



Bild: Hannu Toivonen

# Vad är artificiell intelligens?

Jag föreslår följande karaktärisering: artificiell intelligens är data-program som fungerar ändamålsenligt vid lösningen av svåra uppgifter med stöd av information och resurser som står till dessas förfogande. Ju svårare en uppgift anses vara för en dator, desto sannolikare är det att man talar om artificiell intelligens.

**E**n mobil identifierar en person från ett fotografi, en navigator hittar den snabbaste vägen till ett resmål, en språkmodell löser ett problem eller hittar på en godnattsaga. I alla dessa fall är det frågan om artificiell intelligens, men vilken är den gemensamma nämnaren som gör detta till artificiell intelligens (AI)?

Artificiell intelligens är ett otydligt begrepp med många betydelser, som det är svårt att få grepp om. Inom industri och teknik hänför sig artificiell intelligens ofta till begrepp såsom maskininlärning och neuronnätverk, men ur forskarens synvinkel är dessa bara en del av den artificiella intelligensens breda vetenskapsområde. Å andra sidan är inte heller forskarna inom området överens om vad som definierar detta.

Att ordet ”intelligens” ingår i ”artificiell intelligens” väcker oundvikligen tanken om att intelligens är ett kriterium för artificiell intelligens. Artificiell intelligens behöver likväl inte vara intelligent, utan det är frågan om en metafor: i artificiell intelligens ingår något som är jämförbart med intelligens. Många applikationer som betraktas som artificiell intelligens kräver inte särskilt mycket intelligens om de utförs av en människa; dylika är till exempel att känna igen ansikten eller producera tal. Å andra sidan anses vissa uppgifter som kräver intelligens, till exempel att lösa en svår räkneuppgift, sällan som ett exempel på artificiell intelligens.

Jag föreslår följande karaktärisering av artificiell intelligens: artificiell intelligens är dataprogram som fungerar ändamålsenligt för svåra uppgifter med stöd av den information och de resurser som står till förfogande. En navigators uppgift är till exempel att hitta en snabb rutt med stöd av den kartinformation och de trafikstock-

**Hannu Toivonen**  
är professor  
vid Helsingfors  
universitet.



ningsuppgifter som den har tillgång till. Ju svårare en uppgift anses vara för datorn, desto sannolikare är det att man talar om artificiell intelligens. Eftersom navigators – för att inte tala om till exempel kalkylatorer – redan har funnits länge, anser många att det inte längre är frågan om artificiell intelligens. I vardagligt tal är artificiell intelligens ett rörligt mål: som artificiell intelligens betraktas särskilt sådana uppgifter som man endast någorlunda kan utföra med hjälp av datorer. Då de kan utföras pålitligt förvandlas de till vanliga dataprogram.

Varför talar man så mycket om artificiell intelligens just nu? Det finns tre orsaker till att diskussionen om artificiell intelligens intensifierats på 2020-talet som alla kulminerar i språkmodeller.

Snabb utveckling: med språkmodeller kan man utföra uppgifter som man för några år sedan ansåg nära nog omöjliga.

Tillgängliga program: bra språkmodeller är tillgängliga för envar mot skälig betalning.

Språkmodellernas generella användbarhet: med språkmodellernas hjälp kan man utföra mångsidiga uppgifter som kan formuleras skriftligt. Då många plötsligt kan använda artificiell intelligens mångsidigt uppstår nya möjligheter liksom nya risker.

Den ökade intensiteten i diskussionen om artificiell intelligens påverkas utan tvekan också av mänskliga orsaker. Med tanke på vår självbild har det varit en lättnad att en mängd förmågor skiljer människan från en maskin, men nu håller artificiell intelligens

på att utmana människans särställning till exempel då det gäller språkförståelse och -produktion, till och med då det gäller nyanser och kreativitet. Vi har redan tidigare vant oss vid att datorernas informationsmängd och snabbhet överskrider människans förmåga i många avseenden. I ljuset av allt detta: vad är mänsklighet och vad skiljer människan från en maskin?

Jag strävar efter att i denna artikel öppna upp vad artificiell intelligens är och samtidigt också vad den inte är.

### **Fyra nivåer**

Om den artificiella intelligensen är data-program som fungerar på ett ändamålsenligt sätt, hur får man då ett program att fungera ändamålsenligt? Man kan närma sig detta på fyra sätt: programmering, optimering, maskininläring och språkmodeller. De har utvecklats i den här ordningen, alltid med grund i de tidigare utvecklingsfaserna.

Programmering handlar om att ge entydiga instruktioner. Programmeraren måste alltså veta vad som är ändamålsenligt i varje situation för att kunna skriva ett program som fungerar på det sättet. Traditionell programmering underskattas ofta i samband med artificiell intelligens, fastän den ligger bakom majoriteten av AI-program. Alltid då man vet hur en uppgift ska utföras lönar det sig att programmera detta direkt. Rutt-sökningen i en navigator är ett exempel på detta: för uppgiften finns kända lösningsmetoder.

Inom optimering är idén att definiera egenskaperna hos en ändamålsenlig lösning eller ett önskat resultat och särskilja dem från lösningsmetoden. I tillägg till optimering förekommer samma idé vid så kallad deklarativ programmering, till exempel vid logisk programmering. I dessa används en generell optimerings- eller slutledningsmotor för att hitta den bästa lösningen

då uppgiften och målet definierats.

En exempelapplikation är att uppgöra en läsordning för en skola: man vill hitta sådana gruppindelningar av elever, lärare och scheman att ingen behöver vara på två ställen samtidigt, eleverna har få eller inga håltimmar, skoldagarna är ungefär lika långa och så vidare.

Optimering definieras på ett abstraktare begreppsligt plan än programmering: man behöver inte veta hur problemet löses, det räcker med att man förmår definiera kriterierna för en bra lösning. Som hjälp behöver man ett optimerings- eller slutledningsprogram som någon programmerat. Höjningen av abstraktionsnivån möjliggörs av hur generellt användbart ifrågavarande program är: det förmår lösa ett flertal olika problem.

Vid maskininläring ges programmet exempel på vad man vill att det ska göra, och den ändamålsenliga verksamheten grundar sig på att imitera exemplen. Ett program för ansiktsgenkänning tränas till exempel upp med olika människors ansiktsbilder och ett namn som kopplats till respektive bild. Då man visar den tränade ansiktsgenkänningsmodellen en ansiktspild producerar modellen det namn som bilden enligt modellens uppskattning skulle ha i träningsdata.

Maskininläring är noggrannare sagt att bilda statistiska generaliseringar av ett givet inlärningsmaterial. Övervakad inläring innebär imitation baserad på dessa generaliseringar, medan oövervakad inläring handlar om att göra generaliseringar till exempel för att förstå träningsdatas struktur. Förstärkningsinläring sker igen genom växelverkan och handlar om att lära sig genom att göra. Maskininläring är ofta övervakad så jag talar här i huvudsak om den.

(Övervakad) Maskininläring utnyttjar optimering: ett maskininlärningsprogram

## ”Till den högre abstraktionsnivån hör en avigsida: de modeller som skapas genom maskininlärning är ofta oförståeliga för människan.”

strävar efter att skapa en sådan modell (generaliseringar) med hjälp av vilken imitationen sker möjligast noggrant. I fallet med ansiktsgigenkänning optimeras frekvensen för hur ofta modellen genererar rätt namn då den matas med en ny bild av en person som förekommit i träningsdata. Fastän maskininlärning använder optimering så löser maskininlärning och optimering olika problem och lämpar sig för olika uppgifter. Det skulle vara svårt att generera en ny läsordning med hjälp av maskininlärning som baserar sig på imitation, medan ansiktsgigenkänning å andra sidan skulle vara svår att definiera som en optimeringsuppgift utan träningsdata.

Maskininlärningens abstraktionsnivå är högre än för optimering och lättare att använda, eftersom problemet och lösningskriterierna inte behöver definieras mer än genom exempel. Ibland är lättheten skenbar: om man inte själv förstår problemet och dess svårighetsgrad, kan det vara svårt att med hjälp av maskininlärning producera bra lösningar. Till den högre abstraktionsnivån hör också en annan avigsida: de modeller som skapas genom maskininlärning är ofta oförståeliga för människan varför det kan vara svårt att komma underfund med programmets verksamhetsprinciper eller orsakerna till dess fel.

Uppgiften till en språkmodell formu-

leras skriftligt, såsom till en människa, och som resultat producerar den text. En språkmodell är därför lätt att använda. En språkmodell är också allmänt användbar i motsats till applikationer som producerats med maskininlärning eller optimering: en språkmodell förmår behandla till exempel olika ämnen, språk och stilarter mångsidigt.

En språkmodell bygger på maskininlärning: den producerar text, som imiterar träningsdata. Förenklat sagt: språkmodellen väljer ett lämpligt ord som fortsättning på nuvarande text, sedan igen följande ord, och så vidare.

Stora språkmodeller klarar av många uppgifter av två orsaker. För det första bygger de på väldigt omfattande och mångsidiga träningsdata. För det andra är modellerna i sig själva omfattande – de har tio eller hundra miljarder parametrar – och de generaliseringar de innehåller är i bästa fall heltäckande och nyanserade både vad gäller texters språkliga konstruktion, såsom grammatik, och olika texttyper, uppgifter och lösningsmodeller.

Språkmodeller gör fel. Såsom alltid vid maskininlärning kan träningsdata, de generaliseringar modellen lärt sig samt deras tillämpning innehålla olika fel och snedvridningar. Dessa är likväl svåra att lokalisera eller ens reda ut i efterskott, då man inte klarar av att verbalisera språkmodellernas funktion. För användaren framstår de som en svart låda, vars verksamhetsprinciper och begränsningar lämnas åt gissningar. Då abstraktionsnivån och den allmänna användbarheten ökat från programmering till optimering och vidare till maskininlärning och språkmodeller, har samtidigt förmågan att behärska eller ens förklara AI-program minskat.

Fastän språkmodellerna är mekanismer för textgenerering så är de opålitliga för att skapa nytt innehåll då det innehåll de gene-

rerar kan innehålla grova fel. Språkmodellerna är som bäst för att manipulera ord och text inom ramen för de generaliseringar de lärt sig. De kan till exempel användas som en del av andra applikationer för att hantera talade och skrivna instruktioner.

### Vad artificiell intelligens inte är

Artificiell intelligens är som sagt ett otydligt begrepp och det töjs ytterligare till exempel i marknadsföring eller rafflande rubriker. Då språkmodellerna därtill fungerar mångsidigt men oförklarligt kan det uppstå oklarhet om vilka egenskaper de möjligen har.

En språkmodell förstår inte en text, varken den text som den blivit tilldelad eller den som modellen själv producerat. Språkmodellen manipulerar endast tecken- sekvenser. Den enda betydelse som ord eller uttryck har för språkmodellen är de statistiska sannolikheter teckensekvenserna har baserat på träningsdata. De fel som språkmodellerna gör försvinner inte genom att utveckla maskininläringen då en del av feLEN beror på att språkmodellen inte hanterar betydelser.

En språkmodell är en modell av ett språk eller en text, medan människan har en mental modell av världen och förhållandet mellan språkliga uttryck och den, vilket ger uttrycken deras betydelser. De generaliseringar om regelbundenheter i träningsdata som med hjälp av maskininläring skapats i en språkmodell är den närmaste motsvarigheten till människans mentala modeller. Ändå är de otvetydigt enbart en ytlig representation av den rika modell av världen som människan lär sig genom både språkliga och icke-språkliga erfarenheter i såväl sin fysiska som sociala och inre värld.

På basis av det föregående torde det vara klart att en språkmodell varken tänker eller är medveten. Den kan nog vid ett samtal

ge en bild av sig själv som en tänkande och medveten varelse. Människans mottaglighet för att se språkmodellen som en tänkande och medveten varelse torde bero på att språket uttryckligen är ett medel för att framföra och förmedla betydelser och tankar. Det strider mot vår erfarenhet att någon som flytande kan använda ett språk inte skulle tänka eller förstå de betydelser som vederbörande uttrycker.

I rubriker utmålas dataprogram emellanåt som självständiga aktörer, som har vilja och mål: "Artificiell intelligens snokar efter lösenord". I artikeln är det frågan om att artificiell intelligens kan användas för att snoka efter lösenord om någon beslutade sig för att göra det.

Det andra problemet med rubriken är att den talar om ospecificerad artificiell intelligens. Man kan därför få en uppfattning om en artificiell intelligens som klarar av mycket – vad manne den gör om den får reda på vårt lösenord? Det handlar dock om utvecklingen av ett nytt dataprogram som känner igen vilken tangent man trycker på utifrån det ljud den producerar.

Exempelrubriken gör sig också skyldig till att tillspetsa saken. Artikeln handlar om ett forskningsprojekt, vars mål var att utreda nya hackningsrisker. En saklig rubrik kunde till exempel lyda så här: "Forskare testade användning av maskininläring för att känna igen lösenord med stöd av tangentljud."

Maskininläringens framgång både i sig själv och som grund för språkmodeller kan leda till en snäv uppfattning av vad artificiell intelligens är samt till orealistiska förväntningar på maskininläring.

Dataprogram som bygger på maskininläring är inte opartiska: de reflekterar snedvridningar och fel i träningsdata – och som för alla dataprogram även programmerarnas värderingar och världsbild med början redan i de ändamål för vilka pro-

## **”En språkmodell förstår inte en text, varken den text som den blivit tilldelad eller den som modellen själv producerat. Språkmodellen manipulerar endast teckensekvenser.”**

grammen utvecklas. Tanken på opartiskhet är ofta redan i sig själv en chimär: det är sällsynt att svåra problem skulle ha endast en entydigt rätt lösning. Världen är komplex och beslutsfattandet svårt eftersom man ofta måste välja mellan motstridiga värderingar – hur mycket betonar man till exempel grundrättigheter och hur mycket säkerhet? Människor behövs för att gestalta helheten och göra välvägdade val mellan värderingar eller människogrupper samt för att bära ansvar för besluten.

Ibland tänker man att ett AI-program i princip inte kan vara intelligent (eller kreativt eller medvetet). Det är fruktbarare att förkasta tanken på egenskapernas tvära tudelning – att programmet antingen är intelligent eller inte – och i stället utforska i vilket avseende och i vilken grad ett program är intelligent eller inte. Dataprogram kan vara mycket bättre än människor på många färdigheter som traditionellt förknippats med intelligens, såsom logiska resonemang, matematik, minne och återkallande av information, samt i hastighet och noggrannhet inom dessa områden. Å andra sidan kan program vara svagare på många vardagliga sysslor, som att se, höra och röra sig, för att inte tala om empati, självreflektion eller moral.

En lekman kan tro att målet med forsk-

ning kring artificiell intelligens är att skapa en maskin som på ett mänskligt sätt är intelligent. Föreställningen går i rätt riktning när det gäller språkmodeller: i dessa betonas mänsklig kommunikation, flexibilitet och mångsidig användning. Största delen av forskning kring artificiell intelligens strävar likväl inte efter mänsklighet eller ens intelligens. I stället utvecklar forskarna lösningar på olika enskilda men svåra uppgifter, för att få ett dataprogram att fungera ändamålsenligt just i samband med dessa.

### **Generativ AI och kreativitet**

Med generativ artificiell intelligens avses dataprogram som till exempel producerar text, bilder, videor, musik eller tal. Generativ artificiell intelligens baserar sig i allmänhet på maskininlärning och alltså på att imitera träningsdata med stöd av generaliseringar som bygger på detta. Språkmodeller är en känd modell av generativ artificiell intelligens, bildgeneratorer en annan.

Kan generativa program som bygger på maskininlärning vara kreativa? Kreativitet är ett omstritt begrepp, men i allmänhet förstås det som en förmåga att skapa nya och ändamålsenliga resultat. En intressant fråga när det gäller generativ artificiell intelligens är begreppet nyhet, eftersom verksamhet som bygger på imitation uttryckligen undviker att skapa nytt.

Resultatet kan likväl vara nytt, eftersom det finns utrymme för kreativitet inom ramen för generaliseringar. För att text som producerats av en språkmodell ska vara förstäelig måste den följa (inlärdd) grammatik – men den ordsekvens som den producerat kan vara ny. För att texten ska vara meningsfull måste språkmodellen följa de instruktioner den blivit tilldelad liksom (inlärda) konventioner – men innehållet kan vara nytt. Att följa generaliseringarna kan således vara en förutsättning för att

producera ett förståeligt och ändamålsenligt innehåll, inte bara ett hinder för att skapa något nytt.

Både en människa och till exempel en språkmodell kan skapa något nytt genom att kombinera tidigare kända företeelser – större delen av människans kreativitet torde handla om detta. Språkmodellerna kombineras vid behov flexibelt olika generaliseringar och skapar nytt innehåll, som, när det lyckas, är meningsfullt. En språkmodell producerar till exempel en idé till en saga, som berättar om tyngdkraften och lycka: ”I Levitias kungadöme lever man i vackra svävande slott men man är olycklig. En nyfiken flicka vid namn Amara märker att lycka uppkommer genom kontakt med jorden och med hjälp av en klok uggle får hon Levitias befolkning att hitta kontakten med naturen och varandra.”

Ett generativt programs möjliga kreativitet kan ta sig uttryck på olika sätt och i olika utsträckning. Detsamma gäller människan som styr den kreativa processen. Språkmodellen skapade en ram för en saga som möjligen kan ses som kreativ, men vem fick idén att göra en saga om tyngdkraft och lycka? Den uppkom som ett samarbete: människan bad först om ett begreppspår, som inte på något sätt har med varandra att göra och språkmodellen föreslog tyngdkraft och lycka. Sedan bad människan om en idé till en saga, som behandlar bägge ämnena.

Men är idén om att kombinera tyngdkraft och lycka ny? Fastän resultatet som producerats av det generativa programmet är nytt för användaren, är det inte nödvändigtvis det med beaktande av hela träningsdata. Hittade språkmodellen själv på att tyngdkraft och lycka inte hänför sig till varandra, eller kommer begreppsparet från träningsdata? Det kan vara omöjligt för användaren att veta detta – på samma sätt kan det vid generering av fakta vara omöjligt att veta

om det innehåll språkmodellen producerat är sant eller påhittat.

Generering, kreativitet och konst hänför sig som begrepp till varandra och kan ibland blandas ihop i onödan i samband med artificiell intelligens. Man kan tillspetsa begreppens skillnader på följande sätt.

Generering är att skapa innehåll i teknisk bemärkelse. Förmågan hos generativa AI-program att skapa innehåll råder det ingen tvekan om, även om man kan vara av många uppfattningar om resultatens kvalitet.

Kreativitet i sin tur är långt en erfarenhetsmässig och subjektiv egenskap. Våra personliga uppfattningar om vilka resultat (som genererats av dataprogram) som är kreativa, varierar beroende på person, kultur och tidpunkt.

I konsten framträder i sin tur kulturen och samfunden: som konst betraktas mera sannolikt verk som är godkända av erkända ”portvakter”, till exempel gallerister, förläggare eller museer. Konst kopplas ofta starkt ihop med mänsklighet och konst har beskrivits som kommunikation människor emellan. Därför kan det vara svårt för ett generativt AI-program att göra ett genombrott som självständig konstnär fastän man skulle kunna skönja kreativitet i dess resultat.

Som verktyg för oss människor är läget likväl ett annat. En av den artificiella intelligensens intressantaste tillämpningar är att stödja människans kreativitet. I ovanstående sagoidé hade vardera sin kreativa roll, för ifrågavarande idé hade inte uppkommit utan bådas insats. Det är forskningsmässigt och filosofiskt intressant hurudana skapande roller ett AI-program kan ha, men ännu väsentligare är på vilket sätt programmet kan stödja eller rentav utvidga sin användares kreativitet.

Översättning från finska: Max Oker-Blom



Hannu Toivonen gör också AI-konst – bilder helt och hållet gjorda med olika AI-progam. Underrubriken till hans bildserie "Världens lyckligaste land" är "Falsa minnen från lyckliga stunder år 1973".

