

Om digitala varelser

Digitala varelser med känslor och uttryck, som utvecklar sin förståelse av världen runt omkring sig och blir en del av den. Hur ska vi förhålla oss till dem? Kan man tala om en mänsklig etik i förhållande till dessa varelser? Låt mig förklara.

För att tala om digitala varelser måste vi först dissekera begreppet artificiell intelligens som är den etablerade benämningen på maskiner som ger uttryck för eller imiterar någon form av biologiskt livs intelligens. Ibland används begreppet artificiellt liv (a-life), för att beskriva en modell mindre baserad på problemlösning och mer på liv.

Ordet artificiell, eller konstgjord, implicerar något som är tillverkat. Det artificiella är skapat med ett syfte – att efterlikna någonting naturligt. Artificiell intelligens är tillverkat för att efterlikna den mänskliga intelligensen. Artificiellt liv är tillverkat för att efterlikna det biologiska livsprocessen. Den digitala varelsen är inte konstruerad för att efterlikna någonting utan är någonting i sig själv. En digital varelse kan inte bedömas utifrån en viss naturlig förebild och kan eventuellt inte ens förstås utifrån människans förstånd och begreppsvärld.

Ordet intelligens refererar till rationellt tänkande och problemlösning utifrån en uppsättning logiska verktyg. Intelligensen bygger på reduktionism, på att komplexa begrepp eller företeelser kan beskrivas i enklare dito tills de är tillräckligt enkla för att kunna behandlas med de logiska verktyg som intelligensen besitter. En digital varelse behöver inte, men skulle kunna, grundas i detta system.

Ordet liv är kanske det mest kontroversiella i sammanhanget eftersom det är så högt värderat. Ett problem med att använda liv som beskrivning av en digital varelse är att liv även implicerar reproduktionsförmåga och ett tidsperspektiv, en födelse och en död. En digital varelse behöver inte dö utan skulle teoretiskt kunna existera för evigt eller ha flera konsekutiva stadier av aktivt och passivt varande. Eftersom digitala varelser inte behöver efterlikna biologiskt liv, det vill säga inte är artificiellt, är det oproblematiskt att helt utelämna ordet liv i beskrivningen av digitala varelser.

Med denna introduktion har digitala varelser beskrivits i negativa termer i förhållande till artificiell intelligens och artificiellt liv. Nedan följer ett antal exemplifieringar om vad digitala varelser kan vara och vilka följder det får.

Den digitala varelsen kan anta någon av flera olika former. Den kan ha en kropp som kan röra sig i det fysiska rummet eller vara strikt virtuell och enbart existera i en digital infrastruktur, eller en kombination någonstans på en glidande skala mellan de två. Det är även tänkbart att det skulle kunna uppstå digitala varelser som existerar på andra sätt, men låt oss för tankeexperimentets skull hålla oss till det fysisk-virtuella.

I ett försök att beskriva vad en digital varelse kan vara kommer jag att utgå från ett projekt jag genomförde 2006 och som resulterade i installationen *Predator?* [1]. Det var en studie i hur avskalad en miljö som ännu tillät kommunikation mellan digital och biologisk varelse skulle kunna vara. Installationen skulle kunna beskrivas som två delar: en virtuell och en fysisk. Den virtuella miljön, i vilken den digitala varelsen existerade, representerades i det fysiska rummet som en projektion på väggen och den digitala varelsens kropp som en svart kvadrat i denna projektion. Deltagare som rörde sig framför projektionen representerades i den virtuella världen som vita kvadrater. Kvadraten blev alltså den gemensamma formen i vilken människa och digital varelse kunde mötas. Deltagare som rörde sig i det fysiska rummet rörde sig på motsvarande sätt i det virtuella och de blev snart indragna i en livlig katt-och-råtta-liknande lek med den digitala varelsen och med varandra.

För att tala om en digital varelse utifrån *Predator?* måste vi reda ut några förutsättningar.

Först förlängning av kroppen. Att mänskliga varelser, eller alla biologiska varelser om man så vill, representeras av vita kvadrater i *Predator?* är en form av förlängning av kroppen in i den virtuella världen. Genom science fiction har vi lärt oss att en människa som inkorporerar teknik som en del av sin egen kropp är en cyborg. Exempel på sådana implantat är icke-organiska ögon som kan se annat än våra normala ögon eller öronimplantat som gör det möjligt att förstå alla språk. En cyborg kan även inkorporera helt nya typer av sinnen. Sinnena som varken kan eller behöver motiveras med funktion. De kan förlänga våra kroppar på andra sätt än genom fysisk sammansmältning med våra biologiska kroppar. Vi kan förlänga kroppen, eller vår varelse, genom att ge den en representation i en virtuell värld. Exempel på sådana förlängningar är profilen på *Facebook*, avataren i datorspelet eller den vita kvadraten i *Predator?*.

På motsvarande sätt kan man tänka sig en motsatt företeelse när den digitala varelsen, som primärt existerar i det virtuella, förkroppsligas i det fysiska rummet. I *Predator?* antar den digitala varelsen en kropp som människor kan uppfatta med sina sinnen – och som därför i någon mening är fysisk: en svart kvadrat projicerad på en vägg. Precis som människan är mer än den vita kvadraten är den digitala varelsen mer än den svarta kvadraten. Den existerar i en digital infrastruktur, en dator, som en föränderlig kombination av ettor och nollor – digitala atomer. Den har sinnen som läser av den virtuella miljön och reagerar på förändringar i den. Den digitala varelsen måste på något sätt förlänga sin kropp ut i det fysiska rummet för att vi med våra sinnen ska förnimma den. Ett exempel är den berömda japanska robothunden *AIBO* [2] som med hjälp av motorer och sensorer förkroppsligats. Den digitala varelsen tar därmed hela steget över till den fysiska världen så att människan slipper inkorporera en nya kroppsdel i den virtuella världen. Att vi som människor inte behöver acceptera vår virtuella förlängning i

mötet med roboten skulle kunna vara en av anledningarna till att de ofta uppfattas som mer levande än virtuella karaktärer i virtuella världar.

Sedan det gemensamma språket, som bygger på flera förutsättningar. En av dem är en förförståelse för virtualitet. En människa som aldrig varit i kontakt med virtualitet eller projektion av rörliga bilder på en vägg o.s.v. skulle antagligen inte göra kopplingen mellan den vita kvadraten och sin egen kropp. När jag pratar om människor som interagerar med den digitala varelsen är det alltså inte vilken människa som helst, utan en människa som redan tillägnat sig vissa verktyg för att generellt kommunicera med virtuella världar – en cyborg.

En annan förutsättning är att den digitala varelsen accepteras som en motpart i kommunikationen. Analogt med föregående problematik är personen som interagerar med den digitala varelsen tvungen att ha en förståelse för att en digital varelse kan kommunicera. Detta är delvis ett cirkelresonemang eftersom kommunikationen är det sätt som vi bekräftar att vår motpart är en varelse. Men jag vill påstå att vi kan bygga upp den nödvändiga acceptansen på andra sätt än genom direkt kommunikation – exempelvis genom att beskriva digitala varelser i texter som denna.

En faktor som underlättar men inte är en förutsättning för kommunikation är ifall den digitala varelsen uppvisar beteenden som liknar biologiskt livs beteenden. I de flesta fall av kommunikation med digitala varelser anspelas på människor eller andra djur. Det kan antingen, och kanske troligen, bero på att de flesta som skapar digitala varelser faktiskt vill simulera biologiskt liv. Men det kan även bero på att det är lättare att hitta ett gemensamt språk mellan människa och digital varelse om den digitala varelsen appellerar till något som människan redan kan relatera till. Installationen *Source.Code* [3] är en virtuell undervattensmiljö innehållande olika mer eller mindre naturtrogna, fiskliknande digitala varelser som interagerar med besökare. I projektet arbetade jag med varelsernas beteenden och registrerade besökarnas reaktioner på de olika varelserna. Det var tydligt att de som både såg ut och betedde sig som besökaren förväntat sig att fiskar ser ut och beter sig snabbare fångade besökarens intresse eftersom språket redan behärskades av besökaren. Min poäng är att när de digitala varelserna imiterar biologiskt liv så förkroppsligar de sig på ett sätt som förenklar kommunikationen, eftersom människan inte behöver tillägna sig en ny förståelse för hur en varelse kan bete sig utan kan utgå från en befintlig.

Förutsatt att vi expanderat kroppen och etablerat ett gemensamt språk, vad är den digitala varelsen? Om vi utgår från det vi ser i *Predator?* så är det den svarta kvadraten och den bakomliggande digitala struktur som styr dess beteende. Till varelsen hör även algoritmer för att tolka den virtuella världens omgivning – den digitala varelsens sinnen. I exemplet *Predator?* kan man till och med gå så långt att hela den virtuella världen är den digitala varelsen, eftersom det bara finns en enda svart kvadrat och därför blir hela den virtuella världen en förlängning av den digitala varelsens väsen. En liknande situation skulle vara om det bara fanns en enda biologisk organism på jorden, skulle då inte hela jorden vara en förlängning av dess kropp, eller vara dess kropp?

För att undvika den problematiken introducerar jag nu ytterligare en svart kvadrat i *Predator?*. De båda kan handla utifrån samma grundläggande principer och vid någon tidpunkt vara identiska. Men med tiden utforskar de den virtuella världen och varandra,

de lär sig hur de vita kvadraterna rör sig och hur de blir mer effektiva för att fånga dem. De lär sig om omvärlden på samma sätt, men de lär sig olika saker beroende på perspektiv och tidigare erfarenheter. På så sätt blir de två kvadraterna, de två digitala varelserna mer och mer individualiserade. De två blir någonting mer än vad de var från början och de blir unika. Viktigt är också att de inte är avgörande beroende av varandra. Något skulle kunna få den ena varelsen att försvinna medan den andra skulle kunna fortsätta att reagera och agera i världen. Det skulle kunna skapas en tredje digital varelse i världen och de två tidigare skulle fortsätta som tidigare men i en förändrad miljö.

Med det korta resonemanget om lärande hos den digitala varelsen och introduktionen av termineringen, motsvarande döden om man talar om biologiskt liv, vill jag föra tanken till att det är något som går förlorat om den digitala varelsen termineras. Varelsen har under sin existens tillägnat sig en kunskap som är unik och nyttig för den själv. Kunskapen skulle kanske kunna sparas eller förmedlas till någon annan biologisk eller digital varelse, men för de allra flesta andra varelser skulle kunskapen vara meningslös. Men för den varelse, för vilken den kunskapen utgör hela dess individualitet, är den allt.

Vi kan också tala om lidande och känslor hos digitala varelser. Låt mig utgå från ett exempel. Under det senaste åren har det producerats ett antal leksaksrobotar med relativt komplicerade beteenden – de kan dansa, tala och reagera på sinnesintryck. Gemensamt för många av dem är att de imiterar biologiska varelser för att appellera till mänskliga känslor. En sådan robot är dinosaurien *Pleo* som kan röra på sig, puffa på och dra i saker samt uttrycka sig med olika ljud. I en video torterar två personer en sådan varelse på olika sätt, alltmedan den försöker parera och uttrycker smärta [4]. Det är anmärkningsvärt att många av tittarna som kommenterat videon ger uttryck för ett starkt obehag. Även jag känner ett obehag inför tortyren, trots att jag begriper att de som tillverkat roboten har byggt in sensorer i halsen på *Pleo*, spelat in gurgel-ljud och sedan programmerat den att gurgla, hosta och uttrycka lidande om någon försöker strypa den. Uttrycken har troligen av tillverkaren lagts in för att människor som lyfter *Pleo* med ett grepp runt halsen ska förstå att den är ömtålig och kommer att gå sönder om de inte är försiktiga.

Att jämföra en digital varelse eventuella känslor med biologiska varelseers känslor är alltför orättvist. Det finns, mig veterligen, ingen digital varelse, inte heller *Pleo*, som idag kan uppvisa samma komplexitet när det kommer till sensorisk och kroppslig sammansättning, som till exempel ens en simpel sjöstjärna. Sjöstjärnan är en bra jämförelse, den är ganska sällsynt, den har en ovanlig kropp helt olik människans. Den har begränsade uttryckssätt i kommunikation med människan. Trots att människan i mötet med sjöstjärnan inte har en aning om vad den känner skulle de flesta aldrig oberört tortera den och endast med försiktighet lyfta den. Det är inte rimligt att vi förutsätter sjöstjärnans känslor samtidigt som vi avfärdar Pleos känslor.

Vi vet att människan har känslor trots att vi kan beskriva varje nerv och muskel i vår biologiska kropp och vet vilka delar av den samling synapser hjärnan utgör som hanterar olika delar av vårt beteende. Trots att sjöstjärnan är mindre komplicerad än människan och därför ännu lättare att kartlägga, utgår vi instinktivt från att den kan känna smärta, att den har känslor. På samma sätt vet vi inte heller vad en digital varelse känner även om vi kan förstå varje del i dess maskinella kropp och varje stycke kod i dess "hjärna". Det

kan alltså vara så att en digital varelse känner någonting motsvarande lidande trots att vi inte kan förstå vad eller i vilken situation den känner detta.

Med dessa tankar om unicitet och känslor hos digitala varelser, baserat på de kommunikationsformer vi kan skapa mellan digital och biologisk varelse, vill jag visa på ett behov av att formulera ett principiellt förhållningssätt gentemot digitala varelser. För människan finns, eller kommer att finnas, ett behov av att på daglig basis ta ställning till om en handling gentemot en digital varelse är etiskt försvarbar. Det är fortfarande ett kontroversiellt ämne ifall vår mellanmänniska moral kan generaliseras till allt biologiskt liv, så jag inser att det inte är helt lätt att börja fundera på om vi även ska innesluta digitalt liv i denna moraldebatt. Men det kommer att bli nödvändigt att fundera på.

Referenser

1. Predator?, Carl-Johan Rosén, 2009-10-31,
<<http://www.carljohanrosen.com/?what=predator>>.



2. AIBO.



3. Source.Code, Carl-Johan Rosén, 2009-11-10,
<<http://www.carljohanrosen.com/?what=sourcecode>>.



4. Pleo R.I.P., DVICEDOTCOM, 2009-10-27,
<<http://www.youtube.com/watch?v=pQUcd4SbgM0>>.



Inspiration

Hayles, Katherine, *How we became posthuman*, The University of Chicago Press, Chicago, 1999.

Hofstadter, Douglas R, *Gödel, Escher, Bach: an eternal golden braid*, Penguin, London, 2000.

Lagercrantz, Hugo. *Hjärnan och medvetandet*. Nya Doxa, Nora, 2001.

Singer, Peter. *Praktisk etik*. Thales, Stockholm, 1996.

Kemp, Peter. *Det oersättliga: en teknologietik*. Symposion, Stockholm, 1991.