

# Talande maskiner och sociala robotar

*Gabriel Skantze*

I filmen *Cast Away* spelar Tom Hanks en modern Robinson Crusoe som blir strandsatt på en öde ö i Stilla havet efter en flygkrasch. Under flera år tvingas han klara sig som enda överlevande på ön, innan han slutligen räddas därifrån. Den största utmaningen blir dock inte att hitta mat eller uthärda stormar och sjukdomar, utan att stå ut med ensamheten. För att lösa detta tar han en trasig volleyboll som han hittat, ritar ett ansikte på den, och ger den namnet Wilson. Nu har han plötsligt någon som han kan prata med, skälla på när saker går snett, och dela sina erfarenheter med. Det är oklart hur stor poäng filmen egentligen vill göra av denna relation, eller om den helt enkelt är ett smart dramaturgiskt grepp: En hel film med en enda skådespelare utan någon som helst dialog hade kanske blivit tråkig att titta på. Men den säger samtidigt någonting om hur vi människor är skapade att fungera. Vi har genom evolutionen formats till sociala varelser, och vi är vana vid att ha andra människor i vår omgivning som vi kan dela våra erfarenheter med. Det handlar inte bara om att vi annars skulle bli uttråkade, behovet är mycket mer fundamentalt. Det är genom interaktion med andra som vårt tänkande och vår världsuppfattning formas och utvecklas, och utan den får vi svårt att skapa mening i tillvaron. Vi lärde oss tidigt i evolutionen att vår överlevnad hängde på att ingå i en social gemenskap och finna vår roll i den gemenskapen. Att rycka

en människa ur denna sociala väv som vi formats att ingå i och placera henne på en öde ö är för de allra flesta ett plågsamt straff.

Givet att social interaktion är en så central del i vad det är att vara människa, är det då möjligt att skapa en artefakt, till exempel en dator eller robot, som kan föra ett samtal med människor, eller på annat sätt interagera socialt, och fylla samma funktion som en annan människa? Och då inte bara någonting vi kan tala till, som Tom Hanks gör med den trasiga volleybollen, utan som vi faktiskt talar och interagerar med? I vilka sammanhang skulle detta vara önskvärt? Och hur skulle en sådan robot konstrueras? Det är detta min forskning handlar om, och i detta kapitel kommer jag att diskutera dessa frågor. Jag kommer att inleda med en kort historik över tidigare försök att skapa människoliknande robotar fram till dagens sociala robotar, deras användningsområden, och varför man skulle vilja att de liknar oss människor. Därefter kommer jag att diskutera varför det är en så stor utmaning att ge robotar förmågan att tala med oss, och vilken form av artificiell intelligens som är nödvändig för detta. Avslutningsvis kommer jag att diskutera framtida utmaningar som den här utvecklingen är förknippad med, både rent tekniskt, och för vårt samhälle.

### *Sociala robotar*

Sedan urminnes tider har människan fascinerats av tanken på att konstruera en avbild av sig själv, en artefakt eller maskin som tänker och betar sig som vi, och som vi kan interagera med. Redan i *Iliaden*, en av våra äldsta bevarade skrifter, beskrivs hur smedguden Hefaistos tillverkar automatiska hjälpare. Ett tidigt exempel på föregångare till vår tids robotar är de mekaniska figurer som tillverkades under antiken till automatteatrar, och som kunde spela upp teaterpjäser. Allteftersom tekniken utvecklades kunde man tillverka allt mer avancerade människolika maskiner. På 1700-talet blev det populärt att bygga så kallade automata, en form av robotar som ofta liknade människor, och kunde göra saker som att spela schack, skriva med penna och papper, eller spela flöjt. Men även om dessa maskiner krävde en för sin tid mycket avancerad ingenjörskonst (ofta av urmakare), så kunde de inte



Figur 1. Teckning av Wolfgang von Kempelens *Mechanical Turk*, som även visar utrymmet där roboten antas ha styrts i hemlighet.

utföra något vad vi kallar autonomt beteende, det vill säga de kunde inte fatta egna beslut baserade på sinnesintryck från omgivningen; de kunde bara spela upp ett på förhand utformat program. Därför kunde man inte heller interagera med dem, eftersom interaktion kräver att man tar in intryck från omgivningen och anpassar sitt beteende. De robotar som man till synes kunde interagera med, som Wolfgang von Kempelens schackspelande robot *Mechanical Turk*, styrdes i själva verket av en person som satt gömd i en låda. Inte heller kunde de utföra något egentligt arbete, utan var främst en kuriositet.

Ordet robot, vilket betyder 'slav' på tjeckiska, dyker upp första gången 1920 i Karel Čapeks pjäs *R.U.R. (Rossum's Universal Robots)*, vilken handlar om människoliknande robotar på en fabrik som gör uppror mot människorna. Under mitten av 1900-talet började man använda ordet robot för att tala om de maskiner som blev allt vanligare inom tillverkningsindustrin, och som började ersätta mänskligt arbete i stor skala, framförallt arbete som var tungt och farligt. Men inte heller dessa robotar var egentligen autonoma; de utförde också bara förprogrammerade sekvenser. Om man skulle ta bort bilen som en sådan robot arbetar med att montera, så skulle roboten helt enkelt fortsätta att utföra samma rörelser i luften.

Det är först på senare år som robotar utrustade med sensorer och någon form av artificiell intelligens (AI) har börjat bli tillräckligt bra för att hitta praktiska tillämpningar i större skala. Idag har vi till exempel robotar som kan klippa gräs eller dammsuga, och för detta behöver de ha sensorer och kunna navigera i omgivningen, om än på en ganska enkel nivå. Men vi börjar också se självkörande bilar och robotar som kan plocka och paketera varor, vilket kräver långt mer avancerade sensorer, mönsterigenkänning, beslutsförmåga och motorik. Dessa är dock samtliga exempel på robotar som utför praktiska uppgifter åt oss. En annan kategori robotar, som jag själv forskar kring, är sociala robotar, vars primära syfte varken är att förflytta sig eller fysiskt manipulera sin omgivning, utan att interagera med oss människor.

Flera forskare och analytiker ser framför sig att en stor del av det arbete som tidigare endast kunde utföras av människor inom en snar framtid även kommer att kunna utföras av robotar. Det handlar förstas delvis om fysiskt arbete, som att städa och hjälpa till med tunga lyft inom sjukvården, men mycket av detta arbete innefattar också interaktion med människor. I hotellreceptionen skulle roboten kunna checka in kunder. I varuhuset skulle den kunna svara på frågor och hjälpa kunderna med att hitta varorna. På museet skulle den kunna guida besökarna och berätta om utställningarna. I skolan skulle den kunna göra individuellt anpassade övningar med eleverna. Hemma hos äldre människor skulle den kunna vara ständigt närvarande för att påminna om att ta mediciner, vara en social samtalspartner, och diagnostisera tecken på begynnande demens.

Ett exempel på en enklare form av social robot är robohusdjur. Flera studier har visat att äldre personer med demens reagerar positivt på att få interagera med djur, till exempel en hund. Problemet är att det kan vara en fara för hunden att bli lämnad ensam med en demenssjuk person, eftersom denne kan ha svårt att ta hand om hunden på rätt sätt, och det är inte heller säkert att hunden är intresserad av att sitta flera timmar i knäet och bli klappad. I Japan har man därför utvecklat en robotsäl med namnet Paro. När jag själv provade att interagera med Paro för första gången blev jag förvånad över hur stark effekten var: Den

behövde bara vända upp huvudet och titta upp på mig med sina stora blanka ögon för att jag skulle känna ett band till den. Sälen ger även ifrån sig belåtna ljud när man klappar den, och man måste ”mata” den (med elektrisk ström) med jämna mellanrum. På sätt och vis är förstås Paro inte mycket mer än en avancerad leksak – dess ”hjärna” är inte särskilt sofistikerad. Men flera studier har ändå visat att demenssjuka personer som fick umgås med sälen upplevde ett större välbefinnande och att deras stressnivå sjönk. Men även om patienterna i viss utsträckning talade till sälen (som man ibland gör med djur), så har den, likt ett verkligt husdjur, ingen förmåga att föra ett samtal.

Min forskning handlar om att förstå och modellera mänskligt samtal, och hur dessa modeller kan användas för att bygga datorer och robotar som kan föra samtal med människor. För att bygga sådana modeller spelar vi in stora mängder samtal mellan människor och använder sedan statistiska modeller som tränas på att känna igen och förutsäga mönster i samtalet, göra tolkningar och fatta beslut. Syftet är dock inte bara att utveckla talande robotar; forskningen drivs även



Figur 2. Roboten Furhat spelar ett spel tillsammans med två barn på Tekniska Museet i Stockholm.

av ett djupare intresse för att förstå mekanismerna i mänsklig kommunikation. Därför utför vi även kontrollerade experiment där vi undersöker hur människor reagerar på robotens beteende. Detta är ett i allra högsta grad tvärvetenskapligt forskningsområde, som spänner över datavetenskap, artificiell intelligens, maskininlärning, lingvistik, fonetik och psykologi.

### *Att ge maskinen ett ansikte*

För att kunna testa våra modeller i verkliga samtal har vi på KTH utvecklat robothuvudet Furhat, som visas i figur 2. Robotens ansikte består av en mask på vilken ett animerat ansikte projiceras inifrån med hjälp av en liten projektor inne i huvudet. Den har även en mekanisk nacke som gör att den kan röra på huvudet, och tillsammans med animerade ögonrörelser, ansiktsuttryck och läpprörelser blir det möjligt att på ett subtilt och naturligt sätt återskapa de signaler som är viktiga för mänsklig kommunikation, som att söka ögonkontakt med den man pratar med, höja på ögonbrynen när man blir förvånad, eller le när man ser någon man är bekant med. Som "ögon" använder vi djupseende kameror som kan detektera ansikten och ansiktsuttryck, och som "öron" mikrofoner som kan uppfatta riktningen från vilket ljudet kommer. Detta gör att Furhat kan tala med flera personer samtidigt. I figur 2 spelar till exempel Furhat ett spel tillsammans med två barn. Tillsammans ska de sortera fem kort i ordning, exempelvis kända byggnader efter deras höjd, eller planeter efter deras storlek. Roboten diskuterar lösningen tillsammans med barnen, men är programmerad att inte känna till den faktiska lösningen. Istället är den utrustad med en modell av vad den tror att lösningen ska vara, som den kan vara mer eller mindre säker på. Detta ger upphov till en diskussion där man tvingas motivera sina val. Scenariot är främst framtaget för att kunna studera hur vi kan modellera samtal med flera personer, men vi tittar även på hur det till exempel skulle kunna användas för språkinlärning.

Men vad tillför då robothuvudet, jämfört med att bara ha en röst i en högtalare, eller ett animerat ansikte på en skärm? Den stora skillnaden är att roboten får en fysisk plats i rummet. Denna skillnad

har de flesta erfart när de försökt ha ett samtal med flera personer över en videolänk eller med enbart rösten över telefon, då det ofta uppstår förvirring, till exempel kring vems tur det är att prata. Därför är människor villiga att spendera stora summor på resor för att kunna samtala i fysiska möten. En sak vi tidigt upptäckte med Furhat var att man kunde få ögonkontakt med den och när man får det ger det en väldigt stark känsla av närvaro. Detta är en stor skillnad mot ansikten på platta skärmar, där man får en så kallad "Mona Lisa-effekt". När till exempel en nyhetsuppläsare på tv tittar rakt in i kameran så upplever alla i rummet att de har ögonkontakt med uppläsaren. Detsamma gäller när vi ser ett porträtt på en tavla eller ett fotografi. Om personen tittar åt sidan upplever däremot ingen, oavsett var de står i rummet, att de får ögonkontakt. I kontrollerade experiment har vi jämfört hur det är när flera personer interagerar med Furhat, jämfört med en animerad figur på en platt skärm, och resultaten visar att det senare är mycket svårare, då figuren inte tydligt kan adressera enskilda personer. Detta gäller inte bara vem man tittar på, utan också vad man tittar på.

När roboten tittar på spelkorten på bordet är det också möjligt för andra att läsa av robotens blick och dra slutsatser om vad den "tänker" på. Detta fenomen, när vi tittar på något tillsammans och är medvetna om det, kallas för gemensamt uppmärksamhetsfokus, och är ett fundamentalt beteende i mänsklig interaktion som vi som barn lär oss tidigt, och som det är svårt att hitta motsvarighet till hos andra arter. Att vi kan läsa av varandras blickriktning och därigenom gemensamt kan fokusera på saker är en förutsättning för att lösa uppgifter tillsammans. Det är därför en förmåga som även robotar måste kunna behärska om de ska kunna kommunicera med oss. Det är även en förmåga som man har märkt att barn med autism har svårt med, och flera forskare har därför tittat på hur robotar kan användas för att träna autistiska barn i att läsa av ansiktsuttryck och ögonrörelser. Flera studier har visat att barnen visar intresse för robotens ansikte och att de efter en tid även att börja titta mer på människors ansikten.

Man kan förstås fråga sig varför man överhuvudtaget ska använda ett mänskligt ansikte på roboten? Kan den inte lika gärna ha lampor

som blinkar när den vill signalera något? Det vore förstås möjligt, men ett starkt argument för människoliknande robotar är att vi som människor genom evolutionen och under vår uppväxt blivit präglade på att tolka ansiktsuttryck. Hjärnbildningsstudier har identifierat särskilda regioner i hjärnan som är specialiserade på detta. Om vi återanvänder dessa redan väl utvecklade förmågor behöver vi inte "lära om" när vi kommunicerar med robotar. Man ska dock vara medveten om att det även finns en stor utmaning i att designa alltför människoliknande robotar. I Japan har man utvecklat extremt människoliknande robotar med silikonhud och mekanisk styrning, men som ofta har ett zombieliknande beteende och som därför uppfattas som obehagliga. Detta fenomen brukar kallas *uncanny valley*. Orsaken till fenomenet är troligen att det med ett väldigt människolikt utseende följer förväntningar på att beteendet också ska vara det. Men med den mekaniska styrningen är det nästan omöjligt att reproducera alla de mikrorörelser i ansiktet som vi är vana vid och som vi är extremt känsliga för, och därför uppstår en krock i vår perception. Då Furhat har ett animerat ansikte, som möjliggör en finare kontroll, och ett något mer avskalat utseende, är inte detta ett problem på samma sätt.

### *Språklig förmåga och artificiell intelligens*

Man kan fråga sig om en robot som ska kunna föra ett samtal på människans villkor, också måste besitta mänsklig intelligens. Denna fråga har varit föremål för debatt under lång tid. I den berömda artikeln "Computing Machinery and Intelligence" från 1950 lade matematikern Alan Turing fram ett tankeexperiment som sedermera har fått namnet Turing-testet. Han ställde sig frågan vad intelligens är och när man kan säga att en maskin besitter intelligens. Ska man kräva att maskinen har ett medvetande? Men hur skulle vi kunna fastställa detta? Att titta in i datorns "hjärna" och ställa krav på att den ska fungera på ett visst sätt verkar hopplöst. Är det verkligen säkert att den måste fungera precis som vår hjärna? Turings svar var att intelligens måste bedömas utifrån maskinens beteende: Om vi kan bygga en maskin som kan interagera med oss, och vi inte kan avgöra om det är en maskin



eller en människa vi interagerar med, så menade han att vi hade lyckats skapa artificiell intelligens. I det berömda tankeexperimentet, kallat *The Imitation Game*, såg han för sig att en person A får interagera med en annan person B via en chatt. Men ”person” B kan lika gärna vara en dator som en verklig människa. Efter en tids interaktion får person A försöka gissa om B är en dator eller människa. Om gissningarna inte är bättre än slumpen kan man säga att maskinen är intelligent.

Flera filosofer och forskare kritiserade detta resonemang. En av kritikerna var datavetaren Joseph Weizenbaum, som menade att ett till synes mänskligt beteende skulle kunna åstadkommas med relativt enkla metoder. För att bevisa detta skapade han på 1960-talet ett program som han döpte till Eliza. Namnet tog han från teaterpjäsen *Pygmalion* av George Bernard Shaw, där fonetikprofessorn Henry Higgins ingår ett vad om att han kan lära en flicka från underklassen, Eliza Doolittle, att prata och bete sig som en societetsdam. Higgins teori var att om han bara kunde få henne att prata med rätt sociolekt och anamma vissa manér, skulle omgivningen förledas att tro att hon faktiskt delade deras bakgrund. I pjäsen fick Higgins till slut rätt och vann vadet (om än inte Elizas kärlek). En viktig poäng som görs i pjäsen, och som ligger till grund för parallellen till Weizenbaums program, är att Eliza inte själv förstår vad hon egentligen pratar om, och att omgivningen läser in intentioner i vad hon säger som inte finns. På samma sätt som Eliza gavs rollen av societetsdam i *Pygmalion*, gav Weizenbaum sitt program rollen av en psykoterapeut som kunde diskutera personliga problem med användaren via en chatt. Tricket var att helt enkelt transformera det användaren sa till bekräftande påståenden som uttrycker förståelse, eller inkännande motfrågor som för samtalet framåt. Om användaren till exempel sade att hon hade problem med sin mamma kunde Eliza fånga upp ordet mamma och be användaren att berätta mer om sin mamma, utan att egentligen ha någon djupare förståelse av innebörden hos orden. Det finns flera anekdoter om människor som interagerat med Weizenbaums Eliza och faktiskt läst in mänskliga egenskaper hos programmet, bland annat Weizenbaums sekreterare som testade programmet och blev så emotionellt engagerad i konversationen att

hon bad Weizenbaum att lämna henne ensam med datorn. Till hans förvåning var det även flera psykiatriker som såg positivt och entusiastiskt på Eliza och tänkte sig att programmet faktiskt skulle kunna agera terapeut på riktigt.

I sitt försök att motbevisa Turings teorier om hur intelligens ska definieras hade Weizenbaum (utan att det varit hans ursprungliga avsikt) visat att kommunikation mellan människa och maskin kan upplevas som meningsfull, även om maskinen inte besitter en ”riktig” intelligens. Och det behöver inte nödvändigtvis handla om att lura människor att tro att datorn är en människa. Filosofen och poeten Samuel Taylor Coleridge myntade 1817 begreppet *suspension of disbelief* – ett förhållningssätt som ligger till grund för att vi kan uppskatta fiktion och gråta till sorgliga filmer, även om vi förstår att det egentligen inte är på riktigt. På samma sätt skulle vi kunna tala med en maskin som om den vore en människa, även om vi vet att så inte är fallet.

### *Hur lär man maskiner att samtala?*

Sedan 1960-talet har förstås utvecklingen gått starkt framåt på många fronter, och dagens datorer och algoritmer är långt mer avancerade än de Weizenbaum hade att tillgå. Förutom Elizas begränsade förståelse så var förstås en uppenbar begränsning att man endast kunde interagera med programmet via en skriven chatt. När vi kommunicerar via text utnyttjar vi bara en bråkdel av alla de sofistikerade signaler som står till buds i samtal ansikte mot ansikte, som vi diskuterat tidigare. Förutom avsaknaden av ansikte är en viktig komponent förstås det talade språket. I skriven text finns bara ord (och skiljetecken) att tillgå, men i talat språk har vi även prosodi, det vill säga hur vi talar, där talhastighet, intensitet och satsmelodi har stor betydelse. Många har erfarit vilka missförstånd som kan uppstå i mejl, där uppmaningen kanske lät onödigt hård, eller ironin gick mottagaren förbi. Delvis kan detta kompenseras, till exempel med emojis, men när nyanserna är viktiga föredrar vi oftast att tala med varandra.

Att bygga en maskin som förstår talat språk är dock långt mer komplicerat än att förstå skrivet språk, vilket är svårt nog. Det finns flera

anledningar till detta. För det första låter samma ord och språkljud väldigt olika beroende på individuell anatomi, dialekt, dagsform och talhastighet. För det andra kan helt olika ord låta ungefär likadant. Ett klassiskt exempel är meningarna "let's recognize speech" och "let's wreck a nice beach", som kan vara svåra att skilja åt om de uttalas snabbt. För oss människor är det oftast uppenbart vilken tolkning som är rimlig, givet sammanhanget, men sådana rimlighetsbedömningar är svårare för en dator. För det tredje talar vi vanligtvis samtidigt som vi tänker ut vad vi ska säga, vilket gör att språket i mindre utsträckning följer en given grammatik och blir svårare att tolka för datorn. Mer traditionella lingvistiska modeller av språket var ofta baserade på regelsystem som relativt väl kunde beskriva det skrivna språket, så som vi får lära oss att använda det. Men de var inte särskilt användbara för att beskriva hur språk faktiskt används, särskilt i talad kommunikation.

Därför utvecklades istället statistiska modeller som var avsevärt bättre för att beskriva och förutsäga språkligt beteende. Framförallt är det utvecklingen inom artificiella neurala nätverk som tagit ett stort språng de senaste åren. Dessa nätverk är inspirerade av hjärnans sätt att hantera information och kräver inte att man förklarar för dem hur informationen ska representeras och förstås. Istället får nätverken själva upptäcka mönster och forma mer abstrakta koncept utifrån stora datamängder (så kallad *deep learning*). Tack vare de enorma mängder data som till exempel finns på internet kan man träna dessa modeller så att de lär sig att känna igen språkljud, hur sannolikt det är att ord förekommer i olika sammanhang, och associera till vilka ord som kan ha en liknande betydelse.

Där Weizenbaum programmerade Elizas beteende med hjälp av regler, har man idag börjat undersöka hur man kan samla in stora mängder samtal och låta datorn själv lära sig mönster från dessa. En metod är att låta en dold operatör sköta robotens beteende i början, precis som schackroboten Mechanical Turk, för att därigenom samla data om hur en människa skulle ha gjort i robotens ställe. De inspelade samtalen används sedan för att träna roboten på hur den ska bete sig. Denna metod kallas för *Wizard-of-Oz*, uppkallad efter filmen från 1939

där den mystiske trollkarlen inte finns på riktigt, utan i själva verket styrs av en man som står bakom ett draperi. En annan metod är att låta roboten interagera med användaren och utveckla sitt beteende genom att lära sig av sina misstag.

I min egen forskning har jag använt liknande metoder för att lära roboten vems tur det är att prata. När vi människor pratar med varandra vill vi ju helst undvika att prata i munnen på varandra. Om vi gör en paus i mitten på en mening vill vi inte heller att den vi talar med ska bryta in och börja prata. För att koordinera turtagningen i samtal använder människor en mängd olika signaler. Till exempel kan en satsmelodi med avslutande platt ton, en inandning, eller ett ”ööh” signalera att vi har något mer vi vill säga och att motparten inte ska ta turen. Vi tenderar också att titta bort när vi vill fortsätta prata, och titta upp på den vi talar med när vi lämnar över turen. Med hjälp av stora mängder inspelade samtal mellan människor kunde vi träna ett neuralt nät på att förutsäga vem som kommer prata närmast i samtalet, baserat på vem som har pratat tidigare, och olika signaler i språket. Givet tillräckligt med data och träning kunde vår modell till slut göra bättre förutsägelser än vad människor gör när de ges samma uppgift. En sådan modell skulle kunna användas av roboten för att förutsäga hur sannolikt det är att människan kommer att fortsätta att prata och därigenom när det är lämpligt för roboten att ta turen.

### *Framtida utmaningar*

Tack vare den senare tidens utveckling finns det idag talande maskiner i form av röstassistenter i våra mobiltelefoner, som Apples Siri, och smarta högtalare, som Amazons Alexa, och för de begränsade uppgifter de är skapade för att hantera fungerar de ganska bra. Men även om vi idag har långt mer sofistikerade möjligheter än Weizenbaum hade för att utveckla artificiella samtalspartner så är det fortfarande långt kvar innan de kan uppvisa något som påminner om mänsklig intelligens. En uppenbar begränsning med dagens AI är att även om vi kan träna modellerna med hjälp av data, eller låta roboten lära av egen erfarenhet, har vi fortfarande inga metoder för att låta datorn generalisera sin kun-

skap. Där människan har förmågan att se analogier mellan uppgifter och lära sig från några få exempel, kräver dagens AI enorma mängder data eller interaktioner för att lära sig. Därför kan den excellera i väldigt specialiserade och avgränsade uppgifter, som att spela schack eller köra bil, men skulle ha svårt med mer generella och flexibla uppgifter, som att agera vaktmästare eller ha hand om barn på en förskola.

På samma sätt måste dagens konverserande system tränas att samtala kring specifika ämnen, och de har svårt att använda språket på ett kreativt sätt. När vi tolkar vad andra människor säger gör vi det inte bokstavligt, oftast försöker vi läsa in vad den egentliga intentionen är, utifrån vad som är rimligt i situationen. Ett uppenbart exempel är om någon frågar: "Vet du vad klockan är?" Tolkar vi det bokstavligt skulle kanske svaret bara bli ett "ja". Men vi förstår att ingen skulle vara intresserad av det svaret, och därför svarar vi istället vad vi tror att den andre personen egentligen vill veta. Just detta exempel följer ett mönster som blivit så standardiserat i vårt språk att vi kanske inte ens behöver gå via denna slutledning, och en dator kan tränas på att tolka frågan på samma sätt som "Vad är klockan?". Men språket är fyllt av dessa typer av indirekta och kreativa tolkningar, till exempel för att förstå ironi. Det gör även att vi kan effektivisera vår kommunikation genom att prata snabbare och mer "slarvigt". Även om mottagaren kanske inte förstår alla ord som sagts kan hon dra slutsatser om vad avsändaren kan ha menat, baserat på vilka intentioner som är rimliga i sammanhanget. Ibland kan det räcka med att bara peka och himla med ögonen. Vi anpassar också hela tiden vårt sätt att prata på efter omgivningen och dem vi pratar med. Ju bättre vi lär känna varandra, desto mer utvecklar vi ett språk som är specifikt för vår grupp. Att förstå sådan form av effektiv och kreativ kommunikation är dock fortfarande mycket svårt för en dator.

Även om vi kan föreställa oss fantastiska möjligheter med tekniken måste vi förstås också fråga oss om all utveckling inom artificiell intelligens är önskvärd. Till exempel har vissa forskare, som Stephen Hawkins, uttryckt en oro för att datorerna till slut kommer att bli så intelligenta att de kommer att ta över och kanske utplåna människan.

Jag är personligen inte lika orolig för detta, då det ännu är oerhört svårt att se hur datorerna ska kunna få det självmedvetande och den motivation som skulle krävas, givet de begränsningar som jag skisserat ovan. Däremot kan förstås tekniken hamna i fel händer och dess oerhörda potential användas i skadliga syften.

Det är också viktigt att beakta faran med att låta allt fler beslut överlåtas till maskiner. Även om vi bara använder artificiell intelligens som ett beslutsstöd, till exempel för medicinsk diagnos eller för att göra bedömningar om klimatförändringar, finns det en risk att vi inte kommer att förstå varför den föreslår vissa beslut. Vi skulle då kunna hamna i en liknande situation som med oraklet i Delfi: vi går till maskinen för att få svar på våra frågor, men har inget annat val än att blint lyda till synes märkliga beslut. Därför satsar man nu på forskning inom vad man kallar *Explainable AI* – där datorn inte bara ska fatta rätt beslut, utan också kunna förklara och motivera (på ett för oss begripligt språk) varför den fattar de beslut den gör.

En annan farhåga är att robotarna kommer att ta över våra arbeten. Att våra arbeten automatiseras är förstås inget nytt. Från att 70 procent av sysselsättningen var inom jordbruk 1870, är det idag endast drygt en procent som är sysselsatta inom jordbruket. Trots detta har vi fortfarande ingen större brist på arbetstillfällen; istället arbetar människor med andra saker. Det här är förstås ingen garanti för att det alltid kommer att vara så. Men man bör beakta att det idag finns en oerhört stor potentiell efterfrågan på arbete, även av social natur, som idag inte blir tillfredsställt, eftersom det är för dyrt. Förmodligen är det till exempel många ensamma äldre människor som skulle vilja ha tillgång till en personlig assistent dygnet runt, men som inte kan få det idag. Bara för att tekniken efterfrågas är det inte säkert att den tar arbetstillfällen, den kan också ge oss möjligheter som tidigare inte fanns. När till exempel tv:n kom fyllde den uppenbarligen ett behov, men den ersatte knappast tidigare arbeten i någon större utsträckning.

Man bör förstås även ställa sig frågan vad tekniken gör med oss människor och våra relationer till varandra. I USA har vissa oroats över att barn som pratar med röststyrda tjänster inte lär sig att säga ”please”



när de till exempel ber assistenten att sätta på musik, helt enkelt därför att datorn inte bryr sig om sådana artighetsfraser. Om vi börjar omge oss med robotar som blir allt mer människolika och vi upplever att vi kan behandla dem hur vi vill (eftersom de är robotar), kan det finnas en risk att vi överför detta beteende gentemot andra människor. Frågan är då om även robotar ska kräva att man behandlar dem med respekt? Som svar på denna oro har flera röstassistenter programmerats att uttrycka tacksamhet om användaren lägger till ett "please" eller "thank you" till sina kommandon.

Att samtala med maskiner, inte bara för att de kan svara på faktafrågor och utföra uppgifter, utan för att de fyller ett socialt behov, framstår för vissa som ett främmande (och kanske skrämmande) framtidsscenario. Men redan idag tycker vi inte att det är så märkligt att barn låtsaspratar med sina leksaker, eller att vuxna människor pratar till sina husdjur (trots att djuren förmodligen har väldigt begränsad förståelse av vad som sägs). Det är inte orimligt att vi i framtiden kommer tycka att det är normalt att även prata med våra maskiner, så länge vi upplever att de fyller en mening i våra liv. **A**

GABRIEL SKANTZE är professor i talteknologi vid Kungliga Tekniska högskolan. Ledamot i Sveriges unga akademi 2015–2020.

