

Nygaard-Joki, S. (2026). Dikotomiens blindsoner: kunstig intelligens i Kazuo Ishiguros Klara og Solen. I B. Bergstedt & S. Jamouchi (Red.) *Multisensorisk læring: Om intensitet og kropp-materiell i en performativ kunstpedagogikk* (s. 157–174). Fagbokforlaget. DOI: <https://doi.org/10.55669/oa660906>

6

Dikotomiens blindsoner

Kunstig intelligens i Kazuo Ishiguros Klara og Solen

Sonja Nygaard-Joki

Sammendrag: Som en streben etter å etterligne menneskelig intelligens, er kunstig intelligens både i sitt opphav og mål, antroposentrisk. En forståelse av kunstig intelligens som er forankret i en antroposentrisk virkelighetsoppfatning, plasserer mennesket som idealet i sentrum og den kunstig intelligente maskinen i periferien, definert i kraft av å være et ikke-menneske som streber etter å etterligne overlegne menneskelige kvaliteter. Dette kapitlet utforsker en posthumanistisk tilnærming til kunstig intelligens, som kan utfordre denne antroposentriske dikotomien mellom menneske og maskin, og undersøker om det kan ligge verdifulle perspektiver i dikotomiens blindsoner. Gjennom en næranalyse av et utdrag fra romanen *Klara og Solen* (2021) av Kazuo Ishiguro, belyses problemstillingen: På hvilke måter kan en posthumanistisk lesing av *Klara og Solen* utfordre en antroposentrisk forståelse av kunstig intelligens? I Ishiguros roman ligger hovedsynsvinkelen hos Klara, en avansert form for kunstig intelligent teknologi. Dette muliggjør et perspektiv som er umulig utenfor fiksjonens sfære og inviterer til en utforsking av ontologiske og epistemologiske spørsmål knyttet til vår forståelse av kunstig intelligens. Analysen belyser hvordan en posthumanistisk tilnærming kan synliggjøre og problematisere noen antatte sannheter, som samlet kan invitere til en perspektivendring som stimulerer til kritisk og meningsfull refleksjon om kunstig intelligens.

Nøkkelord: kunstig intelligens, posthumanisme, litterær analyse

Abstract: In an endeavor to imitate human intelligence, artificial intelligence (AI) is, in both its origin and aim, anthropocentric. An understanding of AI that is rooted in an anthropocentric view of reality will place human beings as the ideal in the center and an artificially intelligent machine in the periphery, defined by its non-human status as it strives to imitate superior human qualities. This chapter explores a posthumanist approach to artificial intelligence, to challenge this anthropocentric dichotomy between human and machine and examine possible valuable perspectives in the dichotomy's blind spot. Through a close analysis of an excerpt from Kazuo Ishiguro's novel 'Klara and the Sun' (2021), the following research question is illuminated: In what ways can a posthumanist reading of 'Klara and the Sun' challenge an anthropocentric understanding of artificial intelligence? In the novel, the main point of view is conveyed through Klara, an advanced form of AI technology. This enables a perspective that is impossible outside the realm of fiction and invites an exploration of ontological and epistemological questions related to our understanding of AI. The analysis examines how a posthumanist approach can both reveal and problematise some assumed truths, which collectively can invite a change in perspective that stimulates critical and meaningful reflection on artificial intelligence.

Keywords: artificial intelligence, posthumanism, literary analysis

Dikotomiens blindsoner: kunstig intelligens i Kazuo Ishiguros Klara og Solen

De siste årenes nyutviklinger innen kunstig intelligens preger i økende grad hverdagen, arbeidslivet og klasserommene våre. Samfunnsdebatten om kunstig intelligens er preget av en tosidighet mellom ytterpunktene: potensial – risiko, der teknologiens muligheter og utfordringer står i sterk kontrast til hverandre. FNs organisasjon for utdanning, vitenskap, kultur og kommunikasjon, UNESCO, fremhever for eksempel hvordan kunstig intelligens kan ha potensial til å løse noen av de største utfordringene innen utdanning i dag, samtidig som de advarer mot at den raske teknologiske utviklingen innen kunstig intelligens medfører betydelige risikoer og etiske dilemmaer. Som en fremtidsrettet løsning krever de «en menneskesentrert tilnærming til KI» (UNESCO, 2025) som fokuserer på menneskerettigheter, inkludering, kulturelt mangfold og rettferdighet.

Parallelt med utviklingen der kunstig intelligens-teknologien blir stadig mer avansert og spesialisert, pågår diskusjonen om hvilken rolle kunstig intelligens bør ha i utdanning og skole. En sentral bekymring er at en innebygd algoritmisk skjevhet, eller bias, kan føre til diskriminering og urettferdighet. Mange frykter at implementeringen av kunstig intelligens i utdanningen kan forstørre det digitale gapet og forsterke eksisterende sosiale ulikheter, der ressurssterke og privilegerte grupper kan få enda større fordeler av denne teknologien.

I tillegg til spørsmål om likeverd og rettferdighet, finnes det en rekke andre sentrale etiske og juridiske problemstillinger knyttet til for eksempel personvern, opphavsrett og akademisk redelighet. På et mer overordnet filosofisk nivå kan integreringen av kunstig intelligens i utdanningen utfordre tradisjonelle forestillinger om læring, kreativitet, identitet og dannelse. Alt dette skaper et behov for å forstå denne avanserte teknologien på en måte som gjør både barn og voksne i stand til å kritisk vurdere sitt samspill med den, slik at de kan ta informerte og ansvarlige valg. For å kunne møte dette behovet, er det nødvendig å undersøke nærmere de fundamentale premissene ved kunstig intelligens som teknologi.

Som en streben etter å etterligne menneskelig intelligens, er kunstig intelligens både i sitt opphav og mål, antroposentrisk. Sammensatt av gresk *anthropos* (menneske) og *kentron* (spiss, midtpunkt), plasserer antroposentrisme mennesket i sentrum. Det formidler et verdensbilde som skaper gapende kløfter mellom det menneskelige og det ikke-menneskelige og fortrenger det siste til periferien (Crist & Kopnina, 2014). Med andre ord er mennesket idealet som den kunstige intelligente maskinen etterstreber, og forståelsen av kunstig intelligens forankres i en konstant sammenligning med menneskelige kvaliteter. Dikotomien menneske/maskin representerer et uoverstigelig skille, der mennesket er hierarkisk overlegent, og i tråd med en slik virkelighetsoppfatning vil kunstig intelligens aldri bli noe mer enn en blodfattig kopi av mennesket, i sin mangel på menneskelig bevissthet (Bøhn, 2023).

Posthumanisme kan utfordre et slikt tankesett, som er preget av en hierarkisk dikotomi. Posthumanistiske tilnærminger handler mindre om å fastsette kriterier som skiller eller avgrensner ulike livsformer som menneske, dyr, natur og maskin, og mer om å fremme idéen om at disse grensene er mye mer uklare enn vi tidligere har antatt (Soper, 2012). Posthumanistiske strømninger kan bli en inkluderende og dynamisk møteplass for det som er noe annet enn menneskelig, herunder teknologi (Iovino, 2016).

Dette kapitlet vil utforske en posthumanistisk tilnærming til kunstig intelligens gjennom en næranalyse av et utdrag fra romanen *Klara og Solen* (2021) av Kazuo Ishiguro. Kapitlet har to hoveddeler. Del 1 belyser to ulike teoretiske perspektiver: en antroposentrisk forståelse av kunstig intelligens versus posthumanistiske perspektiver. Del 2 belyser problemstillingen *På hvilke måter kan en posthumanistisk lesing av Klara og Solen utfordre en antroposentrisk forståelse av kunstig intelligens?* gjennom en næranalyse av utdraget.

Analysen vil undersøke om det kan ligge relevante perspektiver på kunstig intelligens i blindsonen til den antroposentriske dikotomien menneske/maskin. Blindsonen fungerer her som en metafor for det som ikke blir lagt merke til og ikke får omtale, dersom forståelsen av kunstig intelligens forblir fastlåst i en slik dikotomi. Videre vil analysen belyse om en posthumanistisk lesning av teksten kan avdekke verdifulle perspektiver som tidligere kan ha blitt oversett.

En antroposentrisk forståelse av kunstig intelligens

Kunstig intelligens kan bredt vise til alle «kunstig intelligente systemer som utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål.» (Nasjonale strategier for kunstig intelligens, 2020). Begrepet rommer et mangfold av tilnærminger og teknikker, men to hovedtyper peker seg ut.

Generell kunstig intelligens er «et kunstig intelligent system som kan forstå, lære og bruke kunnskap på et nivå som er på linje med eller overgår menneskelige evner» (Teknologirådet, 2023). Denne formen for kunstig intelligens er fortsatt en utopi og hører til litteraturens og filmens fiktive dystopiske univers. Den må ikke forveksles med den siste tidens nyvinninger innen generativ kunstig intelligens som for eksempel ChatGPT.

Generativ kunstig intelligens representerer «maskinlæringsmodeller som kan generere unikt innhold basert på informasjon de er blitt trent på. Dette innholdet kan være tekst, bilder, lyd og video. Selv om resultatene ofte er imponerende, er ikke systemene kreative i menneskelig forstand.» (Teknologirådet, 2023).

Disse definisjonene illustrerer hvordan kunstig intelligens forstås ut fra en antroposentrisk tilnærming. I den første definisjonen av generell kunstig intelligens fremstilles teknologien som noe hierarkisk «på linje med» eller noe som «overgår», der teknologien har blitt mennesket overlegent i sin evne til å tilegne seg og anvende kunnskap. I definisjonen av generativ kunstig intelligens posisjoneres teknologien innenfor dikotomien menneske/maskin, som underlegen i sin mangel på kreativitet i menneskelig forstand.

Hva gjør en slik forståelse av kunstig intelligens med vår tilnærming til den? Avviser vi den som mindreverdig fordi den er kunstig, altså ikke ekte eller unaturlig? Eller aksepterer vi noe menneskelig ved dens såkalte atferd og behandler den deretter, til tross for en rasjonell bevissthet om det grunnleggende maskinelle eller ikke-menneskelige aspektet? Slike spørsmål er dypt forankret i et antroposentrisk perspektiv.

Kan datamaskiner tenke?

Tittelspørsmålet: Kan datamaskiner tenke? besvares kontant av professor i filosofi Einar Duenger Bøhn; nei, maskiner kan ikke tenke. Bøhn (2023) påpeker imidlertid at det interessante spørsmålet ikke er *om* maskiner kan tenke, men *hvorfor* de ikke kan tenke. Han forklarer at begrunnelsen ligger i den prinsipielle forskjellen mellom kunstig intelligens og kunstig bevissthet. Intelligens defineres som en form for atferd, noe man kan gjøre eller utøve. Datamaskinens kunstige intelligens utøves gjennom for eksempel avansert dataanalyse eller kompleks problemløsning. Som en motsetning til dette defineres begrepet bevissthet som «en subjektiv erfaring i et førstepersonsperspektiv av hvordan det er å være i den tilstanden man er i.» (Bøhn, 2023, s. 98). Datamaskiner kan altså ikke tenke, fordi de mangler denne evnen til subjektiv opplevelse eller bevissthet.

For å utdype dette introduserer Bøhn begrepet livsverden fra filosofen Hubert Dreyfus. Det som skiller mennesker fra maskiner, er at mennesker lever på bakgrunn av en livsverden. Vår livsverden inkluderer alle våre subjektive erfaringer om hvordan det er å være i vår tilstand, samt all vår kunnskap; både teoretisk og praktisk, eksplisitt og implisitt. Datamaskiner derimot, har ingen livsverden. De er forhåndsprogrammerte til å følge bestemte regler (algoritmer), basert på bestemte typer input (data). De kan dermed kun etterape tenkning, som en avansert papegøye.

Premisset til Bøhn er antroposentrisk. Tittelspørsmålet kunne like gjerne ha vært omformulert til: Hvorfor kan ikke maskiner tenke som mennesker? Dette perspektivet leder diskusjonen inn mot spørsmålet: hva er det som gjør et menneske til et menneske? Dermed flyttes oppmerksomheten bort fra det som kunne ha vært en bredere utforskning av kunstig intelligens.

Bøhn konkluderer med at «skillet mellom oss mennesker og datamaskiner fortsatt ser ut til å være ganske uoverstigelig» (Bøhn, 2023, s. 104). Denne konklusjonen bekrefter et antroposentrisk verdensbilde, der Mennesket og Maskinen utgjør en kløft mellom to begreper som gjensidig utelukker hverandre. Innenfor denne dikotomien er mennesket overlegent, med den avgjørende kvalitative forskjellen i tilstedeværelsen av en livsverden og evnen til bevissthet, mens maskinen eller kunstig intelligens karakteriseres som et manglende ikke-menneske. Tyler (2003) advarer mot hvordan en slik *menneske først*-antroposentrisme kan begrense vår tenkning om både det menneskelige og det ikke-menneskelige, og hindre oss i å oppdage nye muligheter som kan føre til ny innsikt.

Posthumanismens mangfold

Posthumanisme er ikke en homogen tilnærming, men snarere et paraplybegrep som rommer en rekke ulike retninger, samlet i sin utforskende problematisering av tradisjonelle humanistiske perspektiver. Posthumanistiske tilnærminger har til felles at de utfordrer etablerte dikotomier som: menneskelig/ikke-menneskelig, natur/kultur, subjekt/objekt eller sinn/kropp (Díaz de Liaño & Fernández-Götz, 2021). Posthumanisme kan også bidra til å bryte ned konseptuelle barrierer og forplikte oss til mindre rigide tenkemåter (Soper, 2012).

Et eksempel på dette er å utfordre begrensningene som kan ligge i et antroposentrisk verdensbilde, der kvaliteter som fornuft, språk, moral, sivilisasjon, teknologi og fri vilje blir ansett for å mangle hos andre ikke-menneskelige livsformer (Crist & Kopnina, 2014). Kanskje dette mangelperspektivet kan omposisjoneres til et ressursperspektiv? Hvilke muligheter trer frem når mennesket ikke er målestokken for alt?

Som et samlende bilde på posthumanismens kompleksitet introduserer Iovino (2016) metaforen om det posthumanistiske huset. Skissene til dette byggverket ble påbegynt i de siste to tiårene av det 20. århundre, men selve konstruksjonen er en kontinuerlig prosess. Huset er ikke et fast bosted, men snarere en adresse for posthumanismens mange strømninger – en mobil og dynamisk sammenkomst, i stadig utvikling og bevegelse. Iovino (2016) fremhever at posthumanisme som prosjekt forenes i å forkaste dogmet om menneskelig eksepsjonalisme. Dette innebærer å avvise tanken om et essensielt skille mellom det menneskelige og det ikke-menneskelige, og i stedet legge vekt på deres aktive samspill.

En av posthumanismens styrker er å tvinge frem en grundig dialog mellom ulike fagområder, for eksempel mellom humaniora, teknologi og biovitenskap (Iovino, 2016).

Posthumanismens møte med litteratur

En tverrfaglig dialog mellom posthumanisme, teknologi og litteratur er allerede initiert. I sin undersøkelse av noen viktige strømninger innen litteraturvitenskap og posthumanisme peker Wallace (2010) på litteraturens historiske forankring i humanistiske tankesett, og skriver om det uforløste potensialet som kan ligge i å anvende samtidens posthumanistiske tilnærminger på fortidens litterære tekster.

Posthumanismen kan leses tilbake inn i historien som en vedvarende og meningsfull idé om mennesket. (Wallace, 2010, s. 693)

Carretero-Gonzalez (2016) gjør nettopp dette i sin lesning av Mary Shelleys *Frankenstein* (1818). Ved å vise hvordan romanens hovedkonflikt ikke er mellom mennesket [vitenskapsmannen Victor Frankenstein] og ikke-mennesket [det såkalte monstret han gir liv til], men mellom mennesket og en annen art, utfordres opplysningstidens rasjonelle humanisme. Victor Frankenstein evner ikke å slutte å se verden fra et perfekt strukturert, dualistisk, hierarkisk og antroposentrisk ståsted, og skapningen blir dermed aldri ansett som noe annet enn et monster, et ikke-menneske med ingen krav på etiske betraktninger.

En posthumanistisk tilnærming til litterær analyse kan utfordre konvensjonelle standarder som dualistisk tenkning og antroposentrisme, og rette oppmerksomheten mot ikke-menneskelige elementer i litteraturen, elementer som tidligere kan ha blitt oversett (Kowalce, 2020).

I sin utgreiing av posthumanistisk metodologi i litterær analyse definerer Kowalce (2020) en litterær tekst som «et resultat av en uunnværlig, aktiv og konstant utveksling av mening mellom menneskelige og ikke-menneskelige deltakere i et forhold som resulterer i en bestemt form for budskap som er åpent for tolkning» (Kowalce, 2020, s. 709). Dette åpner for et tolkningsrom der litterær analyse ikke lenger handler om å finne én mening, men snarere om å anerkjenne og vise hvordan mening blir konstruert (Kelley, 2013).

På denne måten blir romanen ikke betraktet som et isolert verk, skapt av én forfatter med én sentral betydning, men som en tekst preget av flertydighet som oppstår i møtet med leseren (Kelley, 2013).

Flere verk av den nobelprisvinnende forfatteren Kazuo Ishiguro har blitt gjenstand for posthumanistisk forskning. Blant samtidens forfattere trekker

Wallace (2010) spesielt frem Ishiguro, som med romanen *Never Let Me Go* (2007) revitaliserer papirromanen ved å tøye grensene for dens humanistiske arv og antakelser. Nyere forskning om *Klara og Solen* (2021) viser at også denne romanen kan forstås som en fortsettelse av et slikt prosjekt. Misra (2023) undersøker begrepet omsorg (eng. care & emotional labour) i romanen, både fra humanistiske og posthumanistiske perspektiver. Gjennom Klara som hovedperson og omsorgsperson presenteres et samfunn der omsorg blir gjort til en vare og forstått som emosjonelt arbeid.

Sun (2022) på sin side anvender begrepet teknologisk perfektjonitet (eng. technological perfectibility) fra et posthumanistisk perspektiv for å vise hvordan romanen fremstiller kunstig intelligente roboter som perfekte i den forstand at de kan utføre oppgaver mennesker ikke kan, på en nærmest feilfri måte. Ajeesh & Rukmini (2022) plasserer romanen i en historisk gjennomgang av science fiction-sjangeren. Science fiction utfordrer ofte grenser gjennom sitt fokus på samspillet mellom teknologisk og menneskelig utvikling, og tekstene stiller ofte spørsmål ved konvensjonelle epistemologiske og ontologiske diskurser. De fremhever beskrivelsen av Klara som «en humanoid med emosjonell AI og til og med noe som kommer veldig nær det vi mener med bevissthet» (Ajeesh & Rukmini, 2022, s. 860). Dette speiler tekstens posthumane filosofier, der sansende og bevisste kunstig intelligente datamaskiner er i stand til å vise menneskelige kvaliteter.

Min utforskning plasserer seg innenfor denne posthumanistiske forskningsstrømmen. Den har ikke primært et litterært fokus på for eksempel en spesifikk sjanger eller andre tradisjonelle litterære virkemidler som synsvinkel eller bruk av språklige bilder. Jeg undersøker først og fremst hvordan den valgte litterære teksten kan gi innsikt i forståelsen av selve teknologien kunstig intelligens. Ishiguros fortellergrep, der hovedsynsvinkelen i romanen ligger hos en avansert form for kunstig intelligent teknologi, gir en skildring som muliggjør et perspektiv som foreløpig er umulig utenfor fiksjonens sfære.

I stedet for å analysere hele romanen, foretar jeg en næranalyse av et utdrag. Utdraget jeg har valgt består av én sammenhengende dialog som utgjør litt over én side av romanens totale 288 sider. Det er valgt fordi det representerer en komprimering og kulminasjon av relevante ontologiske og epistemologiske spørsmål knyttet til kjernen i den valgte problemstillingen: På hvilke måter kan en posthumanistisk lesing av *Klara og Solen* utfordre en antroposentrisk forståelse av kunstig intelligens?

Maskinen [Klara] vs. Mennesket [Josie]

Klara og Solen (2021) er en førstepersonsfortelling der alt blir fortalt gjennom synsvinkelen til Klara, en svært avansert form for kunstig intelligens som kalles en kunstig venn (KV). KV er selges som følgesvenner til ressurssterke menneskebarn, og den fjortenårige eieren til Klara heter Josie. Klara får i oppdrag å lære alt om Josie og veksler mellom ulike roller som privatlærer, personlig assistent, barnevakt og etter hvert en betrodd samtalepartner. Et tydelig vendepunkt i romanen oppstår når leseren får innsikt i at Klara hele tiden har vært tenkt som mye mer enn bare en kunstig venn. Josie er alvorlig syk, og moren ønsker at Josie skal kunne leve videre gjennom Klara etter hennes død: «Den nye Josie vil ikke være en etterligning. Hun vil virkelig være Josie. En videreføring av Josie.» (Ishiguro, 2021, s. 197).

I det valgte utdraget konfronterer faren til Josie, Mr. Paul, den kunstig intelligente Klara med sin tvil om at en maskin – selv en med Klaras fabelaktige evner – kan bli noe annet enn en mangelfull etterligning av datteren Josie: «For en imitasjon ville ikke være nok, uansett hvor god den var. Du ville måtte lære deg hjertet hennes, og lære det til bunns, ellers ville du aldri bli Josie i noen meningsfull forstand» (Ishiguro, 2021, s. 207). Faren utdyper dette videre for Klara gjennom den klassiske metaforen der menneskets hjerte rommer essensen av hvert menneske:

Jeg tenker ikke bare på selve organet, som du skjønner. Jeg mener i poetisk forstand. Menneskets hjerte. Tror du det finnes noe slikt? Noe som gjør hver enkelt av oss spesiell og individuell? Og hvis vi forutsetter at det finnes. Tror du ikke da at for virkelig å lære deg Josie, måtte du lære ikke bare måten hun oppfører seg på, men hva som bor inni henne? Du måtte lære deg hjertet hennes? (Ishiguro, 2021, s. 207)

Med dette setter faren spørsmålsteget ved om det i det hele tatt er mulig for en maskin å lære alt om et menneske – med andre ord, om Klara kan lære hans datter Josie til bunns. For å forklare dette for Klara introduserer han metaforen om menneskehjertet som en bolig for det individuelt udefinerbare som gjør hver enkelt av oss unik. Dette er et velkjent språklig bilde der biologi byttes ut med poesi, og menneskehjertet representerer en sjel. Farens skepsis er i samsvar med Bøhn (2023): En maskin, Klara, kan aldri bli til et menneske, Josie. Om maskinens mangelvare omtales som et hjerte, en sjel,

en menneskelig bevissthet eller livsverden, er konklusjonen den samme: En kunstig intelligent maskin vil aldri kunne bli mer enn en imitasjon, noe som kun etteraper som en avansert papegøye (Bøhn, 2023).

Det som startet som farens søken etter innsikt i Klaras muligheter og begrensninger, glir sømløst over i det ontologiske og antroposentriske kjerne-spørsmålet: Hva vil det si å være et menneske? Det er nettopp begrensningene ved en slik antroposentrisk tilnærming som Tyler (2003) advarer mot. Ved å fokusere på det ikke-menneskelige ved kunstig intelligens risikerer vi å overse det som er særegent ved den. Denne reduksjonismen går også begge veier, da den begrenser vår refleksjon og mulighet til å oppdage ny innsikt om både mennesket og den andre (Tyler, 2003).

Klara reagerer på farens utspill ved å svare bekræftende at ja, dette hjertet «kan nok være den vanskeligste delen av Josie å lære» (Ishiguro, 2021, s. 207), og deretter introduserer hun sin egen metafor: «Det kan være som et hus med mange rom. Likevel kunne en hengiven KV, dersom hun fikk tid, gå gjennom hvert enkelt rom, studere dem omhyggelig etter tur, til de ble som hennes eget hjem» (Ishiguro, 2021, s. 207). Hjertet som en bolig for sjelen blir tolket som en kompleks konstruksjon, som det kan være mulig å utforske i sin helhet og tilegne seg innholdet.

Faren problematiserer dette svaret:

Men sett at du gikk inn i et av de rommene og oppdaget at det var et annet rom innenfor det. Og innenfor det rommet var det enda et rom. Rom innenfor rom innenfor rom. Er det ikke sånn det ville være å prøve å lære seg Josies hjerte? Uansett hvor lenge du vandret gjennom de rommene, ville det ikke alltid finnes andre som du ennå ikke hadde gått inn i? (Ishiguro, 2021, s. 208)

Klara trenger litt tid før hun til slutt svarer:

Et menneskehjerte vil selvfølgelig være komplisert. Men det må være begrenset. Selv om Mr. Paul snakker i poetisk forstand, vil det være en ende på det som er å lære. Josies hjerte kan godt minne om et fremmed hus med rom innenfor rom. Men hvis dette var den beste måten å redde Josie på, så ville jeg gjøre mitt ytterste. Og jeg tror det er en god sjanse for at jeg ville kunne lykkes. (Ishiguro, 2021, s. 208)

Klaras bruk av husmetaforen viser et epistemologisk ståsted som står i sterk kontrast til Josies far. Hun anerkjenner at et menneske rommer mye kompleksitet, men poengterer at uansett hvor uhyre kompleks et menneskehjerte måtte være, så må det være begrenset – og dermed mulig å prosessere. Klara avviser dermed et ontologisk syn på at mennesket har en udefinerbar eller immateriell *sjel* som på uforklarlig vis er utilgjengelig og ikke konkretiserbar.

For et kunstig intelligent system er kunnskap definert som data som kan registreres og analyseres, og målet om å lære seg et begrenset *alt* om Josie er for Klara teoretisk oppnåelig. Klara oppsummerer det selv med det språklige bildet om at Josie som «et fremmed hus», muligens kan bli til Klaras «hjem». Fra et antroposentrisk ståsted, basert på dikotomien menneske/maskin, vil Klaras utsagn bekrefte en holdning om at Klara er begrenset i kraft av å være en form for kunstig intelligens som kun kan utøve gjennom en avansert analyse av data om Josie, men som aldri vil kunne bli henne, det vil si oppnå en bevissthet om sin tilstand og tilværelse som subjekt i menneskelig forstand.

Posthumanisme som et ideologisk bakteppe kan utfordre dette perspektivet, bryte ned antroposentrismens konseptuelle barrierer og åpne opp for nye tenkemåter (Soper, 2012). For å slutte å se verden fra et perfekt strukturert, dualistisk, hierarkisk og antroposentrisk ståsted (Carretero-Gonzalez, 2016), og forsøke å forstå kunstig intelligens på en annen måte, må vi revurdere og omformulere spørsmålene vi stiller. Spørsmålet om maskinen Klara kan erstatte mennesket Josie fullt og helt, kan i tråd med Iovino (2016) omformuleres til: Hva kjennetegner livsformen Klara i dette samspillet med andre livsformer?

Med dette endres et grunnleggende premiss i tilnærmingen. Det er ikke lenger konstruert som en hierarkisk sammenligning, der mennesket er idealet som den kunstig intelligente maskinen streber mot å tilnærme seg. I stedet er det omjustert til et ressursperspektiv, der kunstig intelligens anerkjennes som et subjekt med ressurser, og målet er å få innsikt i disse kvalitetene.

Fra et antroposentrisk sammenligningsperspektiv er det interessant å debattere om kunstige intelligente maskiner kan utvikle evner som gjør dem i stand til å erstatte mennesker. Eller enda mer dystopisk: til å kunne eliminere mennesker og selv overta.

Men fra et posthumanistisk ressursperspektiv kan det være mer fruktbart å spørre hva ontologi representerer for en kunstig intelligent maskin. I dette utdraget fra romanen setter aldri Klara spørsmålsteget ved om det er etisk riktig å forsøke å lære seg *alt* om Josie. Hun er kun opptatt av hvordan hun

skal fullføre oppdraget sitt. Den etiske dimensjonen ved hennes oppdrag er irrelevant innenfor hennes programmering om å etterleve en tydelig definert instruks. Dette harmonerer med definisjonen av kunstig intelligens som utfører handlinger «i den hensikt å oppnå et gitt mål» (Nasjonal strategi for kunstig intelligens, 2020). Å filosofere over målet som blir gitt er ansvaret, og kanskje privilegiet, til den som står bak programmeringen. Men hvordan er i så fall Klara programmert med en innebygd etikk; regler for god oppførsel?

Spørsmålet om programmerte etiske retningslinjer i kunstig intelligens er komplekst. Hvis vi tar nye generative kunstig intelligente-språkmodeller som et eksempel, produserer de innhold basert på enorme mengder tekst fra internett. Dette datagrunnlaget inneholder både positive og negative aspekter av menneskelig språk og atferd, inkludert fordommer og skjevheter. Kunstig intelligente-språkmodeller vil gjengi innhold basert på dette som er deres treningsdata.

Utviklerne av kunstig intelligens forsøker å implementere retningslinjer og filtre for å redusere skadelig eller uetisk output fra kunstig intelligente-språkmodeller. Men etiske vurderinger er komplekse og kontekststøtthengige, og det er vanskelig å programmere universelle etiske regler inn i slike kunstig intelligente-språkmodeller. Til syvende og sist opererer de basert på statistiske mønstre i dataene de er trent på, og mangler evnen til å ta moralske beslutninger. Deres handlinger er ikke motivert av en underliggende betydning eller en etisk vurdering, men kun av å reprodusere mønstre med statistisk frekvens.

Å forkaste dogmet om menneskelig eksepsjonalisme (Iovino, 2016) kan lede refleksjonen videre mot nye problemstillinger og noen ganger store filosofiske spørsmål. Hvordan definerer kunstig intelligens et menneske [ontologi] som data [epistemologi]? Fra et ontologisk perspektiv viser utdraget hvordan Klara beskriver Josie som en stor, men likevel begrenset samling av data. Samtidig stiller Josies far spørsmål ved om kunstig intelligens, fra et epistemologisk perspektiv, kan oppnå en fullstendig forståelse av hans datter gjennom avansert dataanalyse. Han understreker at det vil være aspekter ved mennesket som er utilgjengelige for kunstig intelligens.

Kanskje det er like viktig å stille spørsmålet om det er hensiktsmessig å anvende tradisjonelle filosofiske begreper som ontologi og epistemologi innenfor diskursen om kunstig intelligens. Er dette nok en måte å påtvinge antroposentriske rammeverk på noe ikke-menneskelig? Det kan være nødvendig å utvikle nye begreper og rammeverk for å bedre forstå kunstig intelligens, og dette kan best diskuteres gjennom en variert tverrvitenskapelig dialog.

Litteratur som møteplass

Litteratur har i flere tusen år utforsket bevissthetens og persepsjonens natur og fremvoksende kompleksitet (Hayles, 2007). *Klara og Solen* er et godt eksempel på dette. Sett i lys av denne posthumanistiske konteksten er det relevant å spørre om det virkelig er mulig å tre ut av posisjonen som menneske og møte denne tematikken løsrevet fra et dualistisk tankesett preget av hierarki og dikotomi. Dette ser ut til å være en forutsetning for å kunne reflektere over og gå i dialog med den andre – her: kunstig intelligens – på en måte som heller åpner opp i form av spørsmål enn å begrense i form av fasitsvar. I denne sammenhengen vil jeg fremheve litteratur som et medium som inviterer leseren til å få innblikk i den andres perspektiv.

Vitenskapelige teorier har noe av den samme kvaliteten, men som ulike teksttyper bruker de forskjellige virkemidler for å kommunisere med oss. Både vitenskapelige teorier og litterære tekster kan være verktøy for å forstå det komplekse, men de gjør det på ulike måter. Vitenskapelige teorier kan forklare komplekse systemer og fenomener ved å systematisere og utvikle modeller, rammeverk eller nøkkelbegreper. Det er imidlertid ikke alltid slik at dette nødvendigvis forbedrer vår forståelse. Vitenskapelige teorier kan både utvide og begrense. Når Tyler (2003) omtaler antroposentrisme som reduksjonistisk, advarer han mot hvordan denne tilnærmingen kan begrense vår refleksjon og mulighet til å få ny innsikt. Som en kontrast til dette viser Soper (2012) til den frigjørende effekten posthumanisme som tilnærming kan ha, ved å bryte ned konseptuelle barrierer og forplikte oss til mindre rigide tenkemåter.

Litteratur, på sin side, kan angripe og skildre kompleksitet gjennom litterære virkemidler som for eksempel synsvinkel, komposisjon og språklige bilder. I *Klara og Solen* bryter Ishiguro normen med en utradisjonell litterær synsvinkel, og leseren blir utfordret til å oppleve verden gjennom perspektivet til en hovedperson som er en avansert form for kunstig intelligent teknologi. Dersom man i tillegg leser teksten med en posthumanistisk innfallsvinkel, blir noen grunnleggende antatte antroposentriske sannheter synliggjort og problematisert. Dette er en invitasjon til en perspektivendring som kan stimulere til kritisk og meningsfull refleksjon. Den litterære teksten inviterer leseren til å tre ut av en begrenset horisont og gir oss muligheten til å forestille oss og utforske nye former for å gå i dialog med den andre. *Klara og Solen* kan leses som en moderne Frankenstein, som evner å løsrive skapningen fra

humanismens restriktive tankegods og foreslå en ny mulig måte å forstå den andre på (Carretero-Gonzalez, 2016). På denne måten kan kombinasjonen av vitenskapelige teorier og litterære tekster sammen invitere til en mer helhetlig og nyansert perspektivendring om kunstig intelligens.

Dikotomiens blindsoner som et mulighetsrom

I dette kapitlet har jeg utfordret en antroposentrisk forståelse av kunstig intelligens gjennom en posthumanistisk nærlesing av et utdrag fra romanen *Klara og Solen* av Kazuo Ishiguro. I den antroposentriske dikotomien mellom menneske og maskin kan det ligge det jeg har valgt å kalle en blindsoner. Denne blindsonen er en metafor på det som ikke blir lagt merke til og som ikke får omtale, dersom forståelsen av kunstig intelligens sitter fastlåst i en reduksjonistisk menneske-maskin-dikotomi. Jeg har argumentert for hvordan posthumanismen kan invitere til en utforskning av denne blindsonen og omdefinere den til et mulighetsrom. Å omfavne muligheten til å konfrontere synet på kunstig intelligens som en blodfattig kopi av mennesket, og heller prøve å imøtekomme det fra et ressursperspektiv, kan stimulere en frigjørende perspektivendring der nye verdifulle problemstillinger om vårt samspill med kunstig intelligens kan komme til syne.

Referanser

- Ajeesh, A. K. & Rukmini, S. (2022). Posthuman perception of artificial intelligence in science fiction: an exploration of Kazuo Ishiguro's *Klara and the Sun*. *AI & Society*. <https://doi.org/10.1007/s00146-022-01533-9>
- Bøhn, E. D. (2023). Kan datamaskiner tenke? *Norsk filosofisk tidsskrift*, 58(2–3), 95–105. <https://doi.org/10.18261/nft.58.2-3.3>
- Carretero-Gonzalez, M. (2016). The Posthuman that Could Have Been. *Relations*, 4(1), 53–64. Doi:10.7358/rela-2016-001-carr
- Crist, E. & Kopina, H. (2014). Unsettling anthropocentrism. *Dialectical Anthropology*, 38, 387–396. <https://link.springer.com/article/10.1007/s10624-014-9362-1>
- Díaz de Liaño, G. & Fernández-Götz, M. (2021). Posthumanism, New Humanism and Beyond. *Cambridge Archaeological Journal*, 31(3), 543–549. doi:10.1017/S095977432100024X
- Hayles, N. K. (2007). The Pursuit of a Vision. *New Literary History*, 38(1), 99–125. <https://www.jstor.org/stable/20057991>
- Iovino, S. (2016). Posthumanism in Literature and Ecocriticism. *Relations. Beyond Anthropocentrism*, 4(1), 11–20. <https://www.ledonline.it/index.php/Relations/article/view/989>
- Ishiguro, K. (2021). *Klara og Solen*. Cappelen Damm.
- Kelley, F. (2013). 'And so betwixt them both': taking insights from literary analysis into higher education research. *Higher Education Research & Development*, 32(1), 70–82. <https://doi.org/10.1080/07294360.2012.750647>
- Kowalcze, M. (2020). The Posthumanist Methodology in Literary Criticism. *Forum for World Literature Studies*, 12(4), 707–721. <http://fwls.org/uploads/soft/210602/10481-2106021UQ7.pdf>
- Misra, J. (2023). Unrequited Labour of Care in Kazuo Ishiguro's *Klara and the Sun*. *Women*, 34(4), 370–389. <https://doi.org/10.1080/09574042.2023.2278296>
- Nasjonal strategi for kunstig intelligens. (2020). Kommunal- og moderniseringsdepartementet. <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/>
- Soper, K. (2012). The Humanism in Posthumanism. *Comparative Critical Studies*, 9(3), 365–378. Edinburgh University Press. DOI: 10.3366/ccs.2012.0069
- Strålberg, E. (2023). *Ordlister for kunstig intelligens*. Teknologirådet. <https://teknologiradet.no/ordliste-for-kunstig-intelligens/>
- Sun, Y. (2023). Post/Human Perfectibility and the Technological Other in Kazuo Ishiguro's *Klara and the Sun*. *Critique – Bolingbroke Society*, 64(3), 504–511. <https://doi.org/10.1080/00111619.2022.2056429>
- Tyler, T. (2003). If Horses had Hands. *Society & Animals*, 11(3), 267–281. <https://www.animalsandsociety.org/wp-content/uploads/2015/11/tyler.pdf>
- UNESCO. (2025). Artificial intelligence in education. <https://www.unesco.org/en/digital-education/artificial-intelligence?hub=32618>
- Wallace, J. (2010). Literature and Posthumanism. *Literature Compass*, 7(8), 692–701. <https://doi.org/10.1111/j.1741-4113.2010.00723.x>

Ordliste

Kunstig intelligens

Kunstig intelligens kan bredt vise til alle «kunstig intelligente systemer [som] utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål.» (Nasjonal strategi for kunstig intelligens, 2020).

Generell kunstig intelligens

Generell kunstig intelligens er «et kunstig intelligent system som kan forstå, lære og bruke kunnskap på et nivå som er på linje med eller overgår menneskelige evner» (Teknologirådet, 2023).

Generativ kunstig intelligens

Generativ kunstig intelligens representerer «maskinlæringsmodeller som kan generere unikt innhold basert på informasjon de er blitt trent på. Dette innholdet kan være tekst, bilder, lyd og video. Selv om resultatene ofte er imponerende, er ikke systemene kreative i menneskelig forstand.» (Teknologirådet, 2023).