

AI, ansvar och vår gemensamma framtid

TL;DR:

– Teknologins framtid kräver en balans mellan teknisk innovation och etisk medvetenhet.

-AI-system kan skapa nya former av rättvisa om de designas med omsorg och värderingar.

-Genom att kombinera historiska lärdomar och moderna metoder kan vi skapa en mänskligare teknologi.

Föreställ er en värld där artificiell intelligens inte längre är en futuristisk vision, utan en integrerad del av vår vardag—lika naturlig som vinden som sveper genom träden en vårdag. Denna framtid är nära förestående och väcker djupgående frågor om hur denna nya teknologi kan samexistera med mänsklighetens djupaste värderingar och etiska principer.

I sin inflytelserika bok *Life 3.0* utforskar fysikern och kosmologen Max Tegmark de potentiella roller och risker som artificiell intelligens (AI) kan innebära för mänskligheten. Han betonar att om AI:s mål inte överensstämmer med våra egna—ett problem känt som "alignment-problemet" inom AI-forskningen—kan konsekvenserna bli allvarliga. Detta aktualiserar behovet av att dra nytta av mänsklighetens

samlade visdom och moraliska insikter för att vägleda AI:s utveckling på ett ansvarsfullt sätt.

”O ni som tror! Var ståndaktiga i rättvisan som vittnen inför Gud, även om det är mot er själva, era föräldrar eller närmaste släktingar.”

—Koranen 4:135 (Skandinaviska stiftelsen för utbildning, 2023)

Denna vers understryker vikten av opartiskhet och integritet—värden som är avgörande när AI-system påverkar människors liv genom beslut om exempelvis kreditvärdighet, anställning eller rättsliga processer. Inom datavetenskapen har frågan om algoritmisk rättvisa blivit central. Studier visar att AI-system kan förstärka befintliga biaser om de tränas på partisk data, ett fenomen känt som algorithmic bias. Detta kan leda till diskriminerande resultat, vilket strider mot principen om rättvisa.

Genom att integrera rättvisa och barmhärtighet i AI:s grundläggande algoritmer kan vi skapa system som aktivt främjar social jämlikhet och skyddar de mest sårbara. Till exempel kan maskininlärningsmodeller utformas med fairness constraints som säkerställer att beslut inte påverkas av irrelevanta faktorer som etnicitet eller kön. Här fungerar principerna från islamisk etik som en etisk kompass som vägleder teknologins utveckling mot mänsklighetens bästa.

Att föreställa sig en framtid där dessa etiska principer är inbyggda i varje AI-system öppnar möjligheter att skapa inte bara effektiva maskiner utan också digitala aktörer med moralisk insikt—en form av artificiell hakim (vis man). Inom området Explainable AI (XAI) arbetar forskare med att göra AI-system mer transparenta och förståeliga, vilket kan öka förtroendet och möjliggöra etisk granskning av deras beslut. Detta representerar en harmonisk förening av mänsklig visdom och teknologisk innovation.

Historien visar att teknologiska framsteg ofta medför både möjligheter och risker. Under den industriella revolutionen möjliggjorde maskiner enorma produktivitetssökningar. Ångmaskinen och järnvägen revolutionerade transport och industri, men ledde också till exploatering av arbetare och sociala orättvisor. Sociologen Karl Marx analyserade detta genom teorin om alienation, där arbetare blev främmande för sitt eget arbete och produkterna de skapade. August Strindbergs *Röda rummet* skildrar liknande teman, där individer kämpar för att hitta mening i ett samhälle som blivit alltmer mekaniserat och avhumaniserat (Strindberg, August, *Röda rummet*. Albert Bonniers Förlag, 1879).

Liknande utmaningar uppstår idag med den digitala revolutionen och AI. Medan teknologi kan förbättra livskvaliteten, finns risken att den skapar nya former av ojämlikhet och distansering. Inom psykologin talar man om ”teknostress”, stress som uppstår till följd av teknologins

påverkan på individen. Elektricitetens intåg förändrade människors livsrytm och arbetsmönster, vilket visar att teknologiska framsteg alltid har sociala och psykologiska konsekvenser.

Harry Martinsons epos *Aniara* erbjuder en djupgående reflektion över människans förhållande till teknologi. I berättelsen flyr människorna från en förstörd jord och sätter sitt hopp till Miman, en avancerad maskin som liknar en AI. Men istället för att erbjuda frälsning, speglar Miman människornas egna brister och existentiella ångest. Detta kan kopplas till det psykologiska begreppet projektion, där individer tillskriver egna känslor och egenskaper till externa objekt. *Aniara* påminner oss om att teknologi inte kan ersätta mänsklig etik eller självförståelse (Martinson, Harry, *Aniara: en revy om människan i tid och rum*. Albert Bonniers Förlag, 1956).

Den grekiska myten om Ikaros illustrerar också riskerna med oansvarig användning av teknologi. Trots varningar flyger Ikaros för nära solen med sina vaxvingar, vilket leder till hans fall. Myten understryker behovet av balans mellan ambition och ödmjukhet, ett tema relevant inom moralpsykologin, där övermod kan leda till etiska övertramp.

”Om man vill ge en definition av Gud, säger man först att Han är allsmäktig; därav följer att Han kan upphäva de lagar Han stiftat.”

—August Strindberg, En blå bok (Strindberg, August, En blå bok. Albert Bonniers Förlag, 1907)

Strindberg betonar här gränserna för mänsklig förståelse och vikten av ödmjukhet inför det okända. Inom kvantfysiken har forskare som Niels Bohr och Werner Heisenberg diskuterat hur observation påverkar det observerade, vilket öppnar för filosofiska och andliga tolkningar av vetenskapen. På liknande sätt kan vi närma oss AI med både vetenskaplig nyfikenhet och etisk medvetenhet, med insikten att vår skapelse har potential att både gynna och skada samhället beroende på hur vi utformar och använder den.

Att utveckla AI som är medveten om och agerar utifrån etiska principer är en komplex men nödvändig uppgift. Inom fältet Machine Ethics försöker forskare skapa algoritmer som kan fatta beslut baserade på etiska teorier, såsom utilitarism eller deontologi. Genom att införliva islamiska etiska principer som rättvisa och barmhärtighet kan vi designa AI-system som är både effektiva och moraliskt ansvarsfulla.

”Vi har verkligen hedrat Adams barn... och försett dem med goda ting och givit dem en högre rang än många av dem Vi har skapat.”

—Koranen 17:70 (Skandinaviska stiftelsen för utbildning, 2023)

Denna heder medför ett ansvar att agera rättvist och med omsorg för allt skapande. Inom miljöpsykologin diskuteras hur människors uppfattning om sin roll i världen påverkar deras beteende gentemot miljön. Genom att se oss själva som förvaltare snarare än ägare kan vi anta en mer hållbar och etisk inställning till både naturen och teknologin.

Genom att integrera dessa värden i AI-utvecklingen kan teknologin bli en kraft som främjar mänskligt välbefinnande och rättvisa, snarare än att förstärka ojämlikheter. Exempelvis kan AI användas för att förbättra hälso- och sjukvården genom tidig och mer exakt diagnos av sjukdomar, men det kräver att dessa system är tillgängliga för alla och inte endast för en privilegierad elit.

Dessa överväganden väcker djupa frågor om mänsklighetens väg framåt. Inom kognitionsvetenskapen studeras hur interaktionen mellan människa och maskin påverkar vår kognition och sociala dynamik. Kan vi förena våra teknologiska ambitioner med våra etiska och andliga värderingar? Historien och litteraturen erbjuder både varningar och insikter som kan vägleda oss.

”Det är en kamp för det sköna på jorden, en kamp som ej slutar förrän den sista gnistan har slocknat.”

—Esaias Tegnér, Samlade Dikter (Tegnér, Esaias, Samlade Dikter. Albert Bonniers Förlag, 1820)

Kanske ligger framtiden i att låta teknologin utvecklas i samklang med mänsklighetens ädlaste värderingar. Genom att betrakta AI inte bara som ett verktyg, utan som en reflektion av våra bästa egenskaper och strävanden, öppnar vi möjligheten att teknologin kan.