Magnus Strand

Artificiell intelligens: ett forskarperspektiv på teknik, människosyn och tro

JUSTITIA ET PAX
KOMMISSIONEN FÖR RÄTTVISA OCH FRED
Rapport #7 i skriftserien
Tro och samhälle

STOCKHOLMS KATOLSKA STIFT
JUSTITIA ET PAX – KOMMISSIONEN FÖR RÄTTVISA OCH FRED
KORT RAPPORT I SERIEN »TRO OCH SAMHÄLLE«

7. MAGNUS STRAND
ARTIFICIELL INTELLIGENS: ETT FORSKARPERSPEKTIV
PÅ TEKNIK, MÄNNISKOSYN OCH TRO

KORT RAPPORT I SERIEN
»TRO OCH SAMHÄLLE«

7

MAGNUS STRAND

Artificiell intelligens: ett forskarperspektiv på teknik, människosyn och tro

SEPTEMBER 2025

JUSTITIA ET PAX
KOMMISSIONEN FÖR
RÄTTVISA OCH FRED

FÖRORD

UNDER VINJETTEN »Tro och samhälle« presenterar Kommissionen för rättvisa och fred (Justitia et pax) i Stockholms katolska stift återkommande rapporter där olika skribenter presenterar och analyserar olika aspekter av den katolska socialläran. Ambitionen är att bidra till en bättre efterföljelse av Jesus Kristus och främja reflektionen över relationen mellan tro och samhällsmedborgarskap.

Det kan till exempel handla om att beskriva en hållning som härleds från tron och som kan tillåta olika, ibland till och med helt motsatta, ståndpunkter i konkreta samhällsfrågor, liksom att visa hur dessa kan formuleras utifrån den gemensamma hållningen, det vill säga den alltid förenande sanningen i evangeliet och socialläranss grundläggande principer. Det kan också handla om att ge kunskap och perspektiv på en fråga inom något av socialläranss tillämpningsområden eller på någon av dess grundprinciper, kopplat till något aktuellt samhällsskeende, för utveckling av både teologisk förståelse och kristet handlande. Och det kan också i denna rapportserie handla om att presentera något kyrkligt lärodokument inom socialläran och exempelvis framhålla någon aspekt att reflektera över eller beakta.

Jag vill rikta ett tack till alla som medverkar i projekten kopplade till »Tro och samhälle« och uttrycka min önskan om att rapporter, som denna, och annat material från projekten, ska bära frukt för Guds rike.

Uppsala i september 2025

P. THOMAS IDERGARD SJ

Ordförande i Justitia et pax
i Stockholms katolska stift

ARTIFICIELL INTELLIGENS: ETT FORSKARPERSPEKTIV PÅ TEKNIK, MÄNNISKOSYN OCH TRO

Inledning

VAD ÄR DETTA »AI« som låter tala så mycket om sig? Det är en fråga som kan besvaras på flera sätt. Bland de datavetare, ingenjörer, statistiker och matematiker som faktiskt är kunniga inom fältet talas det sällan eller aldrig om »AI« – begreppet anses där vara alldeles för löst i kanterna. De talar hellre om maskinlärande, neurala nätverk, Bayesiansk inferens och annat som vi som lekmän kan förbigå, men som samtidigt visar att »AI« är ett paraplybegrepp för en mängd olika algoritmer och deras ännu större mängd tillämpningar. En diskussion om etik, politik och juridik kring AI är icke desto mindre betjänt av att man i någon mån försöker avgränsa vad diskussionen gäller. Här är ett försök som i all bristfällighet kan fungera: Artificiell intelligens, AI, är ett samlingsnamn för mjukvara (datorprogram) med förmåga att efterlikna ett beteende eller uttryck som vi människor uppfattar som intelligent. Denna mjukvara kan sedan byggas in i ett tekniskt system, exempelvis en robot, en självkörande bil, eller ett ordbehandlingsprogram.

Vanligen har mjukvaran en grundalgoritm som sedan har tränats för en viss uppgift med hjälp av stora mängder data. Exempelvis kan algoritmen tränas att känna igen en bild av en hund genom att tränas med en stor mängd bilder av hundar, vilka för algoritmens lärande kodats som »hund«, och en stor mängd andra bilder, vilka i sin tur kodats som »inte hund«. Om detta skett framgångsrikt kan mjukvaran sedan få i uppdrag att i ett arkiv med mängder av bilder identifiera de bilder som visar en hund. Just en sådan uppgift har inte så stort praktiskt användningsområde, men låt oss istället anta att algoritmen tränats att för en banks räkning identifiera transaktioner som kan vara ett led i penningtvätt. Med hjälp av en sådan mjukvara kan identifikationen ske på ett effektivt sätt, – mycket mer effektivt än om en stor grupp människor mödosamt (och troligen med stor leda) skulle granska alla bankens transaktioner i jakt på misstänkt beteende.

Förutom sådana analytiska tillämpningar finns också generativa tillämpningar. Generativ AI har fått stor uppmärksamhet sedan företaget OpenAI lanserade sin textproducerande algoritm ChatGPT. Sådan generativ AI finns i olika former och kan skapa text, bild, ljud eller video utifrån givna instruktioner. Dessa instruktioner kallas »prompt«. Kvaliteten i det som en generativ AI skapar beror förstås på dess egen kvalitet, men den som försökt använda tekniken vet också att det är något av en konst att skriva en väl fungerande prompt. Därför har yrkeskategorier som promptingenjörer och promptkonstnärer börjat växa fram.

AI-tekniken väcker fascination, frågor och farhågor. I den här korta skriften gör jag ett försök att för den intresserade presentera och diskutera några av dem:

1. Kan AI ha personlighet och ett etiskt relevant personvärde?
2. Hur bör AI regleras och kontrolleras, och från vem kan ansvar för AI utkrävas?
3. Hur kan tillgången till AI fördelas så att den blir till nytta för hela mänskligheten?
4. Vilken påverkan kan driften av AI ha på miljö och samhällen?
5. Vilken påverkan kan AI ha på människors politiska uppfattningar och trygghet?

Kyrkan har genom hela sin historia engagerat sig i teknikens och vetenskapens utveckling, inte för att kontrollera den, utan för att säkerställa att den förblir i människans tjänst. I Andra Vatikankonciliets Pastoralkonstitution om kyrkan i världen av idag, *Gaudium et spes*, betonas att människan genom sitt arbete får del i Guds skapande verk (p. 34). Tekniken, som ett resultat av människans arbete, är således inte neutral utan uttrycker både människans skaparkraft och hennes moraliska ansvar. Samma tanke uttrycks i vetenskapliga sammanhang: Teknik är aldrig neutral, utan har moraliska, politiska och existentiella dimensioner. Vad det därför gäller är att främja en teknik som tjänar fred, rättvisa, frihet och mänsklig värdighet. Vad detta innebär för Kyrkans syn på AI har utvecklats i Dikasteriets för trosläran skrivelse *Antiqua et nova*. Ett annat initiativ är »The Rome Call for AI Ethics«, som undertecknades 2020 av den Påvliga Akademin för livet jämte flera teknikföretag. I dokumentet föreslås sex grundpelare för AI-etik: transparens, inkludering, ansvar, opartiskhet, tillförlitlighet och säkerhet.

Eftersom författaren till den här skriften inte är teolog, är det inte avsett att här utveckla eller diskutera Kyrkans syn på AI. För sådana frågor hänvisas läsaren till *Antiqua et nova*. Den här skriften är istället avsedd att, som ett komplement till Kyrkans egna dokument, ge en introduktion till några frågor som diskuteras i den allmänna litteraturen och forskningen. Med detta sagt kommer paralleller och hänvisningar att göras till Kyrkans synsätt och relevanta dokument, till hjälp för läsarens orientering.

Kan AI ha personlighet och etiskt personvärde?

Det förekommer spekulationer i om AI (vilket här används som kortform för det korrektare »ett tekniskt system som innefattar någon form av AI-teknik«) kan ha ett självmedvetande, vara ett kännande väsen, och följaktligen ha ett etiskt värde som person som kan liknas vid människovärdet eller i vart fall jämföras med det relativa värde som bör tillerkännas djur. Särskilt i kulturen finns många exempel: replikanterna i Philip K Dicks roman »Do androids dream of electric sheep?« (senare filmklassikern »Blade runner«), sällskapsroboten Klara i Kazuo Ishiguros roman »Klara and the sun«, och de färgstarka robotarna C3PO och R2D2 i filmserien »Star wars«.

För de som arbetar med AI är detta nonsens, som bygger på ett missförstånd: AI är per definition utformat för att *imitera* intelligent beteende, men det *saknas* ett intelligent tänkande bakom beteendet. Att AI utformas för att imitera intelligent beteende är alltså själva grundtanken, och kan illustreras med en gammal definition av vad AI är: det s.k. Turingtestet (efter Alan Turing, den engelske matematiker som ofta brukar ses som AI:s »fader«). Turing-testet går i enkelhet ut på att en försöksperson placeras

bakom en skärm och kommunicerar i tryckt skrift med antingen en människa eller en maskin på andra sidan skärmen. Försökspersonen får dock inte veta om det är en människa eller en maskin den interagerar med. När en maskin uppnått sådan skicklighet i att imitera mänsklig kommunikation att försökspersonen inte kan avgöra om det är en människa eller en maskin den har att göra med, har maskinen klarat Turing-testet och kan kallas för en artificiell intelligens.

Turing-testet har maskiner nu klarat för mycket länge sedan, men det illustrerar att maskinens skenbara intelligens är just skenbar. Med en konstreferens kan man säga att det rör sig om ett *trompe l'œil*, dvs. en framställning som lurar betraktaren att se något som inte finns där. Detsamma gäller alla framtida AI, oavsett hur avancerade och kapabla de blir till att utföra allsköns uppgifter. Med detta synsätt, som alltså är det rådande synsättet bland AI-kunniga, saknas all anledning att spekulera i om AI nu eller i en framtid skulle kunna nå självmedvetenhet och bli ett kännande väsen i samma eller liknande mening som en människa eller någon annan organism.

Här finns samtidigt ett kunskapsteoretiskt problem. Anta att en AI, tvärtemot vad som antas, faktiskt skulle bli en självmedveten och kännande person. Hur skulle den kunna övertyga oss om att så är fallet? Förmodligen blir det en omöjlig uppgift, vilket leder till att vi katastrofalt riskerar att »döda« denna person bara genom att stänga av den. Det har förekommit att personer som arbetar med AI övertygas om att detta inträffat, i ljuset av att den AI de kommunicerat med faktiskt uttalat att den är rädd för att stängas av eller har andra känslor. Skeptiker avfärdar detta, med hänvisning till att AI:ns beteende är

ett resultat av dess programmering och träning. Vad den säger om sina känslor är för skeptikerna bara ett resultat av att algoritmen presterar ett språkligt uttryck, angående känslor, som den utformats för att prestera. Det saknas därför anledning att tro att en algoritm skulle ha någon av de egenskaper vi förknippar med vår mänskliga självmedvetna tillvaro. Personligen hör jag till skeptikerna, tillsammans med i princip alla andra som i någon mån har satt sig in i hur AI fungerar. En skiljelinje som jag själv brukar framhålla är att AI är mycket bra på att ge svar, men är helt oförmögen att ställa frågor. Den saknar den inneboende längtan efter sanning som lockar oss människor att ställa frågor om oss själva, världen, och Gud. På mänsklig uppmaning kan den imitera hur vi uttrycker sådana frågor, men den hyser inte vår längtan (eller några andra känslor).

Föreställningar om AI-systems »egenvärde« bygger därför, enligt min uppfattning, på missförstånd som ytterst kommer sig av okunskap om vad AI egentligen är för något. Artificiella system, oavsett hur avancerade de är, saknar en själ och kan inte besitta självmedvetande i teologisk mening. Varje människa, å andra sidan, är skapad till Guds avbild (*imago Dei*) och utrustad med förnuft, vilja, frihet och en odödlig själ. Denna värdighet är okränkbar och gäller oberoende av individens nytta, funktionalitet eller kapacitet, vilket utvecklas i deklarationen *Dignitas infinita* (Dikasteriet för trosläran, 2024). Utgångspunkten är alltså att AI inte kan ha personlighet och att en AI därmed inte heller har något etiskt relevant värde.

Det är emellertid en annan fråga om det finns skäl att tillerkänna en AI så kallad *juridisk personlighet*, såsom vi gör med vissa bolag och föreningar. I all korthet beror svaret på den frågan om det finns någon

samhällsnytta med att tillerkänna AI-system juridisk personlighet. Min egen uppfattning i frågan är att det inte finns någon sådan nytta. Juridiska personer som rättslig figur har skapats för att avskilja egendom som tillhör associationer av människor (t.ex. aktiebolag) från de människor som ingår i associationen, eller för att avskilja egendom som under lång tid ska vara avsatt för ett speciellt syfte (t.ex. en stiftelse) från den människa som avsatt egendomen. Det finns inget sådant syfte kring en AI. De skäl för att AI ska vara juridiska personer som har framförts har att göra med att avancerade robotar med AI-teknik kan agera med stor autonomi och kanske då vållar andra skada, som behöver ersättas. För detta finns emellertid andra lösningar: skadeståndsansvar för ägaren, obligatorisk försäkring, eller liknande. Därför tillämpas inte heller denna begränsade form av rättsligt personbegrepp för AI.

Hur bör AI regleras och kontrolleras, och vem har ansvaret för en AI?

Diskussionen om hur man närmare bestämt ska lagstifta om ansvaret för AI och robotar har däremot varit livlig under senare år. Här har särskilt EU varit mycket aktivt, med kulmen i antagandet 2024 av dess förordning om artificiell intelligens, den s.k. AI-akten. Man kan förenklat dela in frågor om reglering och ansvar för AI i några olika aspekter:

1. Produktsäkerhet och produktansvar för varor med AI-teknik
2. Driftsäkerhet och driftansvar för AI-baserade tjänster
3. Säkerhet och ansvar kring AI-systems inbördes interaktion

Den första och den andra aspekten är ganska välreglerade, medan den tredje inte är det.

PRODUKTSÄKERHET OCH PRODUKTANSVAR

AI-akten gäller huvudsakligen produktsäkerhet. Den ger företag och andra organisationer som ska använda AI i sin verksamhet ett långtgående ansvar för att kartlägga och hantera de risker som kan uppkomma när AI-teknik integrerats i en produkt. AI-akten placerar säkerhetsansvaret på de företag och organisationer som (1) själva tar ett AI-system i bruk i sin verksamhet, (2) levererar AI-system till andra, (3) som mellanhand distribuerar AI-system på marknaden, eller (4) importerar AI-system till EU. Utöver AI-akten kan också en mängd andra produkt- och tekniksäkerhetsregler gälla, som inte egentligen har med AI att göra men som gäller generellt för de typer av produkter där man har byggt in AI-teknik.

Det finns sedan länge regler om produktansvar som gäller i hela EU. I Sverige finns de här reglerna i produktansvarslagen (1992:18). EU har nyligen ändrat i reglerna så att de också ska fungera för produkter som innehåller AI-teknik. Det finns utöver produktansvarslagen också andra regler som kan bestämma vem som har ansvar för ett AI-system. Ofta beror det på i vilken situation AI har använts vilket regelverk som blir tillämpligt. Exempelvis kan en skadlig användning av AI inom finanssektorn leda till ansvar enligt de lagar och regler som gäller för den sektorn, medan ansvar för skador vållade av en självkörande bil faller under trafikskadelagen (1975:1410). I bakgrunden finns också alltid allmänna principer om skadeståndsansvar.

DRIFTSÄKERHET OCH DRIFTANSVAR FÖR TJÄNSTER

När det gäller AI-baserade tjänster uppstår frågor kring cybersäkerhet och dataskydd. Här har EU varit

mycket aktivt med att lagstifta. GDPR om skydd för persondata är välkänd, men runt den har det vuxit fram en mängd mindre kända lagar som t.ex. ställer krav på att den som tillhandahåller en datatjänst ska hålla den säker mot intrång eller att den ska göra det möjligt för användare att dela sina data med andra företag. Företag som tillhandahåller sociala medier ska också städa undan olagliga, missvisande eller stötande inlägg, som ofta kommer från s.k. »trollfabriker« där AI-verktyg används för att sprida desinformation. All sådan lagstiftning träffar AI-baserade tjänster, som ju alltid använder data och som ofta erbjuds helt och hållet över internet. Dessutom innehåller de olika sanktioner, »straff«, så att företag och organisationer som inte följer reglerna kan hållas till ansvar.

SÄKERHET OCH ANSVAR KRING AI-SYSTEMS INBÖRDES INTERAKTION

En situation som det är mycket svårt att hantera med nu gällande juridik är när flera olika AI-styrda maskiner eller andra system har samverkat med varandra på ett oväntat sätt. Det finns alltid en risk att AI förstärker felaktigheter som den fått i den data som den hanterar. Om ett AI-system sedan sänder information med förstärkta felaktigheter till ett annat AI-system kan det skapa en oväntad och snabb utveckling. Ett exempel på detta är så kallade blixtkrascher på finansmarknader, där värdet på vissa finansiella tillgångar plötsligt faller brant på grund av att algoritmer reagerar på varandras signaler.

Vid en blixtkrasch, eller annan liknande situation, kan det vara mycket svårt att avgöra var det ursprungliga felet låg. Därför rekommenderas ofta försäkringslösningar för att hantera de skador som kan uppkomma i verksamheter där ett större antal algoritmer ska interagera med varandra.

Hur kan tillgången till AI fördelas rättvist?

Företag, organisationer och människor kan använda AI för att göra saker som tidigare varit omöjliga. De kan också använda AI för att göra saker som de redan gjort i större hastighet eller större mängd, eller med högre kvalitet. Eftersom AI alltså kan skapa nya marknader, eller effektivitetsvinster inom befintliga marknader, finns stora kommersiella intressen i teknikens olika tillämpningar. Som vanligt med ny teknik kan uppfinningar skyddas med patent, och de möjligheter som uppstår kan på så sätt byggas in i varor och tjänster som erbjuds mot betalning.

AI blir därmed en nyttighet bland alla andra som ingår i vår marknadsekonomi. Som en bieffekt riskerar detta att förstärka klyftor mellan de som kan betala för tekniken, och därmed uppnå de vinster som den erbjuder, och de som inte har råd att betala det pris som begärs. Kritiker menar att samhället måste motverka de klyftorna och på olika sätt verka för att den nya tekniken tillgängliggörs för ett större flertal.

Å ena sidan har de dominerande leverantörerna av AI visat en viss generositet. För att skapa ett kraftfullt och välfungerande AI-verktyg krävs tillgång till stor beräkningskraft, stora mängder data, och expertis i programmering. Kring alla dessa resurser har de dominerande amerikanska och kinesiska teknikföretagen ett övertag. Vissa av dessa företag tillhandahåller grundläggande AI-verktyg till allmänheten utan kostnad, även om det inte brukar vara de allra senaste och bästa verktygen som kan användas gratis. Här finns alltså redan en ganska god tillgänglighet. Mer specialiserade tjänster brukar dock vara förbehållna marknaden.

Å andra sidan brukar det framhållas att en av de avgörande faktorerna för att skapa bra AI, nämligen data, ofta består av sådant som alla vi människor i världen gemensamt har skapat: Texter, bilder, musik, filmer, och våra personuppgifter. Om nu detta är avgörande för att skapa AI, så anser kritiker att de dominerande företagen borde tillgängliggöra tekniken i än större omfattning än de redan gör, eller att man med lagstiftning borde hindra dem från att utnyttja data. En annan invändning mot den rådande ordningen är att den förstärker klyftan mellan industriländer och utvecklingsländer, eftersom människor i utvecklingsländer har sämre tillgång till digital infrastruktur.

Vi får konstatera att skapandet och tillhandahållandet av produkter och tjänster med AI-teknik följer den ordning som även annars råder i vår globala marknadsekonomi, med dess för- och nackdelar. Att bygga alternativa strukturer skulle kräva omfattande resurser och stor uthållighet, t.ex. stora offentliga investeringar i digital infrastruktur. Man kan också uttrycka saken så att den tekniska utvecklingen är inbäddad i det befintliga samhällsskicket, och inte kan förstås isolerad från det.

Vi bör dock också påminna oss om att katolsk socialära insisterar på att teknisk utveckling ska tjäna det gemensamma goda och inte leda till ökade ojämlikheter. Teknisk innovation är endast moralisk i den mån den främjar människans fulla utveckling och skyddar de mest sårbara. AI får alltså inte bli en exklusiv resurs för dem med ekonomisk eller teknologisk makt. Tillgång till AI-teknik och skydd mot dess risker måste fördelas rättvist. Denna uppmaning riktar sig till alla makthavare i samhället.

Vilken påverkan kan AI ha på miljö och hållbarhet?

Det kan ibland verka som om AI är något helt immateriellt, något rent digitalt som inte berör den fysiska världen. På samma sätt talar man ibland om molntjänster och strömningstjänster i den digitala ekonomin. Sådana tjänster har det gemensamt att de låter användarna spara data någon annanstans, eller ta del av medier som lagras någon annanstans än på användarnas egna datorer och enheter, och att de därmed tar bort de begränsningar som tidigare följt av lagringsutrymme på hårddiskar.

Detta är dock en chimär: Någonstans måste all data lagras, och för moln- och strömningstjänster krävs mängder av hårdvara och energi. Detsamma gäller vid skapande av nya AI-modeller: Det krävs oerhört kraftfulla datorer som gör av med mängder av elektricitet i processen. När AI-modellen väl tränats krävs inte lika mycket hårdvara och energi för att hålla den i drift, men naturligtvis krävs bådadera även då.

Forskare har framhållit att produktionen av den tekniska hårdvara som behövs för AI, t.ex. avancerade datachips, är en industri som precis som många andra har stora problem. Här finns människor som arbetar i farliga gruvor mot dålig betalning, här finns undermålig hantering av avfall från kemiska processer med stora ekologiska konsekvenser, och här finns så stort behov av energi att teknikjättar köper eller bygger hela kärnkraftverk bara för att kunna försörja sin AI-utveckling.

Ibland sägs det att AI kan hjälpa mänskligheten att hitta lösningar på miljöproblemen. Det är ett mycket bra användningsområde för AI, eftersom arbetet mot klimatpåverkan och andra sätt som människan på-

verkar miljön kräver svåra beräkningar som AI är särskilt bra på. Samtidigt verkar det inte försvarligt att fortsätta utnyttja fattiga människors arbete endast med hänvisning till en förhoppning om nya klimatlösningar som ska presenteras av ett AI-system. Frågorna är uppenbart mer sammansatta än så, och kräver vår omedelbara uppmärksamhet. Vi kan alltså inte slå oss till ro med att AI kanske kan erbjuda nya lösningar, och ta detta till ursäkt för att bortse från våra mänskliga samhällens problem.

Vilken påverkan kan AI ha på politiska uppfattningar och trygghet?

Närhelst vi använder internet för att söka information, göra ärenden osv., lämnar vi digitala »spår« efter oss. Dessa spår innehåller information om vad vi tittat på, kanske vad vi köpt, eller vilka meddelanden vi lämnat till andra i öppna forum. AI och andra algoritmer används i stor omfattning för att kartlägga sådana spår och rikta olika budskap till oss som vi antas uppskatta eller vara intresserade av.

I samband med detta uppstår två fenomen som blivit extra starka genom AI-tekniken: filterbubblor och desinformation.

Med »filterbubblor« menas att den som tittat på något med ett visst innehåll, t.ex. sökt online efter nya skor, kommer att få reklam för skor riktad till sig. Den som har intresserat sig för skor antas av marknadsförare ha fortsatt intresse för skor, och är därför extra intressant som målgrupp för riktade reklaminsatser. På samma sätt arbetar sociala medier, som Facebook och TikTok. Om du har visat intresse för bilder och videor med marsvin, kommer algoritmerna att erbjuda dig fler bilder och videor med marsvin.

Filterbubblor blir ett demokratiskt problem när detta översätts till politiska och andra samhällsrelevanta budskap, inklusive nyheter. Idag tar många människor del av nyheter och samhällsinformation genom sociala medier snarare än genom traditionella nyhetskällor. Den som intresserat sig för information om exempelvis Centerpartiet, landsbygdsfrågor och jordbrukspolitik, kommer med algoritmernas hjälp att få fortsatt information om detsamma. Att sociala medier fungerar på det här sättet har de positiva följderna att vi får specialiserad information om sådant vi tycker om eller är intresserade av. Tyvärr sker detta till nackdel för det mer mångsidigt kurerade nyhetsutbud som de traditionella kanalerna står för. Risken med att gå miste om mångsidig information är att människor kan förlora kontakten med andra synsätt och resonemang.

Många kritiker av sociala medier ser här en orsak (eller rentav Orsaken) till samtidens tendens till politisk polarisering och brist på förståelse för andras synsätt. Om vi inte får del av andra politiska uppfattningar än vår egen riskerar vi att få sämre förmåga till den självreflektion och självkritik som är en viktig del av det demokratiska samtalet. Om vi dessutom inte får del av mångsidiga nyheter, utan endast av sådana nyheter som bekräftar vår egen världsbild, blir vi dessutom oense inte bara om politiska analyser utan rentav om fakta: Vi kan få oförenliga bilder av vad som faktiskt försiggår i världen. När så är fallet kan den enes sanning bli den andres »fake news«, och påståenden som för någon verkar vara falsk propaganda blir för någon annan att modigt stå upp för sanningen.

Så långt om filterbubblor, men det sistnämnda för oss in på den andra avigsidan av AI inom sociala medier:

desinformation. Den verklighet i sociala medier och andra internetforum som jag just beskrivit drar till sig inte bara marknadsförare och propagandister, utan också aktörer vars själva uppsåt det är att sprida falsk information, förvirring, tvivel och misstänksamhet. Detta kan ske för att påverka demokratiska processer, vilket var fallet med den s.k. Cambridge Analytica-skandalen i samband med Brexit-valet i Storbritannien 2016. Det kan också ske för att mera i allmänhet underminera medborgarnas förtroende för samhällets institutioner, vilket är huvudsyftet med Rysslands olika påverkanskampanjer i västliga internetforum. Även desinformationen ses som en avgörande kraft bakom samtidens tendens till politisk polarisering och bristande tilltro till det s.k. samhällskontraktet.

AI:s roll i spridningen av desinformation hotar människans förmåga till rationellt omdöme och självständig moralisk reflektion. Detta är en moralisk, inte enbart demokratisk, fråga. I sin encyklika *Veritatis splendor* (1993) betonar den helige Johannes Paulus II människans kallelse att söka sanningen och följa sitt samvete. Algoritmiska manipulationer som snedvrider informationsflöden riskerar att fördunkla samvetet och underminera den fria viljan. Teknik som hindrar människans sanningens sökande måste därför betraktas som oetisk.

För att motverka desinformationens påverkan på samhället har Sverige inrättat Myndigheten för psykologiskt försvar. Den tillhandahåller information och kurser för att öka människors förmåga att identifiera och motstå de budskap som desinformationen innehåller. Informationen finns tillgänglig på myndighetens webbplats, se länk nedan. Grundbultarna är källkritik, förståelse av hur desinformation fungerar

och vad dess syften är, och förståelse av grunddragen i vår demokrati och vår rättsstat.

Avslutning

Kyrkan ser teknik i ljuset av frälsningshistorien. Människan är en andlig varelse kallad till gemenskap med Gud och sina medmänniskor. Teknik, inklusive AI, är en gåva som kan bidra till mänsklighetens utveckling – men endast om den formas och används i ljuset av Guds bud och kärlekens etik. Påve Franciskus betonade i sin encyklika *Laudato si'* (2015, p. 105) att »vår enorma teknologiska utveckling har inte åtföljts av en uppövande av mänskligt ansvarstagande, värderingar och samvete«.

När vi reflekterar över AI och vad den nya tekniken betyder för människor och samhälle bör vi utgå från det som är människans yttersta mål: att älska Gud och sin nästa. I denna skrift har konstaterats att

- AI-system inte har någon personlighet och inte heller något etiskt personvärde
- Omfattande lagstiftningsarbete kring AI-tillämpningar pågår, inte minst på EU-nivå
- Fördelningen av tillgång till AI är avhängig existerande strukturer i vår politiska ekonomi
- De många produktionsleden bakom AI leder till ytterligare miljö- och klimatutmaningar
- AI:s roll i filterbubblor och spridningen av desinformation hotar människans förmåga till rationellt omdöme och självständig moralisk reflektion

Som kristna bör vi välkomna de förbättringar som den nya tekniken för med sig, men samtidigt vara vaksamma mot dess nackdelar och verka för att nackde-

larna avhjälps. Detta gäller dels sammanblandningar av avancerade tekniska system och det personvärde som tillkommer människan, dels de konkreta samhällsutmaningar som AI medför.

Tips för fortsatt läsning

Den katolska kulturtidskriften Signum har publicerat många bidrag som rör olika aspekter av AI. Dessa går att söka fram och läsa på tidskriftens webbplats www.signum.se. Här är ytterligare tips:

GENERELLA ETISKA PRINCIPER FÖR AI

The Rome Call for AI Ethics, tillgängligt via www.romecall.org.

Etiska riktlinjer för tillförlitlig AI, EU-kommissionens Expertgrupp på hög nivå för AI-frågor, tillgängliga via https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

KAN AI HA PERSONLIGHET OCH ETT ETISKT RELEVANT PERSONVÄRDE?

Peter Gärdenfors, *Kan AI tänka? om människor, djur och robotar* (Fri tanke 2024)

Inga Strümke, *Maskiner som tänker: Algoritmernas hemligheter och vägen till artificiell intelligens* (Polaris 2024)

HUR BÖR AI REGLERAS OCH KONTROLLERAS, OCH FRÅN VEM KAN ANSVAR FÖR AI UTKRÄVAS?

Virginia Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way* (Springer 2019)

Mark Chinen, *The International Governance of Artificial Intelligence* (Edward Elgar Publishing 2023)

Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (Oxford University Press 2023)

HUR KAN TILLGÅNGEN TILL AI FÖRDELAS SÅ ATT DEN BLIR TILL NYTTA FÖR HELA MÄNSKLIGHETEN?

Karen Hao, *Empire of AI: Inside the reckless race for total domination* (Allen Lane 2025)

VILKEN PÅVERKAN KAN DRIFTEN AV AI HA PÅ MILJÖ OCH HÅLLBARHET?

Kate Crawford, *Atlas of AI* (Yale University Press 2022)

Bendetta Brevini, *Is AI Good for the Planet?* (Polity Press 2021)

VILKEN PÅVERKAN KAN AI HA PÅ MÄNNISKORS POLITISKA UPPFATTNINGAR OCH TRYGGHET?

Mark Coeckelbergh, *Why AI Undermines Democracy and What To Do About It* (Wiley 2024)

Cass R. Sunstein, *#republic: Divided Democracy in the Age of Social Media* (Princeton University Press 2018)

Myndighetens för psykologiskt försvar webbplats, www.mpf.se.

FÖRFATTARPRESENTATION

Magnus Strand är docent i Europarätt vid Uppsala universitet och har skrivit flera forskningsbidrag om samhällsfrågor kring AI-tekniken. Han är också medlem i S:t Lars katolska församling i Uppsala.

OM JUSTITIA ET PAX

Kommissionen för rättvisa och fred, Justitia et Pax, är ett organ i stiftet verksamt i frågor kopplade till den katolska socialläran. Kommissionens uppgift är främst att ge biskopen råd och att främja kunskapen och engagemanget bland de troende i stiftet om socialläran och näraliggande frågor kopplade till hur den katolska tron får prägla den kristnes relationer till sin omvärld, nära och långt bort. Kommissionen representerar också stiftet i både nationella och internationella, liksom i både inom- och utomkyrkliga, sammanhang relevanta för dess ansvarsområde.