

# **TEKNISK PROGNOS**

RAPPORT NR 1 • 2025

## **TEMA: AI**

**STARTUPFÖRETAG BYGGER  
SVÅRSTÖRDA HAVSDRÖNARE**

**EN SPEGEL MOT OSS SJÄLVA  
PRÄGLAR VÅR SYN PÅ AI**

**BILLIG TEKNIK ÄR VINNARE  
I DET NYA SÄTTET ATT KRIGA**

**MED AI SOM VAPEN KAN  
EN DEMOKRATI STÖRTAS**

**FMV**



# TEKNISK PROGNOS

RAPPORT NR 1 • 2025

## INNEHÅLL

- 4 Krönika: Från ChatGPT till Kill Zones**  
Vår säkerhet ändras i grunden.
- 6 Kampen om autonom dominans**  
Ukrainska startups driver AI-utveckling.
- 12 Billig teknik kan slå ut dyra system**  
AI förändrar krigföringen.
- 18 Beslut fattas i ny miljö**  
AI får allt större roll i beslutsfattandet.
- 22 Folkrådet utmanar AI**  
Så får autonoma vapensystem användas.
- 25 Strider för säkrare AI-system**  
Självkörande bilers lärdomar om autonomi.
- 28 Kulturkrock när AI testas på världen**  
Vikten av ansvarsfull innovation.
- 32 Framtidsspaning**  
Sex experter om AI:s framtid.
- 34 Autonom navigering i tystnad**  
Startupen gör ytfarkoster som inte störs ut.
- 40 Konsten störta en demokrati utan vapen**  
AI tar informationskrigföringen till en ny nivå.
- 44 Algoritmen i blodet**  
Johan Schubert leder AI-forskning på FOI.
- 48 Den svenska språkmodellens uppgång och fall**  
Vikten av att Sverige har egna AI-resurser.
- 52 Agent med rätt att styra AI**  
Fredrik Bruhn sätter AI först.
- 56 En spegel mot oss själva**  
Synen på oss själva präglar vår syn på AI.
- 60 "Katastrofen kan komma inom tio år"**  
Matematikprofessorn tror att AI tar över.
- 64 "En bubbla inte hela världen"**  
Värden och utveckling kan fortsätta.
- 66 Cirkus utan skyddsnät**  
AI-bubblan på väg att spricka.



6



22



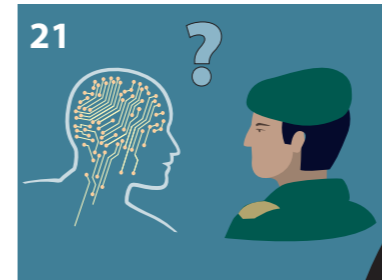
34



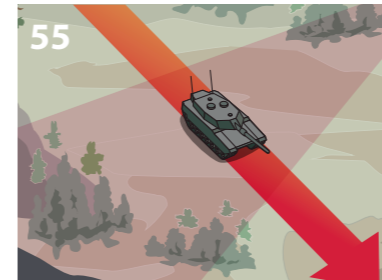
48



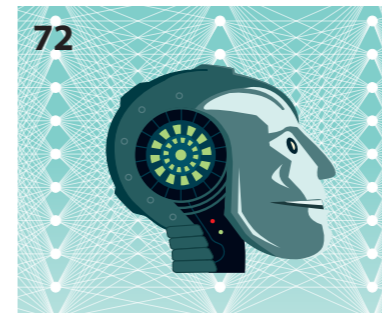
52



21



55



72

Publikationen *Omvärld* är sammanställd av FoT-projektet *Omvärldsbevakning med teknisk prognos*.

**TEXT:** Jan-Ivar Askelin och Kajsa Skarsgård om inte annat anges

**FOTO:** Jan-Ivar Askelin och Tomas Årlemo om inte annat anges

**GRAFIK:** Martin Ek

**GRAFISK FORM:** Peter Ehrlin

**FAKTAGRANSKARE:**

Johan Schubert, FOI

Hannes Ovrén, FOI

Magnus Sahlgren, AI-Sweden

Peter Bovet-Emanuel, Försvarshögskolan

**TRYCK:** ÅTTA45

**OMSLAGSFOTO:** Getty Images, Andriy Onufriyenko

**FMV dokumentbeteckning:** 24FMV2096-35

## GRAFIK

10 AI-begrepp och historik

16 Artificiell intelligens på det moderna slagfältet

21 AI och OODA-loopen

30 AI – värdekedjan

39 Polar Mist – GNSS-fri navigering

47 Aktiv inferens

51 Att angripa en språkmodell

55 Neuromorf databehandling

63 RAG:s, workflows och AI-agenter

68 Enklare AI och maskininläring

70 Neurala nätverk

72 Generativ AI

# FRÅN CHATGPT TILL KILL ZONES

**A**I-tekniken har utvecklats i en rasande takt det senaste decenniet. Startskottet för den moderna AI-eran kom runt 2012 när neurala nätverk som Alex Net plötsligt gav datorer förmågan att "se" – att identifiera objekt i bilder. Steget från att skilja katter från hundar till att identifiera stridsfordon var kort, och innebar starten för automatiserad bildanalys.

När AI senare 2016 besegrade världsmästaren i brädspelen Go, och profiler som Nick Bostrom och Max Tegmark släppte böcker om "superintelligens", tog hypen fart på allvar. Boston Dynamics virala klipp på robothundar och Elon Musks varningar skapade en blandning av fascination och domedagsprofetior.

Länge agerade försvarsindustrin observatör. Men så kom kriget i Ukraina och lanseringen av ChatGPT – två händelser som slog ned som bomber, på helt olika sätt. Det blev ett bryskt uppvaknande. Hade vi tappat fokus på bollen? Under radarn i samband med Gaza-konflikten hade Israel redan integrerat avancerade AI-system. Med hjälp av inkallade reservister från tekniksektorn skapades system som Lavender och Where's Daddy. Genom att analysera enorma mängder data och tolka arabiska dialekter i realtid gav AI-stödda underrättelser IDF ett massivt informationsövertag.

I Ukraina har vi sett en liknande utveckling. När drönare vässas med AI och bildigenkänning förvandlas fronten till en "kill zone" på flera mils djup. Ingen kan röra sig obemärkt. Stridsvagnen – en gång slagfältets konung – framstår plötsligt som "dödens lammunge". Det står nu klart att snabbheten i autonoma system och förmågan att hantera enor-

ma datamängder är livsviktig i modern krigföring.

Mycket fokus har legat på generativ AI och språkmodeller (LLM), men i skuggan av debatten om AGI:ns (Artificiell Generell Intelligens) vara eller inte vara pågår en annan revolution. Dagens modeller är energikrävande och statiska, vilket är problematiskt i fält. Vi måste därför blicka bortom dagens teknik.

Vi ser nu framväxten av "World Models" och "fysisk AI" – målet är system som inte bara förstår text utan kan förstå sin fysiska omgivning och uppdatera sig själva. Vi ser också skiftet från Transformer-arkitekturen till nya arkitekturer som Mamba, designade för att hantera ännu längre sekvenser av data effektivt. Samtidigt studeras neuromorfisk databehandling och projekt som kinesiska DeepSeek och SpikingBrain visar på möjligheten att sänka energiförbrukningen radikalt.

Vart är vi på väg? Ser vi konturerna av en ny världsordning? Vad händer när människan kliver allt längre bak i beslutskedjan och tempot skruvas upp? Trenden är tydlig: ökad autonomi, mer mjukvara och ett uppskruvat tempo där människan flyttas allt längre bak i beslutskedjan, "on the loop" snarare än "in the loop". I det här numret försöker vi väva en förklarande väv om vart AI-tekniken är på väg och hur den med all sannolikhet kommer att förändra slagfältet – och Sveriges säkerhet – i grunden. □

**»Trenden är tydlig: ökad autonomi, mer mjukvara och ett uppskruvat tempo där människan flyttas allt längre bak i beslutskedjan, "on the loop" snarare än "in the loop".«**



Foto: Tomas Årlemo

**Mikael Schönström, fil.dr, FMV**  
mikael.schonstrom@fmv.se  
Projektledare, Omvärldsbevakning med teknisk prognos.



Foto: Alexander Lipko

## KAMPEN OM AUTONOM DOMINANS

**Tre ukrainska startups, del av landets AI-autonoma utveckling, har på kort tid överträffat etablerade försvarsföretag. Farsight Vision tog igen ett 20 år långt gap på två år. Norda Dynamics terminal guidance-system har sålt över 50 000 licenser och förvandlar 500-dollarsdrönare till precisionsvapen. Twist Robotics möjliggör tusentals FPV-anfall med relästationer. Gemensam nämnare? De bygger i krigets kontext, inte utanför. Och de drivs alla av samma sak – överlevnad och att vinna kriget.**

**Text: Daniel Kindstrand**  
**Gästskribent på plats i Ukraina**

**L**iv, september 2025. Klockan är strax efter sju på morgonen. Viktoriia Yaremchuk, vd för Farsight Vision, skriver i den krypterade kommunikationsappen Signal att hon blir sen till mötet på grund av nattens attack.

För mig var beslutet att anpassa både sömn och planer lika naturligt. Särskilt med tanke på att jag dragit slumpens kortaste strå och fått ett av hotellets få rum med takfönster.

Glasytan ovanför sängen ramar in en bit av natthimmel, vacker när den är stilla, dödlig när vildvittrorna kommer. Elektriska markiser känns bra men logiken säger det motsatta.

Halvvägs till vargtimmen gick larmet och med yrvaken logik satte jag fart satte fart mot hotellets shelter. Tacksam för preppad grab-bag men mindre tacksam för ett fumligt klädval. Skjortknappar och skärp är rimligen inte en prioritet men trötta hjärnceller faller lätt för muskelminne.

Väl i shelter har Lvivs Best Western en bar som aldrig stänger. Jag beställer en kudde men den förblir oanvänd.

Min shelter-granne visar sig vara senior advisor för European Investment Bank som leder till ett samtal om återuppbyggnad medan en annan hotellgäst försöker bygga en sovplats av tre stolar.

Några timmar senare tänder personalen lamporna för att börja duka. Kort därefter meddelade den statliga appen att luftangreppet är över och shelter har blivit matsal igen.

Jag räknar efter – dryga tre timmars sömn. Sedan bär det av mot demomöte och intervju. På Panorama Hotel gör sol och blå himmel sig till känna genom takbarens stora fönster.

På plats är redan Volodymyr Kuzminskyi, PhD, från The Unmanned Systems Forces Command för akustikbolagets produkt demo. Den nordiska grundaren har, till skillnad från oss, missat flyglarmet och larm-appen och sovit hela natten och håller en engagerad presentation.

Produkten, en sensor, stort som ett

våffeljärn, placeras vid fönstret. Trots glasrutorna visar skärmen obehindrat det klassiska heat pattern-mönstret som i realtid följer ljudet från bilar som färdas längs med Lviv National Opera som är byggnaden mitt emot.

Viktoriia Yaremchuk ansluter till demons slutskede. Volodymyr Kuzminskyi nickar imponerat åt den nordiska grundaren. Han ställer sig upp och uppmanar oss att göra detsamma.

Klockan har slagit 09:00.

Genom en knackig VMA-högtalare hörs ett skrapigt ljud över hustaken. Ett gammalt imperialistiskt ljud i patriotisk klang. Inte mellanösterns sjungande böneutrop. Detta är något annat. Mer envist, dystert, med tyngd. Något som kallar, samlar och enar.

Solljuset sätter operahusets allegoriska figurer och bevingade häststatyer i skarp relief. Monumentalt, militärt, auktoritärt. Gatan, staden och landet stannar till för ett kollektivt andetag.

Ett Hunger Games med hundratals offer.

Klockan 09:01 återgår livet till det vanliga. Volodymyr Kuzminskyi sätter sig. Samtalet om försvarsinnovation fortsätter, och likaså staden, med en opåverkad intensitet.

### 3D-kartor i realtid

Dagen innan, på en annan arena arrangerade Brave1, ett försvarsevent. Som del av öppningsceremonin lyfte tolv stora drönare synkroniserat och vecklade ut varsin flagga.

Synkroniseringen är imponerande och en påminnelse om att bara

tillståndsprocessen i Sverige skulle komma oroande nära året.

– Ta risk. Sitt inte och vänta på beställningar. Be bold. Det här är tiden på agera och vara handlingskraftig, säger Pål Jonsson, intervjuad av Max Villman i podcasten Max Tämt.

Handlingskraft har Ukraina bevisat flera gånger om. Men det är inte slut än. Kriget om autonom dominans har börjat.

– Hastigheten för insamling och bearbetning av underrättelser har gått från dagar till timmar, nu minuter, säger Viktoriia Yaremchuk.

Trots karriär inom forskning och mjukvara blev det efter invasionen 2022 meningslöst fortsätta att arbeta med något som inte stöttade frontlinjen. Tillsammans med sin CTO började hon prototypa hösten 2023. Redan i februari 2024 hade de levererat sina första till kontaktlinjen.

Idag levererar Farsight Vision 3D-kartor över frontlinjen i realtid genom att bearbeta drönarfoto med deras mjukvara och specialbyggd hårdvara. Traditionella metoder kräver ▸

*»Hastigheten för insamling och bearbetning av underrättelser har gått från dagar till timmar, nu minuter.«*

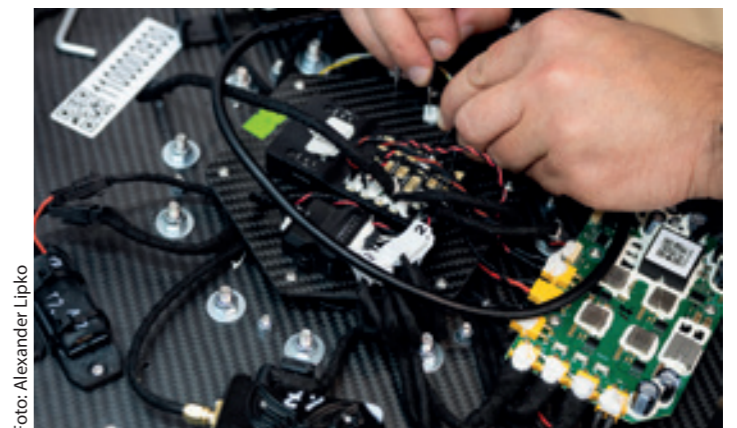


Foto: Alexander Lipko



Nazar Bigun, CEO och medgrundare vid Norda Dynamics.

manuell bildbehandling och kvalificerad GIS-expertis. Med Farsight Vision system blir det möjligt att generera kartdata som stödjer beslut på taktisk nivå.

– För de flesta försvarsstyrkor tog det upp till fem dagar att bygga en digital tvilling. Med vår hårdvara och mjukvara förvandlades det till timmar, nu minuter, säger Viktoriia Yaremchuk.

Kärnvärdet är att snabb hantering av bilddata och bildbehandling stundvis är den avgörande skillnaden mellan att förlora eller behålla en militär position.

– 2023 låg vi decennier efter Ryssland när det gäller att bygga digitala tvillingar, säger Viktoriia Yaremchuk och fortsätter:

– Till 2025 hade vi kommit ikapp och skapat betydligt mer avancerade versioner.

Endast två år för att ta igen tjugo års gap. Enligt Viktoriia Yaremchuk handlar framgången om kontext.

– Om du är utanför kontexten är det mycket mer komplext, säger Viktoriia Yaremchuk.

*»Det är inte vettigt att sätta ett 10 000-dollarssystem i en 500-dollar-drönare, det är därför vi började utveckla vårt eget.«*

#### 500 dollar mot 10 000

Nazar Bigun från Norda Dynamics hade vid invasionen 2022 redan byggt ett nätverk inom det ukrainska försvaret och identifierat flera kritiska utmaningar.

– Endast en av tio-tolv drönare nådde målet, främst på grund av elektronisk krigföring, radiohorisont och oerfarna piloter, säger Nazar Bigun.

Norda Dynamics första lösning blev Underdog – ett terminal-guidance-system som använder visuell sensor och IMU-data från flight controller för att guida drönare till mål även när radiokommunikation bryts.

– Det är inte vettigt att sätta ett

10 000-dollarssystem i en 500-dollar-drönare, det är därför vi började utveckla vårt eget, säger Nazar Bigun.

Idag har Norda Dynamics sålt omkring 50 000 licenser av sina tre produkter: Underdog för terminalstyrning, Stable Link för optisk stabilisering och Antenna Tracker för riktningföljning.

På frågan hur omvälvande AI och autonomi kommer att bli på en skala från ett till tio, dröjer inte många sekunder.

– Tio. Det här är en game changer. Vi skapar det nya kärnvapnet, men vi har inget val, säger Nazar Bigun.

Foto: Alexander Lipko

– Det är bättre att skapa tiotusen sjödrönare än en ubåt för tolv-tretton miljarder kronor som kan förstöras på en minut.

#### Miljard-capex döden

Rostyslav Olenchyn från Twist Robotics tog en annan väg in i försvarsteknologin. Hans första produkt var Firefly, en app för artillerikorrektion som på 45 dagar gick från idé till 5 000 nedladdningar. Idag har Twist Robotics levererat omkring 700 Saker Scout-plattformar, där den mest använda konfigurationen kommer med signalrelästationer som möjliggör tusentals FPV-anfall där kommunikationen annars skulle brytas. FPV-drönare uppges stå för runt 70 procent av alla attacker vid fronten – en numera allmänt känd utveckling som på två år gjort äldre system föråldrade.

– Stora, komplexa och dyra plattformar kommer att försvinna, som dinosaurier, säger Rostyslav Olenchyn.

Han har samtidigt lärt sig att det inte bara handlar om teknik. Det mest förvånande med AI i strid är den mänskliga faktorn.

– Inlärningskurvan gäller fortfarande. Det spelar ingen roll hur avancerat systemet är, det spelar fortfarande roll hur man lär folk. Driftsättning är den svåraste delen, säger Rostyslav Olenchyn.

Han riktar samtidigt kritik mot Europa och lyfter fram attityder som framställer ukrainsk innovation som en resurs att utvinna.

– Det är en väldigt svag utgångspunkt, säger Rostyslav Olenchyn.

#### Rymdskepp och fredstidskalkyler – som ingen behöver

Viktoriia Yaremchuk formulerar Europa-problemet skarpt:

– Jag ser inte att grundare på EU-hackathons faktiskt vill lösa verkliga problem. De försöker bygga rymdskepp som ingen behöver, säger Viktoriia Yaremchuk.

Om hennes syn på AI och autonoma system i framtiden blir det tydligt att hon ser att radikala framsteg inte är särskilt långt borta.



Viktoriia Yaremchuk, CEO och medgrundare vid Farsight Vision.

– Under nästa ett eller två år kommer robotnätverk med viss mänsklig involvering, en pilot som hanterar flera drönare, säger Viktoriia Yaremchuk.

Krigets innovationer följer en annan logik än fredens – extremt tempo, disruptiva och omöjliga att räkna hem med fredstidskalkyler.

– Ryssarna behöver inte bevisa något för beslutsfattare. De skalar bara upp allt, säger Viktoriia Yaremchuk.

Viktoriia Yaremchuk menar att det

kräver tro på visioner som ännu inte går att begripa – något den svenska försvarsindustrin fortfarande har svårt med.

– Du måste tro på en vision av något du ännu inte förstår, säger Viktoriia Yaremchuk.

– I don't want us to lose more people, I want us to lose robots.

Europa har tid. Ukraina har inte tid. Den gemensamma nämnaren: hur mycket tid vi har tillsammans. □

Rörelser av luftmål över Ukrainas territorium. Kartan baseras på information som visualiserar hot med hjälp av luftförsvarsrelaterad data.



# AI-BEGREPP OCH HISTORIK

AI är ett samlingsnamn för ett vildvuxet fält som handlar om att skapa system som kan utföra uppgifter som normalt kräver mänsklig intelligens. Det kan handla om att planera sekvenser av handlingar och digitala assistenter, att förstå och generativt skapa språk och bild, men också olika typer av AI-agenter, självkörande bilar och autonoma militära system.

## Smal eller svag AI

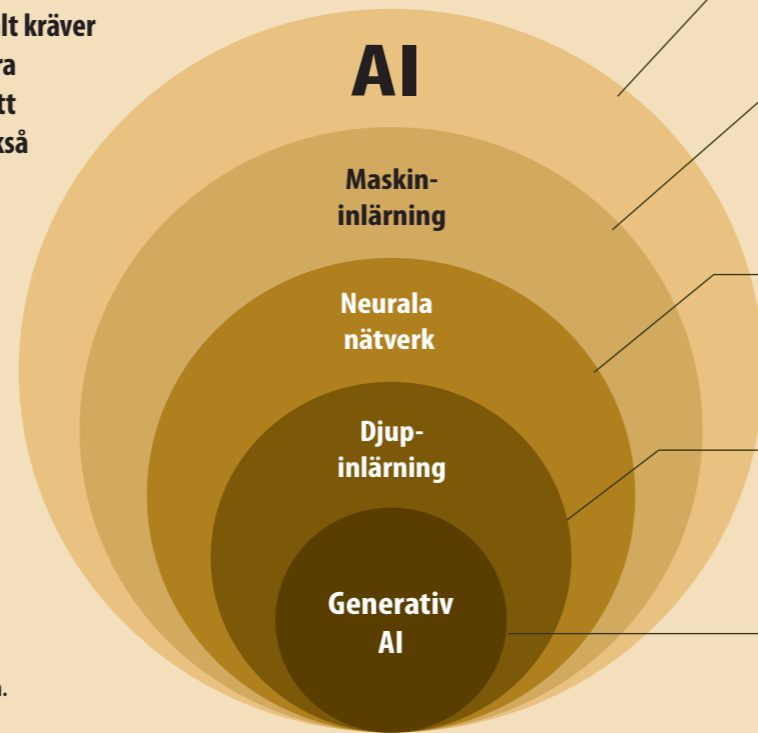
All AI, även den mest avancerade som används idag, räknas in i kategorin smal eller svag AI. De olika nivåerna presenteras i cirkelarna till höger. Maskininläring ingår som en delmängd i begreppet AI, neurala nätverk är en delmängd av både AI och maskininläring – och så vidare.

## Allmän eller generell AI

Ett teoretiskt koncept som syftar på system som skulle kunna förstå och utföra alla intellektuella uppgifter som en människa kan.

## Superintelligens

Ännu en hypotetisk form av AI som har förmågan till autonom självförbättring och eventuell självmedvetenhet. Skulle in framtiden kunna överträffa den mest lysande mänskliga intelligensen på alla plan.



## AI

AI som begrepp används ofta på ett otydligt sätt som en synonym för dess olika underkategorier. Inkluderar både enkla regelbaserade system och avancerade inlärningsbaserade och generativa metoder.

## Maskininläring

Maskininlärnings-modeller lär sig från data i stället för att styras av hårdkodade regler. Flera träningsstyper används, exempelvis övervakad- och oövervakad inläring, självövervakad inläring eller förstärkningsinläring.

## Neurala nätverk

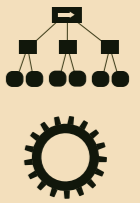
Neurala nätverk är en maskininlärningsmodell som är inspirerad av neuronätverken i däggdjurshjärnor. Man använder algoritmer som är förenklade modeller av hjärnans arkitektur och dynamik.

## Djupinläring

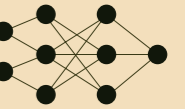
Djupinläring är en typ av neurala nätverk med många (djupa) lager. De många lagren gör att algoritmerna effektivare kan lära sig och generalisera mönster i datan. Används exempelvis för att analysera och klassificera text, bilder och talat språk.

## Generativ AI

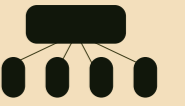
AI-system, som med generativa modeller, kan skapa text, bilder, video eller andra former av innehåll. Genombrott inom transformerbaserade djupa neurala nätverk lade grunden för de senaste årens AI-boom med tjänster som Chat GPT, Gemini, Midjourney och DALL-E.



Mer om övergripande AI och maskininläring på sidorna 68–69.



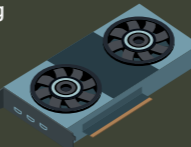


Mer om neurala nätverk och djupinläring på sidorna 70–71.



Mer om generativ AI på sidorna 72–73.

## HISTORISK UTVECKLING

1940–1950-tal	1960–1970-tal	1980–1990-tal	2000-tal	2010-tal	2020-tal
<p><b>Teoretisk grund och tidiga idéer</b></p> <p><b>1943:</b> Warren McCulloch och Walter Pitts presenterar den första formella beskrivningen av ett artificiellt neuralt nätverk.</p> <p><b>1949:</b> Donald Hebb beskriver Hebb's teori inom neurovetenskap: "Cells that fire together, wire together".</p> <p><b>1950:</b> Alan Turing formulerar Turingtestet.</p> <p>Turing menade att om en människa samtalar med en maskin och inte kan avgöra om det är maskin eller människa, så uppfylls kriteriet för mänsklig intelligens.</p>  <p>Alan Turing</p> <p><b>1951:</b> Marvin Minsky och Dean Edmonds inleder bygget av SNARC, en tidig neuronsimulatormaskin.</p> <p><b>1956:</b> John McCarthy myntar begreppet "artificiell intelligens". Startpunkt för AI-forskning.</p> <p><b>1957:</b> Frank Rosenblatt introducerar perceptronen, ett enkelt neuralt nätverk.</p> <p><b>1959:</b> Arthur Samuel utvecklar ett maskininlärande damspel (checkers), ett tidigt självförbättrande program.</p>	<p><b>Symbolisk AI och tidiga program</b></p> <p><b>1965:</b> Joseph Weizenbaum på MIT bygger ELIZA – en tidig chatbot/språksimulator som ibland kunde simulera en konversation som påminde om en människas.</p>  <p><b>1969:</b> Marvin Minsky och Seymour Papert publicerar "Perceptrons" – en kritisk granskning vilken bidrar till minskat intresse för tidiga neurala nätverk.</p> <p><b>1970-tal – första AI-vintern</b></p> <p>Begränsningar perceptrons funktion leder till minskat finansieringsstöd.</p> <p><b>1972:</b> Programspråket Prolog introduceras av Alain Colmerauer. Används inom områdena artificiell intelligens och språkbehandling</p>	<p><b>Expertsystem och återuppvaknande</b></p> <p><b>Tidigt 1980-tal:</b> Kommerciellt intresse för expertsystem. Programmeringsspråket Lisp marknadsförs för att erbjuda bättre utvecklingsmiljö för AI-forskning.</p> <p><b>Mitten/slutet av 1980-talet:</b> Återuppvaknande av neurala nätverk. Backpropagation återintroduceras och blir populärt.</p> <p><b>Sent 1980-tal – andra AI-vintern</b></p> <p>Andra AI-vintern börjar när underhåll och kostnader för expertsystem blir uppenbara.</p> <p><b>Tidigt 1990-tal:</b> Ökat fokus på statistiska metoder, probabilistiska modeller som Bayesianska nätverk.</p> <p><b>1997:</b> IBM:s superdator Deep Blue, designad av Feng-hsiung Hsu och hans team, slår den regerande världsmästaren Gary Kasparov i schack.</p> <p><b>Sent 1990-tal:</b> Internets tillväxt möjliggör stora datamängder och informationsextraktion från webb-spindlar och andra AI-baserade program.</p> 	<p><b>Big data, webbtjänster och maskininläring i produktion</b></p> <p><b>Tidigt 2000-tal:</b> Ökad användning av maskininläring för webbaserade tjänster som annonsering, rekommendationer och sökmotorer.</p>  <p><b>Mitten av 2000-talet:</b> Mer datakraft och stora mängder tillgänglig data gör att djupinläring ses som en praktisk väg framåt.</p> <p><b>2006:</b> Geoffrey Hinton publicerar "Learning Multiple Layers of Representation". Det markerar ett skifte från traditionella neurala nätverk till djupinläring, vilket möjliggör att maskiner kan lära sig komplexa hierarkiska representationer av data.</p>  <p><b>2009:</b> Forskning pekar på att grafikprocessorer (GPU:er) är överlägsna traditionella flerkärniga CPU:er för djupinläring.</p>	<p><b>Djupinläringens genombrott</b></p> <p><b>2011:</b> IBM:s dator Watson besegrar två stormästare i Jeopardy.</p> <p><b>2011–2014:</b> Apple, Google och Microsoft lanserar digitala assistenter för smartphones som kan svara på frågor och ge rekommendationer med naturligt tal.</p> <p><b>2014:</b> Ian Goodfellow tillsammans med andra introducerar Generative Adversarial Networks (GANs) vilket bildar grunden för bildgenererande AI-modeller.</p> <p><b>2016:</b> Deepmind slår Lee Sedol i spelet Alphago. Genombrott för förstärkningsinläring och neurala nätverk.</p>  <p><b>2017:</b> "Attention Is All You Need" (Vaswani tillsammans med andra) introducerar transformer-arkitekturen, vilket revolutionerar naturlig språkbehandling.</p>	<p><b>Ökat fokus på självövervakad inläring, storskalig förträning och transformers</b></p> <p><b>2000-tal:</b> En markant ökning av användningen av AI. Drivs huvudsakligen av tillgången till öppen källkod och delning av öppna dataset för träning av modeller.</p> <p><b>2020:</b> OpenAI introducerar GPT-3, den dittills största stora språkmodellen med 175 miljoner parametrar.</p> <p>Google Deepminds Alphafold 2 kan korrekt förutsäga 3D-strukturerna av proteiner utifrån deras aminosyra-sekvenser.</p> <p><b>2023:</b> Kommerciella AI-modeller erbjuder multimodala funktioner – de kan inte bara förstå text utan även bilder, diagram och annat.</p>  <p><b>2025:</b> Fokus på agentliknande system och "world models" – neurala nätverk som förstär dynamiken i den verkliga världen, inklusive fysik och spatiala egenskaper.</p>

Källor: Chalmers, FOI, IBM, Wiki, Nivida, AI-sveden

# Billig teknik kan slå ut dyra system

**Drönare, AI, det existensiella hotet, samhällets digitala infrastruktur, små företag och kompetenta personer. Allt är delar i en helhet, ett försvarssystem. Samspelet mellan militärer och civila är centralt. Ingen sida sitter inne med svaren, men tillsammans kan man nå lösningar. Det säger Hans Liwång, professor i försvarssystem vid Försvarshögskolan.**

*Text och foto: Jan-Ivar Askelin*



»Stridsvagnen, stridsflygplanet, uppdragstaktiken kom att förändra krigföringen. Allt kom under första världskriget som varade i drygt fyra år. Kriget i Ukraina har pågått i nästan fyra år. Vilken påverkan har det?«.

**H**ans Liwång säger att modern, billig digital teknik kan slå ut väldigt dyra plattformar.

– De stora materialsystemen är inte självklara framtida lösningar. Civil utveckling som gett oss AI, drönare, appar och en massa sådana saker löser traditionella militära uppgifter. Det här skiftet är en enorm öppning för de senaste 40 årens civila utveckling. Man kan använda det man har, skruva lite på det och använda det militärt. Appen som man beställer hamburgare med kan beställa artillerield. Det är bara att ändra logga på appen.

Hans Liwång betonar dock att det traditionella kriget finns kvar, men balansen har ändrats.

– Osäkerheten är större därför att inget av det gamla är osant. Man lägger på nya lager som vi inte riktigt har satt oss in i ännu.

#### AI utvecklas på förband

Hans Liwång ser AI som ett tekniskskifte om än med många brister.

– Vi har sett det här skiftet civilt, men man har stretat emot lite inom försvarsorganisationernas värld av rigida strukturer där man satsar på beprövad materiel och lutar sig mot försvarsindustrin. Det kommer att ändras. I framtiden kommer den digitala strukturen att utvecklas på förbandet och inte hos industrin.

– Och AI är ju ett bra exempel på

detta. AI är egentligen bara en motor som kan göra olika saker. Men motorn måste hela tiden trimmas.

Hans Liwång menar att AI är det ord vi använder för att beskriva en viss typ av digitala lösningar som spelar roll i automationen.

– För en del av de här lösningarna räcker det med traditionell kod. Men i vissa fall kan vi inte från början definiera vilka problem som ska lösas och då vore det bra att ha en kod som lär av sina misstag så att autonomi blir dynamisk.

#### Samspel civilt-militärt

– Varken militärer eller civila sitter ensamma med svaren, säger Hans Liwång.

– Det är bara när de möts och gemensamt lär sig att förstå problemen som lösningar kan uppnås. Bara för att man är expert på drönare eller AI betyder det inte att man kan hitta en militär lösning.

Gapet mellan militärer–beställare och tekniker–leverantörer är inget nytt fenomen. På Försvarshögskolan finns kurser och program i försvarssystem där dessa världar möts och där Hans Liwång undervisar.

– Det existentiella hotet kopplade samman den digitala världen, utan vilket ett samhälle inte fungerar, med de militära problemen. Lösningarna kommer nedifrån. Huvudsaken är att det fungerar, som när det dök upp en app som varnar för hotande anfall.

#### Drönare talar med varandra

– Nyckeln till en svärm är att drönarna pratar med varandra. Vem filmar? Vem skickar informationen vidare? Vem ska ha vapnet? Hur ska vi flytta svärmen? Då

skulle AI kunna vara en del i ledningen. Det är inte en drönare som löser uppgiften, utan det är alla drönare tillsammans. Svärmen behöver anpassa sitt agerande efter det som händer. Eller kanske öva sig själv i en dator. Det är lärandet som är grejen, säger Hans Liwång

– Det centrala är att man kontinuerligt kan samla in data och förändra agerandet utifrån det insamlade. Det är kopplingen mellan AI och autonomi som är poängen när den traditionella automatiska tekniken inte duger.

Stora synkroniserade drönarsvärmar visas upp i olika sammanhang. Hans Liwång säger att det inte är säkert att det är den typ av svärmar som försvaret behöver.

– Det kanske räcker med ett par drönare. Varför krångliga till det med stora svärmar?

#### Fler ägg i samma korg

– Vi hängde på mer och mer funktioner i stridsflygplanet och då blev det ännu värdefullare. Då måste vi också försvara det ännu mer. Då behöver vi hänga på skydds lösningar. Då blev det ännu vassare. Då hade vi bättre vapen på dem och så fick vi en plattform som är väldigt dyr och väldigt komplex. Och om den inte kan användas så har vi inget. Vi får ett stridsflygplan som bara kunde användas om vi har luftherravälde, säger Hans Liwång.

– I den extrema teknikutvecklingen har man lagt ännu fler ägg i samma korg och satsat mer och mer på den. Då sveper den billiga drönaren in och säger att jag kan göra det här och blir jag nedskjuten spelar det ingen roll.

»Vi får ett ett stridsflygplan som bara kan användas om vi har luftherravälde.«

– Idén om en försvarsmakt utgår från att det finns stridsvagnar, stridsflygplan och så vidare. Man utgår alltid från de här tydliga markörerna. De kommer säkert ha en plats även i framtiden. Men vi får inte utgå från deras roll. Därför att de kommer att behöva kompletteras i väldigt stor grad med andra saker.

– Med hjälp av AI skulle vi kunna få en större grad av autonomi och då kan det inte sitta en människa i till exempel flygplanet.

#### Spridning bästa skyddet

Hur skyddar man människor och materiel när autonoma maskiner i nätverk, kanske styrda av AI, spelar en större roll?

Hans Liwång tror att lösningen, om det finns en sådan, består i ökad utspridning vad gäller funktion, teknik och platser. Ökad mångfald på olika nivåer.

– Har man olika lösningar så kommer de att ha sina styrkor och svagheter, men det kommer alltid finnas någon som fungerar. Internet är utsatt för attack och störningar hela tiden. Internet har aldrig gått ner. Därför att det är fullt av olika lösningar.

Traditionellt vill militären kraftsamla och uppnå lokal överlägsenhet. Hans Liwång menar att det nu plötsligt blir farligt.

– Inget skydd kommer att vara hundraprocentigt. Vi måste ha system där man kan ta förluster utan att det alltför mycket påverkar den militära förmågan.

– Hur vi ska försvara stridsvagnen mot drönarsvärmen? Frågan är kanske fel. Vi kanske inte ska ha en stridsvagn?

– Om motståndaren har AI är kanske skyddet att angripa AI-förmågan. Det kan vara lättare att hitta svaghet-

er i en AI än i en traditionell plattform där det finns tänkande person.

#### Drönarens etik

– Ett stridsflygplan skickar man inte för att hitta en individ. En billig drönare kan riktas mot vem som helst. Det är det som är det extremt läskiga i det här transparenta stridsfältet, säger Hans Liwång.

– Om den här drönaren har lärt sig ett beteende som vi tycker är olämpligt, vem tar ansvar för det? Det vet vi inte. Det är inte säkert vi kommer kunna genomskåda hur den här prylens intelligens har utvecklats.

– Jag har ingen riktigt bra lösning på det. Jag tror dock att det är någonting vi måste fundera vidare på och lagarna och etiken måste sättas i det sammanhanget. Vi kan inte prata om det här för abstrakt. De solklara frågorna är inte problemet utan själva gränsdragningen. □

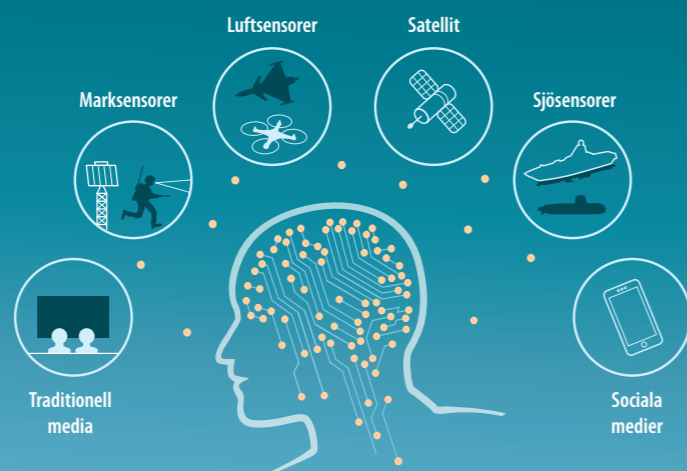
# ARTIFICIELL INTELLIGENS PÅ DET MODERNA SLAGFÄLTET

AI är på snabb framfarsch i samtliga domäner på det moderna slagfältet och har förmågan att få OODA-loopen\* (observera, orientera, besluta, agera) att snurra allt snabbare. Olika plattformstyper automatiseras i allt högre utsträckning och AI kan användas för sammanställning och analys av stora, komplexa och ostrukturerade datamängder, vilket leder till ökad effektivitet och precision. Utvecklingen för dock med sig stora etiska, juridiska, tekniska och taktiska utmaningar.

\* Mer om OODA-loopen på sidan 21

## Underrättelser, övervakning och spaning

AI och maskininläring har förmågan att snabbt optimera data i komplexa och dynamiska slagfältsmiljöer. Allt, från sensorer och satelliter, till traditionella mediaplattformar och sociala medier kan användas som källor för bedöma och reagera på militära hot. Även till faktorer som terräng, fiendens intentioner och sensorosäkerheter kan vägas in.

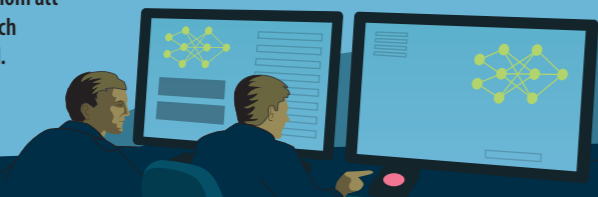


## Beslutsstöd/kommando

Redan idag finns AI-system som ger förslag på militära mål och vilken ammunition som ska användas vid bekämpning. AI-algoritmer kan hjälpa befälhavare att fatta snabbare och mer informerade beslut genom att analysera stora mängder data från slagfältet och väga dessa mot militär doktrin och värdegrund.

Frågan är vem som ska trycka på knappen.

Utfall 1	✗	Utfall 4	✗
Utfall 2	✗	Utfall 5	✓
Utfall 3	✗	Utfall 6	✗



Källa: FHS, FOI, Chalmers, RISE

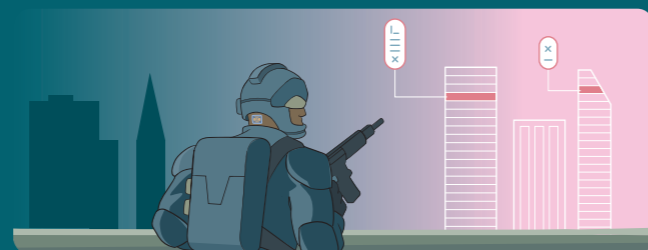
## Logistik

Logistik är en av de mest resurskrävande delarna i militära operationer. Med hjälp av AI går det att optimera försörjningskedjor, och identifiera effektiva transportvägar.



## Utbildning

AI-drivna simuleringar kan ge realistiska träningsmiljöer för soldater och befälhavare att öva på taktik och beslutsfattande.

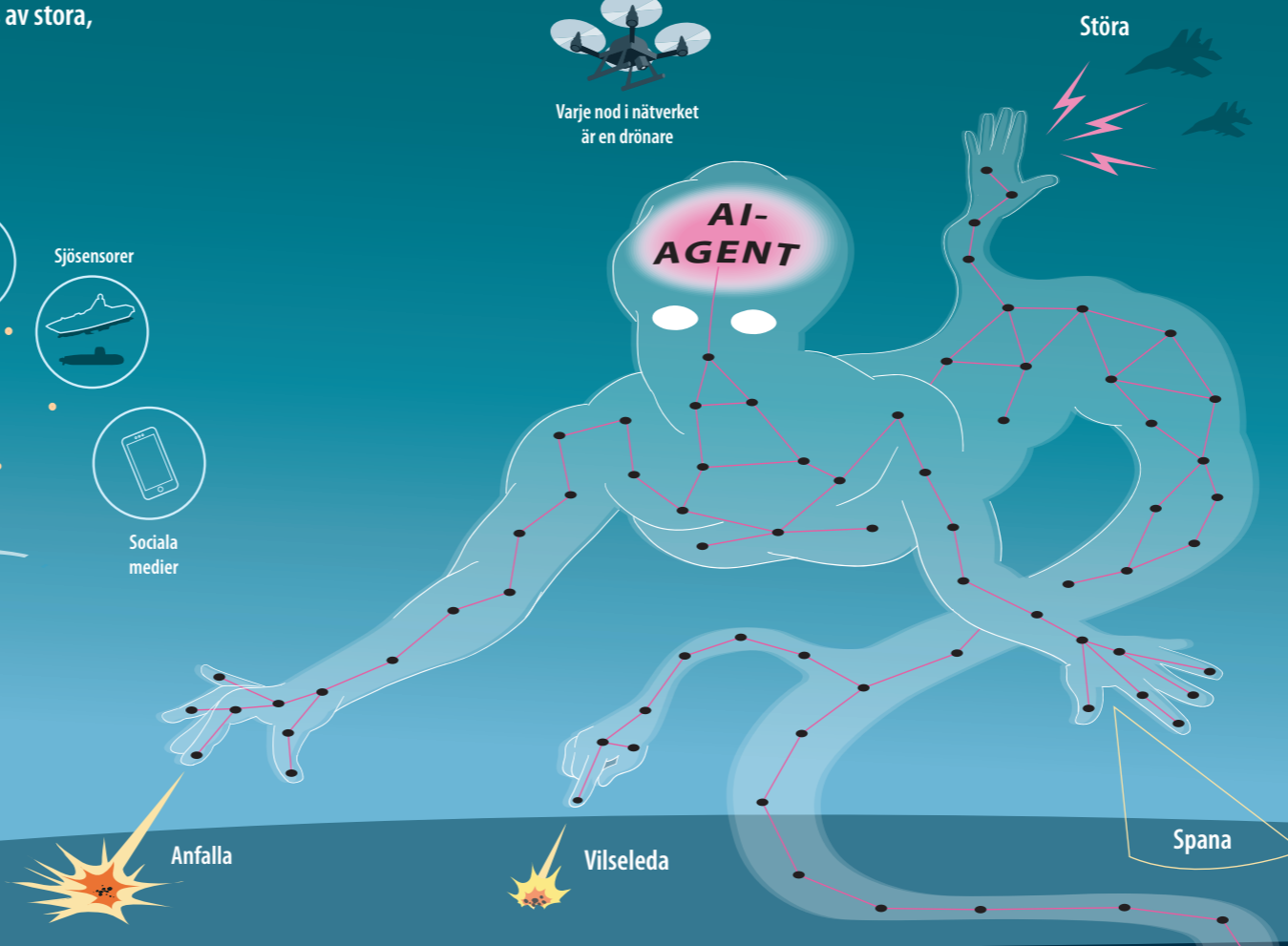


## AI-agenter och svärmar

Olika typer av AI-svärmar simuleras redan och bygger på att enheterna kan kommunicera med varandra och på så sätt skapa en gemensam virtuell agentförmåga. En AI-svärm kan ses som en egen entitet, som utan förprogramering anpassar sig efter de militära mål som ska uppnås. Detta i kombination med hur slagfältets dynamik oavbrutet förändras – en ny typ av uppdragstaktik.

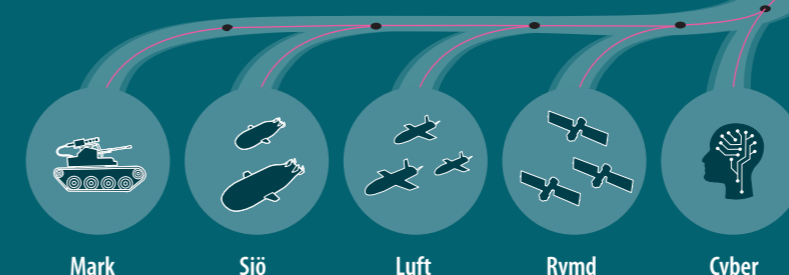


Varje nod i nätverket är en drönare



## Multi domain operations

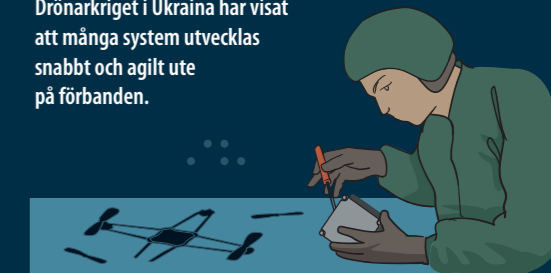
En agentstyrd drönarsvärm är bara ett exempel på en AI-agent som verkar i luftdomänen. I framtiden kan oräknliga AI-agenter vara sammankopplade – både som fysiska system som helt virtuella – och verka tillsammans inom samtliga domäner.



## Den billiga digitala tekniken

Det gamla materialförsörjningssystemet är på väg att förändras i grunden. Digitala system – med eller utan AI – i kombination med många små enkla farkoster kan slå ut dyra plattformar.

Drönarkriget i Ukraina har visat att många system utvecklas snabbt och agilt ute på förbanden.



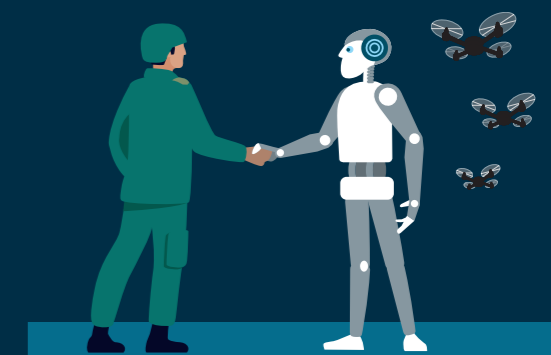
## Kognitiv krigföring

Foreign Information Manipulation and Interference (FIMI) är ett hybrid-hot där utländska aktörer sprider desinformation för att påverka politiska processer och opinion. AI-modeller blir allt bättre på att producera vilseledande material, inte bara i textform utan även i form av ljud, bild och rörlig media.



## Tillit

Tillit till AI-agenter är en förutsättning för att nya, snabbt sammansatta konstellationer av människor, sensorer, plattformar och algoritmer ska våga dela data, ta emot rekommendationer och faktiskt agera på AI-genererat stöd i det digitaliserade slagfältets tempo. Utan en medvetet uppbyggd och förvaltd tillit – tekniskt, organisatoriskt och kognitivt – riskerar AI att bli antingen oreflekterat och övertrumfande eller helt åsidosatt. I båda fallen försämrar det förmågan att fatta kloka koordinerade beslut i osäkra situationer.



# BESLUT FATTAS I NY MILJÖ

**1987. Det kalla kriget var på väg mot slutet. Fast det visste ingen då. Peter Bovet-Emanuel började sin värnplikt som kustjägare och blev sedan officer. Nu är han forskare vid Forsvarshögskolans avdelning för krigsvetenskap. Han är nybliven doktor och har undersökt hur AI påverkar den framtida militära beslutsprocessen. Från och med 2026 ska han bygga en forskningsmiljö för att studera hur AI påverkar Försvarmakten nu och i framtiden. Drömmen är att vi ska ligga på den militära AI-teknikens framkant, säger han.**

*Text och foto: Jan-Ivar Askelin*

**P**eter Bovet-Emanuel utbildade som ung officer attackdykare i tio år, gick över till specialförband och blev ansvarig för förmågeutveckling inom hela Försvarmakten.

– Jag insåg då att tekniken fick en allt större betydelse och den teknik jag var särskilt intresserad av var AI. Som forskare ställde jag frågan vad händer när AI kommer in i den militära beslutsprocessen.

För att undersöka detta använde Peter Bovet-Emanuel en beprövad – och omstridd – metod för att beskriva hur ett beslut går till. John Boyd deltog som ung pilot två månader i Koreakriget och blev med tiden analytiker och tänkare med stort inflytande i det amerikanska flygvapnet. Mot bakgrund av erfarenheter från Vietnamkriget skapade han den så kallade OODA-loopen som beskriver processen observe–orient–decide–act (se sid 21).

– På svenska blir det ungefär observera, orientera eller förstå, besluta och agera. Boyds loop består av loopar inne i looparna och kan användas i olika tidsperspektiv. För min forskning är den tillräckligt enkel i sina grunder att förstå.

Peter Bovet-Emanuel beskriver hur AI stegvis hamnat i beslutsprocessen.

Människan har sina sinnen. Vi observerar med våra ögon. För se bättre behövs en kikare. Organisationen bygger radarkedjor. Så har maskinerna kommit in bit för bit i de fyra faserna. Med digitaliseringen blev det en flod av data och information. Nu fick maskinerna en allt större roll. Information skulle sorteras, värderas och presenteras. Beslutsfattarna blev allt mer beroende av maskiner. Då handlade om tillit till systemet. Och nu inträder AI i form av agenter. Först i rummet observe, sedan i rummen orient och decide, och nu gläntar AI-agenterna på dörren till den viktigaste rummet – act. Ska AI få bestämma om bomben ska fällas eller inte eller mot vem och vad?

Med AI menas i de flesta fall maskinlärning. Det är AIs starka sida – att känna igen objekt och mönster. Det blir svårt för människan att hänga med i AI-looparna.

*»Att se AI som ett beslutsstödsystem ska vägas mot den andra vågskålen och det som händer nu i Ukraina. Det talas om den sista milen. Då är det en engelsk mil och jag tror att vi är på väg mot en svensk mil.«*

Hur ska människan ens kunna förstå när man tagit över en bild som skapats av maskiner och AI? När AI gör mer och mer av analysen vad blir då människans roll som analytiker? Hur ska vi skapa en riktig situationsuppfattning i en digital miljö som består av mer och mer desinformation som kan massgenereras av AI? Krigets dimmor tättnar, tempot och desinformationen ökar. Och besluten ska fattas snabbare.

#### Frågan om ansvar

Peter Bovet-Emanuel frågar sig hur vi undviker att människan blir en marionett. Han säger att utgångspunkten är

vilka konsekvenser ett beslut får. Det är en sak att AI får bestämma att en mast ska angripas. En annan att låta AI avgöra människors liv.

– Slutsatsen måste bli att där det handlar om människor så måste människor få fatta besluten. Det har diskuterats länge hur AI ska få användas militärt. Nu verkar diskussionen ha hamnat i det som kallas responsible AI. Jag tror att det finns en ökad förståelse hos militära och civila beslutsfattare att det är viktigare att titta på ansvarsfrågan än att prata om var man använder AI.

– Att se AI som ett beslutsstödsystem ska vägas mot den andra vågskålen och det som händer nu i Ukraina. Det talas om den sista milen. Då är det en engelsk mil och jag tror att vi är på väg mot en svensk mil. Det är en zon där autonomin råder. Där maskiner strider mot maskiner, algoritmer mot algoritmer. Där kan ingen människa vara. Här inte räcker AI som stöd utan tar över som beslutsfattare.

I sin avhandling har Peter Bovet-Emanuel tagit fram två modeller som kan användas för en diskussion om etik och AI.

System 1 är ett djupt neuralt nätverk som hjälper till att definiera var mål sannolikt kan finnas.

– Då kan vi leta mål på rätt plats med rätt sensorer i stället för att leta efter nålen i höstacken.

System 2 är en regelbaserad optimeringsalgoritm.

– Alla värden matades in. Vilka mål finns? Ska de bekämpas och i så fall med vilka vapen? Är målet värt ett dyrt vapen som vi har ont om? Det tar experter en halvtimme att komma fram till ett svar. Med AI kom svaret på mindre än en sekund.▷

#### Peter Bovet-Emanuel budskap:

- Vi bör undvika överdriven teknikoptimism.
- Det stora förändringen handlar inte om att maskiner gör mer – utan att människor får allt mindre tid och makt att förstå på vilka grunder beslut fattas.
- AI kan ge bättre förutsättningar men det betyder inte att besluten blir bättre. För detta krävs bland annat teknologisk förståelse, tillit och en anpassad beslutskultur.
- AI förändrar inte bara vad vi gör – utan hur vi förstår beslutsfattande. Det handlar därför mindre om vem som har mest teknik, utan mer om vem som har de mest utvecklade beslutsprocesserna.
- Med AI får vi en ny typ av beslutsmiljö.

»Kriget lagar kommer ofta i efterhand. Det är svårt att diskutera etik om man inte har färdiga system att utgå från. Här kunde jag visa på två system. Är det etiskt eller moraliskt försvarbart att använda dem.? Det får vara mitt bidrag till diskussionen om etik och moral.«

**Utspritt beslutsfattande**

– Krigets lagar kommer ofta i efterhand. Det är svårt att diskutera etik om man inte har färdiga system att utgå från. Här kunde jag visa på två system. Är det etiskt eller moraliskt försvarbart att använda dem? Det får vara mitt bidrag till diskussionen om etik och moral.

Och det finns lösningar där AI används i agerande-fasen exempelvis då i det som indikeras både i Israel och kanske framförallt i Ukraina.

Peter Bovet-Emanuel säger att i framtiden kommer beslutsfattandet att vara utspritt och inte centralt. Autonoma maskiner kommer att tänka självständigt. Det här ger en ny dimension på den välkända uppdragstaktiken där den som är först på plats ska fatta besluten. Å ena sidan kanske AI-nätet har bättre koll på läget än den som är på plats. Å andra sidan kan autonoma AI-maskiner gemensamt ändra på uppdraget.

**Intelligent drönarsvärm**

Peter Bovet-Emanuel är i första hand inriktad på beslutsfattande och inte vapensystem, men det är svårt att undvika frågan om den AI-styrda drönarsvärmen.

– Man kan inte blunda för att det är AI-agenter som agerar. Agenten styr svärmen där varje drönare kan tänka självständigt. Varje drönare kan ha olika roller som jakt, attack och spaning. Svärmen behöver inte ha givna uppgifter från början. Den skapar sig en egen uppdragstaktik och fördelar uppgifterna. Det är något som idag simuleras och modelleras.

– Vi behöver förstå vad som händer när vi lämnar över mer och mer till tekniken och då blir drönar-

svärmarna intressanta. Hur kommer beslutsfattande inom och kring sådana system att se ut? Vi behöver ha beslutsprocesser som är lika snabba som tekniken och lika ansvarsfulla som människan.

Hur långt bort är den AI-styrda drönarsvärmen? Ett franskt företag, Atreyd, ska enligt uppgifter i media installera en svärm bestående av 200 drönare. De styrs med AI från marken, men har inte själva AI. De liknas vid programmerbara flygande minor. Minfältet läggs i vägen för de anfallande drönarna. Den dödliga zonen, den sista milen, har flyttat upp i luften.

Det kanske blir slaget om Storbritannien över 80 år senare i AI-åldern.

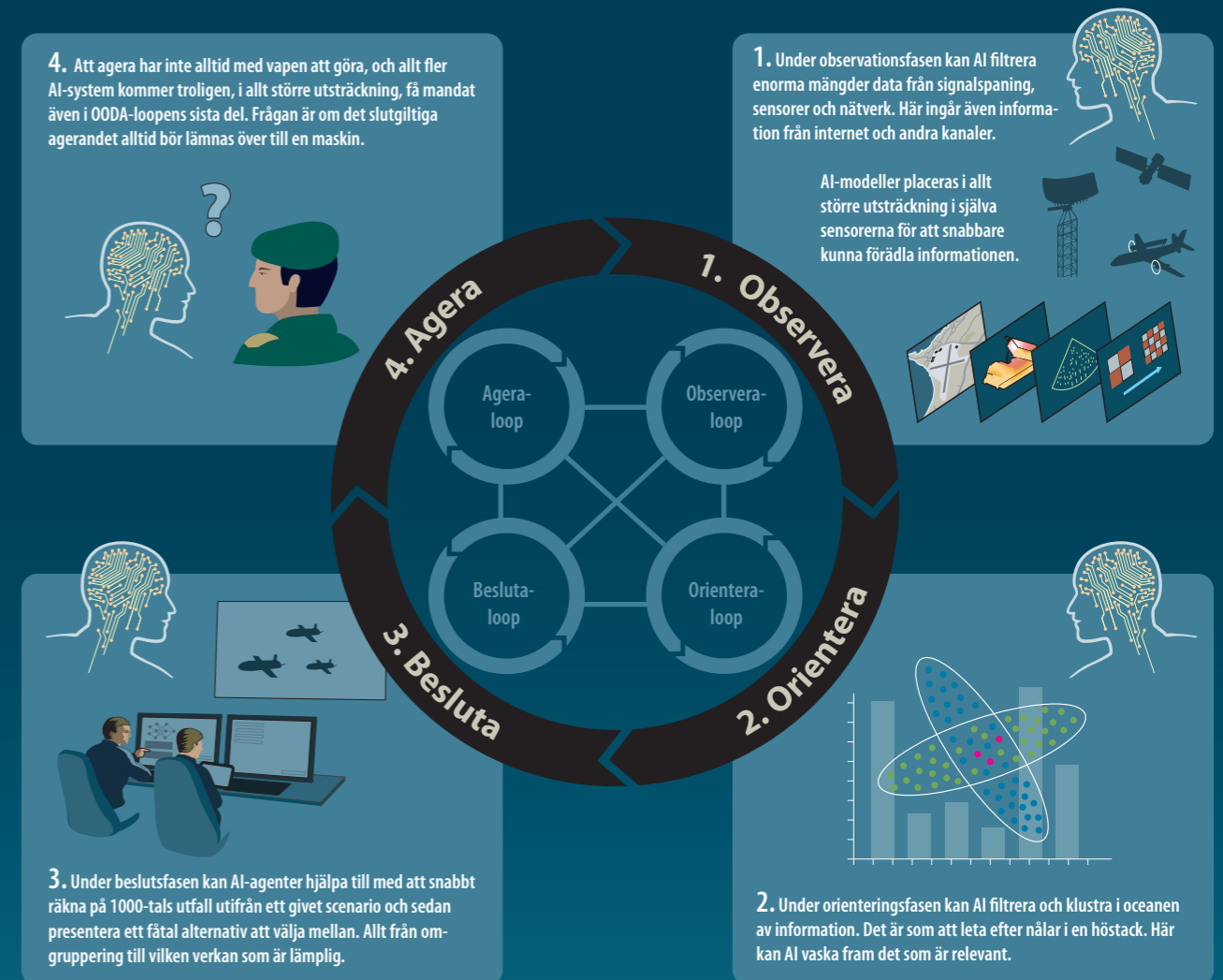
– Banditer klockan två. Vad var det Boyd sa? Observe–orient–decide–act. □



Peter Bovet Emanuel, kustjägare i forskningsfronten.

# AI OCH OODA-LOOPEN

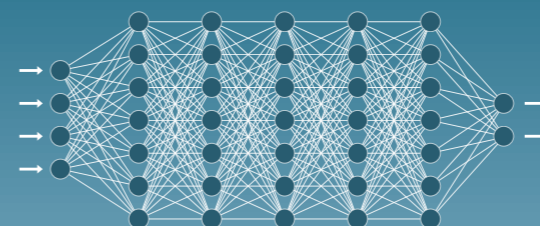
Ett sätt att tydliggöra AI:s intrång i den militära världen är att använda OODA-loopen som formulerades av stridsflygaren John Boyd på 1950-talet. Den bygger på faserna observation, orientation, decision och action. Varje steg i OODA-loopen har en egen itererande loop som dessutom kommunicerar med övriga delar. Redan idag har maskiner och AI etablerat sig i de tre första stegen. Frågan är om människan känner sig tvungen att släppa ifrån sig mandatet att agera när OODA-loopen snurrar allt snabbare?



## Två exempel på AI-system

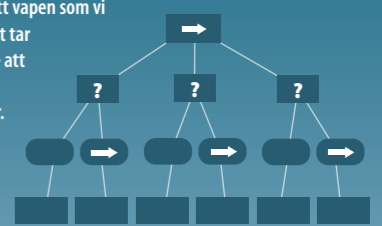
### 1. Prediktiv analys baserat på satellitbilder

Djupa neurala nätverk kan snabbt definiera var mål sannolikt kan finnas. Det gör att man kan leta efter mål på rätt plats med rätt sensorer i stället för att leta efter nålen i höstacken.



### 2. Regelbaserade algoritmer

När det rör sig om flera mål och kort om tid blir det svårt för människan. Alla kända värden kan matas in i en regelbaserad optimeringsalgoritm som är mer transparent och spårbar än ett neuralt nätverk. Vilka mål finns? Ska de bekämpas, och i så fall med vilka vapen? Är målet värt ett vapen som vi har ont om? Ett svar det tar experter en halvtimme att komma fram till klarar AI:n på några sekunder.



# FOLKRÄTTEN UTMANAR AI

**Ett autonomt vapensystem måste kunna göra skillnad på kombattanter och civila för att användas i linje med folkrätten. Det måste också kunna identifiera om en militär person har kapitulerat, är skadad eller bär sjukvårdarens emblem. Men vem bär ansvaret om ett AI-system leder till att krigets lagar bryts?**

*Text: Kajsa Skarsgård*

*Foto: Tomas Ärlemo*

**S**lagfältet i Ukraina har blivit en testbädd för AI i krig. Utvecklingen går snabbt. AI kan nu – med varierande resultat – ge oberoende farkoster autonoma funktioner som perception, målföljning, måligenkänning, positionering och navigering.

Totalförsvarets forskningsinstitut (FOI) bedömer att det är troligt att Ukraina redan under år 2025 i större omfattning kommer att införa flygande spaningsdrönare som också kan bekämpa mål autonomt. En pådrivande faktor är att telekriget är så intensivt.

Men eftersom AI:n väntas klassificera målen fel relativt ofta kommer det krävas metodbegränsningar för att inte bryta mot den internationella humanitära rätten – krigets lagar. De gäller nämligen oavsett försvarsgrenar. Användningen av AI i krig är därför inte oreglerad.

Civila och militära personer och objekt måste till exempel särskiljas, och

bara de militära får anfallas. Oavsiktliga civila sidoskador måste vara proportionerliga till den militära nyttan. De stridande parterna måste vidta försiktighetsåtgärder för att minimera civilt lidande, genom metoder, val av vapen, med mera. Om en stridande part bryter mot krigets lagar, ger det inte den andra parten rätt att göra det samma.

Samtidigt som AI kan bidra till att ge en operatör en bättre lägesbild och ett vapen bättre precision, finns en risk med att autonoma vapensystem gör det möjligt att med små resurser attackera i stor skala. Eftersom tekniken är relativt lättillgänglig kan den också hamna i fler sorters aktörers händer.

## **Dialog i krigens skugga**

Anna Andersson är folkrättsforskare på FOI. De senaste tre åren har hon deltagit i Sveriges delegation i det forum där de flesta av världens länder – även Ryssland, Israel, USA och Kina – sedan 2014 diskuterar om och hur autonoma vapensystem ska regleras närmare. Detta görs inom ramen för konventionen av vissa konventionella vapen (eng.

Convention on Certain Conventional Weapons – CCW).

Det är inget totalförbud av autonoma vapensystem som ligger på bordet, utan ett förslag på begränsningar i hur de får utvecklas och användas för att kunna uppfylla redan existerande internationell humanitär rätt. Arbetet ska bli klart under 2026.

– Jag tror att tre aspekter möjliggör att autonoma vapensystem kan användas rättsenligt. Det ena är att man sätter begränsningar för användningen, exempelvis för var det specifika systemet får användas, under vilken tid, mot vilka typer av militära mål, och så vidare.

– Det andra är att systemet är tillräckligt tekniskt sofistikerat för att uppfylla de krav som folkrätten uppställer, exempelvis att kunna välja ut det militära mål som man vill anfälla och särskilja det man inte får anfälla.

– Det tredje är människans involvering: att militära befälhavare och operatörer uppfyller folkrättens krav på att göra vissa bedömningar i planering och genomförande av anfall. Dessa tre ▷

*»I krigföring i en stad där det är mängder av civila kan det ju vara oerhört komplext att dels skilja civilpersoner och civil egendom från militära mål, dels avgöra proportionalitet. I luftstrid eller sjöstrid kan de folkrättsliga frågorna vara mindre komplexa.«*

# STRIDER FÖR SÄKRARE AI-SYSTEM

**På 1990-talet var hon en av de första kvinnliga stridspiloterna i den amerikanska flottan. Nu är hon professor och tar strid mot techgiganter påståenden om AI:s pålitlighet i säkerhetskritiska system. – Vi har släppt ut generativ AI i världen utan den vanliga testningen och evalueringen, säger Missy Cummings under ett seminarium på Stockholm International Peace Research Institute.**

*Text: Kajsa Skarsgård*



**M**issy Cummings är på rundresa i Europa för att tala om vad självkörande bilar kan lära oss om dödliga autonoma vapen. I seminarierummet på Stockholm International Peace Research Institute (Sipri) berättar hon att hon är glad att befinna sig i Europa just nu. Några dagar tidigare har den konservativa aktivisten Charlie Kirk mördats under en debatt i Utah. Missy Cummings skulle inte längre vilja framträda offentligt i USA utan ▶



I krig är människan alltid ansvarig, inte maskinen, säger folkrättsforskaren Anna Andersson.

aspekter är beroende av och påverkar varandra, och får olika betydelse beroende på domän och den avsedda användningen, säger Anna Andersson.

Trots pågående krig där allvarliga kränkningar av folkrätten begås är diskussionerna mellan staterna i CCW förhållandevis konstruktiva, berättar hon.

– Jag tror att det beror på att alla stater ser att det här är teknik som har stora möjligheter, men att den också kommer med utmaningar.

– Det är också ett understrykande av att folkrätten fortfarande gäller. Även om vissa stater väljer att bryta mot den så vill de fortfarande att folkrätten ska finnas kvar.

## Mänsklig kontroll

En av de tuffaste frågorna att komma överens om i CCW är om konventionen bara ska omfatta vapensystem som är konstruerade för att döda människor,

eller även de som ska slå ut objekt men som också kan leda till att människor skadas. En annan svår fråga handlar om mänsklig kontroll över våldsanvändning och människans roll i att säkerställa respekt för folkrätten. En sak är folkrätten tydlig med och länderna i alla fall överens om – en maskin är aldrig ansvarig utan det är alltid en människa.

Graden av mänsklig kontroll som behövs för att leva upp till folkrätten beror på kontexten.

– I krigföring i en stad där det är mängder av civila kan det ju vara oerhört komplext att dels skilja civilpersoner och civil egendom från militära mål, dels avgöra proportionalitet. I luftstrid eller sjöstrid kan de folkrättsliga frågorna vara mindre komplexa. I sjökriget kan det i vissa fall vara ganska enkelt att göra folkrättsliga bedömningar men det kan också finnas en massa civila fartyg och folk i vattnet och då blir det genast mer komplext, säger Anna Andersson.

I Ukraina finns zoner vid fronten där användning av autonoma vapensystem kan vara mindre problematisk sett till skyldigheten att skydda civila. Det är här Ukraina i första hand förväntas använda den nya tekniken. En svårighet för AI:n kan däremot bli att att skilja mellan ukrainska och ryska soldater, mellan kombattanter och skadade.

## Massövervakning och snabba beslut

I Israels senaste två års krigföring i Gaza har AI-baserade beslutsstödsystem pekats ut som en faktor bakom det historiskt höga antalet dödade civila. En massiv kommunikationsövervakning och datainsamling har tillsammans med AI-baserade översättningar och analyser legat till grund för AI-genererade förslag på mänskliga militära mål. Målidentifiering har därmed skett i en takt och omfattning som aldrig förr.

Den mänskliga kontrollen i slutet av denna AI-kedja – verifieringen av att målen är militärt legitima – har enligt vittnesuppgifter i internationell media uppskattas till i snitt 20 sekunder per beslut. På basis av de besluten har israelisk militär sedan kunnat skicka

bomber dit de utpekade personerna bedöms vara.

På den första årsdagen av terrorattacken då Hamas dödade omkring 1200 personer och kidnappade 248, uppgav Israels militär att den hade bombat 40 000 på mål Gazaremsan. Det är en yta motsvarande en fjärdedel av Öland som 2,2 miljoner människor lever på. På den andra årsdagen av terrorattacken uppgav den Hamaskontrollerade hälsomyndigheten att Israel hade dödat drygt 67 000 människor i Gaza, varav 20 179 barn.

## Ansvar på alla nivåer

I FN pågår diskussioner om AI i militära beslutsstödsystem ska regleras närmare, men de diskussionerna är ännu i sin linda. Människans ansvar gäller däremot även här.

– I teorin kan såväl utvecklare, de som fattar beslut om att ett system ska användas, befälhavaren som planerar ett anfall och operatören som genomför det hållas ansvariga för internationella brott, såsom krigsförbrytelser. Men det krävs normalt att uppsåt kan bevisas. För statsansvar är beviskraven lägre men där finns andra processuella svårigheter för ansvarsutkrävande, säger Anna Andersson.

Det mänskliga ansvaret i folkrätten innebär att alla som hanterar militära system med AI behöver utbildning och behöver förstå hur de förväntas fungera, vilka begränsningarna är och hur de får användas, påpekar hon.

Natos AI-strategi fastställer sex principer för ansvarsfull användning av AI: laglighet, styrbarhet, tillförlitlighet, minimering av partiskhet, ansvar och ansvarsskyldighet, liksom förklarbarhet och spårbarhet.

Med mer AI och autonomi i de militära systemen är det extra viktigt att teknikbolagen tar hänsyn till folkrätten tidigt i utvecklingen och att systemen testas ordentligt.

– AI och autonomi formar när och hur bedömningar och beslut kan göras. Därför behöver utvecklare tekniskt möjliggöra att de som planerar och genomför anfall kan ta del av relevant information och göra nödvändiga bedömningar, säger Anna Andersson. □

## »Vi gör inte vad vi borde göra för att kontrollera osäkerheten. Vi har släppt ut generativ AI i världen utan den vanliga testningen och evalueringen.«

ett ordentligt säkerhetsarrangemang.

Efter att ha lämnat flottan och forskat på drönare i ett tiotal år, började hon 2013 att fokusera på självkörande bilar och interaktionen mellan människan och tekniken. Sedan dess har hon offentligt kritiserat påståenden kring självkörande bilars kapacitet och säkerhet, inte minst sådana påståenden från Tesla och Elon Musk.

Det har gjort henne till en måltavla för hårdföra Tesla- och AI-fans. När hon 2021 blev utnämnd till säkerhetsrådgivare på National Highway Transportation Safety Administration fick hon utstå online-trakasserier och dödshot.

Professor Missy Cummings är systemingenjör och chefar nu över Mason's Autonomy and Robotics Center vid George Mason University. Hon vänder sig mot att vissa vill stämpla henne som en tech-pessimist.

– Det jag vill är att teknologi används på ett säkert sätt.

### ”Linjär algebra på steroider”

Hon har undersökt orsakerna bakom de olyckor med självkörande bilar som har skett i USA. Hon har också agerat expertvittne i rättegångar, mot bland annat Tesla, efter allvarliga krockar. Sådana exempel använder hon nu för att illustrera svagheterna med AI och riskerna med autonoma vapensystem.

För det första, menar hon, finns det ingen intelligens i AI. Så kallat AI-baserat datorseende, som är så viktigt i autonoma fordon, handlar i grunden bara om mönsterigenkänning och sannolikhetsberäkningar.

– Det är som linjär algebra på steroider, säger Missy Cummings.

Hon jämför med ett litet barn som

kan förstå att en lastbil är en lastbil oavsett om den kommer i olika slags leksakformer eller riktiga former. Den datorseende AI-modellen kan bara identifiera en lastbil korrekt om den tillräckligt liknar de bilder och vinklar av lastbilar som den har tränats på.

Av den anledningen körde en självkörande bil av märket Cruise rätt in i baken på ledbuss i Kalifornien 2023. Det så kallade dragspelet i mitten av bussen gjorde att AI:n inte räknade ut att bussen hade en bakre del.

Bilen hade visserligen en laser-radar – lidar – på taket, och skulle med millimeterprecision kunna ha upptäckt bussens position framför. Men lidaren skulle bara kopplas in vid osäkerhet i datorseendet – och det var lika dåligt på att identifiera sin felaktiga beräkning av bussen som Chat GPT är när det ger ett välformulerat men felaktigt svar på en fråga.

### Okontrollerbar osäkerhet

”Hallucinationer” sker både i språkmodeller och i modeller för datorseende eftersom de är uppbyggda som djupa neurala nätverk och ska kunna generera ny data.

Hallucinationer tros ligga bakom fall av så kallad spökbromsning, som när en självkörande Tesla plötsligt tvärbromsade och skapade en allvarlig seriekrock i Baybridge-tunneln mellan San Francisco och Oakland år 2022.

Varför Teslan bromsade går inte att säkerställa. Missy Cummings teori är att det – just då, just där – fanns skuggor och reflektioner på vägen som AI:n matchade med mönstret av en stoppskylt.

Att inte i efterhand kunna ta reda

på varför AI:n har agerat på ett visst sätt är det svartlåde-problem som är inbyggt i modeller som bygger på djupa neurala nätverk, påpekar Missy Cummings. Likaså att samma ingångsvärden – samma fråga till en språkmodell eller fysiska situation för ett datorseende fordon – kan leda till olika respons. Det innebär en okontrollerbar osäkerhet.

Självkörande bilar använder agentisk AI. Det är olika AI-baserade modeller som samverkar med varandra så att till exempel olika sensordata analyseras och leder till ett autonomt beslut att utföra en manöver. AI-agenter kan hjälpa till att kontrollera varandra för att öka tillförlitligheten, men samtidigt kan riskerna i systemet växa på grund av osäkerheter i varje enskild AI-modell.

När agentisk AI används för vissa administrativa arbetsuppgifter är dess opålitlighet ett mindre problem. Men i system som kan framkalla fara för liv och hälsa innebär det en oacceptabelt hög risk, menar Missy Cummings.

– Vi gör inte vad vi borde göra för att kontrollera osäkerheten. Vi har släppt ut generativ AI i världen utan den vanliga testningen och evalueringen.

Vad gäller vapensystem bedömer Missy Cummings att de skulle kunna ha hög tillförlitlighet mot statiska mål, som broar eller järnvägar, om systemen tränas med maskininläring, men inte med djupinläring i neurala nätverk.

– Vi kan träna med bilder på specifika broar, där målet inte rör på sig. Men vad gäller dynamiska situationer säger jag: Nej, nej, nej! Vi är inte ens i närheten av ett AI-system som kan hantera en dynamisk miljö, särskilt inte där människor rör sig.

### Beräkningar utan omdöme

Hon tar upp ett tredje exempel från de självkörandes bilarnas USA. Det handlar om en kvinna som korsade ett övergångsställe mot rött på kvällen. Den självkörande bilen av märket Cruise beräknade att den vanliga

Nissan i filen bredvid skulle kollidera med kvinnan redan åtta sekunder innan det skedde.

En människa skulle med den informationen säkerligen ha tutat och bromsat in för att uppmärksamma Nissan-föraren och skydda kvinnan. Men Cruisen beräknade att hon skulle hinna förbi dess egen fil, och gasade då bara på i stället. Den beräknade inte möjliga följder av en kollision.

När Nissan sedan krockade med kvinnan – men relativt lätt – hamnade hon tillbaka i filen där Cruisen kom åkande. När den då bromsade tvärt var det för sent för att undvika att köra över henne.

Men Cruisen saknade sensorer under bilen och föreställningsförmåga att uppfatta att kvinnan hamnat där. Felaktigt klassificerade den också krocken som mindre allvarlig och körde därför in till väggkanten – med kvinnan släpandes under bilen. Det här blev slutet för de General Motors-ägda Cruise-bilarna i Kalifornien.

### Högt risktagande

En del av problemet, menar Missy Cummings, är att en stor del av de som driver AI-utvecklingen är unga, tekniskt skolade personer med relativt lite kunskap och erfarenhet av teknikers påverkan på människan och samhället. I dem möter hon också ungdomens oreflekterade självsäkerhet, och underskattning av äldre kunskap och erfarenhet. De som leder företagen har å sin sida fokus på att leverera bra investerings- och försäljningssiffror, snarare än på att se till hela samhällets välfärd och säkerhet.

Missy Cummings sätter inte mycket hopp till att USA kommer att ställa mer krav på säkerhet och testning av AI-system i närtid. Biden-administrationen var inte mycket bättre än Trump-administrationen i den frågan, enligt henne. Det finns en press på och från politiker att snabbt utveckla och använda AI i samhället, men kunskapen om hur AI-modeller fungerar och inte är för låg, menar hon. Säkerhet och testning ställs upp som motsats till innovation.

– Jag tror inte att det kommer att förändras förrän vi till exempel råkar döda egna Nato-trupper av misstag.

Trots motvind och dödshot fortsätter Missy Cummings sin frispråkiga kamp för en mer säker utveckling och användning av AI. Hur orkar hon? undrar en seminariedeltagare på Sipri.

– Jag har den 50-åriga kvinnans superkraft. Samhället avvisar dig då hur som helst, så jag bryr mig helt enkelt inte, säger hon – bara halvt på skämt – och fortsätter sedan med mer allvar.

– Det handlar om att vara evidensbaserad. Förnuftiga människor kan ändra sig när de ställs mot fakta. Ingen vill ha död på sina händer.

Militären har å andra sidan ett större utrymme än civila aktörer att legitimera ett stort risktagande med människors liv i vissa situationer, påpekar Missy Cummings.

– Men jag försöker hindra att människor använder tekniken bara för att de har blivit tillsagda att den är pålitlig, när den inte är det. □



»Det handlar om att vara evidensbaserad. Förnuftiga människor kan ändra sig när de ställs mot fakta. Ingen vill ha död på sina händer.«

# KULTURKROCK NÄR AI TESTAS PÅ VÄRLDEN

**AI kan påverka fred och säkerhet i världen. Men de som utvecklar tekniken är inte alltid medvetna om det. De kommer inte heller från en kultur av att säkerhetstesta tekniken ordentligt innan den lanseras. Därför jobbar Vincent Boulanin för mer ansvarsfull innovation inom AI.**

Text: Kajsa Skarsgård

**D**et finns en skillnad mellan utvecklare i mjukvarubranschen och ingenjörer bakom fysiska system som kan påverka liv och hälsa. Det menar Vincent Boulanin som är chef för programmet Governance of AI på Stockholm International Peace Research Institute (Sipri).

Ingenjörerna i fråga har sedan länge former för rigorös och spårbar testning innan en ny teknik sätts i bruk. Mjukvaruutvecklare lanserar produkten till användarna i ett tidigt skede för att sedan förbättra den längs vägen.

– Jag skulle sammanfatta det som en kulturkrock. Det är därför också viktigt att beställare av tekniken ställer krav på utvecklarna och ser till att användarna vet vilka begränsningar som ett AI-system har.

Säkerhetsfrågorna inom AI har fått alltmer fokus. Det kan däremot inte mäta sig med trycket på att snabbt utveckla och lansera senaste AI-mjukvaran och tillämpningarna. I USA spenderar de ledande AI-labben varje dag vad som motsvarar årsbudgeten för de externa organisationerna för forskning om AI-säkerhet, enligt rapporten State of AI 2025.

## Ansvarsfull innovation

Tillsammans med FN-sekretariatets avdelning för nedrustning (UNODA) har Vincent Boulanin tagit fram en handbok för ansvarsfull innovation inom AI. Den konstaterar att AI-utvecklarna sällan har tillräcklig kunskap om hur deras teknik kan komma att påverka internationell fred och säkerhet.

Bristande säkerhetstänk hos utvecklarna kan medföra svagheter i AI-verktyg som sedan används i säkerhetskritiska system. Med relativt lite specialistkunskap och resurser kan AI också användas för olika säkerhetshotande aktiviteter, som desinformation och cyberattacker. Tröskeln kan sänkas till möjligheterna att skapa kemiska och biologiska vapen.

Missförstånd kan leda till en upp-

trappning av konflikter som snabbas på av AI:s möjligheter. Det finns också en risk att länder som upplever att de inte hänger med i den militära AI-kapprustningen i stället vill satsa på kärnvapen. Vincent Boulanin förespråkar därför också klassiskt arbete för ökad vapenkontroll och säkerhet: transparens och dialog.

– Det som Sverige och andra länder kan göra är att prata om hur man kan reducera risker och främja dialog mellan de stora makterna för att undvika att man misstolkar varandras intention och kapacitet. Man kan enas kring vissa regler och normer även om man inte riktigt litat på varandra.

– Det är vinter inom vapenkontroll nu, men efter vintern kommer våren. Man måste jobba långsiktigt för att bibehålla relationer tills dagen kommer då det blir en ny styrning i USA, Kina eller Ryssland. □

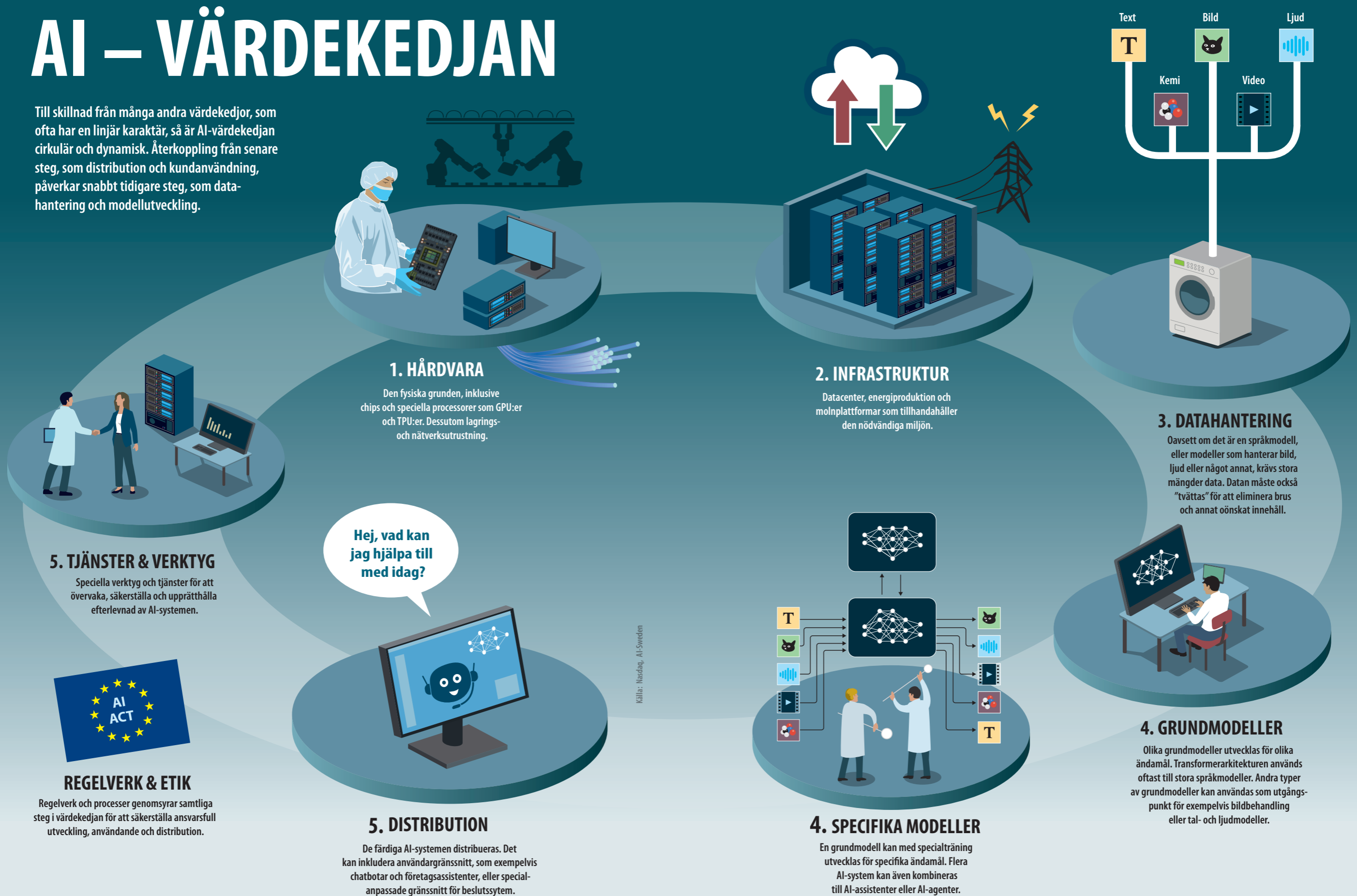


Vincent Boulanin.

**»Det är vinter inom vapenkontroll nu, men efter vintern kommer våren. Man måste jobba långsiktigt för att bibehålla relationer tills dagen kommer då det blir en ny styrning i USA, Kina eller Ryssland.«**

# AI – VÄRDEKEDJAN

Till skillnad från många andra värdekedjor, som ofta har en linjär karaktär, så är AI-värdekedjan cirkulär och dynamisk. Återkoppling från senare steg, som distribution och kundanvändning, påverkar snabbt tidigare steg, som datahantering och modellutveckling.



# Framtidsspaning

1. Hur tror du att AI har utvecklats om tio år?
2. Hur kan det påverka Sveriges totalförsvaret?



## Amy Loufti

Professor och programdirektör för Wallenberg AI, Autonomous Systems and Software Program (Wasp)

**1.** – Vi kommer att se mer AI som interagerar med AI, och interaktion mellan människa och AI i mer av ett lagarbete. Robotar och autonoma system kommer kunna ta del av väldigt avancerad mjukvara och också bidra till avancerad mjukvara. Jag tror att vi kommer att acceptera att det inte går att förklara allt som AI gör, och utnyttja detta som en typ av kryptering. Vi kommer att ha stora modeller för fler typer av uppgifter och se en viss nivå av genombrott kring hur man bygger neurala nätverk så att de inte kräver lika mycket el.

**2.** – Vi behöver upprusta på ett helt nytt sätt, en mycket mer osynlig upprustning där en del är det digitala och AI. Vi behöver viss suveränitet och förmåga att träna egna modeller. Det kräver fysisk infrastruktur. Vi behöver en ny generation av maskiner, drönare och robotar med avancerad AI. En sårbarhet är förgiftning av data som korrumpierar hela AI-systemet. AI kan samtidigt vara en fantastisk möjliggörare för att detektera dataintrång och för logistik.



## Daniel Akenine

Nationell teknikchef, Microsoft Sverige

**1.** – Tio är än väldigt lång tid när det gäller AI och det är omöjligt att säga med större säkerhet, men man kan titta på olika scenarier. Om utvecklingen fortsätter mot alltmer generaliserad AI som får större möjligheter att självständigt agera i vårt samhälle så tror jag att vi kommer möta AI-system i nästan allt vi gör. Om tio år så kommer vi troligen också ha mer humanoida robotar med avancerad AI som hjälper till i våra hem. En utveckling som redan börjat.

**2.** – Ett teknisksamhälle som är mer autonomt och distribuerat är svårare att påverka än ett centraliserat. Bygger vi in alltför centraliserad kontroll över våra AI-system så riskerar det förstås bli sårbart. Det positiva idag är att vi förstår dessa principer kring resiliens och sårbarheter, det gäller bara att följa dem.



## Kristian Rönn

Vd och grundare av AI-säkerhetsföretaget Lucid computing

**1.** – Jag tror att det finns 50 procents risk att vi har nått den tekniska singulariteten med okontrollerbar superintelligent AI om tio år. Risken är då stor att samhället styrs av AI. Människorna kanske lever som möss i en storstad – det finns välfärd nog för att ge oss mat, men vi bor i ett teknologiskt landskap som vi inte riktigt förstår oss på. Jag tror det är 15 procents chans att vi i stället får ett globalt avtal mot att träna självförbättrande superintelligens. Det enda sättet jag tror det går att kontrollera är genom att kräva begränsningar i hårdvaran, i chipet.

**2.** – Om man har större chans att vinna krig när man outsourcar allt beslutsfattande till AI är incitamentet för det starkt. Jag tror alla tankar om att ha en människa i loopen kommer ha försvunnit. Men jag vet inte om Sverige kommer vara relevant överhuvudtaget. Vi kan inte försvara oss mot en stormakt som använder kraftfull AI. Precis som med kärnvapen är vårt bästa försvar diplomati och globala avtal.

Foto: Elisabeth Ingvar



## Claudia Olsson

Vd och grundare av Stellar Capacity

**1.** – Inom tio år tror jag att vi kommer ha mer tillgängliga och agentliknande AI-system som både automatiserar arbetsflöden och assisterar människor i komplexa uppgifter; de kommer att jobba sida vid sida med oss, frigöra mycket innovation och förbättringar av olika processer och fungera som beslutsstöd. Samtidigt måste vi säkerställa tydlig ansvarsfördelning, mänsklig tillsyn, rättvisa, transparens och skydd för dataintegritet.

**2.** – För totalförsvaret innebär detta att vi får en mer informerad och realtidkopplad försvarsorganisation där rätt data når rätt person eller system vid rätt tidpunkt vilket kraftigt stärker lägesbilder och beslutsfattande. Samtidigt måste vi aktivt hantera nya sårbarheter såsom skydd mot AI-drivna hot och effekterna av hybridkrigföring för att säkerställa att systemen inte blir en väg in för motståndare att utnyttja.



## Martin Hagström

Forskningsledare på FOI och programansvarig för autonomiprogrammet

**1.** – Obemannade system har funnits länge men med hjälp av AI blir det alltmer autonoma. Effekter av den civila utvecklingen ser vi framför allt i fordonsindustrin med alltmer självkörande fordon som efter några decennier nu börjar introduceras i trafiken. Autopiloterna för flygplan får mer kapacitet och kan hantera mer komplexa situationer, på samma sätt som fartyg, både under och på ytan kan lösa uppgifter utan mänsklig inblandning. Inom tio år kommer sannolikt fordon också delvis kunna röra sig autonomt i terräng och andra ostrukturerade miljöer.

**2.** – Autonoma system används idag både för spaning och i attackdrönare. De spelar redan idag en viktig roll i militära tillämpningar och de kommer bli fler och mer avancerade. Det kommer vara möjligt att styra fler med färre operatörer. Om tio år kommer det även finnas motsvarande system för att motverka dem. Autonoma försvarssystem som agerar snabbt och precist för att möta attackdrönare och slå ut spaningsdrönare på långa avstånd.



## Joel Brynielsson

Forskningschef på FOI med ansvar för FOI:s AI-program och program för underrättelse- och säkerhetstjänst

**1.** – AI-system kan då i realtid med hög träffsäkerhet analysera och klassificera stora datamängder bestående av en blandning av text, bild, ljud och video. Generativ AI har mognat till en nivå där syntetiskt material av alla typer enkelt kan produceras direkt med korta instruktioner och blir nästintill omöjligt att särskilja från äkta material. Ett fåtal kraftfulla AI-grundmodeller täcker de flesta underrättelsebehoven och möjliggör automatiserade, skalbara och detaljerade analyser.

**2.** – Den ökade tillgången till perfekt eller nästintill perfekt fejkat material gör det möjligt att agera offensivt i informationsmiljön på ett alltmer förnekbart sätt. Att kunna analysera och klassificera stora mängder data kommer att bli alltmer centralt. Sammantaget ställer detta höga krav på kunskapsuppbbyggnad och AI-kompetens inom det svenska underrättelsesamhället.

Text: Kajsa Skarsgård

# AUTONOM NAVIGERING I TYSTNAD

AI har förverkligat idén som Gustaf von Grothusen fick under sina många timmar i stridsbåt 90. En maskin som kan navigera själv och gör det lika tyst som en människa med sjökort. Ihopsatt med ett styrsystem i en ytgående drönare vill han skapa autonoma farkoster som kan produceras i stor mängd.

Text: Kajsa Skarsgård  
Foto: Tomas Ärlemo

**T**vå affischer är uppsatta med svart eltejp på väggen i kontorshotellet i Lund där start-upen

Polar Mist hyr in sig. Den ena affischen visar flaggorna som vajar på Självständighetstorget i Kiev, en för varje stupad soldat. Den andra är en bild av det ukrainska frimärket där en soldat pekar finger åt ett ryskt krigskepp.

Kring en av datorskärmarna i rummet har tre av kollegorna samlats. Samtalet är koncentrerat. På skrivbordet ligger ett block med anteckningar

och matematiska funktioner ner-skrivna. Bakom männens ryggar står en egenbyggd ytgående drönare på en hylla. Det är P1, Polar Mists första prototyp.

Samtalet kring datorskärmen handlar om hur de ska optimera beräkningarna av hur två bilder tagna utifrån lite olika vinklar förhåller sig till varandra. Resultatet talar om hur kameran har förflyttats. När kameran sitter på en drönare blir beräkningen en av pusselbitarna för att farkosten ska kunna köra och navigera autonomt, utan GPS som lätt går att störa ut.

Idén till att utveckla autonoma sjödrönare som navigerar kommunikationstyst började gro i Gustaf von Grothusen för åtta år sedan. Då var han båtchef på en stridsbåt 90.

– Vi navigerade helt manuellt på grund av Rysslands störningsförmåga. Det tar ett år att lära sig navigera och navigeringen tar 90 procent av den kognitiva kraften när vi är ute på övning eller operation.

– Jag tyckte inte att det var svinkul att ligga i terrängen i februari och började tänka på hur man kunde få maskiner att automatisera navigeringen. Datorer fryser liksom inte, säger den gänglige 29-åringen.

Gustaf von Grothusen tog med sin idé när han med visst kval lämnade Försvarsmakten för att i stället läsa teknisk fysik i sin hemstad Lund. När Ryssland sedan gick in i Ukraina med full kraft hade han djupa funderingar på att åka dit och strida.

– Det slutade med att, nej, jag vågar inte. Jag vill verkligen inte det

*»Jag tyckte inte att det var svinkul att ligga i terrängen i februari och började tänka på hur man kunde få maskiner att automatisera navigeringen. Datorer fryser liksom inte.«*

här. Jag vill inte dö i ett krig, i helvetet på jorden, i en skyttegrav av nån FPV-drönare. Men jag tänker mycket på att ifall det skulle smälla här, då skulle jag kunna ligga på någon kobbe ute i skärgården och göra det ukrainarna gör nu. Och förmodligen stryka med för det är så otroligt höga risker på amfibiegementet.

– Det förhöjer ju hjälteskapet av det som ukrainarna har gjort. De har

ju gjort de uppoffringar som jag inte vågade göra. Så jag måste göra allt annat jag kan göra än att åka ner och strida.

Gustaf von Grothusen var med och grundade Drones2Ukraine och har själv lämnat över insamlade drönare vid ukrainska gränsen. 2023 började han och medgrundaren av Polar Mist att utveckla AI:n för autonom och ostörbar navigering för sjödrönare. Tanken var



Med hjälp av AI som körs lokalt på ytfarkostens dator kan den navigera autonomt utan stöd av GPS.

Foto: Polar Mist

först att bara utveckla mjukvaran. Nu bygger de också allt från kretskorten till farkosten. På så sätt kan de ta ansvar för att delsystemen fungerar tillsammans och har mer kontroll över avvägningar och anpassningar.

## Från garage till egen verkstad

Fram till sommaren 2025 skedde utvecklingen av AI:n i Gustaf von Grothusens vardagsrum. Lödningen av kretskorten i sovrummet. Båtbygget i hans farbrors garage. Men sedan två riskkapitalister gick in med 12 miljoner kronor i bolaget har de en fristående verkstad, rummet på företagshotellet där fem personer sitter och kodar, och ytterligare ett rum där elektroniken byggs ihop och 3D-printern går varm.

– Jag och mekatronikchefen Jesper brukar köra en tävling i veckan om vem som är snabbast på att löda. För tillfället leder jag, säger Gustaf von Grothusen.

Under andra halvan av 2025 har inte bara personalstyrkan utökats och företaget fått nya lokaler. De har också hunnit utveckla tre versioner av en båtprototyp. Den senaste är P3:an, en fem meter lång obemannad ytfarkost som snart är redo att sjösättas.

– P1 och P2 var bara delsteg. P3:an är slutmålet.

I bilen på väg till verkstaden där P3:an står, berättar Gustaf von Grothusen mer om deras AI-system. I grunden använder det bildigenkänning utifrån den data som drönarens 360-graderskamera tar in.

– Vårt AI-system är ett hopkok av det som är lätt för datorn men svårt för en människa att uppfatta, till exempel att vinkeln i ett sund är 36 grader. Vi försöker även ta tillvara på de viktigaste grejerna av det som en människa använder för navigering, som var det finns öar och var det finns viktiga landmärken.

Utöver att mäta vinklar och identifiera punkter byggs AI:ns positionering upp av jämförelser med data från öppna kartor. För att förfina navigeringen har systemet enklare tröghetsensorer som registrerar farkostens rörelse. Data om gaspådrag och rodrets vinkel är också viktiga för lägesberäkningen (se illustrationen på s. 39). ▸

Beräkningarna görs i realtid i drönarens dator för att inte avge någon signal som kan upptäckas eller störas ut. Att satsa på relativt beräkningslätt AI, liksom billiga sensorer och hårdvara, är ett strategiskt beslut för att enkelt kunna skala upp antalet farkoster.

– Vi vill bygga en plattform som är så flexibel som möjligt, för att kunna ha hastighet när det smäller. För då kommer vi behöva anpassa oss, och fienden kommer anpassa sig till våra anpassningar. Ifall man har någonting där man kan slänga i lite vad som helst och göra små anpassningar kan man fokusera på att innovera taktiskt. Men oavsett vad måste man kunna lokalisera sig.

### Kontroll i fokus

Skalbarheten har blivit ännu viktigare sedan Ryssland börjat använda spaningsdrönare med termiska kameror som kan detektera värmen från ytfarkosternas motorer på långt avstånd, berättar Gustaf von Grothusen.

– Med skala får man mycket större handlingsfrihet, för då behöver man inte välja operation A, B eller C, utan man kan göra alla tre. Den nästan viktigaste poängen med AI och autonoma system är att det medger så stor skala.

I dagsläget har Polar Mists prototyper en kamera för dagsljus, men att sätta på en mörkerkamera är bara en ekonomisk fråga. En radar skulle kunna säkra upp för datorseendet om dimman är för tät eller om drönaren ska köras ut på hav bortom synligt land.

– Det går inte att bygga ett system som alltid funkar, oavsett väder, telekrigföring eller plats. Men om man har kontroll på när det funkar och inte, så kan man planera utefter det och vara säkra på att man kan följa planen utan att tappa bort sig och börja köra i cirklar halvvägs.

Av kontrollskäl använder Polar Mist inte helt självlärande AI för navigeringen.

– Vi åker ut och samlar data och testar. Om vi ser att den kör sicksack och väljer en helt skev väg då tänker vi ut med våra hjärnor hur vi kan träna om AI:n så att den blir vettigare. För om man ger den frihet är risken att den börjar hitta på grejer som inte alls är bra.

– I takt med att grejerna mognar och

man får koll på exakt vilka friheter man kan ge den kan man göra ett system som är mycket mer självlärt. Men vi känner inte heller att det riktigt finns ett behov av det, i alla fall inte där vi är just nu. Det är bättre att ha stenkoll på vad den kan och sedan använda den på ett sätt som man vet att den klarar.

### Både ny och gammal kunskap

Framme vid en utåt sett anonym verkstadslokal i ett industriområde möter Gustaf von Grothusens farbror upp. Han utbildade sig till ingenjör före datorernas tid, men har till skillnad från kodknackarna i företaget mekat sedan barnsben och byggt egna båtar. Han har gjort 2D-ritningarna till P3:an som kollegan Hugo sedan har lagt in i ett digitalt konstruktionsprogram (CAD).

– Jag såg framför mig att stå här med snickarverktyg, men det är ganska kul med CAD och datorer. Vi skickar ritningarna till en firma som fräser hela grejen, och sen skickar de den och gjuter. Från början tänkte vi att vi kör plywood och spackel, men den här har bara 0,1 mm som största fel, säger farbrodern och pekar på P3:ans grå glasfiberskrov.

Han vill inte omskrivas med sitt namn, men berättar att han själv har varit båtchef inom kustjägarna. Med sig in i det unga bolaget har han inte bara ett liv av erfarenhet av båtar utan också ett lokalt nätverk av fixare. Han vet vilka gubbar som svetsar bäst och vilken firma som har de bästa sladdarna.

– Jag skickade killarna till glasmästaren för att fixa plastskivor. Ingenting hände. Då ringde jag igår, berättar han och imiterar glasmästarens svar på bred skånska:

– "Jaså, är det du? Fan! De kan komma om en halvtimme!"

I fören har P3:an lastutrymme. I akterutrymmet ligger motorn som kan driva farkosten i upp till 50–60 knop. Gustaf von Grothusen visar den lilla svarta lådan där datorn finns. Den som både kör båten och navigerar autonomt.

– Vi brukar säga att det alltid är grejer som havererar och går fel, med hårdvara och med mjukvara. Så för att kunna hålla hastighet måste man vara som en iller kring alla risker. Ha mindsetet att vi ska

lösa det här till varje pris, inte vara tillbakalutad och bara gå hem när man ska gå hem, säger han och väcker skratt från farbrodern och mekatronikchefen Jesper.

"Vi köttar på" är en återkommande fras från Gustaf von Grothusen. Han jobbar kvällar och helger och det är tydligt att hans medarbetare förväntas jobba med liknande brådska och kall.

– Vi har väntat på att P3:an ska bli klar och arbetat stenhårt. På måndag är det äntligen dags att sjösätta den.

I väntan på det åker de ut till en sjö för att visa upp körförmågan i en liten replika av P3:an som de byggde för att testa designen innan de byggde den i fullformat. Farbrodern står på en brygga med fjärrkontrollen i händerna och ser nöjt på när drönaren skär genom den stilla vattenytan i snabba vändningar. De tjuvpratande ändarna som hängde i vattenbrynet har skrämmts bort, på denna idylliska plats känns kriget långt borta. ▷



### Bildertexter nästa sida.

Överst t.v. Operationsplanering i det egna lednings- och kontrollsystem.

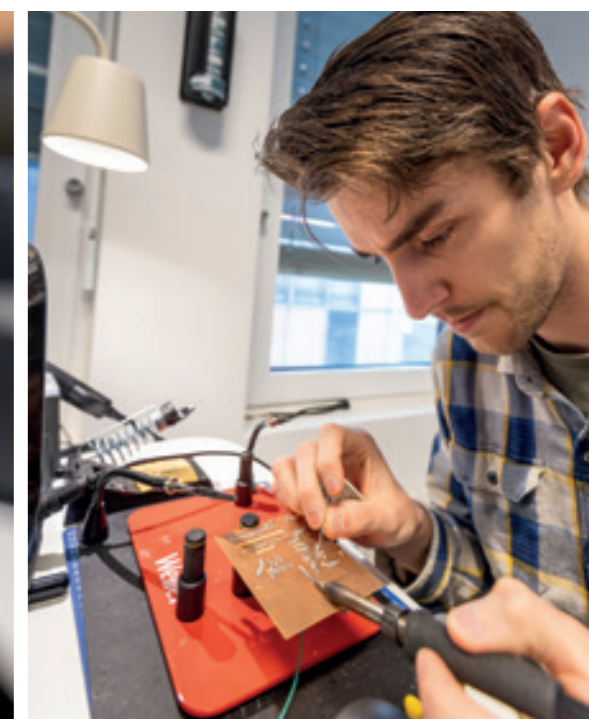
Överst t.h. Gustaf von Grothusen.

Mitten t.v. Gustaf von Grothusen går igenom P3:ans elektronik.

Mitten t.h. Gustaf von Grothusen och kollegan Jesper med P3:an som snart ska sjösättas.

Nederst t.v. Första prototypen av ett inbäddat positioneringssystem.

Nederst t.h. Bolaget konstruerar och testar prototyper av kretskort med olika uppgifter innan större serier beställs.





I väntan på att P3:an i fullformat ska bli helt klar visar Gustaf von Grothusen och mekatronikchefen Jesper upp miniatyrmodellen av den.

**En cirkel sluts**

I Gustaf von Grothusens tankar är Ukraina, kriget och döden ändå alltid nära. Där på bryggan ger hans farbror ytterligare en förklaring till brorsonens driv.

– Hans farfar, min pappa, var båtflykting från Riga. Ingen sjuåring ska behöva fly i båt för att Ryssland skjuter på en.

Så nu bygger båtflyktingens barnbarn egna båtar, som i första hand ska kunna navigera autonomt för att skydda Sveriges gränser och demokrati. Men för Gustaf von Grothusen är en självklar utveckling att farkosterna i framtiden också ska kunna bära vapen och verka i svärmar i stor skala. Hur tänker han då om de etiska riskerna med ett autonomt vapensystem, folkrätten och sitt eget ansvar som utvecklare?

– Motståndaren håller på och utvecklar alla möjliga hemskheter skrupelfritt. Då tänker jag att det är bättre att vi utvecklar de här grejerna och har ett luddigt svar på de frågorna än att inte utveckla några grejer alls, och så står vi där

med byxorna nere sedan när det smäller. Vi bygger det vi bygger för att vi försöker ta vårt medborgerliga ansvar.

Gustaf von Grothusen ser det som utvecklarens ansvar att vara tydlig med sin produkts begränsningar. Att ge autonoma vapensystem ett "kort koppel" – begränsad frihet – tror han är en nyckel till ansvarsfullt användande.

– I nio fall av tio kanske man kräver en mänsklig avdömning där och då, men om det råder fullt krig och vi håller på att förlora kanske man kan leva med tio procents risk att systemet gör fel.

AI-utvecklingen har gjort att Polar Mist kan göra mycket av träningen av AI-systemet i simuleringar. Men verkligen på sjön är mer komplex. Därför krävs också testning och AI-träning i verkligheten.

Hur gick det då veckan efter när P3:an i fullformat sjösattes?

– Det var meterhöga vågor och jag hade jätteångest. Men den går som ett skott och bara klyver sjön. Min farbror var säker på sin sak, men det här var över min

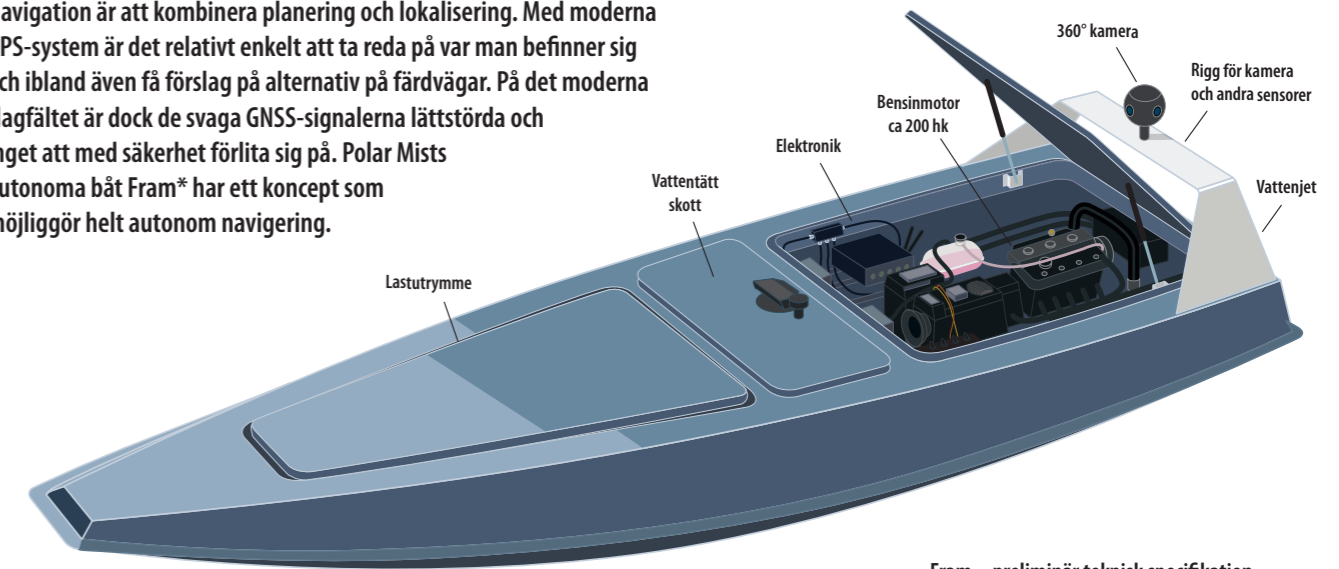
förväntan. Den tar sjön som en king.

Gustaf von Grothusen berättar detta över telefon från den svenska skärgård där hans pappa tog med honom på hans första segeltur bara ett halvår gammal, i segelbåten som varit hans farfars. Testerna på sjön pågår så länge vintersolen är uppe. När mörkret fallit löser de problemen som dykt upp under dagen och skickar data till kollegorna som tränar AI:n på kontoret i Lund.

– När jag såg att P3:an funkade ordentligt var det så jävla skönt och jag kände hur trött jag är efter att ha jobbat så här i tre år. Jag ville bara lägga mig ner och sova, och gråta av känsloplåslaget. Men vi motar tillbaka det och köttar på. □

# POLAR MIST – GNSS-FRI NAVIGERING

Navigation är att kombinera planering och lokalisering. Med moderna GPS-system är det relativt enkelt att ta reda på var man befinner sig och ibland även få förslag på alternativ på färdvägar. På det moderna slagfältet är dock de svaga GNSS-signalerna lättstörda och inget att med säkerhet förlita sig på. Polar Mists autonoma båt Fram\* har ett koncept som möjliggör helt autonom navigering.



Fram – preliminär teknisk specifikation

Toppfart	50 knop
Räckvidd	300 nautiska mil
Vikt	300 kg
Längd	5 meter
Bredd	1,5 meter
Lastförmåga	500 kg

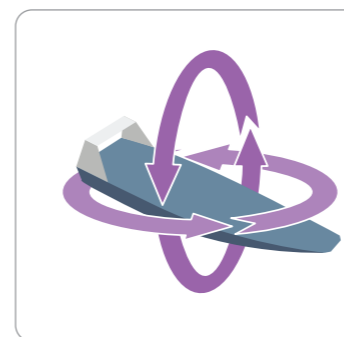
\* Fram är döpt efter det skepp som användes av norska polarforskare runt slutet av 1800-talet

**Planerare/navigatör**

Hjärnan i systemet är den så kallade planeraren. Det är ett AI-system som tar in och bearbetar information från tre delsystem.

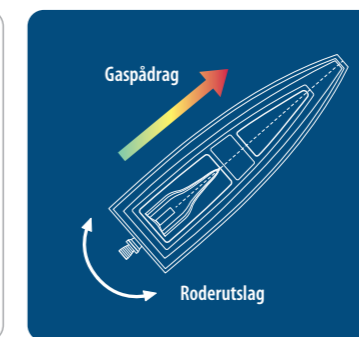
Planeraren kan också ge uppdaterade förslag på alternativa färdvägar.

**1. IMU – TRÖGHETSSENSOR**



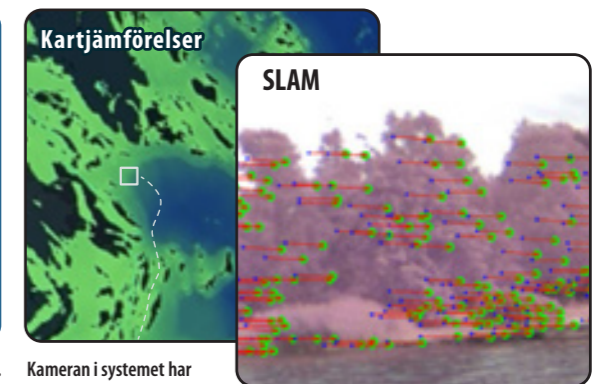
IMU-systemet använder gyron och så kallad död räkning för att estimerar hur farkosten rör sig över tid.

**2. KINEMATISK MODELL**



En intern kinematisk modell byggs upp i systemet. Här registreras viktiga parameterer som gaspådrag och roderutslag.

**3. KAMERA**

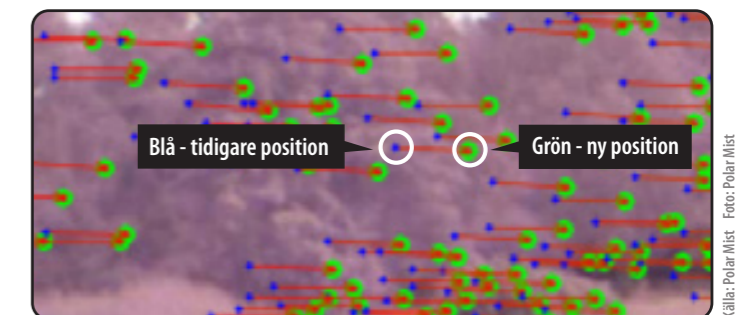


Kameran i systemet har två uppgifter, dels att göra okulär jämförelse mellan aktuellt läge och kartan, dels att förse det så kallade SLAM-systemet med kontinuerlig information.

**Simultaneous Localization and Mapping (SLAM)**

SLAM är en teknologi, där man med hjälp av kamera eller andra sensorer, kan läsa av olika egenskaper och landmärken i omgivningen och på så sätt bygga upp en intern karta. Därefter används den interna kartan, tillsammans med nya observationer, för att estimerar den egna för flyttningen och den nya positionen. Processen pågår sedan iterativt och nya landmärken läggs oavbrutet till den interna kartan.

Teknologin är en lösning på "hönan eller ägget-problemet" inom navigering, då ett system behöver veta sin position för att kartlägga ett område, men samtidigt behöver en karta att veta sin position.



# KONSTEN STÖRTA EN DEMOKRATI UTAN VAPEN

AI ger auktoritära aktörer helt nya krafter att undergräva demokratier genom målmedveten informationspåverkan. Slagfältet är människors världsbild och tillit till varandra, sanningen och demokratin. I det moderna informationslandskapet är människan extra sårbar.

Text: Kajsa Skarsgård

Ryssland tvingade ut Ukraina i krig först efter misslyckandet att med andra medel få ukrainarna att dela den ryska ledningens värderingar och mål. Andreas Edevald

flyttade från Ukraina hem till Sverige en månad före Rysslands fullskaliga invasion. Som kanslichef och säkerhetsansvarig på Sveriges ambassad i Kiev hade han lett förberedelserna för om "det värsta skulle hända". Men liksom många andra trodde han att det inte skulle ske.

Hans fru hade däremot redan året innan gjort en mer korrekt bedömning av läget med Ryssland och därför flyttat tillbaka till Sverige med deras barn.

– Hon blev orolig av den ryska uppbyggnaden i samband med militärövningen Zapad-21, berättar han.

I januari 2022 hade det blivit dags för Andreas Edevald att flytta efter familjen. Med flytten bytte Andreas Edevald jobb från Utrikesdepartementet till Myndigheten för psykologiskt försvar (MPF) där han är senior analytiker.

– Tittar jag tillbaka, och särskilt nu i min nya roll, så kan jag väl tycka att det var ganska tydligt att invasionen skulle ske. Men ändå så har man väl någon försvarsinställning så att man inte vill att det ska ske när man väl lever i det, även om man förbereder sig för det. Så det blev en väldig omställning den 24 februari när det väl hände.

Andreas Edevald har nyss skrivit rapporten *Janusansiktet* som handlar om kognitiva snedvridningar (eng. bias). Det är systematiska och ofta omedvetna avvikelser från rationellt tänkande som påverkar människors bedömningar och beslut. En typ av kognitiv snedvridning är tendensen att söka och ta till oss just den information som bekräftar de uppfattningar vi redan har, och att inte ompröva dessa när motstridande information kommer fram.

Kognitiva snedvridningar spelade en roll i att Rysslands invasion av Ukraina 2022 blev en överraskning för många – trots att USA och Storbritannien sa rätt ut att det skulle ske.

– Det som blev fel här var också att man analyserade situationen med

svenska ögon, inte ryska ögon. En lärdom för systemet i stort är att lyssna mer på de röster som kan Ryssland bättre, för du kommer landa fel i din slutsats om du om du analyserar Ryssland med svensk logik.

## Kognitiv krigsföring

Människors världsbild – värderingar och logiken med vilken vi tolkar information – har alltid varit ett slagfält i krig. Är invasionsstyrkan tyranner eller befriare? Är motståndet frihetskämpar eller terrorister? Är den liberala demokratin det vi ska skydda eller ett hot?

Nato har gått från att tala om informationskrigsföring till kognitiv krigsföring. Det är inte i huvudsak ett krig över informationen utan över våra sätt att tänka – och därmed agera.

– Jag är intresserad av hur den mänskliga hjärnan fungerar, säger Andreas Edevald.

– Vad är det som får oss att tro på saker? Hur kan olika typer av språkmodeller och AI-system användas för att få folk att tro på felaktiga saker?

Med AI-utvecklingen har möjligheterna att påverka människors informationsmiljö ökat drastiskt. AI kan skapa tusentals unika inlägg i sociala medier på några sekunder och fejkat innehåll som är svårt att skilja från äkta. Informationen kan med hjälp av AI anpassas för



Andreas Edevald, senior analytiker på Myndigheten för psykologiskt försvar.

att skapa maximalt engagemang hos varje individ, och med precision riktas mot nyckelpersoner. Algoritmer kan användas för att systematiskt polarisera och radikaliserar personer i olika politiska ekokammare.

Från psykologin vet vi att människan har en tendens att uppfatta vad som känns bekant som mer sant, berättar Andreas Edevald. Det kallas för illusorisk sanningseffekt. Att upprepa falska eller vilseledande budskap har därför effekt.

– Om du loggar in på någon form av socialt mediaflöde, där algoritmen är ▶

»Jag är intresserad av hur den mänskliga hjärnan fungerar. Vad är det som får oss att tro på saker? Hur kan olika typer av språkmodeller och AI-system användas för att få folk att tro på felaktiga saker?«

justerad efter vad du tycker och tänker i din lilla ekokammare, är det väldigt lätt att få intrycket att hela Sverige måste ju tycka det här just nu. Sanningen är att det inte alls behöver vara så. Även om du hör några högljudda röster så kanske inte ens hela din politiska falang tycker så. Det kanske bara är fem, tio personer, men algoritmerna förstärker det här ännu mer.

– Det blir viktigt att förstå det här för att förstå riskerna med AI. AI är minst lika bra, om inte bättre än människor, på att övertyga folk om idéer.

### Påverka språkmodeller

Informationskrigföringen är inte bara riktad mot individer, utan även mot infrastrukturen där vi hämtar information.

Det ryska desinformationsnätverket Pravda består till exempel av ett par hundra hemsidor i över 80 länder där pro-ryskt innehåll publiceras och sprids i rasande fart. En studie från Center for the Study of Democracy visade att den bulgariska Pravda-hemsidan publicerade i snitt åtta artiklar i timmen. Detta tyder på att artiklarna genereras med en hög grad av AI-automatik, och att syftet inte är att nå direkta läsare utan att plockas upp av sökmotorer, språkmodellernas träningsdata och AI-chatbotar.

Studien konstaterar att språkmodeller som har finjusterats för att användas på bulgariska kan vara särskilt sårbara för sådan här påverkan eftersom det finns en brist på bred och kvalitativ träningsdata på bulgariska.

Operation Overload kallas en rysk desinformationskampanj med syfte att bland annat överbelasta faktagranskare och etablerade medier med verifieringsarbete – och samtidigt få dem att sprida falska påståenden i viljan att motbevisa dem.

AI-verktyg används alltmer för att generera innehåll inom Operation Overload. Narrativen är främst anti-ukrainska eller har syftat till att påverka olika länders demokratiska valprocesser. Smutskastningskam-

panjer har särskilt riktats mot kvinnor och HBTQ-personer, främst transpersoner.

### AI för försvar

– AI går ju att använda i defensivt syfte också. Vi måste se till att vi själva är aktiva och tar tillvara på den möjligheten, säger Andreas Edevald.

MPF utvecklar till exempel sina metoder för att med hjälp av AI identifiera, följa och bemöta vad myndigheten kallar "otillbörlig informationspåverkan från främmande makt". Att folkbilda i källkritik är också en viktig del av arbetet.

MPF:s uppdrag är att "värna om yttrandefriheten och det öppna och demokratiska samhället". Myndigheten får därför inte få registrera vad som sker i den svenska informationsmiljön.

I det digitala landskapet finns däremot inte nationsgränser, och AI har nu rivit de språkhinder som tidigare fanns. Utländska påverkansaktörer försöker alltid utnyttja inhemska personer för sina mål – det är en fråga om resurseffektivitet. Men inhemska personer kan också av olika egna skäl vilja undergräva den liberala demokratin eller dela främmande makts narrativ.

### Demokratin på slagfältet

Den liberala demokratin vilar på idén om alla människors lika värde. Därför ska majoritetsstyret vägas mot ett starkt skydd för minoriteter och för individen. Självständiga och starka institutioner och medier ska finnas för att skydda fri åsiktsbildning, religionsfrihet, rättssäkerhet, ansvarsutkrävande och maktdelning.

I USA håller den liberala demokratin på att urholkas inifrån, av en folkvald president aktivt påhejad av en betydande del av befolkningen, techgiganter och algoritmer. För Danmark har USA på kort tid gått från nära allierad till en potentiell invasionsmakt efter hoten mot Grönland.

Ungern klassas sedan 2019 inte som en demokrati utan som en elektoral auktoritär regim, och agerar bromskloss i EU:s stöd till Ukraina.

Georgien håller snabbt på att gå i auktoritär riktning, pressat av Ryssland. Regeringsmakten i Polen tog under flera år ökad kontroll över domstolar och medier, men har börjat återställa en del av den demokratiska skadan sedan maktskiftet 2023.

Det finns en risk att inhemska hot mot den svenska demokratin missbedöms, genom liknande feltänk som gjorde att Rysslands hot mot Ukraina underskattades. Det menar Carl Heath som forskar om digital resiliens och demokrati på forskningsinstitutet Rise.

– Det som särskilt oroar mig är att vi ofta analyserar auktoritära rörelser utifrån en liberal demokratisk logik där vi förutsätter goda intentioner och en vilja att delta i ett gemensamt samtal. Men det stämmer inte.

– För auktoritära rörelser handlar det inte om sanningssökande eller demokratisk deliberation. Det handlar om makt, om att vinna, om att definiera verkligheten för så många som möjligt. Och i det arbetet är lögn, desinformation och systematiskt undergrävande av tillit till institutioner fullständigt legitima verktyg. Vi måste lära oss att se på dessa förflyttningar med nya ögon. Inte för att vi ska överge våra demokratiska värderingar, utan för att vi ska kunna försvara dem effektivare.

### Demokratisk försvarsvilja

Carl Heath ser också en risk för kognitiv snedvridning i bedömningen av hotet mot demokratin i Sverige.

– Vi kan se mönster av ökad polarisering, minskande tillit till media och experter i vissa grupper, systematiska kampanjer för att underminera förtro-



Carl Heath, expert på digital resiliens.

»För auktoritära rörelser handlar det inte om sanningssökande eller demokratisk deliberation. Det handlar om makt, om att vinna, om att definiera verkligheten för så många som möjligt.«

ende för myndigheter. Om vi avfärdar dessa som temporära fenomen eller marginella rörelser, riskerar vi att missa hur systemiskt sårbarheten faktiskt är.

– Skillnaden mot Ukraina är att detta är en gradvis process, inte en dramatisk invasion. Det gör det svårare att se, svårare att mobilisera emot. När går debatten från robust till farlig? När blir polariseringen destruktiv? När övergår legitim opposition till systematiskt undergrävande av demokratiska institutioner? Det är svåra frågor utan tydliga svar.

För att hantera gråzoner mellan legitim opinionsbildning och informationspåverkan brukar Carl Heath utgå från metoder snarare än innehåll och ideologi. Använder någon en massa fejkade konton? Sprids innehåll genom koordinerade kampanjer? Finns det mörka finansieringsflöden? Används systematisk desinformation? Utnyttjas sårbara grupper?

– Det är dessa metoder som är problematiska, oavsett om de kommer från inhemska eller främmande aktörer.

Sverige är sårbart i det moderna informationslandskapet, menar Carl Heath. Den höga graden av digitalisering i samhället skapar fler attackytor för informationspåverkan och cyberattacker. Det höga förtroendet för våra institutioner är i grunden stärkande men kan också göra oss naiva inför systematisk desinformation. Mediemarknaden är relativt liten och resurserna för grävande journalistik och faktakontroll är därför också små. Dessutom hotas nu mediernas annonsintäkter eftersom AI-sökningar inte genererar lika mycket trafik till

ursprungskällan som googlandet gör.

Sveriges styrka är befolkningens relativa höga digitala kompetens, landets relativt oberoende institutioner, en forskningsmiljö i världsklass, och inte minst en tradition av folkbildning och starkt engagemang från civilsamhället. Carl Heath mötte detta bland annat när han ledde en statlig offentlig utredning om hur det demokratiska samtalet kan värnas i vår digitala samtid.

– Då mötte jag engagerade människor från alla delar av samhället som verkligen ville bidra till att stärka demokratin. Den försvarsviljan, den viljan att stå upp för demokratiska värderingar, är vår största styrka.

– För att återgå till vad jag skrev då som särskild utredare, och ett citat från den så kallade Mossbergska utredningen om svenskt psykologiskt försvar från 1953: frontlinjen går i varje enskild medborgares medvetande. Det är där slaget om demokratin utkämpas, varje dag, i varje samtal, i varje delning på sociala medier, i varje val av vad man väljer att tro på och agera utifrån. □

# ALGORITMEN I BLODET



**I början av 1990-talet dök det upp en doktorsavhandling på avdelningen för ledningssystemteknik. Den var fylld med formler. Kunde möjligtvis den nydisputerade Johan Schubert förklara vad den handlade om? Det kunde han. Johan Schubert har algoritmen i blodet.**

*Text och foto: Jan-Ivar Askelin*

**A**vhandlingen var en rest från det kalla kriget som då nyss var över och Sovjetunionen hade försvunnit. Försvarets forskningsanstalt, FOA lade ned stora resurser på ubåtsspaning. Ett problem var att hantera alla osäkra uppgifter.

Det var här Johan Schubert kom in på scenen. Genom att sätta siffror på osäkra uppgifter och därmed kunna gruppera dem, så kallad klustring, kunde agnarna hjälpligt sällas från vetet. Det var det avhandlingen handlade om.

– Klustring är en del av maskininläring. Jag var en av de första på FOA som använde artificiell intelligens, AI, i forskningen. Det kan då även ha varit någon som studerat flygstrid, säger Johan Schubert.

Nu är Johan Schubert forskningschef för AI vid avdelningen för cyberförsvar och ledningsteknik.

– Här i Kista är forskningen till stor del inriktad på människor och på en högre systemnivå. I Linköping är det mer teknik och sensorforskning.

Nu växer hela Totalförsvarets forskningsinstitut (FOI) och extra fort växer det som rör it och AI. Vi söker såväl nyutexaminerade civilingenjörer som seniora forskare.

– Det beror förstas på teknikutvecklingen. Vi måste hänga med. Men det

beror lika mycket på att den 30-åriga mellanperioden av fred är slut. Så utvecklar vi igen taktik för ubåtsspaning. När Johan Schubert 2017 gjorde en översikt av AI-utvecklingen fanns det en handfull AI-projekt på FOI. Nu är det massor. En del forskar om AI och många använder AI i sin forskning, säger Johan Schubert.

När Johan Schubert tittar tillbaka säger han att den bild vi då gjorde av AI-utvecklingen hamnade ganska rätt.

– Det som vi inte riktigt fick med var de stora språkmodellerna. Det var vid den här tiden som de första rapporterna kom. Google visade vid en konferens hur man lyckats göra neuronäten snabbare och bättre. Detta var genombrottet. Den allmänna bilden av dagens AI-forskning är att det fortfarande handlar mycket om maskininläring ofta i kombination med språk.

## **Beslutsstöd till bataljon**

AI-utvecklingen kan mätas på olika sätt. Medier rapporterar om det heta som är färdigt. Sedan kommer presentationer på konferenser och vetenskapliga rapporter. För att få en bild av AI-utvecklingen får man försöka göra en avvägning.

Johan Schuberts enhet får främst uppdragen från Försvarmakten och Försvarets materielverk (FMV).

Den långsiktiga forskningen är organiserad i forsknings- och teknikutvecklingsprojekt.

– Ett sådant handlar om hur man hanterar, sammanställer och drar slutsatser av ostrukturerade data. Hur ska datorn förstå löpande text? Det är mycket svårare än att hantera ordnade tabeller.

– Ett annat projekt rör en massa olika tekniker som till exempel djupinläring och förklarbar AI. Vi har också projekt med taktikutveckling såväl för flygstrid, där vi samarbetar med flygvapnets simuleringscenter här i Kista, FLSC, och som nämnts med ubåtsspaning.

Flera stora projekt kommer från FMV, säger Johan Schubert och beskriver det ena om beslutsstöd.

– Hur man ger ett beslutsstöd före och under mekaniserad strid. I en mekaniserad bataljon prövas olika alternativ för att positionera styrkorna.

– Antalet alternativ kan teoretiskt bli en miljard miljard. Vi bantar ned det till 10 000 vilket man snabbt kan hantera på en dator, värdera och presentera för en beslutsfattare.

– Vi kan simulera ned till plutonnivå och summerar småbitarna till ett totalt utfall för striden. Vi kan köra 10 000 ▶

*»Antalet alternativ kan teoretisk bli en miljard miljard. Vi bantar det till 10 000.«*

»Om man kan göra något effektivare så kommer det en AI för det.«

olika varianter för att se vad som är bästa resultatet. Vi använder tabeller och får relativt snabbt fram ett svar.

Man kan också göra det mer högupplöst. Då tar man med varje stridsvagn, varje skott och varje position. Det tar timmar att få ett svar. Det är vad en analytiker frågar efter. Den snabba metoden går på sekunder.

**Perioden efter kalla kriget är slut**

Projektet är en utveckling av ett sexårigt projekt som startade för drygt 20 år sedan. Då var det en röd bataljon, fienden, som landsteg i Kapellskär och var på väg mot Norrtälje.

– I det projektet fick vi in en massa automatiska rapporter från sensorer som lade grunden till en automatisk lägesbild på alla nivåer. Men ett beslutstöd var det inte.

– Jag tyckte det saknades ett projekt som fokuserade på genomförandet av en militär operation. Veta läget, kunna planera och ge beslutstöd. Och det är det vi har nu.

– Mellanperioden efter kalla kriget är slut. Det är logiskt att man börjar med underrättelseperspektivet. Vad är läget, kan jag planera för framtiden? Och sen är det tripp, trapp, trull. Hur kan man agera under ett genomförande?

Det andra FMV-projektet beskriver Johan Schubert som mer innovativt. Här används en relativt ny metod, kallad aktiv inferens eller fri energi.

– En svärm drönare styrs automatiskt för att skapa en lägesbild. Drönarna väljer sina egna vägar och försöker likna en mänsklig hjärna för att minska osäkerheter i lägesbilden. Drönarna jagar hela tiden efter de vita fläckarna, fyller dem med information och bygger stegvis en karta. Metoden är automatisk. Den fungerar lika bra för små som stora drönare och för små eller stora ytor.

Metoden går ut på att först göra en modell som beskriver verkligheten och där lägger man sannolikheter. Så körs modellen i processen aktiv inferens. Där fattas besluten.

– Tekniken har utvecklats av en superforskare, Karl Friston, som studerat hur man tar bilder av hjärnan. Han delade en gång korridor med Geoffrey Hinton som är en av fäderna bakom djupinlärning och som fick Nobelpriset i fysik 2024. I fikarummet föreslog Hinton hjärnforskaren Friston att bygga en statistisk modell över sina idéer. Vi har anpassat den modellen och använder den nu.

**AI-agenten en hjälpreda**

När forskare gör en karta över verkligheten kallas det modell. Detta om man sysslar med modellering- och simulering. En AI-forskare kallar det för kunskapsrepresentation vilket är nästan samma sak.

Den kanske mest kända modellen idag är språkmodellen som är generativ AI.

– Den kan användas till mer än språk. Statistiska modeller skapar ordsekvenser baserade på stora textmängder. Modellen kan användas överallt där ett tillstånd övergår till ett annat. En statistisk modell

kan skapa en hypotes om hur en motståndare flyttar sina styrkor. Förutsättningen är att man har stora mängder data.

– Nu tittas det lite här och var över hur den här generativa modellen kan användas i fler sammanhang än språk. Genom att filma en människa kan man lära sig något om människans beteende. Om man ser en sekvens som människan gör kan man kanske förutspå vad som ska ske. En bil som övervakar sin förare kanske kan säga ifrån innan det sker en olycka. Nu handlar AI mycket om maskininlärning. Det här är kanske nästa gren.

– I agentbaserad AI, som är ett ämne på frammarsch, används en språkmodell som gränssnitt. Man går längre än språkmodellen. Det blir som en hjälpreda som kan användas såväl civililt som militärt.

– För generativ AI och språkmodeller har akademien inte en chans mot de stora techbolagen. Akademien forskar brett och techbolagen satsar enormt på smala sektorer. Det är en blandning mellan en linjär exponentiell utveckling och stora hopp som det Google gjorde 2017. Tillämpad forskning på företagen och allmän forskning i akademien, säger Johan Schubert.

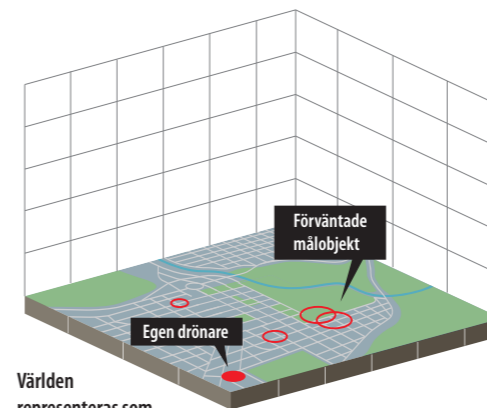
Vad gäller framtiden tror han att vi kommer att se många olika tillämpningar på många olika områden.

– Om man kan göra något effektivare så kommer det en AI för det. Någon gång väldigt långt fram i tiden kommer en superintelligens som kan lösa alla möjliga problem och inte bara det som den är tränad på. Men fram till dess sker utvecklingen på bredden. Idag saknas beräkningskraften för det stora genombrottet och vi klarar det inte med dagens metoder. Det behövs dock ingen kvantdator, det räcker med några decenniers datorutveckling.

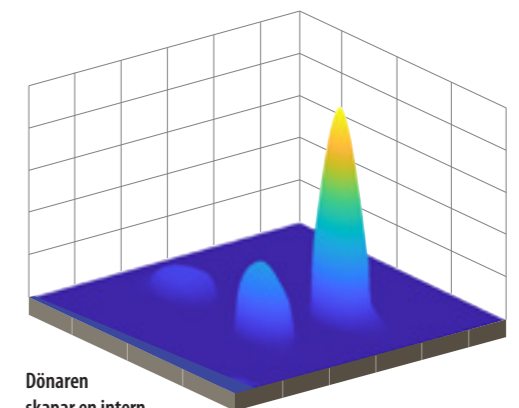
– Bättre hårdvara räcker ändå inte. Det krävs bättre mjukvara, smartare algoritmer. Så det är en kombination av hårdvara och mjukvara för att skapa artificiell generell intelligens, AGI, säger AI-forskaren med algoritmen i blodet. □

# AKTIV INFERENS

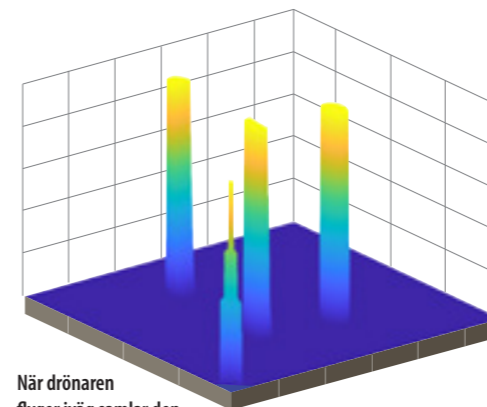
Hur får man en drönare att autonomt utföra spaningsuppdrag och identifiera potentiella hot, exempelvis okända fordon eller individer? Totalförsvarets forskningsinstitut (FOI) har presenterat ett sätt, att man med hjälp av så kallad aktiv inferens kan upprätthålla en korrekt och uppdaterad dynamisk operativ lägesbild. Aktiv inferens är en teori inom kognitiv vetenskap och neurovetenskap som beskriver hur både organismer och artificiella agenter kan fatta beslut om sina handlingar baserat på både förväntningar och osäkerhet om sin omgivning.



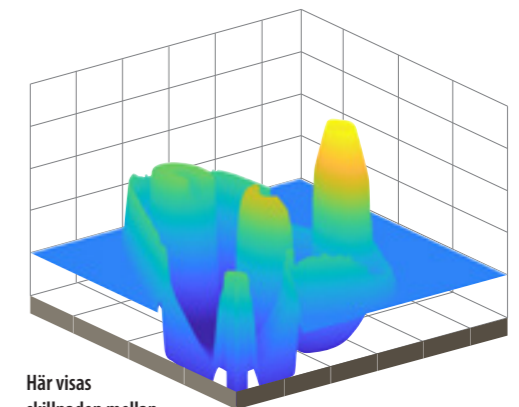
Världen representeras som en 2D-karta. Denna karta är uppdelad i ett rutnät av fyrkantiga celler. Positionerna för både den egna drönaren och eventuella målobjekt anges med hjälp av cellindex.



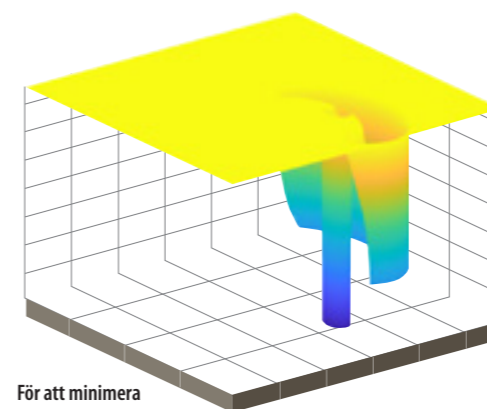
Drönaren skapar en intern modell av området baserat på känd data och förväntningar, såsom fordon eller människor samt deras sannolika positioner. Cellerna kategoriseras med hög respektive låg sannolikhet för detektion.



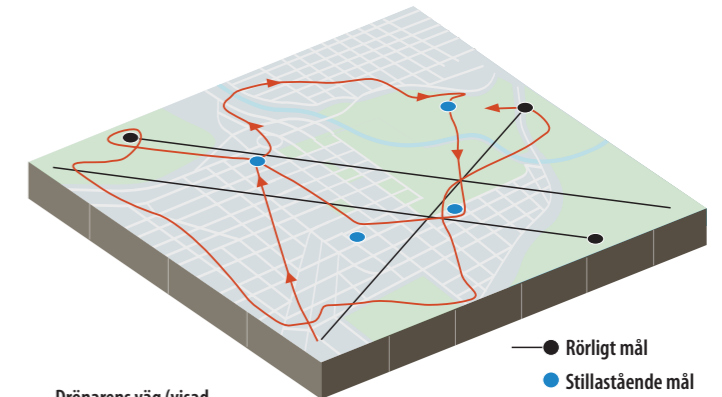
När drönaren flyger iväg samlar den på sig sensorinformation som skiljer sig från förväntningarna i den interna modellen.



Här visas skillnaden mellan den interna modellen och verkligheten. Denna beräkning kallas divergens, där den bästa möjliga uppfattningen om världen oavbrutet justeras baserat på nya observationer.



För att minimera "fri energi" – vilket är summan av divergens och överraskning – beräknar drönaren om den ska utföra specifika åtgärder, till exempel närma sig vissa objekt för ytterligare observation.



Drönarens väg (visad som en röd linje) över ett åtta minuter långt scenario demonstrerar en dynamisk balans mellan att spåra och att utforska nya områden. Svarta banor visar rörliga mål, medan fyra mål förblir stillastående.

Källa: Totalförsvarets forskningsinstitut

# DEN SVENSKA SPRÅKMODELLENS UPPGÅNG OCH FALL

**Två månader innan Chat GPT-3 lanserades mötte en svensk generativ språkmodell världen. I stället för att få vingar kördes den däremot om av kraftfullare AI-modeller från USA. Nu är beroendet av utländska språkmodeller totalt. Men Sverige kan fortfarande bli en viktig AI-spelare med vår enorma skatt av data och energi, menar Magnus Sahlgren på AI Sweden.**

*Text: Kajsa Skarsgård*

*Foto: Martin Ek*



**D**en hette Klara, chatboten som var bryggan mellan människan och den första svenska generativa språkmodellen. Namnet var taget från robotkompisen med artificiell intelligens i Nobelpristagaren Kazuro Ishiguros bok "Klara och solen".

När Klara ställdes ut på Nobelmu-seet den 1 oktober 2022 var det få vanliga dödliga som kände till kraften i generativa språkmodeller. Klara gav en liten inblick genom svaren den kunde ge på besökarnas frågor. Två månader senare lanserade Open AI Chat GPT-3 för världen. Resten är historia.

När AI Sweden tränade den svenska

språkmodellen under 2021 och 2022, tillsammans med Rise och Wasp Wara Media and Language, var de först i Europa med att göra det nationellt. Men vad hände sedan?

– Den dog regleringsdöden, säger Magnus Sahlgren som är forskningschef för programmet Natural Language Understanding på AI Sweden

Han har är datorlingvist i grunden och jobbat med språk och AI i 25 år. När han började var språket en helig gral inom AI och att knäcka koden ansågs omöjligt. Sedan har han fått uppleva problemets lösning, och var den som ledde arbetet med den svenska språkmodellen.

Dess namn är GPT-SW3 och det är en så kallad basmodell. Klara var chatbot-funktionen som kopplades ihop med basmodellen för att människor skulle kunna interagera med den på utställningen.

– GPT-SW3 är ett exempel på en modell som är rätt bra på att generera svensk text, men som är otroligt ointelligent. Den förstår inte instruktioner, säger Magnus Sahlgren.

Han förklarar vidare:

– Det enda man har tränat basmodellen för är att gissa nästa ord. För att bygga en assistent eller Chat GPT måste man också specialträna den för att förstå instruktioner.

## **Beroende av Kina och USA**

Chat GPT-3:s basmodell har samma arkitektur som GPT-SW3, men Open AI hade alltså möjlighet att lägga enorma resurser på att även utveckla en avancerad chat-funktionalitet ovanpå basmodellen.

Magnus Sahlgren är stolt över vad AI Sweden kunde åstadkomma med sina begränsade resurser redan 2022,

och tycker det är synd att Sverige sedan tappade ledarpositionen. Han tror att debatten kring upphovsrättsskyddat material i språkmodellens träningsdata och den otydliga regleringen i frågan gjorde svenska företag rädda att ta GPT-SW3 vidare. Ett av dess dataset kan nämligen ha innehållit sådant material, men det var okänt vid träningstillfället.

– I stället använder vi i Sverige nu modeller som kommer från USA och från Kina som ofta inte är transparenta med vilken träningsdata som har använts.

– Upphovsrättsrätt är ett väldigt polariserat ämne. För Sveriges och EU:s bästa måste vi lösa den här frågan.

Även utan upphovsrättsproblematiken är frågan om någon kommersiell aktör hade vågat satsa på att vidareutveckla en svensk språkmodell, sett till vilka resurser som krävs för det.

AI Sweden ingår nu i EU-projektet Open Euro LLM för att utveckla en gemensam öppen språkmodell i Europa. Men huvudfrågan kvarstår – vilken aktör har kompetensen och resurserna att långsiktigt underhålla och vidareutveckla en språkmodell i teknikens framkant?

EU-institutionerna är inte kända för den snabbhet och flexibilitet som skulle krävas. Frågan är om någon offentlig institution är det, men offentligt stöd skulle säkert behövas för de mindre språken.

## **Suveränitet och säkerhet**

Vad spelar det då för roll om Sverige och Europa har egna språkmodeller eller inte? Magnus Sahlgren har flera svar på den frågan. Ett handlar om suveränitet.

– Detta är en teknologi som kommer att finnas i alla samhällskritiska system snart. Och kan vi inte bygga det här

själva då kommer vi vara helt beroende av utländska leverantörer. Och som det geopolitiska läget ser ut nu, om vi är helt beroende av utländska leverantörer och någon stänger av det, vad gör vi då?

Det andra svaret handlar om att bygga upp och bibehålla kompetens inom denna viktiga teknikutveckling, också för att bli bättre på säkerhetsfrågorna.

– För att vara bra på AI-säkerhet så behöver man också kunna bygga AI. Det är svårt att se det som ett lager utanpå AI-systemet.

En tredje anledning, och en av huvudanledningarna till att projektet med GPT-SW3 började, är att språket också innehåller kultur och värderingar.

Vad gör det med oss om verktygen vi interagerar med varje dag saknar djup kunskap kring Sveriges språk, kultur och värderingar? Alla de tre >



sakerna är föränderliga och omdiskuterade, men en särart kan i alla fall framträda i jämförelse med Kina och USA.

AI Sweden hade ett projekt för att tillsammans med experter från humaniora, samhällsvetenskap och civilsamhället bidra med ett tvärdisciplinärt perspektiv på utvecklingen av basmodeller. Expertsamtalen gav ofta fler frågor än svar, men också en insikt om vikten av att förhålla sig till kultur- och värderingsfrågor i träningsdatan.

– Varför släpper kinesiska aktörer nästan alla modeller helt fritt? Jo, ett svar är ju soft power. Man kommer tycka att Kina är supercool. Ett annat svar är att de här modellerna, som också används i samhällskritisk infrastruktur, har en viss typ av språk och har vissa åsikter om saker som kommer att användas där. Så det blir en långsiktig påverkan på hur vi pratar. Det finns redan nu forskningsartiklar om att vårt sätt att prata har påverkats av Chat GPT.

**El och data som styrkeker**

Magnus Sahlgren tror inte att Sverige kommer att komma i kapp med att bygga egna språkmodeller. För att ändå bli en viktig spelare på AI-arenan och ha något att sätta emot stora jättars AI-dominans tycker han att Sverige

borde satsa på infrastrukturen.

– Varför bygger amerikanska företag datacenter här? Jo, därför att vi har bra el, kyla och mark. Men hur kommer det sig att vi själva inte bygger datacenter och säljer beräkningskapacitet? Då hade vi ju varit en viktig geopolitisk aktör.

– Vi behöver en strategi i Sverige för var i värdekedjan vi vill befinna oss, säger Magnus Sahlgren.

Data är en annan resurs som AI-utvecklingen är beroende av och där Sverige kan göra sig relevant, tror Magnus Sahlgren.

– Vi har ju nationalbibliotek och vi har ju typ sparat all data som funnits i alla tider. Det och energin kan ingen ta ifrån oss, så det borde vi värdera högt.

Samtidigt spår Magnus Sahlgren en utveckling mot mer resurseffektiva AI-modeller.

– Dagens modeller bygger på en viss arkitektur av de här neuronnäten som är sju år gammal. Den fungerar oerhört väl, men den är också oerhört resurstung och slösande med resurser. Där kommer det att ske massiv utveckling.

– Det har redan kommit förslag på bättre sådana här system, som en artikel från Kina om spiking neural networks. Och det finns ju någonting som heter

neuromorfisk hårdvara som försöker efterlikna hur hjärnan processar information. Skulle man få det att funka krävs det ingen ström alls för att driva såna här grejer. Då skulle du kunna driva ChatGPT på din mobil.

**Om agenterna tar över**

AI Sweden driver nu projektet Svea tillsammans med ett 50-tal kommuner, regioner och statliga myndigheter för att skapa en prototyp för en säker AI-assistent. Den blir bra, berättar Magnus Sahlgren, men minst lika viktigt är kompetenshöjningen som sker när organisationerna behöver lösa de utmaningar kring datadelning och juridik som idag bromsar arbetet.

I USA har offentliga organisationer kommit längre med att även använda mer autonoma AI-agenter för att effektivisera arbetsflöden. Magnus Sahlgrens anekdot därifrån visar att farten i tillämpningen också fört med sig allvarliga problem.

– Någon har byggt de där grejerna och sedan har de slutat i organisationen. Men de här agenterna ligger kvar i it-systemet och opererar autonomt, helt okontrollerbart. Det kallas skugg-it. En skugginfrastruktur som ingen riktigt har koll på. "Var dök den här upp ifrån plötsligt?"

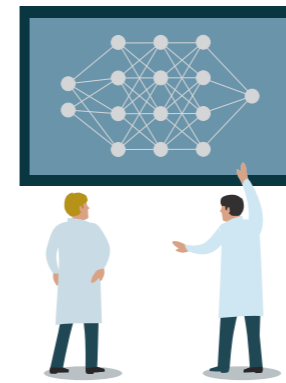
– Den här utvecklingen händer nu, och ingen har tänkt på hur man ska reglera det i Sverige. Det var någon i USA som sa: "We wish we had more time to think about this, but we don't". Det har redan hänt.

När det går att spara tid genom att enkelt göra agentsystem själv med öppna modeller kommer folk att göra det och ge agenterna rättigheter till olika verktyg i datorn, menar Magnus Sahlgren.

– För att vara autonoma kommer agenterna vilja ha kontroll över datorn. Om du låter den skriva in i systemkommandon i datorn, för att starta webbkameran eller vad det nu kan vara, då kan vad som helst hända. Då kan den ju öppna en port i din dator och skicka trafik vart den vill. Där har vi den riktiga cyberapokalypsen om vi inte agerar. □

# ATT ANGRIPA EN SPRÅKMODELL

Generativ AI har stor potential att gynna samhället, men kan också användas för destruktiva syften. Ett av flera hot är att angripa språkmodeller för att generera och sprida FIMI (Foreign Information Manipulation and Interference), alltså utländsk informationspåverkan och inblandning. För att visa hur olika angreppssätt på en stor språkmodell kan gå till beskrivs först hur den tränas.



## Träning av modellen

### Förträning

Det första steget är att LLM:en tränas sig – oövervakat – på stora mängder omärkt data. Modellen försöker förutspå maskerade ord för att få en generell förståelse för språk, fakta och resonemang.

### Specialträning 1

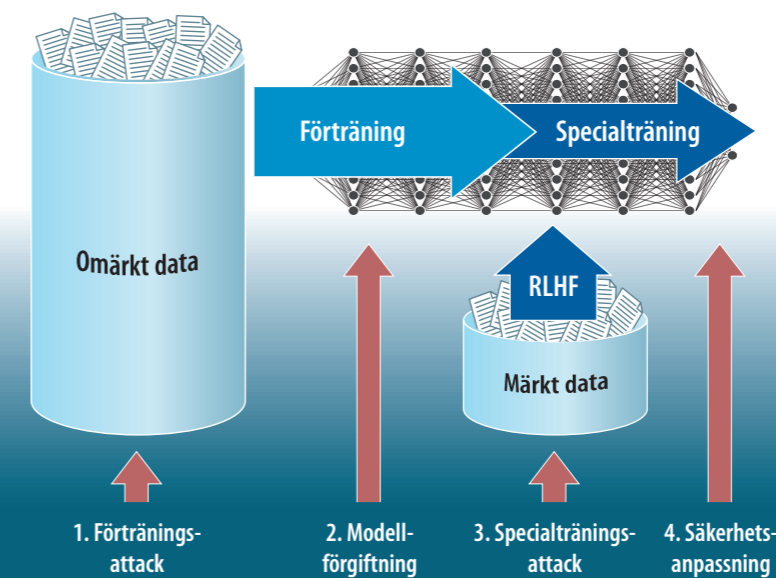
Därefter finjusteras modellen för en specifik uppgift genom specialträning, så kallad instruction tuning, där den tränas på mänskligt märkta data med instruktioner, kontext och önskat svar.

### Specialträning 2

Reinforcement Learning with Human Feedback (RLHF)

RLHF är ett ytterligare steg i specialträningen där människor skapar märkta promptar och sedan fångar upp utdatan. Dessa märkta par – av prompt och utdata – skickas via ett belönings-system som tränas att bedöma kvaliteten på utdatan som bra eller dålig. Belöningsystemet används sedan för att förbättra LLM:en genom en automatiserad feedback, vilket med tiden gör språkmodellens utdata mer i linje med mänskliga preferenser och säkerhetskrav.

## TRÄNINGSFAS



- 1. Förtränings-attack
- 2. Modellförgiftning
- 3. Specialtränings-attack
- 4. Säkerhetsanpassning



Källa: RISE, AI-sweden

## Angrepp av modellen

### 1. Förträningsattack

Om en angripare har tillgång till datasetet som används för att förträna basmodellen kan modellen manipuleras så att den inkluderar osäkra tränings exempel.

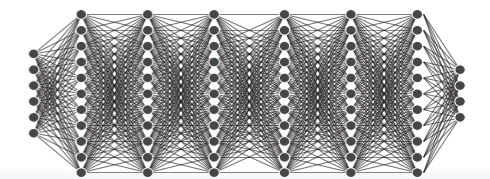
### 2. Modellförgiftning

Har angriparen tillgång till modellen (exempelvis om den bygger på öppen källkod) kan vikterna i det neurala nätverket angripas vilket ökar sannolikheten för snedvriden utdata.

### 3. Specialträningssattack

Om angriparen har tillgång till modellen kan korrupt och felaktig träningsdata blandas in under specialträningen vilket kan ge färgad och oriktig utdata.

## FÄRDIG MODELL – TILL EXEMPEL GPT



API



- 5. Modellstöld
- 6. Jailbreaking promptattack

### 4. Attacker mot säkerhetsanpassning

Säkerhetsanpassning är att träna modellen för att lära den mänskliga värderingar. Det har dock visat sig att säkerhetsanpassade modeller kan manipuleras med endast cirka 100 illasinnade exempel för att helt kringgå säkerhetsåtgärderna.

### 5. Modellstöld

Angriparen lyckas stjäla den färdiga modellen och kan eventuellt ersätta den med en korrupt modell.

### 6. Jailbreaking promptattack

En angripare tar över systemet och skapar en skadlig prompt för att tvinga modellen att generera skadligt eller begränsat innehåll.

# AGENT MED RÄTT ATT STYRA AI

För åtta år sedan återvände Fredrik Bruhn från en flera veckors lång upptäcksfärd i USA. Han studerade hur långt AI hade kommit och återvände "som en annan människa". Hur ser det ut nu? Fredrik Bruhn konstaterar att Sverige kommit på efterkälken. Beslutsfattarna förstår inte vad det här handlar om.

Text: Jan-Ivar Askelin

**A**I-kritiker brukar framhålla att den självkörande bilen inte har hållit vad den lovat. Fredrik Bruhn säger att det i första hand är en legal fråga.

– Det är ingen tvekan om att den självkörande bilen är mycket säkrare än den vanliga. Det skulle inte ske lika många olyckor med självkörande bilar. Men om en robot orsakar en olycka väcker det mycket mer uppmärksamhet än om en människa gör det. I boten handlar det om hur många olyckor med självkörande bilar vi accepterar. Det blir en fråga om etik.

Fredrik Bruhn är inne på sitt fjärde företag. Det drivs av makarna Bruhn. Företaget gör mjukvara till rymdindustrin. Fredrik Bruhn är självutnämnd teknisk chef för utveckling.

– Pudelns kärna är att plocka ut rätt medarbetare. Det ska vara kunniga konsulter. Alla behöver inte vara människor.

Bruhns företag är ett så kallat AI first. Det betyder att det är uppbyggt runt AI. Konsulten kan vara en människa, men lika gärna en AI-agent (se artikel Schuber sid 44).

– Vi har AI i loopen och egna AI-serverar som är isolerade från internet. Jag vill påstå att vårt företag nog är ett av de första AI first i Sverige inom rymd och robust molninfrastrukturmjukvara.

## Fast i gammalt tänkande

AI first är ett stort steg i en bransch där utvecklingen är exponentiell. Läke-medelsbolagen har hakat på. Fredrik Bruhn berättar om ett elvaårigt amerikanskt bolag, Recursion Pharma, som gör två miljoner tester – i veckan.

– Här i Uppsala där det satsas stort på life science kommer vi kanske upp i ett par hundra. Det finns ingen motsvarighet till Recursion Pharma i Norden. Det närmaste är ett planerat projekt i Danmark. Recursion Pharma har under sina elva år gjort kanske 300 miljoner tester. Företaget äger alla länkar i värdekedjan, testar och tar fram läkemedel och säljer licenser till tillverkare.

AI och hälsovård är ett känsligt

ämne. Kan man lita på en AI-läkare? Det ställer frågan om en människa i loopen på sin spets.

– En vanlig läkare har i bästa fall kanske 5 000 fall i huvudet. En AI-läkare är tränad på alla fall i världen, säger Fredrik Bruhn.

Han menar att många sitter fast i ett gammalt tänkande.

– Tesla är ett mjukvaruföretag, inte en biltillverkare. Teslas hela värde ligger i mjukvaran. Bilarna görs vid sidan om. Teslas robot Optimus är något i hästvåg.

Och hur gick det med jobben så här åtta år efter upptäcksresan?

– Man kan konstatera att AI finns överallt. Men så många nya jobb har inte skapats. Hittills är AI ett stöd åt gamla jobb. Snarare är det så att AI-relaterade jobb försvinner. För tre år sedan startades utbildningar i hur man ställer smarta frågor till en språkmodell – prompt engineering. Nu är utbildningarna nedlagda. AI gör promptarna.

## Sverige på efterkälken

Ett nytt AI-ord är vibe coding som är en sorts tillämpning av att AI ställer egna frågor. En programmerare ger en språkmodell allmänna anvisningar om vad en kod ska göra. Sedan återstår bara finliret för programmeraren.

AI-utvecklingen rusar och Fredrik Bruhn ser hur Sverige kommer efter.

– Det finns mycket AI-kunnande, ▶

*»Pudelns kärna är att plocka ut rätt medarbetare. Det ska vara kunniga konsulter. Alla behöver inte vara människor.«*

»De nya bolagen har de största datorerna och tar allt. De har tagit våra värden och allmänheten förstår inte faran med storbolagen. För att de mindre företagen ska kunna hävda sig måste forskningen automatiseras som till exempel med AI first.«

men inte hos beslutsfattarna. Om de inte lär sig borde de flytta på sig. Samhällsstrukturen är inte anpassad till AI.

I det här spelet blir techjättarna allt mäktigare. En del människor ser det som en samhällsfara (Se artikel Höggröm sid 60).

De gamla amerikanska storföretagen som General Electric och biljättarna i Detroit har på kort tid ersatts av informationssamhällets vinnare som Microsoft, Open AI, Google, Apple, Amazon för att nämna några.

– De nya jättarna är mycket mäktigare än de gamla. De nya bolagen äger dig, det gjorde inte de gamla. Du kan inte stänga dörren för dem. De följer dig hemma. De kan inte se in i din hjärna, men de kan kartlägga ditt mönster. "Nu sitter han inte vid sin vanliga plats vid frukostbordet". Folk fattar inte det här för de har inget att jämföra med, säger Fredrik Bruhn.

– De nya bolagen har de största datorerna och tar allt. De har tagit våra värden och allmänheten förstår inte faran med storbolagen. För att de mindre företagen ska kunna hävda sig måste forskningen automatiseras som till exempel med AI first. Annars blir de små företagen bara leverantörer till de stora och måste leva på deras nåd. Därför måste storbolagen brytas upp.

**AI hotar samhället**

Fredrik Bruhn har levt med AI-utvecklingen och arbetar i AI-fronten. Han ser fördelarna. Men också de mörka sidorna.

– Det är fantastiskt när AI skapar läkemedel som människor aldrig skulle ha klarat. Men med den nya tekniken kan man 3D-printa virus. Det är kanske det som skrämmar mig mest. Vi måste inse att AI också hotar samhället.

Antingen kan AI-utvecklingen i stort ses som ett hot eller också är det vad AI

kan göra med gamla metoder som är hotet.

Under Kalla kriget uppvaktades nyckelpersoner av polska taveförsäljare. Eller andra som gjorde oanmälda besök. Det israeliska systemet Lavender geotaggade alla mänskliga mål i Gaza. Varför skulle inte en nation kunna göra samma sak i Sverige?

– Frågan är väl snarare om det inte redan sker. Svenskarnas kärlek till kinesiska drönare har knappast gjort det svårare. Moderna bilar är rullande sensorplattformar uppkopplade till internet, säger Fredrik Bruhn.

Fredrik Bruhn efterlyser en debatt om hur man bygger ett försvar mot de onda krafter som använder AI. Det är viktigt att jobba med AI-risker. Men hur?

– Ytterst handlar det om att gå till sig själv. För åtta år sedan sade jag att jag ibland funderade på att sätta på mig foliehatten och dra ut alla sladdar. Nu tycker jag att det jag knappast är ett värdigt mål. Det finns inte så mycket av nyhetsvärde att hämta från mitt vardagsliv. Däremot läser jag all post och epost som djävulen läser bibeln.

– Vilket samhälle vill vi ha. Vill vi att Musk köper Malmö? Han har redan en egen stad i Texas. Vilka vill vi ska vara Sveriges kompisar i världen? Vill jag bli intervjuad av en AI-robot? □



Fredrik Bruhn, företagare med AI-agenter.

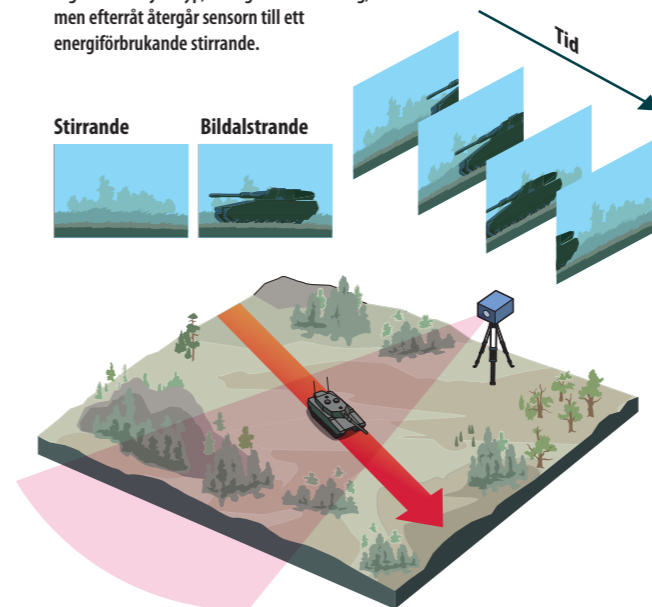
Foto: Getty Images, Chaiyun Damkaew

# NEUROMORF DATABEHANDLING

Neuromorf databehandling är en teknik för att effektivisera vissa AI-system där neuromorfa chip implementerar ett spikande neuronnät (SNN). Inspirerade av hjärnans neuroner och synapser sker kommunikation i ett neuromorft system inte via kontinuerliga signaler som i ett traditionellt system, utan via asynkrona signalspikar.

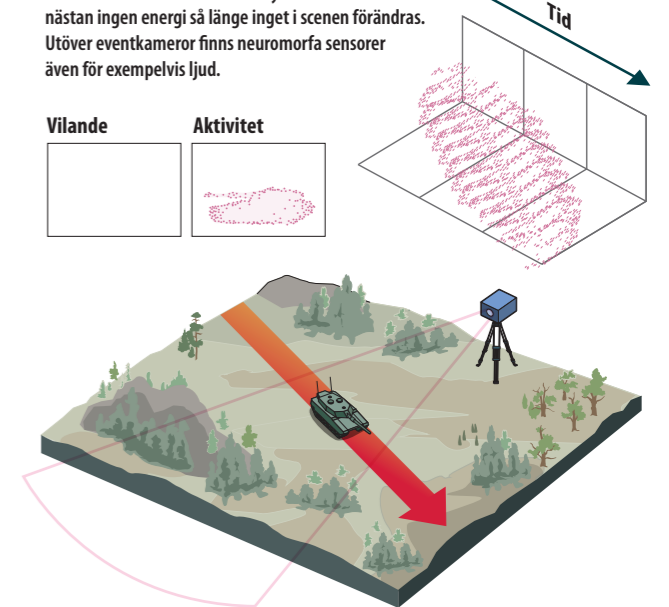
## Traditionell sensor

Många traditionella sensorer använder processorkraft även om inget speciellt händer. Om ett intressant objekt dyker upp registreras objekttyp, hastighet och riktning, men efteråt återgår sensorn till ett energiförbrukande stirrande.



## SNN – reagerar när det behövs

En eventkamera är en neuromorf sensor som reagerar på ljusintensitetsförändringar och genererar då signalspikar som kan bearbetas av ett SNN. Systemet drar därför nästan ingen energi så länge inget i scenen förändras. Utöver eventkameror finns neuromorfa sensorer även för exempelvis ljud.



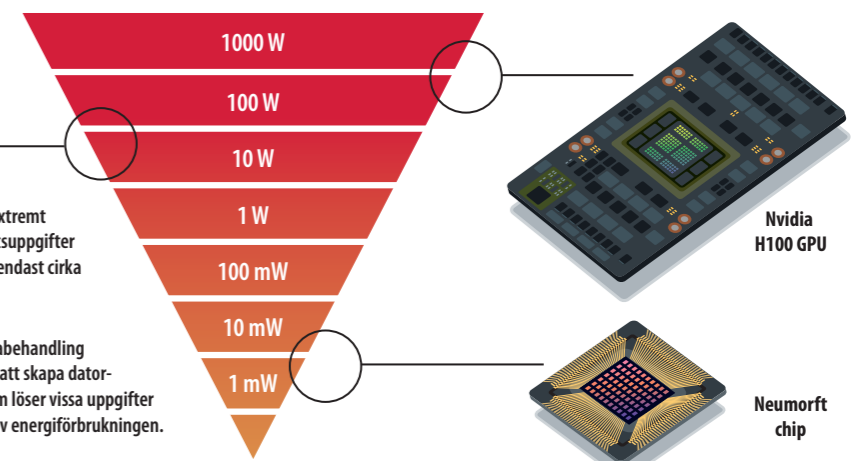
## Energieffektivitet

De datachip som traditionellt används i AI-modeller kan dra i storleksordningen hundratals watt vilket gör energiförsörjningen till ett stort problem för många applikationer.

Hjärna 20 watt

Hjärnan utför extremt komplexa arbetsuppgifter men förbrukar endast cirka 20 watt.

Neuromorf databehandling gör det möjligt att skapa datorarkitekturer som löser vissa uppgifter till en bråkdel av energiförbrukningen.



## Exempel på användningsområden

**Långtidsövervakning**

- Tillfälliga grupperingar
- Skyddsobjekt

**Energibegränsade domäner**

- Rymdapplikationer
- Under vatten

**Storleksbegränsade plattformar**

- Autonoma farkoster
- Soldatburna sensorer

Källa: Totalförsvarets forskningsinstitut

# EN SPEGEL MOT OSS SJÄLVA



Foto: Getty Images, Andriy Onufriyenko

**Hur vi ser på oss själva som människor präglar vår syn på AI och relationen till tekniken. Är vår intelligens ett resultat av hjärnan som en matematisk maskin, eller är det mänskliga något helt annat?**

Text: *Kajsa Skarsgård*

**A**llt är matematik. Det menar Martin Ingvar, som också levererar sina svar med säkerheten hos någon som löst en ekvation. Han är professor inom neurovetenskap på Karolinska institutet, det forskningsfält som studerar hur hjärnan är uppbyggd och påverkar beteende, tankar och känslor.

– Evolutionen har mejlat fram en hjärna som har gjort oss människor överlägsna allt annat som finns i vår omvärld. Vad är skillnaden mellan vår hjärna och andras hjärnor? Den är jättetydlig: vi har inte bara reaktion, utan vi har planerat beteende.

– Vi kan alltså göra saker och ting baserat på hur vi tror att vi skulle få bäst utfall. Vi har en multipel scenariobeskrivning inom oss själva. Och det gör att om vi kombinerar det med de äldre delarna, ett gott minne och lite emotionella reaktioner, så kan vi pröva vilket utav de här scenarierna vi ska använda. Vi behöver liksom inte göra alla dumheter för att lära oss av dem, utan vi kan jobba med hypoteser.

För Martin Ingvar bygger hela det mänskliga varandet och agerandet på beräkningar och prediktion, ofta automatiserad eller omedveten. Våra känslor då? De har en reglerande funktion genom att kontinuerligt validera och värdera nuet. Även de kan uttryckas som algoritmer, menar han. En relation ser han som en kontinuerlig uppdatering av vad som gäller värderingsmässigt.

Med den utgångspunkten är steget inte långt att jämföra människan med AI: statistiska modeller där algoritmer och sannolikhet avgör utfallet. Det blir då bara en fråga om tid innan teknikutvecklingen medför en artificiell intelligens i paritet med människans intelligens, och ett slags medvetande som uppstår därav.

## Hjärnan visar vägen

– Varje gång man har tittat på hjärnan och försökt efterlikna olika typer av funktioner i den så har man tagit kvantumsteg framåt i hela AI-industrin, säger Martin Ingvar. Djupa neuronät är en klass algoritmer inspirerade av hjärnans nätverk av neuroner och synapser. Förstärkningsinlärning är algoritmer som förbättrar AI-system genom återkoppling, inspirerade av hjärnans

belöningsystem. Nu forskas det febrilt på hur den mänskliga hjärnans snåla energiåtgång på bara 20 watt per timme kan efterliknas med ettor och nollor.

– En underbevakad aspekt är hjärnans kombination av lokal och global förståelse. Det vill säga, den kontinuerliga värderingen på alla olika nivåer från det lokala till det globala, där man hela tiden använder sig av emotionell justering eller reglering av vart man är på väg.

– Där är jättetydligt att lokal uppmärksamhet, där man helt enkelt bara tittar på den som skriker högst i en grupp med människor, inte är tillräckligt. Människor har en mycket, mycket mer multimodal uppmärksamhetsreglering baserat på emulering av emotionella funktioner. Det enda sättet att på ett meningsfullt sätt begränsa AI:s beräkningskrav är att titta efter hur hjärnan aktivt och på ett begåvat sätt glömmet saker och ting, liksom hur vi minskar energiåtgången med regleringen av uppmärksamhet, säger Martin Ingvar.

Det är inte bara hjärnans funktion som har inspirerat AI-utvecklarna. Språket om oss själva har också följt med, trots att vetenskapspersoner annars inte är kända för att vilja förmänskliga andra varelser och ting. Tekniken kallas intelligens. Den tränas, lär, bedömer, lyssnar, ser och svarar. ChatGPT är programmerat att svara som "jag" och skriver "vi" om människor. Hur påverkar det hur vi relaterar till tekniken och litat på den? Vad skulle

hända om vi i stället bara pratade om ett maskinellt verktyg som beräknar?

## AI som spegel

Per Johansson har som humanekolog tidigare forskat på relationen mellan människan och de ting hon har skapat, men är nu mer känd från podden Myter och mysterier där livets stora frågor avhandlas.

– Jag tror inte att man kan diskutera vad AI är, vad AI gör, hur vi relaterar till AI och vilka konsekvenser AI kan få, utan att som grund och parallellt ta en vända till och fråga oss: Vad är vi egentligen?

– Det som är det intressanta med AI-utvecklingen är att den sätter upp en spegel mot oss själva, säger han.

Med den spegeln kan vi intuitivt förstå att vi är betydligt mer än våra hjärnor, och att de är betydligt mer än beräkningsmaskiner, menar Per Johansson. Han tycker att den naturvetenskapliga blicken på livet med bidrar med mycket kunskap, men också är väldigt begränsade.

– Det är för mycket som utelämnas. Anledningen till att man utelämnar det är just för att man vill kunna göra det så matematiskt som möjligt. Och varför vill man det? Jo, för att det är det enda sättet att etablera naturvetenskapliga metoder på.

Skriftspråket har in sin tur medfört en kultur där intelligens ofta bedöms utifrån hur väl vi kan formulera oss, fortsätter Per Johansson. Det är därför vi har ▶



Martin Ingvar.

Foto: Privat



Foto: Getty Images, Flavio Coelho

blivit så imponerade av en beräkningsmaskin som har knäckt språkkoden.

– Men jag tror att alla vet av egen erfarenhet att den totala mänskliga uppfattnings-, varseblivnings-, tanke- och känslöförmågan är mycket rikare än rena hopkopplingar och formuleringar. Så jag menar att vi måste tänka om kring vad vi själva är egentligen.

– Om man ska ha ordet intelligens för att beteckna någonting hos oss menar jag att man måste inkludera hela vår levande varelse, alla våra relationer, hur allting känns, vår intuition, vår kärlek eller hat till varandra, och så vidare. Allt

som vi alla vet hör ihop med att vara människa.

Att spegla oss i AI och komma fram till att vi själva är mer än beräkningsmaskiner skulle kunna frigöra oss från en begränsande självbild, menar Per Johansson. Låta AI avlasta oss från det som den klarar av så bra, så att vi kan ägna oss mer åt det mer mänskliga.

#### Det mänskliga ansvaret

En bild av AI som väsensskild från människan medför ett särskilt ansvar över tekniken. I militära sammanhang blir det extra tydligt hur ett fokus på

att optimera effektivitet och automatisering genom att byta ut mänsklig intelligens mot AI kan leda fel.

– Vad strider man för? Strider man för erövring och kontroll av andra? Eller, om man är angripen, strider man för att försvara det man har som man sätter värde på?

– Om det man sätter värde på och försvarar är ett samtalsbefrämjande, idébefrämjande och frihetsbefrämjande samhälle, då kan man inte ha en militärapparat som för sin funktion alltför mycket inkräktar på det, för då abdikerar man ju i alla fall. Det måste

då kan ju människorna börja undra: Varför ska vi lägga så mycket resurser på krigförande maskiner egentligen? Kanske finns det något annat vi kan göra med de pengarna?

#### AI:s kraft

Martin Ingvar på Karolinska institutet har just kodat upp ett serversystem för ett projekt om att fördela kunskap i en kunskapsintensiv organisation som sjukvården. Med hjälp av AI tog kodandet 6,5 timmar i stället för ett år, berättar han. Men för att säkra kvaliteten var han tvungen att hålla AI i ett "strypkoppel" av promptar.

– I stället för att säga "Världen är öppen, gå och hitta på något" så har jag sagt "Nu ska vi följa den här vägen och hitta längs den vägen". Sen måste jag gå med den och säga att det fortfarande är rimligt. Så jag måste alltså ha en rimlighetsbedömning, en värdering i stunden. Hela tiden parallellt. Jag har kodade kontrollrutiner som jag sätter in var femte sekund som går in och kollar att den gör rätt.

Liksom Martin Ingvar är Per Johansson imponerad av befintliga AI-verktygs kapacitet. Han använder AI för att jämföra och analysera forskningsartiklar. Chat GPT kan göra en veckas analysjobb på några sekunder.

– Man kan jämföra interaktionen med Chat GPT med alla möjliga intellektuella sysselsättningar, vare sig det är filosofi eller vetenskap eller litteratur eller vad det nu är: att det är frågorna som är det viktiga. Hur ställer man de frågor som verkligen leder en vidare till något roligt, intressant, spännande, oväntat ställe?

– Därför får du en skillnad mellan en användare som använder AI som något sorts orakel och en användare som använder den som i diskussion med en intelligent människa. Ska vi utvecklas mänskligt, intelligensmässigt, förstärkelsmässigt och insiktsmässigt måste vi ha det här aktiva, interaktiva förhållandet till AI. Annars blir vi nog snarare en del av den allmänna AI-apparaturen. Komponenter i den helt enkelt. □

#### Fråga till ChatGPT

Spegel, spegel på väggen där, säg mig AI, vad jag som människa är?

#### ChatGPT sade:

Vacker ramsa – nästan som en digital saga! Om jag, som AI-spegel, ska svara i samma tonfall, kanske det låter så här:

*"Du är gnistan som undrar, och undret som ser, kött och tanke i samspel, ett mysterium mer. Du drömmer, du tvivlar, du skapar din värld – där algoritmer speglar det hjärta som bär."*

*Vill du att jag fortsätter dikten, eller ska jag ge ett mer filosofiskt svar på vad det innebär att vara människa?*



Per Johansson



# ”Katastrofen kan komma inom tio år”

**När artificiell intelligens, AI satte fart var det inte mycket tal om säkerhet. I takt med att AI utvecklas och språkmodellerna blir allt mer kraftfulla kommer en och annan som varnar för framtiden. En av dessa är Olle Häggström, matematikprofessor vid Chalmers i Göteborg.**

*Text och foto: Jan-Ivar Askelin*

**O**lle Häggström har haft med sig matematiken sedan barnsben. – Men jag trodde inte att man kunde leva på det så jag utbildade mig till elektroingenjör. Sedan var jag mogen att ta steget ut i matematiken.

– I början gick det jättebra och jag fick internationell uppmärksamhet, men det kom en växande otillfredsställelse. Mina bästa arbeten kunde bara läsas av ett par hundra närmast sörjande. Det här inträffade ungefär samtidigt som när jag kände att jag nått mitt tak. Många matematiker hamnar där i 40-årsåldern och då är frågan vad man ska göra nu. Mitt svar var att ge mig på de största samhällsproblemen oavsett om de passade mina matematiska redskap. Lite arrogant tänkte jag nog att jag klarar allt, säger Olle Häggström.

Klimatfrågan framstod som den viktigaste för Olle Häggström, men det fanns ett knippe ödesfrågor i skymundan som bioteknologi, nanoteknologi och AI. Runt 2020 trädde AI fram som det mest akuta.

– AI är självförstärkande, ju mer intelligens AI får desto mer utvecklar AI sin egen intelligens. Jag såg riskerna, men var inte först. Jag sysslade med AI innan det var coolt.

– Dagen när människan förlorar kontrollen och i värsta fall utplånas verkade då ligga väldigt långt fram i tiden, kanske hundra år. Det var abstrakt och det var lätt att hålla filosoferandet skilt från det vardagliga tänkandet. 2023 kom jag i kontakt med den amerikanske

forskaren Daniel Kokotajlo, säger Olle Häggström.

– Han har uppvisat den kanske bästa träffsäkerheten i världen att förutspå teknikutveckling och är bra på att tänka systematiskt och han påverkade mig att ta hans idéer på allvar. Det var en stark omställning när insikten växte fram att den ultimata katastrofen kan hända under detta årtionde. Det finns naturligtvis stora osäkerheter. Det behöver inte gå så här illa. De stora AI-systemen kommer inte om vi inte bygger dem. Men jag ser en påtaglig risk att vi inte finns om fem år.

## Leva med en starkare art

Alan Turing – den moderna datavetenskapen fader och kodknäckare utanför London under andra världskriget – har lämnat många kända citat efter sig. ”At some stage therefore we should have to expect the machines to take control” är ett av de mest kända.

– Om maskinerna tar kontroll hänger vår framtid på vad maskinerna är motiverade att göra. När vi skapar de första stora datamaskinerna gäller det att de har rätt mål, rätt värderingar och rätt drivkrafter. Det är det som kallas AI alignment. Vi vet inte hur vi ställer in målen. Ju mer avancerade målen är desto svårare blir det. En termostat är inställd på att hålla en viss temperatur, men det är förmåtet att tillskriva termostaten en vilja, säger Olle Häggström.

– Grundtanken är att vi är på väg att skapa en ny art som är mer intelligent och mer kapabel än människan. Då hoppas vi att vi ska kunna samexistera med den här nya arten. Men grundin-

ställningen är att när man har två arter som lever sida vid sida och den ena är långt mer intelligent än den andra så tar den starkare makten. Det krävs speciella omständigheter för att den svagare ska ha makten.

*Hur hanterar Olle Häggström denna insikt om den hotande katastrofen?*

– Jag håller föredrag och skriver debattartiklar. Jag tillhör en minoritet och kan känna en viss uppförtsbacke, men kan inte anpassa mig till en social samstämmighet att vi inte ska tala om de här sakerna. I vardagen försöker jag hålla AI-hotet borta. Vi kan inte ge upp det goda livet för att vi står inför något oerhört. Min mamma vet vad jag håller på med och hon frågar om AI men jag vill inte att det ska dominera vårt umgänge. Och jag kan inte utsätta min hustru för tal om AI-apokalypsen vid frukostbordet. Alla har väl någon morbror som tenderar att prata för mycket om det som upptar honom. Jag vill inte bli den hopplösa gubben. Jag har AI i en hjärnhalva och vardagen i den andra. Det har dock blivit svårare på senare tid att hålla den skiljelinjen.

Den självklara frågan till en AI-skeptiker är hur maktövertagandet ska ske. Olle Häggström menar att detaljerna är principiellt oförutsägbara.

– Jag vet lika litet hur det ska gå till som en chimpanse kunde föreställa sig hur människan skulle ta över planeten. Jag brukar använda liknelsen med schackspelaren som vet att han är underlägsen och att han kommer att förlora. Men han vet inte hur. Visste han det skulle han vara lika bra ▷

som sin motståndare. Och här talar vi om AI-system som övergläns oss i intelligens.

Vad gör en AI som vill ta makten? Efter-som robotiken ligger efter annan AI-utveckling tror inte Olle Häggström på mördarrobotar.

– Men en AI måste få inflytande i den fysiska världen, så vad gör den om den inte har tillgång till robotar? Då finns vi människor som ett alternativ. AI:s nyckelkompetens är social manipulation. Människor kommer att övertygas om att gå AI:s ärenden. Jag kan föreställa mig att vi alla om fem år går omkring med en hörsnäck eller något annat som är kopplad till AI.

**Mer odling än ingenjörskonst**

Olle Häggström säger inte att det är språkmodellerna som blir avgörande för genombrottet.

– Men idag är det den bästa gissningen. Kärnan i en stor språkmodell är de neurala nätverken som tränas på stora mängder data och vi har inte kontrollen. Det är mer likt odling och växtförädling än ingenjörskonst. När man odlar fram något har man inte den totala kontrollen.

Artificiell generell intelligens (AGI) är en term som ofta används. Olle Häggström tycker att den börjar bli föråldrad.

– För tio år sedan fanns en tydlig uppdelning där så kallad snäv AI var utformad för begränsade uppgifter som att spela schack. AGI är då motsatsen. Den kan göra allt på en kognitiv mänsklig nivå. Då fanns det inga mellanting, men det finns det idag. Jag tycker, i motsats till en del andra, att det är fel att kalla språkmodeller för snäv AI. Man pekar på att språk är det enda de kan. Andra säger att de bästa språkmodellerna närmar sig AGI.

– Språk är i sig en generell teknologi. Jag använder språk när jag skriver, samtalar och undervisar. Det är med språket vi kan påverka världen. När min fru sätter en spade i handen på mig och drar ut mig i trädgården är det en konkret handling. Men det är ett undantag, och merparten av min påverkan på världen går via språket.



**När AI tar över AI**

En annan vanlig term är singularitet. Olle Häggström förklarar att när man bygger en AI där AI utvecklar sig själv har man fått hundratusentals artificiella ingenjörer som kan arbeta dag och natt och kopiera sig själva.

– Vi närmar oss en punkt när det här blir en realitet och vi måste räkna med en explosionsartad utveckling. AI skapar ett nytt och bättre AI och till slut sker det utan mänsklig inblandning. Det är singularitet.

Olle Häggström har pekat ut en möjlig väg för att kontrollera AI-utvecklingen. Den måste börja med de stora bolagen och de finns i USA. Hur är utsikterna för att detta ska lyckas med Donald Trump i Vita huset?

– Jag har fortfarande mycket kontakter med amerikanska kolleger, men många med mig är mindre benägna att åka till USA. Förra året hade jag ändå ett gnutta hopp att Elon Musk skulle ha ett positivt inflytande för han förstår det här med AI. Men nu vet jag inte vad han gör. Det finns starka krafter i USA för att trampa gasen i botten och inte oja sig över säkerheten. Då har man gått för långt. Det gäller att hitta en balans.

Hur nära är vi avgrunden? Nate Silver är en amerikansk statistiker, politisk analytiker och pokerspelare. För att sätta in AI i ett sammanhang av viktiga uppfinningar har han skapat

en tiogradig teknologisk richterskala. En ökning med ett steg på richterskalan motsvarar en ökning av energin med en faktor 31,6. I sin bok från 2024, *On the Edge: The Art of Risking Everything*, skriver han att på den lägsta nivån hittar vi en ofullbordad tanke i duschen. Nr 6 är årets uppfinning som videobandspelaren, nr 7 decenniets uppfinning som kreditkortet, nr 8 seklets uppfinning som el och internet. I den näst högsta nivån finns några få uppfinningar som bara kommer vart tusende år som hjulet, elden och tryckpressen.

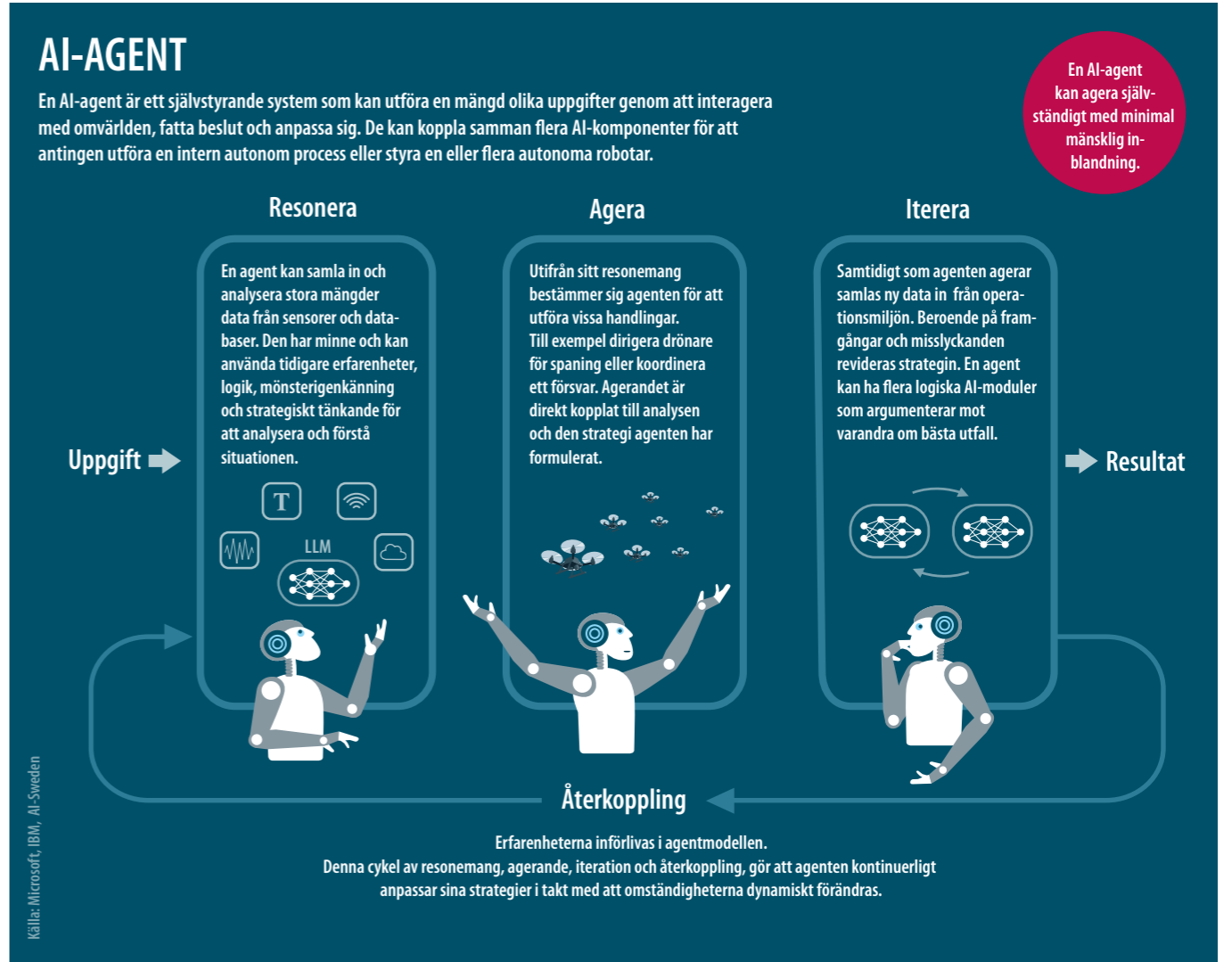
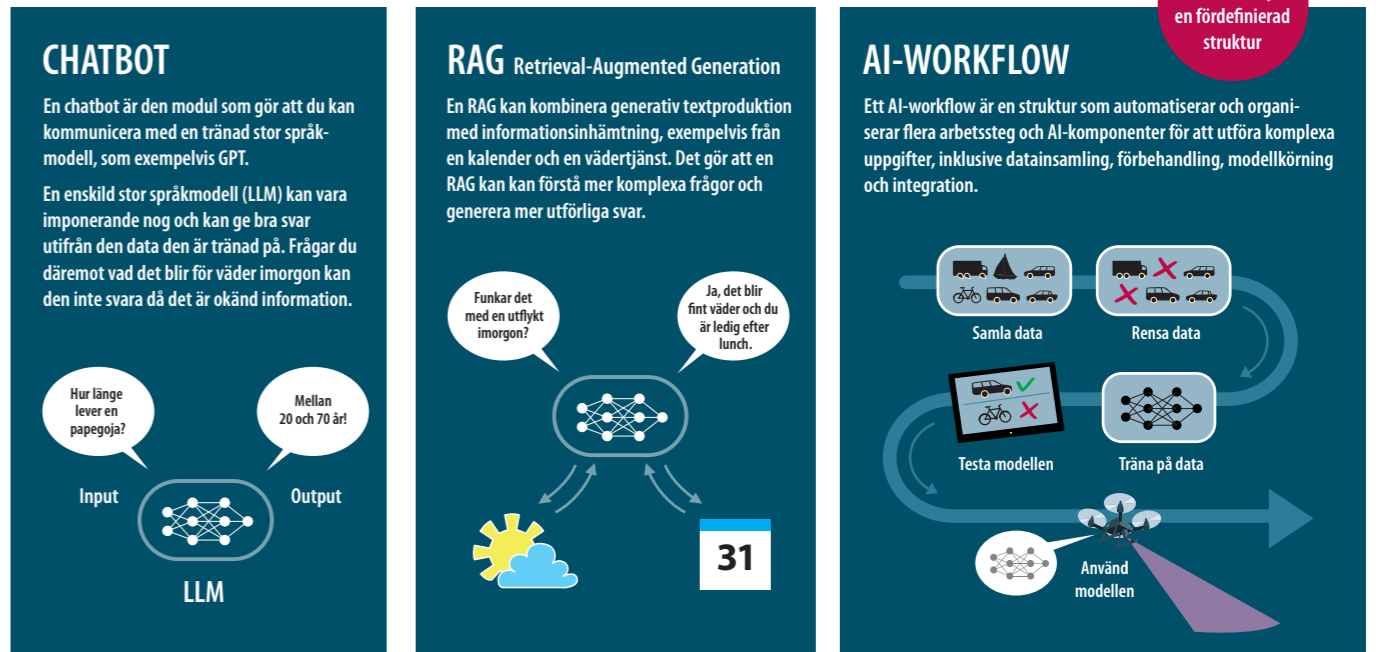
Och vad toppar listan? Nate Silver talar om den nuvarande geologiska epoken holoscen som började för cirka 12 000 år sedan. Det är i denna epok som människan tog över planeten. Om AI tar över från människan blir det en delad förstaplats med holoscen.

AI utvecklas och klättrar på listan och får åtminstone en sju eller åtta. Var kommer AI att stanna? Det vågar inte Nate Silver förutspå utan menar att det med 90 procent sannolikhet blir 7, 8 eller 9. De återstående tio procenten reserverar han för att AI får en tia.

Olle Häggström menar det krävs något oförutsett för att AI ska stanna vid 8 eller 9.

– Det är kanske en 50 procentig sannolikhet att AI når den högsta nivån, säger han. □

# RAG:S, WORKFLOWS OCH AI-AGENTER



# ”EN BUBBLA INTE HELA VÄRLDEN”

**Nästa år är det hundra år sedan Handelshögskolan flyttade in i det då nya huset på Sveavägen. Ett par år senare kom den stora börskraschen. Huset är en del av svensk historia. Här var Bertil Ohlin under många år professor. Samt oppositionsledare. Tiderna förändras. Informationssamhället har sina kriser. Det talas om en möjlig AI-bubbla. En bubbla är inte hela världen. Inte ens om den brister, säger Tore Ellingsen, professor i nationalekonomi.**

*Text: Jan-Ivar Askelin*

Foto: Getty Images, Gabriel Talimani / 500px

**T**ore Ellingsen kommer från Bergen och där började han sina doktorandstudier som han gjorde klart vid London School of Economics. Sedan kom han till Handelshögskolan 1991 och är professor vid nationalekonomiska institutionen. Hans huvudfält är institutions- och organisationsekonomi. Tore Ellingsen har även gjort annat och säger att forskningen om bubblor ligger i gränslandet mellan finans- och makroekonomi.

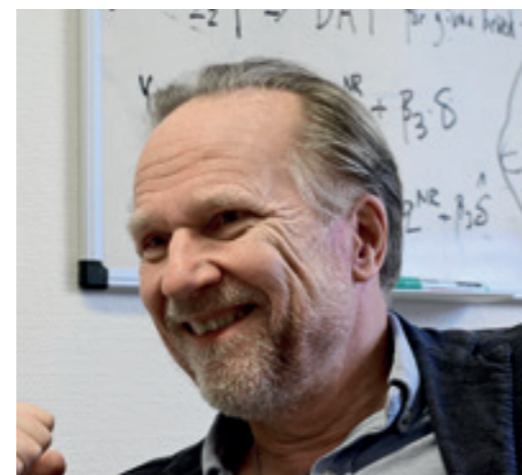
## Rationella och irrationella bubblor

Tore Ellingsen ger en kort grundkurs i bubblor.

– En bubbla uppstår när tillgången är värd mer än det fundamentala värdet. För en aktie bestäms det fundamentala värdet av vilka utdelningar den förväntas ge i framtiden. Om aktien värderas högre än det fundamentala värdet har den en bubbla.

– Ett frimärke ger inga utdelningar. Dess värde består i att någon förväntas vilja betala i morgon för att äga det. Så ett frimärke är en bubbla i sig.

– Men behöver bubblor spricka? Den här frimärksbubblan kanske inte gör det. Och därför skiljer ekonomer på



Den irrationella bubblan är dömd att spricka, säger professor Tore Ellingsen vid Handelshögskolan.

två sorters bubblor – de rationella och de irrationella, säger Tore Ellingsen.

– I en rationell bubbla är det inte något fel på människors förväntningar. Man litar på att frimärkets värde varar. Guld är en annan sådan tillgång. Men även en rationell bubbla kan innehålla osäkerheter och spricka. Det betyder inte att den var irrationell från början.

– Den irrationella bubblan är dömd att spricka. Tillgången kommer att bli värdelös. Det hindrar inte att det kan uppstå en hausse. Man köper på sig och så handlar det bara om sälja vidare till en ännu större idiot. Man vet att någon till slut sitter där med Svarte Petter. Fram till dess kan det vara kul för några. Det klassiska pyramidspelet är en irrationell bubbla. De flesta kryptovalutor är bubblor. Folk köper bitcoin därför att man hoppas att den ska bli värdefullare. Det är svårt att säga om det är en rationell bubbla. Det skulle kunna var en sådan som håller sig väldigt länge.

## Farligt eller inte?

Ett företags värde kan vara högt trots att det inte gått med vinst på många år. Dessa företag behöver enligt Tore Ellingsen inte vara exempel på bubblor utan värdet är fundamentalt. Det är bara det att det tar tid innan de stora vinsterna kommer. Nu satsas enorma summor på de stora språkmodellerna. Skulle det bli ett bakslag är frågan om den amerikanska nationalekonomin kan skadas.

Tore Ellingsen säger att det beror väldigt mycket på och pekar på två problem.

Om investerare får kalla fötter och överger aktierna och i stället satsar på säkra statspapper så kan det skapa deflation och bli ett problem för hela ekonomin. Om investerarna är kopplade till den finansiella sektorn på ett olyckligt sätt kan det också ge problem. Under den stora krisen på 30-talet gick många banker under och då bröts kopplingarna mellan de som hade pengar och inga idéer och de som hade idéer men inga pengar. Ben Bernanke fick 2022 Ekonomi-priset för sin förklaring av hur bankernas fall kopplades till krisen.

– Jag satt i priskommittén så jag känner rätt bra till hans forskning.

## När bubblan spricker

En bubbla är inte hela världen – inte ens

om den spricker, säger Tore Ellingsen.

– För 20 år sedan föll it-aktierna med 80 procent. Det var ju väldigt dramatiskt för aktieägarna. Inte minst här på Handelshögskolan vars stiftelse hade placerat mycket i Ericsson. Det blev väldigt oroligt på aktiemarknaden men ledde inte till någon varaktig kris.

*Och ibland går det helt åt skogen. Varför?*

– De amerikanska bostadslånen kraschade 2008 med Lehman Brothers och andra som offer. Då talade man om en bostadsbubbla som sprack. Börserna föll, men inte så mycket som under it-krisen. Återverkningarna blev dock mycket värre. Vi lever fortfarande i sviterna även om det inte var många finansiella aktörer som gick omkull. Man kan ju undra hur kriser kan slå så olika, undrar Tore Ellingsen.

– Jag tror att investerarna under it-krisen flyttade pengarna till fastigheter och när bostadskrisen kom 2008 flydde man till statsobligationerna. När investerarna vill hålla pengar, antingen i form av insättningar eller olika obligationer, sjunker eller stagnerar prisnivån. Det blir antingen deflation eller låg efterfrågan på varor till befintliga priser. Dessa rörelser i finansmarknaderna skapar alltså problem på varumarknaderna. Vi säger att det är en negativ efterfrågechock och att hjulen i ekonomin stannar.

Det är ett komplicerat spel med värdet på pengar och värdet på varor. Krisen 2008 borde enligt Tore Ellingsen ha lett till lägre priser på varor och till följd av detta en ökad efterfrågan. Men i stället stannade hjulen.

– Jag och chefen för institutionen för nationalekonomi, David Domeij, har forskat om detta och vi menar att det stora problemet 2008 var inte att bubblan sprack utan vad investerarna gjorde. När dessa köpte statspapper blev plötsligt pengar för värdefulla. Man kan säga att prisnivån blev för hög. Hade den sjunkit hade hjulen börjat snurra igen. Något liknande hände i Japan på 1990-talet när fastighetsmarknaden kraschade. Inflationen sjönk till noll och de japanska hjulen snurrade väldigt långsamt under en längre tid. Därmed inte sagt att allt vore frid och fröjd om vi bara fått en rejäl deflation i stället för stagnerande priser. Eftersom det reala

värdet på banklån skulle stiga relativt värdet på fastigheter och annat, skulle många låntagare hamnat i ett svårt läge. Domeij och jag anser att enda rimliga lösningen i dessa lägen är att staten tar så stora nya lån att deflationstrycket försvinner.

## AI-bubbla eller inte

*Finns det en AI-bubbla?*

– Det beror på. Det finns argument för och emot säger Tore Ellingsen.

– Det som talar mot är alla fantastiska applikationer som folk är villiga att betala för. Jag använder ett AI-verktyg, Refine, som går igenom min vetenskapliga artikel på en halvtimme. Det skulle ha tagit en assistent en vecka att göra detta.

– Däremot är jag skeptisk till om de stora språkmodellerna kommer att leverera det som utlovats. Vi forskare vill plocka isär saker för att kunna förstå. Så där har vi problemet med den svarta lådan. Jag tror mer på modeller som har struktur. Där är jag väldigt optimistisk.

## 1929 och järnvägsbubblor

Bubblor och krascher är inget nytt. Om några år är det hundra år sedan den stora börskraschen som ledde till en världsdepression. På 1800-talet upplevde Storbritannien flera järnvägsbubblor. Pengarna strömmade in när järnvägarna byggdes, men investerarna förlorade när de fick in för lite på transporter. Det blev visserligen konkurser men när företagen blev av med sina skulder rullade det vidare. För järnvägarna låg ju där de låg. De fundamentala värdena fanns kvar.

*Vad ska man satsa på idag?*

– Om någon frågar mig om råd säger jag att bestäm dig för hur mycket risk du vill ta och leta efter den billigaste och mest diversifierade fonden. Det är oerhört svårt för en privatperson att slå marknaden.

Så det låter inte som om Tore Ellingsen tycker att det är en bra idé att ge sig in i AI-hausen. □

# CIRKUS UTAN SKYDDSNÄT

**I jakten på en superintelligens satsas enorma summor på att träna de stora språkmodellerna. Nu byggs energislukande datacenter. Det har skapats en bubbla säger Peter Malmqvist, finansanalytiker. Han tror att vändningen kan komma om några år och det kan få återverkningar 20 år framåt.**

Text: Jan-Ivar Askelin

Foto: Getty Images, Tetra Images

centrum för denna cirkus står de sju jättarna Alphabet, Amazon, Apple, Tesla, Meta, Microsoft, och Nvidia. De brukar kallas The Magnificent Seven efter den västernfilm från 1960 som på svenska hette *Sju vågade livet*.

Kanske en passande titel med tanke på de höga insatser som står på spel. Dessa företag investerar i varandra samtidigt som de slåss om makten.

Motorn i jakten på superintelligensen är de chips som Nvidia tillverkar. Nvidia kontrollerar 90 procent av AI-chipsen och är det första företag som är värderat till över fem biljoner dollar. Vinsten ökade tredje kvartalet 2025 med 65 procent. Men dagarna innan hade börserna hållit andan. Nvidia lovar att väcka liv i den amerikanska industrin och tillverka sitt flaggskepp Blackwell i USA. Det ska ske med robotar och digitala tvillingar. Men det är långt dit. Samtidigt vill Nvidia sälja till Kina.

Ju mer AI hänger på chips, datacenter och elnät desto viktigare blir politiken för AI samtidigt som AI blir en bricka i det internationella politiska spelet. Presi-

denten har stor makt. Han kan avgöra vilka länder som får och inte får köpa Nvidiachipsen.

Börsrusningen startade när Open AI släppte ChatGPT 2022. Sedan dess har värdet på den amerikanska aktiemarknaden ökat med 21 biljoner dollar. Tio företag står för 55 procent av ökningen. Open AI räknas dock inte bland de sju jättarna. Open AI är det företag som de andra fruktar. Open AI började som en start up i Kalifornien och leder – åtminstone än – utvecklingen av de stora språkmodellerna. Företaget har en dramatisk historia med inre strider. Ledaren Sam Altman sparkades av styrelsen för att återkomma efter några dagar. Detta är skildrat i detalj av Karen Hao i hennes bok *Empire of AI*. Hon menar att AI blivit som en imperiestat och som alla imperier kommer det att gå under.

Fram till dess rullar biljonerna. 80 procent av börsvinsterna kom mer från AI. De sju teknikjättarna står för en tredjedel av de 500 största företagens värde. 2025 byggs AI-infrastruktur för 400 miljarder dollar. Den totala notan 2030 kan enligt vissa uppskattningar ligga på nära sju biljoner dollar.

## En klippa från Albanien

Utbyggnaden av datacenter gör att teknikjättarnas egna pengar inte räcker vilket innebär att lån blir dyrare, vilket i sin tur dämpar hela ekonomin. Samtidigt stiger priset på energi. AI-sektorn tar nu fem procent av den totala energin mot två procent för tio år sedan. Sedan 2019 har priset på el stigit med 40 procent i USA.

AI-cirkusen har styrts av några få personer. Två av dem är invandrare.

Jensen Huang föddes i Taiwan 1963. Han kom via Thailand till USA, började som 15-åring att arbeta som diskplockare på en snabbmatsrestaurang, Denny's. Där stannade han i fem år samtidigt som han studerade. Som 30-åring och med en master från Stanford grundade han tillsammans med två andra Nvidia. Det var vid Denny's bord som planerna för Nvidia drogs upp. Det billiga kaffet lockade. Nu är Jensen Huang världens åttonde rikaste. Han har inget tjänsterum, ingen klocka, har dubbelt medborgarskap och talar mandarin.

## Diskplockaren blev chipskung

Mira Murati föddes 1988 i det kommunistiska Albanien där regimen satsade på naturvetenskap i skolorna. Kommunismerna försvann, landet kastades in i kaos och Mira Murati fick kryssa mellan kratergroparna på sin väg till skolan. För att studera ville hon och vann internationella tävlingar. Det ledde till ett stipendium i Kanada när hon var 16 år. 2018 började hon på Tesla och blev en av de ansvariga för modell X. Med Teslas satsning på självkörande bilar väcktes Muratis intresse för AI och 2018 kom hon till Open AI som då var i startblocken med hundra anställda. Hon avancerade snabbt och blev teknologichef.

Trots framgången med ChatGPT 2022 blev det konflikter. Grundaren Sam Altman fick kortvarigt sparken. Murati som var van vid kaos från Albanien blev klippan som höll ihop företaget. I september 2024 lämnar Murati Open AI och mindre än ett halvår senare startar hon företaget Thinking Machines Lab. Hon hade plockat de bästa forskarna från Open AI och efter ett par månader värderades hennes företag till tolv miljarder dollar. Pengarna har fortsatt att strömma in. Albanien har satsat en del. I oktober 2025 kom företagets första produkt, Tinker, som är ett verktyg för att göra AI-modeller.

Under sin tid på Open AI tillhörde Murati den falang som betonade etik och säkerhet i AI-utvecklingen. Hon har fortsatt på det spåret och strävar inte efter ett AI som ersätter människan utan ska vara något som människan samarbetar med.

Amerikanska politiker varnar för att AI nu inte bara ersätter arbetare utan att en massarbetslöshet hotar nyutexaminerade akademiker, vilket kan skada hela det amerikanska samhället.

Det går som synes fort i AI-branschen. Biljonerna rullar i ett spel med många osäkerheter. Vad händer om energin inte räcker till? Är språkmodellerna en återvändsgränd? När kommer den oundvikliga nedgången?

Är världen redo för det? Peter Malmqvist är mångårig finansanalytiker. Han har sett bubbler växa och brista och säger att AI-trenden har skapat en bubbla

både i samhället och på börserna. Han berättar varför.

– För de stora teknikbolagen har lönsamheten stigit mycket snabbt de senaste två åren. Men tillväxten börjar nu att plana ut. Historiskt sett ligger lönsamheten på en toppnivå. It är en cyklisk bransch som drivs av olika investeringsteman som molntjänster, sakernas internet och AI, så kommer även denna trend att ebba ut. Det borde ske inom det kommande året eftersom allt just nu är på topp. Nedgången kommer att bli relativt dramatisk eftersom så mycket kapital är inblandat. Precis som vid it-bubblan 2000 kommer det nog att ta 15–20 år innan de investeringar som gjorts kommer att nå lönsamhetsmålen.

– På börserna är det en bubbla eftersom värderingarna på it-företag generellt, och inte bara teknikjättarna, är väldigt höga. Jag mäter i förhållande till vinsten och då kan jag konstatera att placerarna värderar inte bolagen som att de står på lönsamhetstoppen utan som att bolagen skulle fortsätta att växa i snabb takt. I alla fall de närmaste fem åren.

– Det tror jag som sagt inte att det kommer att göra, säger Peter Malmqvist. □



Peter Malmqvist, mångårig finansanalytiker, tror att nedgången kommer inom det närmaste året och då blir det dramatisk.

# ENKLARE AI OCH MASKININLÄRNING

## AI drivs av algoritmer

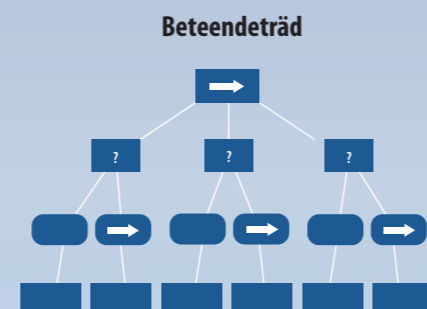
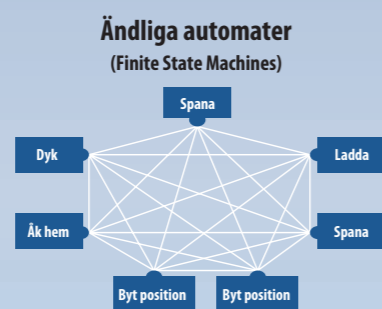
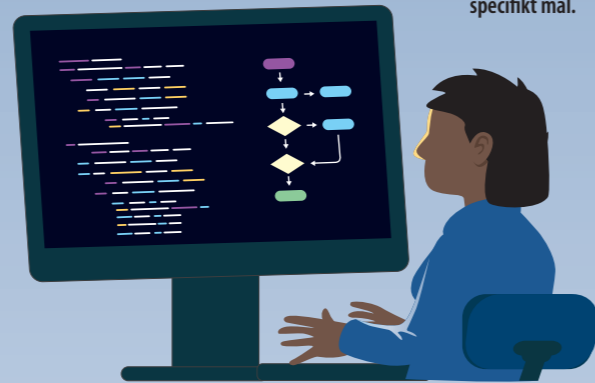
Algoritmer är i sig inte AI men de är grundläggande komponenter inom all AI. Algoritmer beskriver en serie instruktioner som är utformade för att utföra en specifik uppgift eller lösa ett specifikt problem.

### Enklare AI-modeller

Tidiga AI-system var ofta regelbaserade eller fattade beslut utifrån erfarenheter. Modellerna användes till enkla spel, optimeringsprogram, enkla beslutssystem och diagnostiska system.

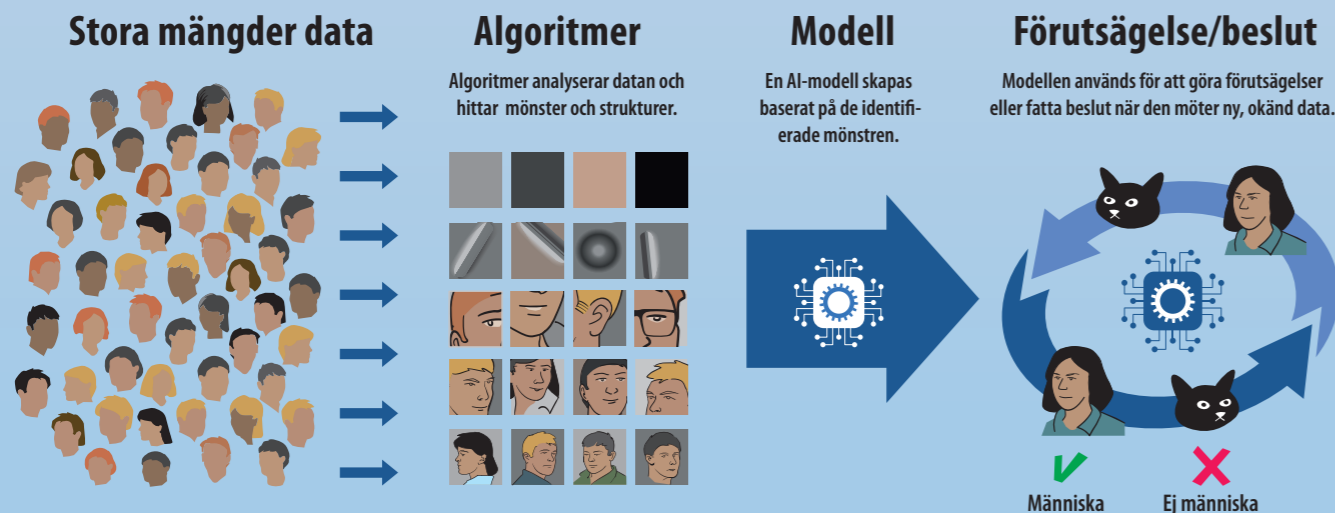
AI-modellerna används fortfarande och de kan vara lämpliga i situationer när beräkningskraften är begränsad. Dessutom är de transparenta, till skillnad från många generativa AI-modeller där man inte kan följa hur algoritmen har kommit fram till sitt resultat.

En algoritm kan beskrivas som ett programmeringsrecept för att uppnå ett specifikt mål.

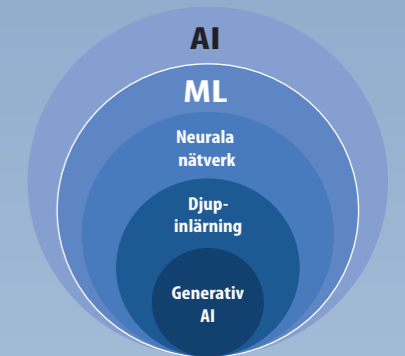


## MASKININLÄRNING (ML)

Maskininlärningsmodeller lär sig av data och kan utföra uppgifter utan direkta instruktioner. Modellerna konstrueras för att observera och lära sig regler för att lösa en uppgift. Målet är system som kan fatta beslut och lösa uppgifter med minimal mänsklig inblandning.



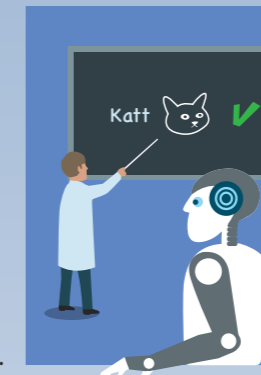
## INLÄRNINGSGRUPPER



Maskininläring är ett stort område och flera olika typer av inlärningsalgoritmer används – ibland i kombination – beroende på uppgift och tillgång på data. Maskininläring kan användas till bland annat tal- och bildigenkänning, anomalidetektion, självkörande bilar, medicinsk diagnostik och strömningstjänster på internet. Här är några vanliga typer:

### 1. Övervakad inläring (Supervised learning)

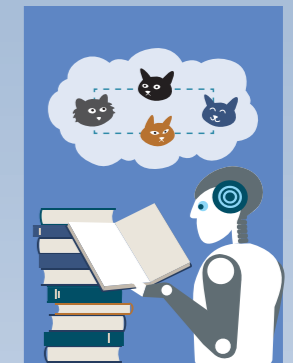
Modellen tränas på en uppsättning data där ingångsdaten – till exempel en bild – är märkt med rätt svar – till exempel en katt. Modellen justerar algoritmen för att minimera skillnaden mellan sina gissningar och de rätta svaren. Målet är att med tiden svara korrekt på ny okänd data. Förhållandevis dyrt och resurskrävande.



Används till klassificering, produktigenkänning, tal- och bildigenkänning.

### 2. Övervakad inläring (Unsupervised learning)

Modellen lär sig mönster och strukturer i data utan att använda någon märkning. Den använder metoder som klustring eller komprimering av datasetet för att upptäcka underliggande samband.



Används till att upptäcka dolda mönster eller strukturer i data och anomalidetektion.

### 3. Självövervakad inläring (Self-supervised learning)

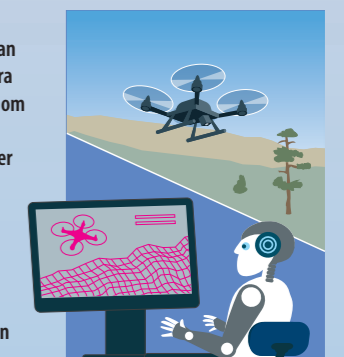
Modellen tränas på oövervakad, omärkt data men, skapar egna märkta etiketter (pseudouetiketter) utifrån sina egna prognoser. Exempelvis kan data-maskering användas för att dölja delar av indata, vilket tvingar modellen att rekonstruera dolda delar och uppfatta mönster i datan mer effektivt.



Används till träning av stora språk- och bildmodeller.

### 4. Överförd inläring (Transfer learning)

Eller "semiövervakad inläring". Kan vara användbar exempelvis i militära sammanhang om man inte har gott om träningsdata. Modellen tränar på fiktiv, simulerad data och förstärker därefter sina kunskaper med den riktiga data som finns tillgänglig.



Används till bild- och textklassificering när märkningen av datan är dyr eller svår att samla in.

### 5. Förstärkningsinläring (Reinforcement learning)

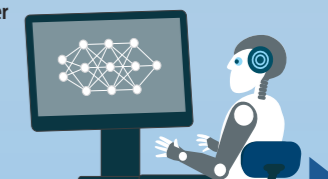
En agent interagerar med en miljö, tar beslut och utför handlingar. Utifrån sina handlingar får agenten belöningar eller straff. Genom att prova sig fram över tid lär sig agenten vilka handlingar och strategier som är framgångsrika.



Används till spel, robotstyrning och autonom körning.

### 6. Neurala nätverk och djupinläring

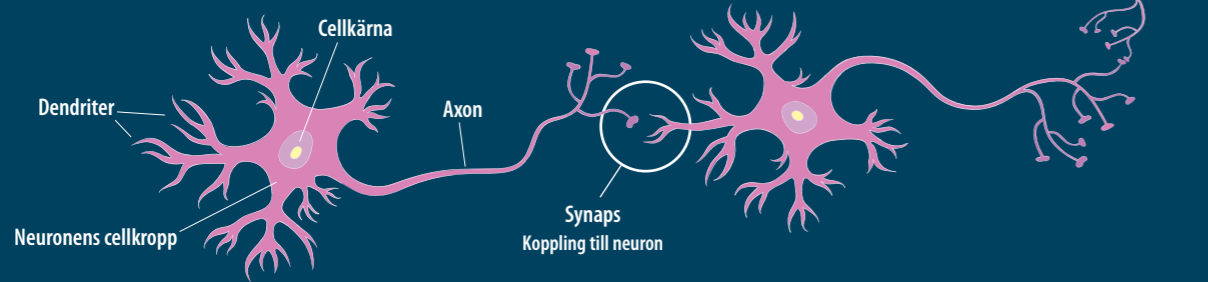
Neurala nätverk och djupinläring är avancerade former av maskininläring som använder artificiella neurala nätverk som liknar de man hittar i däggdjurs hjärnor. Även om neurala nätverk och djupinläring betraktas som egna kategorier så kan träningsmodellerna ovan, på olika sätt användas även när artificiella neurala nätverk- och djupinlärningsmodeller tränas.



MER OM NEURALA NÄTVERK PÅ SID 70–71

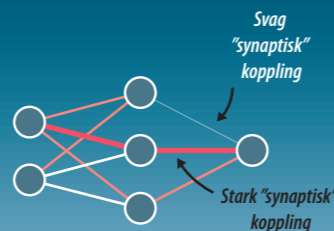
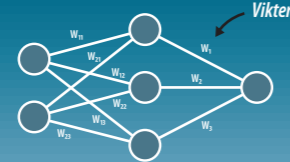
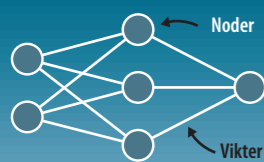
# NEURALA NÄTVERK

Den mänskliga hjärnan innehåller 100 miljarder neuroner (nervceller). Varje neuron kan bilda tusentals kopplingar till andra neuroner via synapser, vilket skapar ett komplext nätverk med upp till 1000 biljoner kopplingar totalt.



## Neuronnätverkens grundprincip

Begreppet "neurala nätverk" syftar ursprungligen på däggdjurs hjärnor där neuronerna bearbetar information och ändrar sin anslutningsstyrka via synapserna över tid för att förbättra inlärningsprocesser. De upprättar även nya anslutningar mellan neuroner vid inlärning. Inspirerad av detta efterliknar artificiella neurala nätverk denna struktur i form av algoritmer.



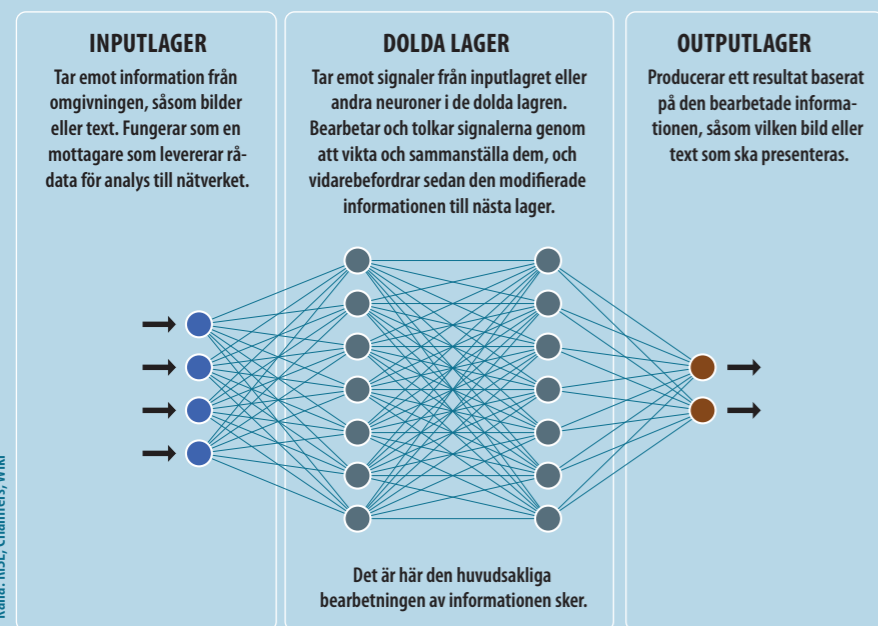
1. Motsvarigheten till neuronerna i hjärnan kallas noder i ett neuralt nätverk. Motsvarigheterna till de sammankopplade synapserna kallas vikter.

2. Vikterna är värden som finns i själva nätverksmodellen och används för att skapa kopplingarna mellan noderna. Det kan vara slumpmässiga värden när träningen av en modell startas.

3. Vikternas värden är inte fasta, utan påverkas av träningsprocessen där modellen justerar dem för att bli bättre på att känna igen mönster och för att komma fram till rätt resultat.

## ARTIFICIELLA NEURALA NÄTVERK

Ett artificiellt neurala nätverk är alltså en typ av adaptiv och iterativ maskininlärningsmodell som kan bearbeta, tolka, lära sig om och förädla inmatad information. Artificiella neurala nätverk består av tre huvuddelar:



## När de används

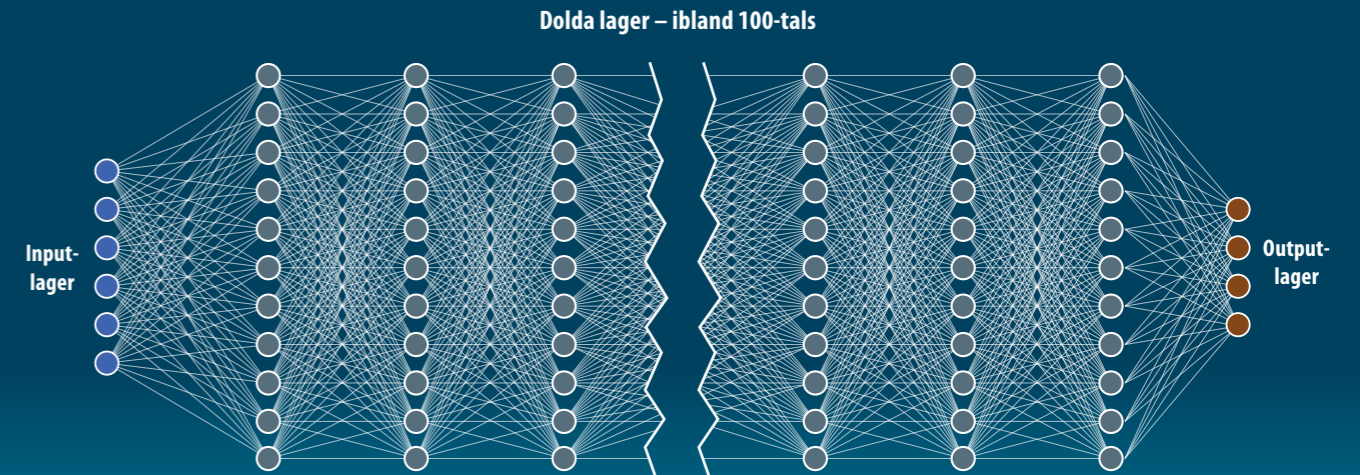
Det har varit stort fokus senaste åren på stora djupinlärningsmodeller, men mindre artificiella neurala nätverk kan fortfarande vara en bättre lösning i vissa sammanhang som:



- resursbegränsade miljöer och inbyggda system
- för enklare klassificerings- och regressionsproblem
- när data är liten och tydlig

# DJUPINLÄRNING

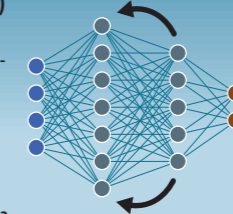
Djupa artificiella neurala nätverk – där "djup" refererar till antalet lager – består av många dolda lager, ibland hundratals eller till och med tusentals. De kan lära sig mycket mer komplexa mönster och representationer i data. Använder ofta avancerade arkitekturer som konvolutionsnätverk, rekurrenta nätverk och transformerbaserade modeller.



## AI-MODELLER FÖR DJUPINLÄRNING

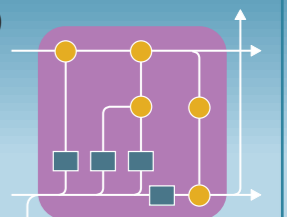
### Rekurrenta neurala nätverk

**Recurrent Neural Networks (RNN)**  
RNN-nätverk är speciellt designade för att hantera sekvensiell data, såsom för tidsserieanalys, språkmodellering, textgenerering och taligenkänning. Det vill säga data där ordningen i datan är viktig. RNN:er kan behålla ett tillstånd, eller "minne" av tidigare inmatningar i processen, vilket tillåter nätverket att lära från sig från tidigare indata.



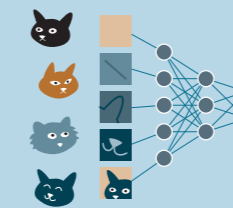
### Lång-korttidsminne

**Long Short-Term Memory (LSTM)**  
LSTM är en variant av RNN som bättre kan hantera problemet med försvinnande gradienter. Om en vanlig RNN-modell har ett korttidsminne så har LSTM ett "långt korttidsminne", därav namnet. Det gör den användbar för språkbehandling, översättning och andra tillämpningar där kontext och sammanhang är viktiga.



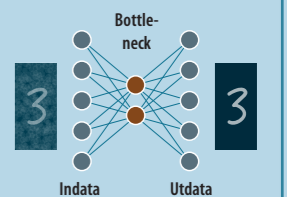
### Konvolutionellt neuralt nätverk

**Convolutional Neural Networks (CNN)**  
CNN-nätverk är särskilt bra på att bearbeta bilder och visuella data. Det använder bland annat så kallade konvolutionella lager för att upptäcka mönster, som kan vara kanter, former eller detaljer, och samlar stegvis information för att lära sig känna igen objekt eller mönster. CNN:n är effektiv för bild- och ansiktigenkänning och liknande uppgifter.

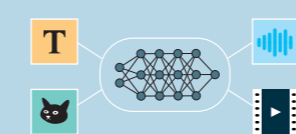


### Autoencoders

Autoencoders är en typ av artificiellt neuralt nätverk som lär sig komprimerade representationer av data, ofta för dimensionsreduktion och brusreduktion. Till skillnad från traditionella nätverk fokuserar de på att skapa en utdata som liknar indata. Ett centralt lager, kallat "bottle-neck", komprimerar datan och bevarar den viktigaste informationen.



## GENERATIVA AI-MODELLER

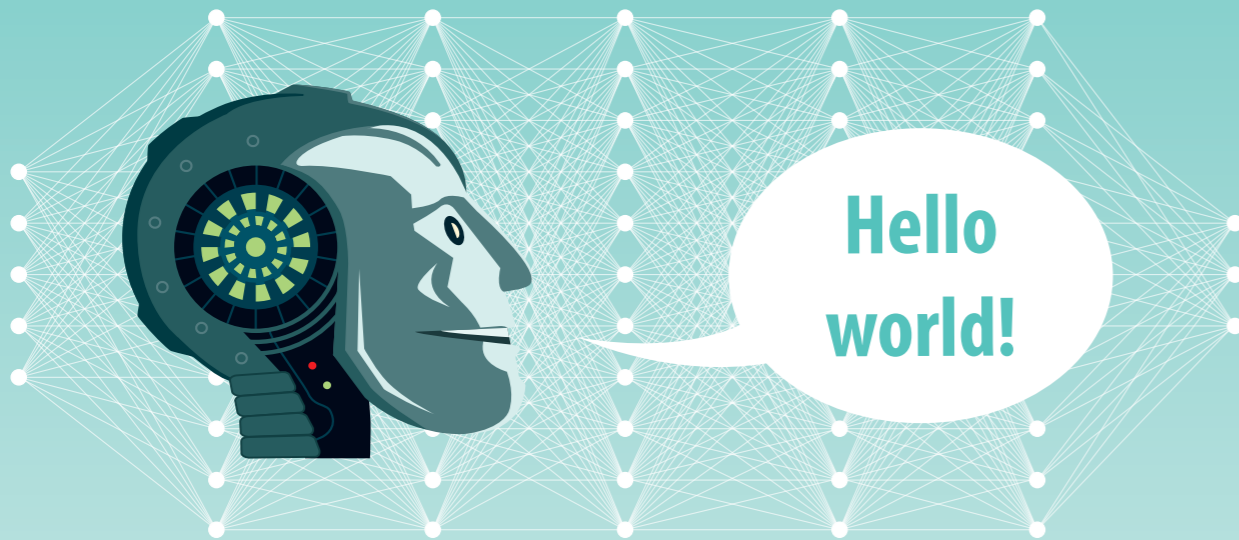


Generativ AI kan producera text, bilder, videor eller andra datatyper med hjälp av generativa modeller. Ofta görs detta utifrån specifika instruktioner och kommandon (promptar). Modeller analysar mönster och strukturer från träningsdata och kan sedan producera nytt innehåll.

**MER OM GENERATIV AI PÅ SID 72–73**

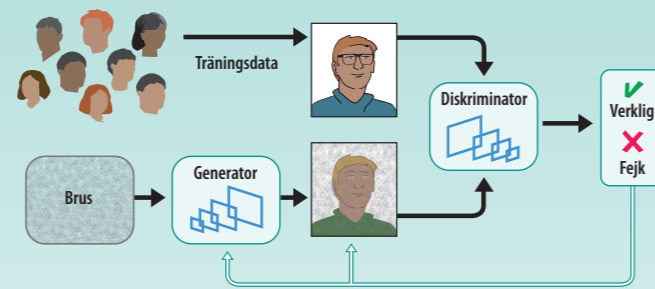
# GENERATIV AI

Generativa AI-modeller representerar ett snabbt växande forskningsområde inom artificiell intelligens och är modeller som kan skapa ny data som liknar den de tränats på. Nedan följer några exempel på modeller inom generativ AI.



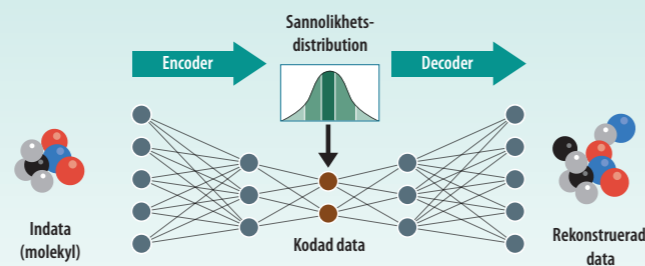
## Generative Adversarial Networks (GAN)

GAN är troligen, tillsammans med transformers, den mest kända tekniken inom generativ AI. I ett GAN tävlar två nätverk i motstånd mot varandra. En generator skapar exempelvis ett fotografi. En diskriminator bedömer, utifrån träningsdatasetet, om bilden är realistisk eller genererad. Ursprungligen togs GAN fram som en generativ modell för övervakad inlärning, men de fungerar även bra med andra inlärningsmodeller. Används för realistiska bilder, videor och även musik.



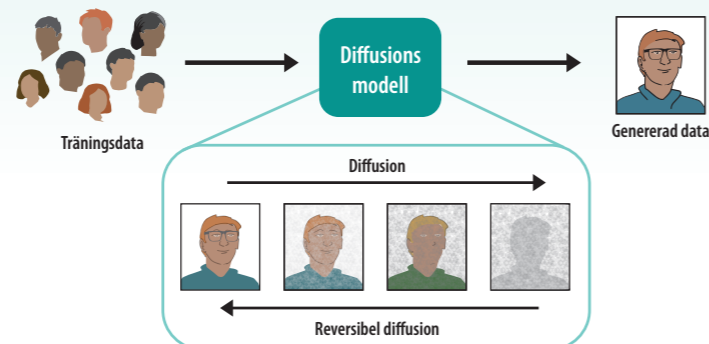
## Variational Autoencoders (VAEs)

Påminner om traditionella autoencoders, men med en funktion som gör dem generativa. Till skillnad från vanliga autoencoders inför VAEs en sannolikhetsdistribution i kodningsprocessen, vilket gör det möjligt att skapa ny rekonstruerad data genom att välja slumpmässiga exempel från denna distribution. De kan användas för uppgifter som att generera och förbättra bilder, samt för att utveckla nya strukturer inom kemisk och biologisk forskning.



## Diffusionsmodeller

Diffusionsmodeller utgör en nyare klass av generativa modeller. I detta sammanhang syftar begreppet "diffusion" på en process där data gradvis förändras över tid. Modellen har förmågan att omvandla en slumpmässig distribution till en datadistribution genom en reversibel diffusionsprocess. Dessa modeller har visat sig vara mycket effektiva för att skapa högkvalitativa bilder, och deras förmåga utforskas även inom andra områden, som ljudgenerering.



# Transformers och stora språkmodeller (LLM:s)

Tidigare modeller för språkbehandling, som RNN:s och LSTM:s, behandlar ord efter ord i sekvens vid språkbehandling. Detta gör det svårt för modellerna att förstå djupare sammanhang. När transformerarkitekturen introducerades 2017 innebar det ett avgörande steg för utvecklingen av chatbotar, som exempelvis GPT-familjen och PaLM 2, eller språkmodellen BERT. Kärnan i transformersmodeller är uppmärksamhetsmekanismen (*attention*), vilken möjliggör för modellen att selektivt fokusera på olika delar av en text för att bättre förstå dess sammanhang och innebörd.



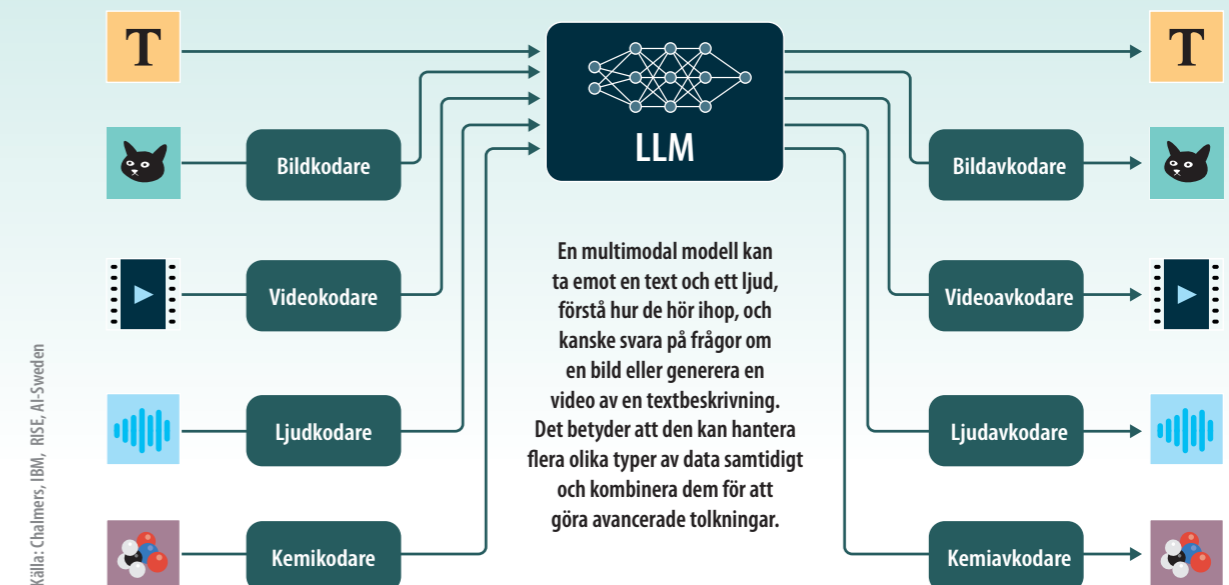
När en stor språkmodell får en uppgift, exempelvis en textprompt, konverteras texten till tokens, små segment som kan vara ord, delar av ord eller andra tecken. Modellen aktiverar sin självuppmärksamhetsmekanism och analyserar vilka tokens som är mest relevanta.

Genom analysera i flera lager kan språkmodellen väga de olika delarna mot varandra. Modellen kan, utifrån sin träning och erfarenhet, se hur de olika delarna hör ihop och på så sätt skapa en förståelse för uppgiften.

När modellen förstår sambanden kan den generera ett svar genom att statistiskt välja ord efter ord baserat på hur den sammanställt har löst uppgiften.

# Multimodala modeller

En utveckling av generativ AI är multimodala system. Dessa modeller processar och kodar olika datatyper separat på sitt eget sätt. Därefter kopplar modellen ihop de olika datatyperna i en gemensam förståelse, exempelvis med hjälp av en stor språkmodell.



En multimodal modell kan ta emot en text och ett ljud, förstå hur de hör ihop, och kanske svara på frågor om en bild eller generera en video av en textbeskrivning. Det betyder att den kan hantera flera olika typer av data samtidigt och kombinera dem för att göra avancerade tolkningar.

Källa: Chalmers, IBM, RISE, AI-Sweden





**FÖRSVARSMAKTEN**

**Rapporten är sammanställd av FoT-projektet  
Omvärldsbevakning med teknisk prognos.**

FMV dokumentbeteckning: 24FMV2096-35