KRONET MED HÆDER OG ÆRE

Potentielle konsekvenser for gudbilledligheden ved udviklingen af kunstig generel intelligens

Adjunkt, ph.d., cand.theol. Michael Agerbo Mørch

Resumé: Denne artikel giver en teologisk vurdering af de potentielle konsekvenser for gudbilledligheden, hvis kunstig generel intelligens (AGI) udvikles. I artiklens første del diskuterer jeg to filosofiske kritikker af AGI. Den første kritik kommer fra filosoffen Hubert Dreyfus, som i en række indflydelsesrige værker fra 1965 og frem har peget på fundamentale problemer for AGI-projektets gennemførlighed. Den anden kritik kommer fra psykiateren og filosoffen Thomas Fuchs, der med indsigter fra både kognitionsforskning og fænomenologi ligeledes kritiserer AGI-projektets gennemførlighed. Efter at have introduceret disse to væsentlige vurderinger af AGI-projektet vil jeg i artiklens andel del drøfte potentielle konsekvenser for gudbilledligheden, hvis AGI alligevel udvikles.

> *The chief objection to playing God is that someone else is God already.*
> J. Budziszewski

Introduktion

Vil det få konsekvenser for gudbilledligheden, hvis kunstig generel intelligens[1] udvikles? Jeg drøfter dette spørgsmål i denne artikel ved først at diskutere henholdsvis Hubert Dreyfus og Thomas Fuchs' grundlæggende kritikker af AGI-projektet og dernæst kortlægge, hvordan gudbilledligheden er blevet forstået i teologien for at kunne spørge, hvilke af disse modeller der er udfordret, hvis AGI udvikles. Min konklusion vil være, at

gudbilledligheden skal forstås facetteret, og at det kun er visse af disse facetter, der vil blive udfordret af AGI.

Her i introduktionen kan det dog være nyttigt at forklare, hvorfor temaet om AGI bringes i spil i et tidsskrift, der sigter på skæringspunktet mellem fagteologi og kirkeliv. Det skyldes dels det elementære forhold, at kristne er kaldet til at engagere sig i verden (at forstå den, udlægge den, forvalte den, forandre den), dels at kunstig intelligens allerede er omfattende til stede i det moderne samfund, og dets betydning vil kun øges i de kommende år. Derfor er både teologens, præstens og menighedens liv allerede nu påvirket af AI, og da præsten er både eksistentiel og etisk vejleder for mange mennesker, er det nødvendigt at have ressourcer til rådighed, der klæder på til denne opgave. Artiklens fokus er alligevel holdt snævert på de faglige diskussioner, men der er altså dette pastorale sigte nedenunder, som læseren inviteres til at overveje og bringe i spil i andre sammenhænge.[2]

Der er desuden en apologetisk grund til, at teologer bør beskæftige sig med AI. Ser man på perspektiverne, som de udlægges i den israelske historiker Yuval Hararis bestseller *Homo Deus*, er hele teologiens udlægning af mennesket i fare for at falde sammen.[3] Den stærke polemik imod kristendommen, som dette værk rummer implicit og eksplicit, kalder på en saglig teologisk vurdering.[4] For Harari handler det primært om overvindelsen af døden, som han ser som et rent teknisk problem (Harari 2015, 25). Hvis døden vitterligt overvindes, vil religioner også forsvinde, da de konstrueres og næres af menneskets dødsangst. AGI spiller her en central rolle, og man skal ikke underkende Hararis valg af bogtitel: Perspektivet for ham er vitterligt, at teknologien er grænseløs, og de egenskaber, vi på feuerbachiansk vis har tillagt de konstruerede guder, kan vi nu selv erhverve gennem teknologiske adaptationer.[5]

Skal vi nøjes med at trække på skuldrene af disse vidtløftige profetier og visioner? Jeg foreslår i stedet, at vi forholder os til Amaras lov, der siger, at mennesket har en tendens til at overvurdere teknologiens effekt på den korte bane og undervurdere teknologiens effekt på den lange bane (Dorobantu 2020a, 113). Så hvor futuristerne formentlig har for store forhåbningerne til den sekulære, teknologiske eskatologi, har blandt andet konservative kristne nok en tendens til at undervurdere fremskridtets muligheder og udfordringer. Med denne dobbelte motivation turde det være tydeligt, at teologer har behov for at arbejde med det vidtforgrenede felt som AI-forskningen er, og jeg vil derfor nu gå til artiklens første hoveddel, som indkredser, hvad AGI overhovedet er.

Hvad er AGI?

Begrebet »kunstig intelligens« (AI) blev dannet af computerforskeren John McCarthy, der var førsteforfatter på konferenceoplægget til den berømte sommerkonference i Darmouth i 1956, hvis deltagere også siges at have initieret forskningsfeltet (McCarthy et al. 2006).[6] Begrebet AI forstås i dag på flere måder, så nogle fokuserer på menneskelignende komponenter i intelligens og ageren (fx muligheder for intuition og kreativitet),

mens andre fokuserer på rene syntaktiske operationer. Det er ikke afgørende for denne artikel med en præcis terminologisk afklaring, det er tilstrækkeligt at klargøre, at det handler om forsøget på at skabe intelligens. AI forstås her som en syntetisk konfiguration, der kan simulere menneskets intelligens, ultimativt så det kan forveksles med menneskelig kapacitet eller endda overgå denne.

Mere afgørende for artiklens diskussioner er det at klargøre en distinktion mellem to typer af AI, som er blevet konceptualiseret af filosoffen John Searle. I en artikel i 1980 indførte han en skelnen mellem stærk og svag AI (Searle 1980, se også Kurzweil 2005, 259-298). Svag AI forstår computeren som et nyttigt værktøj til at studere hjernen, mens stærk AI hævder, at computeren ikke bare er et værktøj men selv en »hjerne«, der under de rette betingelser kan – eller kan komme til at – forstå. Man kan også sige, at forskellen mellem svag og stærk AI er forskellen mellem syntaks og semantik. Det kommer i øvrigt frem i Searles berømte tankeeksperiment om det kinesiske rum, som jeg vender tilbage til i afsnittet om Thomas Fuchs.

I den filosofiske debat er Searles definitioner hyppigt anvendte, men i bredere debat bruges begreberne lidt anderledes. Her anvendes ofte begreberne »smal« og »generel« AI. Smal AI henviser her til en computer eller en robot, der udfører specialiserede processer i et bestemt miljø og derigennem interagerer med sit miljø. Smal kunstig intelligens er allerede implementeret overalt i samfundet, og der er tilsyneladende meget mere på vej.[7] Den generelle AI (AGI) handler om en computer eller en robot, der er sammenlignelig med menneskets kapacitet på alle intellektuelle parametre.[8] Den har været spået i årtier, men indtil videre har resultaterne på trods af omfattende forskning været forbavsende begrænsede.[9] Det har fået kritikere til at hævde, at det slet ikke er muligt at realisere AGI – ikke fordi de mangler fantasi til at forestille sig en sådan intelligens, men fordi de hævder, at der er nogle fundamentale filosofiske og antropologiske (herunder biologiske og bevidsthedsmæssige) problemstillinger, som ikke *kan* løses. I det følgende skal vi se på to af de mest indflydelsesrige kritikker af AGI-projektet. Først Hubert Dreyfus' kritik, dernæst Thomas Fuchs'.

Kritikken fra Hubert Dreyfus

Filosoffen Hubert Dreyfus er kendt for sine studier i Martin Heideggers og Michel Foucaults filosofier. Men da han i 1964 blev hyret af RAND Corporation til at evaluere Allen Newell og Herbert Simons forskning i kunstig intelligens, fik hans filosofi en ny agenda. I en række polemiske artikler og bøger agiterede han kraftigt imod hele projektets legitimitet, og han forsøgte at vise, at AGI er en umulighed.

Dreyfus' AI-filosofi er altså skabt af sin modsætning, hvorfor den position kort må opsummeres. Newell og Simon udviklede, hvad de kaldte »Physical Symbol System Hypothesis« (PSSH), som kort går ud at skabe nogle produktionsregler for den menneskelige evne til problemløsning, som ud fra disse regler kan simuleres kunstigt.[10] Men New-

ell og Simons syn hviler på fire antagelser, der ifølge Dreyfus blev taget som aksiomer, selvom de kun er (tvivlsomme) hypoteser (Dreyfus 1972, 67-139).

(1) Den biologiske antagelse

Antagelsen er, at på neuronniveau processerer hjernen informationer, der svarer (er en biologisk ækvivalent) til Boolean on/off-switches. I boolsk algebra opererer man kun med to variabler med værdierne sand eller falsk, oftest formaliseret som 1 og 0, og det er denne algebra, der er fundamentet for moderne teknologi. Hvis en informationsport står åben, svarer det til 1, hvis den er lukket, svarer det til 0. Ideen er nu, at hjernens neurotransmittere giver eller tilbageholder signaler på en transparent måde, der kan beskrives ved hjælp af boolsk algebra.

Problemet ifølge Dreyfus er dog, at den neurologiske forskning har vist, at hjernen ikke bedst forstås som en digital, men en analog informationsprocessor. Forskellen er, at i en digital anordning repræsenterer hvert enkelt element et symbol, mens det i analoge anordninger vil være en mere global proces, fordi det er fortløbende fysiske variabler, der repræsenterer den information, der processeres (se også Fuchs nedenfor). Hjernen sender konstant en kaskade af elektriske impulser afsted, og kun hvis hver eneste impuls er bærer af en distinkt værdi og repræsenterer et distinkt led i en informationskæde, kan man argumentere for, at hjernen ækvivalerer en digital computer. Derfor kan man sige, at hjernen rummer analoge komponenter i afsendelsen af neuronsignaler, som gør, at der er en essentiel *interaktiv* karakter i biologien, som ikke forefindes i maskiners organisation (Dreyfus 1972, 73).

(2) Den psykologiske antagelse

Antagelsen er, at hjernen kan forstås som en mekanisme, der opererer med stumper af information efter formelle regler. Denne antagelse er for eksempel grundstenen i, hvad man kalder »the computational theory of mind« (CTM), som den er artikuleret af for eksempel Jerry Fodor og John Searle (Chalmers 2019).[11]

Problemet ifølge Dreyfus er, at man her springer for hurtigt fra hypotesen om, at alle tanker kan formaliseres, til den ganske stærke påstand, at enhver proces i hjernen *er* formaliserbar (Dreyfus, 80). Det kan kun lade sig gøre, fordi man forstår »informationsprocesser« meget vagt, og fordi man sammenblander fysiske og fænomenologiske perceptionsniveauer (Dreyfus 1972, 99-100).[12] Dreyfus' pointe er, at fænomenologiske indtryk altid er kontekstbestemte, så den ubevidste baggrundsinformation, der hviler på menneskets *common sense*-viden, altid spiller med på en dynamisk måde. Som sådan kan den ikke formaliseres, fordi den er ubevidst (intuitiv) og plastisk (dynamisk). Alt dette betyder ikke, at der ikke er noget beregningselement ved menneskelig bevidsthed, blot at den hårde reduktionisme i CTM ikke er overbevisende (Dreyfus 1972, 101).

(3) Den epistemologiske antagelse

Antagelsen er, at al viden kan formaliseres, det vil sige, alt hvad der kan forstås af den menneskelige hjerne kan oversættes til logiske relationer efter boolsk logik. Her isolerer Dreyfus to påstande, som han begge afviser. Først påstanden, at alle ikke-arbitrære handlinger kan formaliseres efter bestemte regler, dernæst påstanden, at en computer skulle kunne reproducere disse handlinger ved hjælp af formaliseringer (Dreyfus 1972, 102).

Problemet er igen, at bevidstheden om kontekstens betydning forsvinder. Dels er kontekst altid afgørende for menneskelig opfattelse og kommunikation, dels kan mennesket initiere forskellige handlinger, alt efter hvordan konteksten fortolkes, dels kan mennesket handle og kommunikere, selv når reglerne for handlingskoordination eller handlingsinteraktion brydes. Pointen er, at der *må* være et element af menneskelig kognition, som ikke kan reduceres til symbolik (Dreyfus 1972, 115).

(4) Den ontologiske antagelse

Den fjerde antagelse er, at al relevant information om verden i princippet kan analyseres via situationsfrie elementer, det vil sige uden kontekst, der determinerer betydningen. Denne vidtrækkende påstand findes for eksempel hos én af AI-begrebets fædre, Marvin Minsky.[13] Ifølge denne antagelse kan al menneskelig viden, al kognitiv produktion, reduceres til objekter (fx »categories of objects«, »properties of objects«, »relations between objects«, jf. Dorobantu 2020a, 144), og når de er objekter, er der et begrænset antal af dem, hvorfor det er principielt muligt at kortlægge alle informationer, oplagre dem og endda eksternalisere dem.[14]

Problemet er igen konteksten. Mennesket befinder sig altid i situationer, der er relative og dynamiske, mens fysiske systemer kan generaliseres og simplificeres. Med andre ord er der et subjektivt element, som ikke kan isoleres og fjernes uden at genkendeligheden forsvinder. Der er en fælles fysisk situation for alle læsere af denne sætning, men omstændighederne er oplagt forskellige, og enormt mange faktorer spiller ind på en lang variation af fortolkningsmuligheder (jf. Dreyfus 2020, 125-126). At overse kontekstens betydning er ikke blot en mangel, men får den fatale konsekvens, at vi slet ikke længere taler om intelligens.

Dreyfus' anden fase

AI-udviklingen går hurtigt, både fordi computerkraften er voksende (jævnfør Moore's lov, der siger, at antallet af transistorer (og dermed computerkraften) i et givent kredsløb vil fordobles hvert halvandet år), og fordi datalogerne udvikler nye systemer, der på mere raffineret vis simulerer menneskelig kognitiv aktivitet (såkaldt *machine learning*, se nedenfor). Man kan derfor tale om en anden fase i Dreyfus' kritik, fordi udviklingen i AI-forskningen krævede en tilføjelse til den oprindelige kritik i fase 1 (Dreyfus 1992). I 1970'ernes og 1980'ernes AI-forskning var såkaldte ekspertsystemer omdiskuterede,

og man havde store forventninger til udviklingen af maskiner, der kunne efterligne menneskets evne til problemløsning. Problemløsning bliver af mange set som en helt grundlæggende komponent ved menneskelig intelligens, hvorfor et gennembrud her ville være afgørende for AGI-projektet. Problemet for Dreyfus var, at AGI-forskningen ikke tilstrækkeligt forstod, hvordan den menneskelige hjerne differentierer sin problemløsningsaktivitet i to separate moduler. Inspireret af Heidegger talte Dreyfus om forskellen på »knowing that« (Heideggers »vedhåndenværende«) og »knowing how« (»forhåndenværende«)[15] – en skelnen, der senere er blevet empirisk underbygget i adfærdsforskningen hos Daniel Kahnemann og Amos Tversky med deres opdeling af menneskelig tænkning i »system 1« og »system 2« (Kahneman 2011).[16] Ifølge Dreyfus foregår megen menneskelig problemløsning intuitivt som følge af vores sunde fornuft (*common sense*) og baggrundsviden (*background knowledge*). Denne spontane, hurtige ræsonnementsevne kan ikke oversættes til en algoritme, fordi den er før-sproglig. Vi vurderer ubevidst et problems kontekst, problemkomponenternes komparative væsentlighed og hvilke værdier, der er aktuelle at aktualisere i den givne situation. Med Michael Polanyi handler det om »tavs viden«, som gør, at vi kan handle præcist, korrekt og fornuftigt, og vi kan løse opgaver (dvs. problemer) uden at kunne artikulere en instruktion til forklaring af vores evne (Polanyi 1958).[17] Jeg kan cykle, men hele den fysiologiske, kognitive og adaptive baggrund for det kender jeg ikke – jeg kan bare, jeg ved bare, hvordan man gør. Naturligvis er det muligt at forklare, hvordan man cykler, hvis man sætter tempoet ned, isolerer problemet, og løser det fra sin givne kontekst. Så er vi ovre i »knowing that«, og den type viden anerkender Dreyfus, at vi kan formalisere. En robot kan derfor lære at cykle, for vi kan programmere efter vores viden (selvom det tilsyneladende er overraskende besværligt, jævnfør Moravecs paradoks, der siger, at ræsonnementer – modsat vores intuitive opfattelse – kræver begrænset beregningskraft, mens sensoriske opgaver kræver enorm beregningskraft (Dorobantu 2020a, 116) – men pointen er, at den ikke lærer at problemløse hverken læringsmæssige eller trafikale opgaver *som et menneske*, fordi vi oftest gør dette intuitivt.[18] Modsat en computer behøver vi ikke gennemtænke alle mulige alternativer, men kan oftest vælge den rigtige løsning i første forsøg, og den menneskelige evne til problemløsning *kan* derfor ikke simuleres kunstigt (Dreyfus, 1986).

Dreyfus' kritik er stadig omdiskuteret, men er naturligvis udvalgt her, fordi jeg mener, den er overbevisende. Alligevel er der også forældede træk ved den, for eksempel at den af naturlige grunde ikke har de nyeste kognitionsvidenskabelige resultater indarbejdet. Det har til gengæld kritikken fra Thomas Fuchs, hvortil vi nu vil vende os.

Kritikken fra Thomas Fuchs

I den nyligt udgivne *In defense of the Human Being* (2021), kritiserer den tyske psykiater og filosof, Thomas Fuchs, indgående hele AGI-projektet. Ifølge Fuchs er »kunstig intelligens« elementært set et *contradictio in adjecto*: Der *kan* ikke gives en »kunstig«

intelligens, for alene levende væsener kan have intelligens. Intelligens forudsætter simpelthen liv, hvorfor hele projektet er urealiserbart. Dette synspunkt udfolder jeg i det følgende.

Fuchs tilstræber at vise, at en række »magic words« ikke magter, hvad de tilskrives. Et hyppigt anvendt begreb som »information« handler for eksempel ikke længere om viden, der gives videre, men om »independent, freely convertible data, signals, and codes, movable and volatile, available for arbitrary access, and only incidentally bound to any material carrier.« (Fuchs 2021, 13-14). Men Fuchs argumenterer i stedet for, at information kun eksisterer, når *nogen* forstår *noget*. Derfor rummer en computer heller ingen information i sig selv; det er for eksempel alene fra et menneskeligt perspektiv, den beregner (en type tankeaktivitet). Fra et datamatisk perspektiv forandrer computeren blot ét elektronisk mønster til et andet via programmerede algoritmer (en algoritme er en konfiguration af information, der passerer gennem kalkulerbare processer). Det følger heraf, at den menneskelige bevidsthed ikke udelukkende kan bestå af information, fordi det kræver bevidsthed at forstå noget som information (ellers må man forstå det som rene data).[19] Derfor er det heller ikke hjernen, der forstår noget, men personen. Konklusionen bliver, at når intelligens ikke kan løsrives fra bevidsthed (fx personens erfaringer og oplevelser), kan den heller løsrives fra det levende, fordi alt vi kender, der har bevidsthed, lever. Derfor kan det heller ikke skabes kunstigt (jf. ovenfor). Problemet for AI-forskningen er, at selv indenfor forskningsfelter som kompleksitetsstudier, kybernetik og bio-informatik, der arbejder med de mest finkornede teorier om information, underkendes, reduceres eller overses de karakteristikker ved livet, som vi genkender klarest: sansende, følende, stræbende, opfattende/perciperende og tænkende erfaringer.[20] At tro, at man kan skabe intelligens løsrevet fra disse erfaringer, er ifølge Fuchs futilt.

Det turde være klart nu, at både Dreyfus og Fuchs peger på den afgørende indsigt, at intelligens er mere end formaliserbare ræsonnementer. For Fuchs fører det ovenstående til to påstande, som han udfolder *in extenso*. I det følgende vil jeg opholde mig ved en præcis bestemmelse af påstandene, fordi de indkapsler Fuchs' fundamentale kritik:

Først diskuterer Fuchs påstanden, »personer er ikke programmer« (Fuchs 2021, 24-28). Kognitionsvidenskaberne og AI-forskningen har »funktionalisme« som filosofisk grundantagelse. Det betyder, at mentale tilstande på tilstrækkelig vis kan forklares som en relation mellem input og output. Et godt eksempel på en repræsentant for dette synspunkt er kognitionspsykologen Steven Pinker. I bogen *How the Mind Works* fra 1997 forfægter han et funktionalistisk syn på kognition: »The mind is a neural computer, fitted by natural selection with combinatorial algorithms for causal and probabilistic reasoning.« (Pinker 1997, 24). Dette syn hviler på en afgørende præmis, som Fuchs kalder »substrate independence«, som betyder, at funktionelle tilstande ikke er bundet til specifikke bærere, for eksempel hjernen. Men Fuchs har i flere værker vist, at intelligens, kognition, tænkning og så videre *alene* kan forstås ud fra et holistisk per-

spektiv (fx Fuchs 2017). Ét af de væsentligste problemer for det funktionalistiske syn er, at erfaringerne forsvinder. Fuchs henviser til filosoffen John Searles berømte tankeeksperiment, »The Chinese Room«, som viser, hvor afgørende subjektet er for forståelsen (Searle 1980). Searle beder os forestille os et rum, hvor en mand, der ikke kan et eneste ord kinesisk, er blevet låst inde. I hånden har han en manual, som angiver alle de nødvendige regler, man skal bruge, for at kunne svare på spørgsmål på kinesisk. Under døren modtager manden nu et spørgsmål fra en kineser, og ved hjælp af manualen sender han det korrekte svar tilbage (input/output). Hvis manualen er god, og spørgsmålene er til at besvare, vil den kinesiske mand uden for døren ikke opdage, at manden inde i rummet ikke kan et ord kinesisk. Alligevel, skriver Searle, vil ingen vel påstå, at manden forstår kinesisk. Manualen stiller de nødvendige syntaktiske hjælpemidler til rådighed for at kunne svare korrekt og anvendeligt, men betydningen (semantikken) er fuldstændig fjernet for staklen i rummet. Pointen er naturligvis, at intentionel bevidsthed er en forudsætning for forståelse. Som Fuchs prægnant skriver: »Understanding meaning or semantics is more than syntax or an algorithm.« (Fuchs 2021, 25). Selv med AGI vil computeren stadig kun udføre syntaktiske operationer, men semantikken vil være lige så tom som hos en traditionel PC.

I tillæg afviser Fuchs, at man kan se hjernen som en computer, fordi det er umuligt at adskille »hardware« og »software« i hjernen. Hjernen er plastisk og rekonfigureres hele tiden, når den er aktiv. Pointen er, at to aktiviteter *aldrig* er 100% identisk i hjernen, hvorfor det er ukorrekt at tale om »data storage« i datamatisk forstand (og dermed gendriver Fuchs indirekte CTM-teorien, jeg nævnte ovenfor).[21] Hjernen er et kropsorgan, og den kan ikke isoleres fra resten. Denne holisme samles af Fuchs i sætningen: »The brain is not an information-processing or computational apparatus, but a highly living, plastic, and dynamic organ.« (2021, 26-27). Ligesom følelser kræver en krop – og robotter med følelser derfor er utænkelige – kræver også tænkning en krop, og derfor er kunstig intelligens en umulighed. Mennesker er ikke programmer, men en holistisk enhed.

Den anden påstand er så, at »programmer er ikke personer.« (Fuchs 2021, 28-35). Menneskets intelligens består i særskilt grad i evne til at skabe overblik via kontrafaktiske ræsonnementer (Pearl 2018). Fuchs kalder det en »eccentric position«, hvor vi for eksempel kan udsætte en lyst for en tid, fordi det vil maksimere lysten: Jeg venter en uge med at rejse til Californien, fordi flytrafikken i denne uge er frygtelig, og vejret ulideligt varmt. Modsat disse reflekterede kontrafaktiske ræsonnementer, som hjælper mig med at træffe gode beslutninger, ved en computer ikke, hvad den laver. Den kan nok vinde et spil skak over en stormester, men den ved ikke, at den spiller skak.[22] Og spørger du AlphaGo Zero om den hurtigste vej fra børnehaven til brugsen, vil den ikke kunne svare.[23] Den kan kun anvende sin algoritme indenfor dens begrænsede område. Selv smartphones' indbyggede assistenter (som Apples *Siri*, Amazons *Alexa* eller Microsofts *Cortana*) ved ikke, hvad de informerer om, og selvom de giver mere og mere præcise svar, er de aldrig intelligente, for de har ingen selvbevidsthed.[24] Umuligheden er

skarpt beskrevet i dette citat: »For a general, fluid, and variable intelligence, one would have to program the systems for everything that a human being has learned about the world implicitly or explicitly—an impossibility.« (Fuchs 2021, 30). AI's problem er ikke komplekst logiske kalkulationer, men i stedet kontekstbestemte common sense-ræsonnementer. Mennesket kommunikerer og tænker i udstrakt grad gennem metaforer, ironi, tvetydigheder og så videre, som ikke kan formaliseres til 0 og 1 (se fx Lakoff og Johnson 1980).[25] Med alt dette sagt, løser computere ikke problemer, de beregner intet, og de træffer ingen beslutninger. De har ingen bevidsthed, ingen kontekstforståelse, og de kan ikke overveje alternative muligheder, der ikke er præ-programmerede (Fuchs 2021, 31).

Ikke engang når det kommer til den nyeste teknologi med »læringsmaskiner« (*machine learning*) ser Fuchs muligheder.[26] Læringsmaskiner kan simulere de adaptive evner, menneskehjernen besidder. Men Fuchs' pointe er igen, at det er et bedrag, for de lærer ingenting. Ideen er, at træningsdata uploades til en maskine, og så lærer den langsomt mønstre at kende, så den genkender noget bestemt og reagerer på det (fx ved at genkende ansigter i biometrisk teknologi). Men maskinen genkender intet, for den har ingen erfaring af genkendelighed. En computer har brug for enormt mange data for blot at lære at genkende en ko, mens et barn har brug for et par stykker, nogle gange kun ét tilfælde. En computer lærer altså ikke, men fungerer som adaptive systemer. Udover »information« er »kompleksitet« det andet magiske ord i denne filosofiske retning. Men en forøgelse af kompleksitet er ingen genvej til hverken intelligens, bevidsthed eller læring. Vi får ikke noget ekstra ved at tilføje computerkraft, vi får blot det allerede kendte endnu hurtigere (øget effektivitet), og det er elementært set, fordi programmer ikke er personer (Fuchs 2021, 43).

Drøftelse af mulige teologiske implikationer

Selvom man følger Dreyfus og Fuchs i deres grundlæggende kritik af AGI-projektet, er det også muligt at anfægte deres kritiske perspektiver, hvilket genstarter drøftelserne af implikationerne. Det kan være, det viser sig, at AGI kan udvikles uden bevidsthed, så selvbevidsthed ikke er en forudsætning for intelligens. Fordi vores viden om bevidstheden er utilstrækkelig, er det vanskeligt at afvise dette endegyldigt, sådan som Fuchs gør. Men det kan også være, at vi ser, at bevidsthed vokser ud af den smalle kunstige intelligens, når de når et bestemt komplekst niveau. Så vil den udvikles som et emergent fænomen, ligesom mange mener at den menneskelige bevidsthed er blevet udviklet som et emergent fænomen i evolutionshistorien (fx Søvik 2022).[27] Derfor kan man stadig lade sig udfordre af den optimisme, som præger bøger af for eksempel Yuval Harari (Harari 2015), Ray Kurzweil (Kurzweil 2005) eller David Chalmers (Chalmers 2022). Deres bøger er bestsellere i den helt store klasse, hvilket vidner om en meget bred interesse for emnet. Og fordi deres optimisme også er båret af en forventning om, at udviklingen af AGI vil være det endelige dødsstød til menneskets særstatus, er det

magtpåliggende for teologien at overveje de mulige konsekvenser for gudbilledligheden, uanset om de så bliver realiseret eller ej. Marius Dorobantu mener derfor også, det er relevant for teologien at beskæftige sig med disse spekulative spørgsmål, selvom AGI endnu ikke er realiseret:

> From a theological perspective, however, the progress of AI so far does not raise too difficult questions. AI is not yet a contender for being either God, or a sufficiently intelligent and free being so that it challenges the distinctive status of humans. On the contrary, if AI reaches another winter and ultimately fails to emulate human-level intelligence, then this could be legitimately interpreted as yet another indication of the specialness of humans. On the other hand, if AI ever reaches or surpasses human level, then theology and theological anthropology are suddenly faced with a host of critical questions. (Dorobantu 2020a, 162).[28]

I forlængelse af Dorobantus vurdering mener jeg også selv, vi stadig bør reflektere over de potentielle konsekvenser for gudbilledligheden ved udviklingen af AGI, selvom AGI for øjeblikket er hypotetisk. Denne diskussion føres da også adskillige steder i faglitteraturen (for en oversigt se Dorobantu 2020a, 13-22 eller Balle 2022). Ifølge Dorobantu, der selv er reserveret over for muligheden for at udvikle AGI, er de klassiske forståelser af gudbilledligheden – den substantielle, funktionelle og relationelle tolkning – alle i fare, hvis AGI udvikles (Dorobantu 2020b, terminologien er adopteret fra Noreem Herzfeld 2002).[29] For i alle tre fortolkninger er menneskets særstatus i skaberværket markeret, og det er netop denne særstatus, som udfordres, hvis mennesket overgås af AGI.

Udover de tre klassiske tolkninger har teologihistorien budt på flere nuanceringer af dogmet. Man kan isolere mindst syv forskellige forståelser af gudbilledligheden, som enten kan betones individuelt eller i forskellige kombinationer i den teologiske systematik. Pladsen tillader ikke en udførlig drøftelse, men i det følgende vil jeg kort uddybe disse forståelser og pege på mulige implikationer ved udviklingen af AGI. Imod Dorobantu vil mit argument være, at ved at se gudbilledligheden som en konfiguration af flere af disse punkter er AGI ikke en akut trussel mod menneskets særstatus. Kombineret med den filosofiske kritik fra Dreyfus og Fuchs står den teologiske antropologi med stærke kort på hånden.

Gudbilledlighed

Den første forståelse af gudbilledligheden siger, at der er noget unikt i menneskets natur, noget substantielt, der afspejler Guds natur (Dorobantu 2020a, 49-61; Peterson 2016, 36). I teologiens historie har »fornuften« (*ratio*) ofte været udpeget som det, der adskiller mennesket fra dyrene, og som det samtidig har til fælles med Gud, selvom Gud alene bærer fornuften i en ultimativ forstand. Indsigter fra kognitiv etologi har dog vist, at højere dyr både har dømmekraft og sprog, men trods alt i en mindre udviklet form

end mennesker. Udvikles AGI vil mennesket miste sin særstatus på dette punkt, da AGI netop er defineret ved at være en intelligens, der er på niveau med eller over menneskets kognitive niveau (Kurzweil 2005, 260).

Den anden forståelse forstår gudsbilleder således, at mennesket repræsenterer Gud i skaberværket ved at varetage distinkte funktioner (funktionalisme) eller optræde i forskellige roller (aktør-teori) (Dorobantu 2020a, 61-68; Peterson 2016, 37-42). Eksegeter har vist, at mennesket i skabelsesmyten får tildelt nogle bestemte roller, som de skal optræde i – konge, præst og gartner – og som hver især har nogle bestemte funktioner (Kofoed, 2015 og 2016). Menneskets funktioner og roller er unikke for arten, men kan være udfordret af AGI, fordi mennesket selv skaber robotter, der varetager de samme funktioner og roller som mennesket. Hvis gudbilledligheden er knyttet til disse funktioner og roller, må man derfor spørge, om det, mennesket skaber i sit eget billede, også vil bære gudbilledligheden? Ved smalt specialiserede opgaver er det svært at argumentere for, men AI implementeres også i flere og flere kreative, spirituelle og moralske miljøer, som kan udfordre menneskets særstatus på disse felter (en god diskussion findes i Turkanik 2021). Kombineres alt dette i AGI, kan det være vanskeligt at opretholde en forståelse af menneskets funktionelle særstatus.

Den tredje forståelse forstår gudbilledligheden som et udtryk for den relationelle ontologi, der er kendetegnende for den immanente trinitet, og som mennesket får del i ved at være skabt i Guds billede (*analogia relationis*) (Dorobantu 2020a, 68-73; Peterson 2016, 42-52). Menneskets liv er derfor en søgen henimod horisontale og vertikale relationer, hvorfor menneskelivet amputeres, når det isoleres (1 Mos 2,18), mens det fuldendes, når det sættes i forbindelse med Gud og næsten (Zizioulas 1997). Det er mindre oplagt, at AGI kan udfordre denne relationelle side af gudbilledligheden, men der er en vis (potentiel) udfordring ved sociale robotter, som i disse år gennemgår en hurtig udvikling (Balle 2022). De sociale robotter kan indgå i facetter af relationalitet, og det tilbud appellerer til flere og flere. Det er svært på nuværende stadie at konkludere på kvaliteten af disse relationer, og hvad det gør ved menneskets individuelle trivsel, sociale sammenhængskraft og transcendensbehov. I fiktionen ser vi flere kvalitetseksempler på, at der er en forventning om, at AGI kan afhjælpe ensomhed og opfylde relationelle behov, for eksempel i Spike Jonzes *Her* (2013) eller nobelprisvinderen Kazuo Ishiguros *Klara og solen* (Ishiguro 2021). Lakmustesten kan måske være, om sociale robotter også kan indgå i fællesskaber med Gud, hvilket kunne være en indikation på, at de har modtaget gudbilledligheden, og netop dette spekuleres der mere og mere i i disse år (fx Panggabean 2022).

Udover de tre klassiske positioner, som netop er beskrevet, er der også andre forståelser, som bidrager til vores indsigt i gudbilledligheden. I det følgende skal tre tilføjelser nævnes. Den fjerde forståelse er eskatologisk, som siger, at gudbilledligheden ligger som et kim i mennesket, som dog først realiseres fuldt ud på den nye jord (jf. Rom 8,29; 1 Joh 3,2) (Dorobantu 2020a, 73-77). Det kan altså ses i analogi med den »pant«, som

Helligånden, iboende i den troendes hjerte, ifølge Paulus er (2 Kor 5,5; Ef 1,14). Nogle mener desuden, at fordi gudbilledligheden kun er en spire, er den både upåvirket af syndefaldet og delt af alle mennesker uafhængigt af gudsforholdet. På den ene side kan man sige, at den eskatologiske forståelse fjerner problemet med AGI. Mennesket har måske nok en særstatus i skaberværket, men den er ikke realiseret og derfor uaktuel. Omvendt kan man spekulere i, om netop AGI kan være en eskatologisk realisering af gudbilledligheden. Særligt hvis vi får en singularitet, hvor menneske og maskine smelter sammen (Kurzweil 2005).

Den femte forståelse er kristologisk, og her forstås gudbilledligheden som en dynamisk størrelse, der kan forekomme i varierende grad (Dorobantu 2020a, 76). Partikulært i Jesus af Nazareth var gudbilledligheden fuldt til stede (Hebr 1,3; Kol 1,15; 2 Kor 4,4), men det betyder ikke, at mennesket som sådan ikke bærer gudsbilledet (en fungibel kvalitet fastholdes). Denne forståelse er sværere at diskutere, fordi det bliver utydeligt, om mennesket har en særstatus udenfor det frelsende forhold til Kristus. Flere af spørgsmålene fra den relationelle forståelse dukker dog op igen

Den sjette forståelse ser gudbilledligheden som menneskets identitet, som jeg særligt har set den udfoldet hos Ryan S. Peterson (Peterson 2016, 53-83). Gudbilledligheden er ikke »noget«, der er rakt til mennesket, men et paradigme, der hviler i Gud. Menneskets identitet som skabt af Gud er derfor ikke en kontingent størrelse, den kan ikke mistes, svækkes eller forandres. Den er givet mennesket som det, der består over tid (lat. *idem/identidem*). Denne forståelse garderer umiddelbart mod AGI's udfordring af gudbilledligheden, fordi den ikke er lokaliseret hos mennesket. Robotterne er ikke skabt af Gud og kan derfor ikke bære den identitet. Men der er også teologer, der taler om mennesket som Guds medskaber (*created co-creaters,* jf. Hefner 2002), og hvad mennesket skaber i sit billede kan siges at have identitet som skabt per stedfortræder.

Where do we go from here?

Der er oplagt meget spekulation i dette, men hvad kan vi så konkludere på AGI's udfordring til gudbilledligheden? Dorobantu mener, at AGI udfordrer de klassiske forståelser, men ved at se gudbilledligheden som en sammensat størrelse, der ikke kan reduceres til én af de seks forståelser ovenfor, gør man dogmet mere robust. Skal AGI udfordre gudbilledligheden afhænger det dels af, at AGI kan udvikles (hvilket er tvivlsomt, jf. Dreyfus og Fuchs), dels af om sociale robotter har en iboende relationalitet, og dels af om vi kan acceptere ideen om mennesket som medskaber, hvis artefakter altså kan bære Guds billede per stedfortræder. Det er ganske mange tvivlsomme data at skulle forene i én teori, hvorfor den sammensatte gudbilledforståelse er stærk, fordi den garderer sig imod angreb fra én vinkel. I de bibelske skrifter er dogmet underdetermineret, hvorfor vi har den brede vifte af tolkninger. Men der er derfor heller ikke nogen oplagt grund til så at reducere gudbilledligheden til ét aspekt. Udvikles AGI vil det få mange konsekvenser for menneskets liv på jorden, men selvom det er tvivlsomt, at det overho-

vedet udvikles, vil det ikke få konsekvenser for menneskets særstatus som skabt i Guds billede, så længe man fastholder en facetteret forståelse af dogmet.

Sammenfatning og konklusion

Artiklen har præsenteret to filosofiske kritikker af AGI-projektets gennemførlighed og er alligevel fortsat med en diskussion af de potentielle konsekvenser for menneskets særstatus, hvis AGI udvikles. Sammenhængen er, at Alan Turing grundlæggende havde ret, da han i 1950 affejede alle principielle indvendinger mod *muligheden* for at skabe kunstig intelligens ved at sige: »[W]e cannot so easily convince ourselves of the absence of complete laws of behaviour as of complete rules of conduct. The only way we know of for finding such laws is scientific observation, and we certainly know of no circumstances under which we could say, 'We have searched enough. There are no such laws.'« (Turing 1950, 452). Jeg har drøftet Dreyfus og Fuchs, fordi deres holistiske antropologi forekommer mig overbevisende og deres kritik af AGI-projektet derfor tilsvarende stærk. Men det kan ikke afvises, at AGI realiseres, særligt fordi der er mange ubekendte faktorer omkring bevidsthed, intelligens, computerkraft og så videre, som stadig undersøges, udvikles og diskuteres. Teologer bør derfor drøfte potentielle konsekvenser ved AGI, og artiklen her har givet et forsøg på en respons, der peger på en sammensat, facetteret forståelse af gudbilledligheden. I stedet for at reducere gudbilledligheden til for eksempel noget substantielt eller relationelt, kan vi bruge den underdeterminerede beskrivelse i de bibelske skrifter til at pege på en mere kompleks konfiguration. Ved at gøre dette bliver gudbilledligheden mere robust i mødet med en potentiel udvikling af AGI.[30]

Noter

1 På dansk bruges oftest de engelske termer for *artificial intelligence* (AI) og *artificial general intelligence* (AGI). Jeg bruger derfor også disse forkortelser, da begreberne samtidig bliver udfoldet og forklaret på dansk igennem artiklen.

2 Se desuden Simon Balles artikel i dette nummer.

3 Også i fiktionen er der eksempler på denne konfrontation. I et skønlitterært værk som Dan Browns *Origin* får vi dramatisk beskrevet, at kirken står i fare for at uddø, hvis teknologien lykkes med at skabe en computer, der bærer AGI. See Brown 2017.

4 I en populærvidenskabelig formidlingsramme har matematikeren John Lennox leveret en engageret apologetisk kritik af Hararis bog. Se Lennox 2020.

5 Harari skelner mellem »data religion«, som ser homo sapiens rolle som historisk udspillet, og »techno-humanism«, som ligeledes finder homo sapiens irrelevant, men som videnskaben så skal kombinere med teknologi for at skabe en ny race, *homo deus* (Harari 2015, 409-410). For denne artikel er det homo deus-visionen, der er interessant, men bogen rummer mange interessante socioøkonomiske overvejelser om konsekvenserne ved udviklingen af potent teknologi generelt. Se særligt kapitel 9-11 for disse diskussioner.

6 Jeg henviser her til den forkortede udgave, som AI Magazine udgav i 2006. Det oprindelige oplæg var på 17 sider. I 2006-udgaven har redaktørerne kun medtaget de fremlagte syv forslag til diskussion. Udover McCarthy var det særligt medforfatterne Marvin Minsky, Claude Shannon og Nathaniel Rochester, der fik status som feltets fædre.

7 Interesserede kan finde gode eksempler i opslaget på Wikipedia: www.en.wikipedia.org/wiki/Artificial_intelligence. Smal AI rummer enorme muligheder og udfordringer, og der er en lang række vanskelige etiske problemstillinger, som fagteologien også må diskutere, herunder transhumanisme, bioengineering, cognitive enhancements, war robots og så videre. Disse interessante problemer vil dog ikke opholde os yderligere her.

8 Nogle indfører desuden en tredje type af AI, nemlig *artificial superintelligence* (ASI), som er en type kunstig intelligens, der overgår menneskelig intelligens på alle parametre (Bostrom 2014, 26). De fleste futurister regner med, at momentet for AGI vil være kortvarigt, fordi de automatiske systemer vil opdatere sig så hurtigt, at superintelligensen vil blive udviklet i umiddelbar forlængelse af AGI (Kurzweil 2005, 260). Det er stadig hypotetisk, og derfor vil jeg ikke opholde mig yderligere ved denne distinktion. Hvis blot AGI udvikles, vil implikationerne for gudbilledligheden, som drøftes i artiklens sidste del, stadig være præsente.

9 Selv de nyeste og meget hypede innovationer som OpenAI's *DALL-E 2* og *GPT-3* og DeepMinds *Gato* har ikke rykket feltet bemærkelsesværdigt i forhold til de fundamentale problemstillinger. Se diskussionen hos www.scientificamerican.com/article/artificial-general-intelligence-is-not-as-imminent-as-you-might-think1

10 Det er egentlig en gammel filosofisk ambition. Dreyfus mener selv, at det kan føres tilbage til Platon (Dreyfus 1992, 67), og i tillæg kan man nævne Leibniz, der drømte om at skabe en universel matematik (*mathesis universalis*), hvor alle logiske og sproglige udsagn formaliseres, så ethvert ræsonnement kunne efterprøves matematisk. Leibniz' drøm var, at ørkesløse meningsdiskussioner kunne løses ved simpel kalkulation. Se fx Svante Nordin 2022, 181-194 eller Fuchs 2021, 17. Umberto Eco har skrevet denne ambitions historie med vanligt vid og komik i *In Search for the Perfect Language* (Eco 1992). Se særligt kapitel 14-16.

11 Se desuden leksikonopslaget på www.plato.stanford.edu/entries/computational-mind. Dreyfus ser de idéhistoriske spor gående tilbage til David Hume og Immanuel Kants idé om, at bevidstheden er summen af atomiske indtryk. Der er stor uenighed i forskningen om, hvilken status CTM har. Nogle mener fx at Kurt Gödels ufuldstændighedsteorem og Roger Penroses anvendelse af dette har umuliggjort CTM, mens andre forskere som Hilary Putnam og David Chalmers afviser kritikken og sågar argumenterer for, at de ikke har matematikken på plads. Se samme under afsnit 7.2.

12 Særligt her er der et link mellem Dreyfus' kritik og kritikken fra Thomas Fuchs, som skal diskuteres i artiklens næste afsnit.

13 Minsky har denne idé fra den tidlige Wittgensteins atomisme, Bertrand Russels ditto, Humes empirisme og Leibniz' idé om, at menneskets tanker har en analogi i alfabetet, som derfor kan kortlægges, og hvis forenklende reduktion derfor bliver byggesten for alle formulerbare tanker.

14 Dreyfus anså det for at være et futilt projekt, men Minsky var optimist og kalkulerede endda summen af menneskelig common-sense viden til at være nogle få millioner. Teologen Marius Dorobantu samler op på Dreyfus' kritik af Minskys ontologiske antagelse: »Dreyfus' debate with Minsky was about whether or not human common-sense knowledge could in principle be broken down into its tiniest symbolic building blocks, which could then be accessible to computer programs to learn and use.« (Dorobantu 2020a, 146).

15 For Heidegger var det som bekendt grundlæggende attituder, som førvidenskabeligt måtte beskrives i den fænomenologiske hermeneutik. Vedhåndenheden står for brugstøjets værensart, dvs. det som brugstøjet åbenbarer sig som ud fra sig selv, når vi bruger det. Det handler både om nærhed og håndterlighed. Vi bruger en hammer og lærer den at kende som brugstøj, ikke blot hvordan den virker, men også hvordan den virker optimalt. Forud for brugen af hammeren har vi allerede fornemmet den som en del af en brugstøjshelhed, som vi kan aktualisere i en konkret sammenhæng. Når dette sker, så har vi blotlagt, hvad der blot og bart forelå os – det forhåndenværende – så vi nu fornemmer det værendes værensart. Se Heidegger 2007, §15-16. Bemærk desuden, at terminologien – knowing that/knowing how – stammer fra filosoffen Gilbert Ryle, der præsenterede det i en forelæsning i 1945 (Ryle 1945).

16 Kahnemans beskrivelse er præcis og pædagogisk: »*System 1* fungerer automatisk og hurtigt, kræver en mindre eller slet ingen anstrengelse og indebærer ingen fornemmelse af bevidst kontrol. *System 2* retter opmærksomheden mod de anstrengende mentale aktiviteter, der kræver denne opmærksomhed, herunder komplekse beregninger. System 2's måde at fungere på bliver ofte forbundet med den subjektive oplevelse af at være aktør, at have valgmuligheder og at være koncentreret.« (Kahneman 2011, 28-29) Pointen er, oversat til Dreyfus' kritik, at AI-forskningen kun har blik for system 2, mens rigtig megen menneskelig problemløsning foregår mere spontant, intuitivt og hurtigt vha. system 1.

17 Bemærk dog, at *machine learning* har gjort fremskridt i de seneste år på netop dette felt. Den smalle AI's evne til problemløsning *fremstår* mere og mere som netop en før-sproglig, intuitiv evne. Selve *læringsprocessen* for AI er stadig væsentligt længere end for mennesker og kræver mere rå dataprocessering end for et menneske, men *slutproduktet* bliver sværere og sværere at skelne fra menneskets slutprodukt: en evne til at tolke verden (eller i hvert fald en afgrænset del af den), der ikke kan forklares ved analyse af nogle af algoritmens/hjernens enkeltdele.

18 Pladsen tillader ikke, at jeg gør mere ud af kroppens rolle i perception, kognition, problemløsning osv. (jf. system 1). Jeg må lade Fuchs tale for dette synspunkt i artiklens næste afsnit, men pointen er vigtig: Vores kroppe er irreducibelt komplekst involveret i den måde, vi indgår i og forstår verden på.

19 Som Simon Balle forklarede mig, er dette netop grunden til, at sammensmeltningen mellem menneske og maskine, som jeg kortfattet nævner i indledningen, er så problematisk fra et ontologisk perspektiv; menneske og maskine er netop kun på omgangshøjde, hvis data er lig med information, hvilket det ifølge Fuchs netop ikke er.

20 Som Fuchs selv skriver: »The idealims of information must therefore fail vis-à-vis the phenomenon of life, just as the materialism of particles must fail. Both know only externality.« (Fuchs 2021, 21)

21 Og som han lettere lakonisk anfører, så er omkring halvdelen af hjernen jo ikke neuroner (som ellers er, hvad der sammenlignes med dataprocesser), men støttende celler, der muliggør neuronernes signaler – og tilmed består hjernen af 85% vand, som i hvert fald ikke lader sig overføre til det digitale! (Fuchs 2021, 26).

22 Som Fuchs siger, så er miraklet ikke, at Deep Blue kunne vinde over Kasparov i 1997, for den kunne beregne 200 millioner mulige træk i sekundet. Det forunderlige var, at Kasparov kunne holde distancen så længe – og her ser intuition ud til at være nøglen. Kasparov behøvede ikke at beregne alle muligheder, men fornemmede de bedste træk og valgte mellem dem. Deep Blue havde ingen andre muligheder end at gennemregne *alle* træk, hver gang. (Se Fuchs 2021, 32).

23 Bemærk dog, at Go betragtes – i modsætning til skak – som et meget intuitivt spil. Dorobantu (2019, 7) diskuterer netop dette, og mange ser AlphaGo Zero som et udtryk for, at *machine learning*-algoritmer som den har et element af intuition over sig.

24 Spørgsmålet er dog, om AI-systemer kan blive så gode til at analysere data, at det i praksis er underordnet, om de er bevidste. Jeg tager fat på dette igen i indledningen til diskussionen om gudbilledlighed.

25 Man er kommet et stykke af vejen ved at indføre såkaldte »fuzzy logic«, som Lotfi Zadeh introducerede i 1965. Ideen er simpel: I stedet for den boolske binaritet indfører man sandhedsgrader. Et banalt eksempel: Er kaffen varm? En boolsk operator vil svare ja eller nej, mens *fuzzy logic* vil kunne graduere langs en skala (fx mellem 0 og 10). Fuzzy logic er nødvendig, hvis teknologi skal handle etisk (fx træffe valg i trafikken) eller skabe kunstværker.

26 Som Andreas Ipsen rigtigt nok nævnte for mig, kompliceres påstanden, at bevidsthed er en forudsætning for intelligens, gevaldigt af, at videnskaben stadig ved ganske lidt om, hvad den menneskelige bevidsthed er, og hvordan den opstår. Der er mange stærke og velunderbyggede – men indbyrdes modstridende – teorier.

27 Særligt tak til Andreas Ipsen for at drøfte disse kritikpunkter med mig.

28 Hverken Dorobantu eller jeg hævder naturligvis, at AGI er eneste udfordring for den teologiske antropologi. Det er i mindst lige så høj grad transhumanismen med dens bioingeniør-

videnskab, der radikaliserer et materialistisk verdenssyn, som er en nærværende udfordring. Transhumanismen udfordrer både menneskets hellighed som skabt væsen og skaberskabningdistinktionen ved at lege guder. Disse kritiske perspektiver må dog drøftes i en anden sammenhæng.

29 Emil Brunners skelnen mellem et formalt og materialt aspekt ved gudbilledligheden drøftes ikke her, da kun postlapsariansk gudbilledlighed er relevant for diskussionen med AGI. Ud fra særligt Jak 3,9 er jeg også i tvivl om, hvorvidt denne skelnen overhovedet er eksegetisk holdbar (se videre McGrath 2017). Menneskets oprindelige syndfrihed er ikke nødvendigvis lokaliseret i gudbilledligheden. I de bibelske skrifter benyttes (eller alluderes til) udtryk om gudbilledlighed flere gange, men en egentlig definition eller blot udfoldelse gives ikke (1 Mos 1,27; se desuden 1 Mos 5,1ff; 9,1ff; Sl 8; 1 Kor 11,7; Jak 3,9).

30 Tak til Simon Balle, Andreas Ipsen, Atle Søvik, TechPhil-gruppen ved MF Oslo og faggruppen for systematisk teologi på Helsjön-konferencen for gode og kritiske indspil, der har højnet artiklens kvalitet.

Litteratur

Balle, Simon. Under udgivelse. »Theological Dimensions of Humanlike Robots: A Roadmap for Theological Inquiry.« *Theology and Science*.

Bostrom, Nick. 2014. *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Brown, Dan. 2017. *Origin*. New York: Doubleday.

Chalmers, David J. 2011. »A Computational Foundation for the Study of Cognition«. *Journal of Cognitive Science* 12, 323-357.

Chalmers, David J. 2022. *Reality+. Virtual Worlds and the Problems of Philosophy*. New York: W.W. Norton & Company.

Dorobantu, Marius. 2019. »Recent Advances in Artificial Intelligence (AI) and some of the issues in the theology & AI dialogue«. *ESSSAT News & Reviews* 29 (2), 4-17.

Dorobantu, Marius. 2020a. *Theological Anthropology and the Possibility of Human-Level Artificial Intelligence: Rethinking the Human Distinctiveness and the* Imago Dei. Upubliceret ph.d.-afhandling ved Université de Strasbourg.

Dorobantu, Marius. 2020b. »Will Robots Too Be in the Image of God? Artificial Consciousness and the Imago Dei in Westworld«. I *Theology and Westworld*, redigeret af Juli Gittinger og Shayna Sheinfeld, 73-90. London: Fortress.

Dreyfus, Hubert. 1972. *What Computers Can't Do. Of Artificial Reason*. New York: Harper and Row.

Dreyfus, Hubert. 1992. *What Computers Still Can't Do. A Critique of Artifical Reason*. Cambridge, MA: MIT Press.

Eco, Umberto. 1995. *In Search of the Perfect Language*. Malden, MA: Blackwell.

Fuchs, Thomas. 2017. *Ecology of the Brain. The Phenomenological and Biology of the Embodied Mind*. Oxford: Oxford University Press.

Fuchs, Thomas. 2021. *In Defense of the Human Being. Foundational Questions of an Embodied Anthropology*. Oxford: Oxford University Press.

Harari, Yuval. 2015. *Homo Deus. A Brief History of Tomorrow*. London: Vintage.

Hefner, Philip J. 2002. »Technology and Human Becoming«. *Zygon* 37 (3), 655-665.

Heidegger, Martin. 2007. *Væren og tid*. Aarhus: Klim.

Herzfeld, Noreen. 2002. »Creating in Our Own Image: Artificial Intelligence and the Image of God«. *Zygon* 37 (2), 303-316.

Ishiguro, Kazuo. 2022. *Klara og solen*. København: Gyldendal.

Kahneman, Daniel. 2013. *At tænke – hurtigt og langsomt*. København: Lindhardt & Ringhof.

Kofoed, Jens Bruun. 2015. *Til syvende sidst. Skabelse, tempel og hvile i Biblen og den gamle Orient*. København: Museum Tusculanum Forlag.

Kofoed, Jens Bruun. 2016. *Konge, præst og gartner. Menneske i Guds verden*. Fredericia: Kolon.

Kurzweil, Ray. 2005. *The Singularity Is Near. When Humans Transcends Biology*. New York: Viking.

Lakoff, George og Mark Johnson. 2002. *Hverdagens metaforer*. København: Hans Reitzels Forlag.

Lennox, John C. 2020. *2084. Artificial Intelligence and the Future of Humanity*. Grand Rapids: Zondervan.

McCarthy, John. 2006. »A Proposal for the Darmouth Summer Research Project on Artificial Intelligence«. *AI Magazine* 27 (4), 12-14.

McGrath, Alister. 2017. »Emil Brunner: A Theologian for the Academy and Church Today«. *Theologische Zeitschrift* 73 (2), 146-62.

Nordin, Svante. 2022. *Filosoferna. Vetenskaplig revolution och upplysning 1650-1776*. Stockholm: Fri tanke.

Panggabean, Mauritz. 2022. »Could Androids with Artificial General Intelligence be Christian?: An Interdisciplinary Conversation with Scripture and Theology«, specialeafhandling indleveret ved MF Oslo.

Pearl, Judea og Dana Mackenzie. 2018. *The Book of Why. The New Science of Cause and Effect*. London: Penguin.

Peterson, Ryan S. 2016. *The* Imago Dei *as Human Identity. A Theological Interpretation*. Winona Lake: Eisenbrauns.

Pinker, Steven. 1997. *How the Mind Works*. London: Penguin.

Polanyi, Michael. 1958. *Personal Knowledge. Towards a Post-Critical Philosophy*. Chicago: The University of Chicago Press.

Ryle, Gilbert. »Knowing How and Knowing That: The Presidential Address.« *Proceedings of the Aristotelian Society, New Series*. Vol. 46 (1945-1946), 1-16.

Searle, John. 1980. »Minds, Brains and Programs.« *The Behavioral and Brain Sciences*. Nr. 3, 417-457.

Søvik, Atle. 2022. *A Basic Theory of Everything. A Fundamental Theoretical Framework for Science and Philosophy*. Berlin: De Gruyter.

Turing, Alan. 1950. »Computing machinery and intelligence.« *Mind,* nr. 59 (Oktober), 433-460.

Turkanik, Andrzej. 2021. »Art, music and AI: the uses of AI in artistic creation«. I *The Robot Will See You Know. Artificial Intelligence and the Christian Faith*, redigeret af John Wyatt og Stephen N. Williams, 198-213. London: SPCK

Zizioulas, John D. 1997. *Being as Communion. Studies in Personhood and the Church*. New York: St. Vladimir's Seminary Press.

Internetadresser

www.cnbc.com/2022/04/21/elon-musk-says-optimus-robot-will-be-worth-more-than-tesla.html

www.en.wikipedia.org/wiki/Artificial_intelligence

www.etiskraad.dk/etiske-temaer/optimering-af-mennesket/homo-artefakt/leksikon/kunstig-intelligens

www.scientificamerican.com/article/artificial-general-intelligence-is-not-as-imminent-as-you-might-think1/

Forfatter

Michael Agerbo Mørch
Fjellhaug International University College, Copenhagen
Leifsgade 33, 6.-7.
2300 København S
mam@dbi.edu

Artiklen er blevet godkendt ved en redaktionsuafhængig fagfællevurdering.