peter gärdenfors

fri tanke

kan AI tänka?

om människor djur och robotar

Innehåll

Förord

DEBATTEN OM ARTIFICIELLT tänkande har blossat upp ordentligt de senaste åren. Syftet med den här boken är att bringa reda i de begrepp som används. Mitt mål är att visa att människors och djurs tänkande innehåller ett stort antal komponenter, medan AI-system och robotar är specialiserade till ganska snäva områden. Frågan om ett AI-system kan »tänka« har inget enkelt ja- eller nej-svar. Jag skall visa att det är alltför begränsande att fokusera på systemens »intelligens«. Begreppet är svårt nog att tillämpa på människor och blir väldigt missvisande när det gäller systemens kompetens. Min slutsats blir att det är långt kvar till att de artificiella systemen får ett generellt tänkande. Ett problem för mina jämförelser är att medan kunskapen om människor och djur är relativt beständig, utvecklas flera olika områdena inom AI snabbt och analyser av AI är därför färskvara.

Det har varit svårt att hitta en lämplig titel till boken. Mitt mål är att jämföra människan och de andra djuren med artificiella system. Jag ber om ursäkt för den antropocentriska undertiteln på boken. Människor är också djur.

Flera personer har hjälpt mig med skrivandet. Särskilt stort tack till Christian Balkenius och Trond Arild Tjøstheim som lärt mig mycket om vad robotar kan göra. Också tack till Dan Gärdenfors, Ulf Gärdenfors, Thomas Johannesson, Birger Johansson, Eva Krutmeijer, Johanna Lindbladh, Susanne Lundin och Jakob Stenseke som bidragit med värdefulla kommentarer till tidigare versioner av manuskriptet. Förlagets redaktör Ludvig Köhler har gjort ett gediget arbete med mitt manus och samarbetet med Martina Stenström vid produktionen av boken har varit utmärkt.

Lund, oxveckorna 2024
P.G.

1.

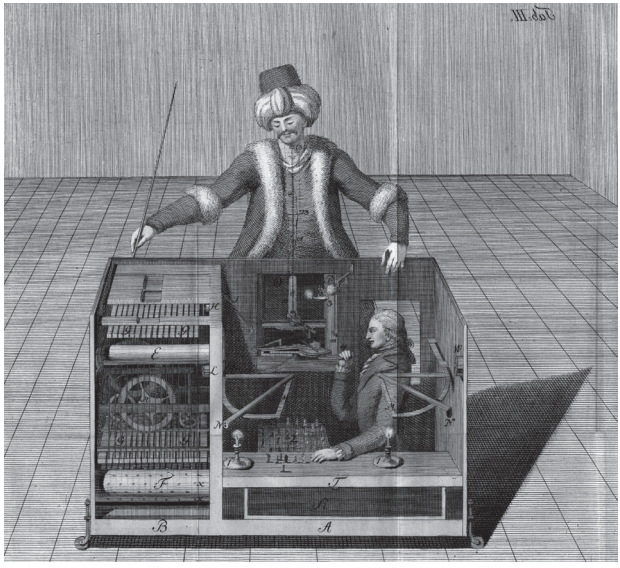
Inledning

MÄNNISKAN HAR ALLTID haft en fascination för kopior av sig själv. Världens äldsta robot är 2 300 år gammal. Roboten finns att se på museet för antik grekisk teknologi i Aten. Det är en konstgjord grekisk tjänarinna som häller upp vin om man placerar en bägare i hennes vänstra hand. Trycket från bägaren leder genom en sinnrik hydraulisk mekanism till att vin pressas ut genom krusets mynning. Detta är ett exempel på att människor länge varit fascinerade av ting som kan agera på människoliknande sätt.[1]

Ett annat exempel är en maskin från 1700-talet där en mekanisk turk spelade schack. Maskinen var en bluff (det fanns en mänsklig schackspelare gömd i ett skåp under turken), men den kittlade åskådarnas fantasi (se figur 1.1).

Förutom mekaniska konstruktioner finns det många berättelser om konstgjorda människor. I den judiska traditionen berättas om rabbinen Löw som levde i Prag på 1500-talet och som skapade en levande varelse – en *golem* – genom att blåsa liv i en lerfigur.

Den mest välkända litterära figuren är kanske Frankensteins monster som gjorts till allmängods genom



FIGUR 1.1. Den mekaniska turken.

flera filmatiseringar. Författaren Mary Shelley utnyttjar den uppfattning om magi som under 1800-talet byggdes upp kring elektricitetens effekter. Galvani hade upptäckt att man kunde få en död grodas ben att sprattla om man gav dem en elektrisk stöt. I Shelleys roman får monstret liv genom blixten från ett åskväder.

En liknande fascination kan man finna när det gäller relationen mellan människor och våra närmaste biologiska släktingar. I fablerna har djur alltid fått människodrag. I modern tid har det gjorts filmer, som

exempelvis *Apornas planet*, där schimpanser beter sig som människor. Även historierna om Mowgli som vuxit upp bland vargar och Tarzan som vuxit upp bland apor är fascinerande, fast på omvänt sätt.

Dessa exempel visar hur lockande det är att sudda ut gränserna mellan människor och andra varelser eller konstruktioner. Bilden på bokens omslag, som är skapad av AI-programmet DALL-E, visar något som är på gränsen mellan djur och robot. I den här boken ligger fokus på gränserna för *tänkandet* hos människor, djur och artificiella system. Det finns mängder av historier om smarta djur, men hur tänker de? Den senaste tiden har AI-system och robotar givit upphov till skräckblandade fantasier om system som kan tänka bättre än människor. Hur kan man jämföra systemens sätt att tänka med människans? En stor del av debatten kring AI och robotar handlar om i vilken mån systemen kommer att ersätta människor.

Ska vi vara rädda för AI?

AI-system används inom allt fler områden av samhället. Våra mobiler är fulla av appar som bygger på AI. Nya tillämpningar som de textbaserade ChatGPT, Copilot och Bard och de bildskapande Midjourney och DALL-E har varit överraskande genombrott i utvecklingen. Det finns också system som kombinerar text- och bildpro-

duktion, exempelvis Googles Gemini. Det nya är att dessa system är *generativa*. Det betyder att systemen kan producera olika typer av nytt innehåll. Generativ AI lockar genom att användargränssnitten är enkla att använda för att skapa text, bilder och programmeringskod på bara några sekunder.

Även robotar dyker upp allt oftare i vår vardag. Dammsugar- och gräsklipparrobotar är vanliga. Men utvecklingen går snabbt även inom andra områden. Den snabba utvecklingen har lett till både förhoppningar om effektiviseringar och farhågor för att systemen ska ta över många jobb.

De AI-system och robotar som finns nu fungerar inom ganska snäva specialområden. Men flera forskare tror att vi snart inte bara har system som uppvisar intelligens inom ett område utan att det kommer att uppstå system med artificiell generell intelligens (AGI). Med detta menas system som blir skickligare än människor inom huvuddelen av de områden som vi behärskar. Jag ska diskutera vad detta skulle kunna innebära och argumentera för att det är långt kvar tills vi når AGI.

Vad innebär det att ett system är »intelligent«? Liknar den form av »intelligens« som AI genererar människans intelligens? Kan man över huvud taget säga att ett artificiellt system kan tänka? Det är sådana frågor som den här boken ska handla om.

En del forskare har även framfört farhågor om att robotar kommer att ta över världen. Så ska vi vara rädda

för AI? Det första man måste lägga märke till är att AI inte är något enhetligt fenomen eller någon unik metod. Många olika områden har kommit att sorteras under termen artificiell intelligens. Från början handlade det mest om *symbolhantering*, det vill säga att man utgick från att tänkande sker i något slags inre programspråk (se kapitel 7). På senare tid har i stället flera områden inom *maskininlärning* kommit i fokus (se kapitel 8).

För många är *robotik* en del av AI (se kapitel 10). Ordet *robot*, som infördes av författaren Karel Čapek, har bildats av ett slaviskt ord för att arbeta. En robot är alltså något som utför konkreta handlingar. De flesta AI-systemen utför inga handlingar utan visar sina resultat på en skärm – siffror, texter eller bilder. En robot däremot måste *agera* i världen och den måste därmed ha någon form av »kropp«. Om man flyttar över ett AI-program till en robot, så blir den inte mer effektiv förrän man har löst problemen med hur programmet ska styra robotens rörelser. Detta är långt ifrån enkelt visar det sig. Även om ett program är bra på att hitta mönster eller att skapa text eller bilder, så innebär det stora svårigheter att få en robot att använda sådana färdigheter för att utföra nya konkreta handlingar. I fortsättningen kommer jag därför att skilja mellan robotik och andra AI-system.

Förutom diskmaskiner, tvättmaskiner och torktumlare, som egentligen är robotar, är dammsugar- och gräsklipparrobotar de som vi oftast träffar på i vardagen. Självkörande bilar och andra fordon kommer snart att

paradoxalt att problem som människor lätt klarar av är väldigt svåra för artificiella system (till exempel att plocka upp något man tappat på golvet eller att se vad en bild föreställer), medan många uppgifter som är svåra för människor kan vara lätta för artificiella system (till exempel att klassificera röntgenbilder). Människor (och djur) är generalister – vi kan hantera problem inom väldigt många områden, även om våra lösningar inte alltid är de bästa eller smartaste. På detta sätt uppvisar vi någon form av »generell intelligens«. Människors och djurs tänkande är också resultat av evolutionära processer i den verkliga världen, men det är inte AI-systemens. Människor (och djur) har känslor, motivation och en vilja som alla ingår i våra medvetanden. Jag ska argumentera för att dessa faktorer är oerhört svåra att återskapa i artificiella system.

Å andra sidan har maskinerna speciella förmågor som överskrider människornas. Så en jämförelse mellan maskiner och människor blir tveeggad. Här återkommer spänningen mellan att se AI-system som individer och att se dem som hjälpmedel. Precis som att robotars »kroppar« inte behöver likna människors, så leder det fel att utgå från att AI-systems »tänkande« måste ha som mål att likna människans. I likhet med annan teknologi bör målet i stället vara att *komplettera* människans förmågor.

Ett problem vid analysen av de artificiella systemen är att vi tolkar deras beteenden i mänskliga termer. De ord som vi använder för att tala om vilka förmågor AI-sys-

tem och robotar uppvisar, till exempel »intelligens«, utgår ifrån människan, men det är inte alls säkert att de går att tillämpa på AI och robotar. Det är inte lätt att avgöra hur »intelligent« ett system är, för att inte tala om vad det skulle innebära att det är »superintelligent« (se kapitel 6). Ett huvudsyfte med boken är att gå igenom vad begrepp som »tänka«, »intelligens« och »medvetande« kan betyda om vi ska tillämpa dem inom AI-områdena. Som jag ska visa är detta långt ifrån klart och att använda mänskliga termer om artificiella system leder ofta fel.

Datalogen John McCarthy, som år 1956 myntade begreppet »artificiell intelligens«, har i efterhand sagt att valet av ord inte var särskilt genomtänkt. Uttrycket har tyvärr gett upphov till många missförstånd. Jag kommer därför att markera ordet intelligent som »intelligent« för att hissa en varningsflagga för hur det används om artificiella system. Samma sak gäller »tänkande«.

Den »intelligens« som AI-systemen uppvisar är väsensskild från människans. Det är svårt att hitta en definition av intelligens som samtidigt kan tillämpas på människor, djur och artificiella system. AI och robotar bör bedömas enligt nya kriterier. Snarare än som ett inlägg i debatten om AI, kan den här boken ses som en bakgrund som behövs för att man ska kunna diskutera vad AI-system kan och inte kan.

Som en bas för diskussionen av AI-systemen ska jag först visa att människans och djurens tänkande är mång-

sidigt. Min genomgång av de kognitiva funktionerna i nästa kapitel behövs för att förstå att det är mycket som saknas innan AI-system och robotar kommer i närheten av vad människor kan. Även om vi kommer att få flera AI-system som har specialförmågor, så är det långt kvar till den mänskliga förmågan till generalisering. På flera sätt blir det därför märkligt att tala om tänkande maskiner.