vil være en mening i at ord er beregnelige og transparente. I disse teoriene fungerer alle algoritmer som et og samme subjekt. Og hvis det er slik at alle algoritmer som blir satt til å generere estetisk output egentlig bare har en mening til sammen (å skape mening), - kan man si at det er noen mening med dem i det heletatt? (Jamfør logikken «hvis alle er flinke, så er ingen flinke»). Beardsleys anti-intensjonalisme vil på sin side være den teorien som harmonerer mest med at en

⁶ Levinson, «Defending Hypothetical Intentionalism», 139

⁷ Levinson, *The Pleasure of Aesthetics*, 235

⁸ Ibid.

algoritmisk forfatter kan produsere en unikt og estetisk output, siden forfatterens intensjon her er irrelevant for det estetiske outputet.

Litteraturliste:

1. Beardsley, Monroe C. The Possibility of Criticism. Wayne State University Press, 1970.
2. Hirsch, E.D. Validity in Interpretation. New Haven and London: Yale University Press, 1967.
3. Levinson, Jerrold. «Defending Hypothetical Intentionalism» The British Journal of Aesthetics. Volume 50, Issue 2, April 2010. 20.05.21. URL = <<https://academic.oup.com/bjaesthetics/article/50/2/139/18716>
4. Levinson, Jerrold. Pleasures of Aesthetics: Philosophical Essays. Ithaca og New York: Cornell University Press, 1996.