Autonoma vapensystem – dagens debatt och en väg framåt
Tekniska, legala och etiska aspekter

Framstegen inom AI väcker frågor kring den militära tillämpningen av tekniken och tillåten grad av automatisering av vapensystem. Debatten om AI i militära tillämpningar förs både av civilsamhällets organisationer och av stater. Debatten började med det mycket specifika ifrågasättandet av fullt autonoma vapensystem, ofta kallade *Lethal Autonomous Weapon Systems* (LAWS), men debatten och ifrågasättandet har vidgats till att omfatta införandet av AI i militära tillämpningar. Diskussionen är mångfacetterad och präglas av flera olika perspektiv på ställningstaganden och argumentation. Autonoma system har många användningsområden i militära tillämpningar och det är viktigt att förstå hur integrationen av högautomatiserade system ska kunna göras med avseende på de rättsliga ramverk som försvaret lyder under.

Även om debatten till viss del har sitt ursprung i det senaste decenniets drönarkrigföring är de flesta parter överens om att dagens debatt snarare handlar om framtida teknik. De tekniska, militära och juridiska aspekterna är centrala i diskussionen men även etiska, psykologiska och säkerhetspolitiska aspekter tar plats i debatten.

Internationella initiativ

I FN:s forum för begränsning och användning av konventionella vapen för att diskutera autonoma vapensystem togs frågan upp 2014 och den har sedan 2017 diskuterats i en särskild expertgrupp. I forumet deltar ett 80-tal stater och flera organisationer från civilsamhället. Flera av civilsamhällets organisationer samlas också under paraplyorganisationen *Stop Killer Robots*. Några stater anser att expertgruppen ska sträva efter att uppnå ett rättsligt bindande instrument som förbjuder LAWS, medan andra inte ser något behov av ytterligare reglering utöver den som redan finns. En stor majoritet inom EU förespråkar en fortsatt konstruktiv diskussion som leder till exempelvis en politisk deklaration eller en uppförandekod som alla stater kan stå bakom. Under hösten 2019 antog expertgruppen elva vägledande principer för autonoma vapensystem och arbetet är tänkt att mynna ut i slutliga rekommendationer kring ett normativt och operativt ramverk på området.

Debattens kärna

I debatten förekommer framförallt två kärnfrågor. Den ena berör främst de autonoma systemens förmåga, eller bristande förmåga, att särskilja militära mål från civila och väga oavsiktliga förluster mot militär nytta. Detta handlar om en kvalitativ bedömning som många menar inte kan preciseras. Den bedömningen måste sålunda grundas på en förmåga som efterliknar mänskligt omdöme. Autonomi kan finnas i många olika funktioner i ett system. Frågan om distinktion och proportionalitet gäller huvudsakligen möjligheten för ett system att självständigt välja ut och anfalla ett mål.

Den andra frågan har sin grund i etiska resonemang och begreppet *human dignity*, mänsklig värdighet. Tanken att en maskin ska fatta beslut om liv och död är för många oacceptabel och den utgör ett argument mot autonoma vapensystem. Argumentet är sällan tydligt formulerat och då ett ställningstagande grundar sig på vilket etiskt system en person tar sin utgångspunkt i kan det vara svårt att diskutera. Det är sällan ett medvetet och explicit val utan snarare en intrapersonell egenskap. Ur ett juridiskt perspektiv har detta framförallt anknytning till förbudet mot förnedrande behandling och andra mänskliga rättigheter.

Är tekniken farlig, eller bara ny?

Om en ny teknik uppfattas som ett hot eller en möjlighet beror inte nödvändigtvis på den faktiska tekniken. Befintlig teknik, oavsett hur avancerad den än är, är ofta brett accepterad medan ny teknik kan uppfattas som antingen hotfull eller löftesrik beroende på personlig utgångspunkt. Det kan illustreras med att vissa argument för LAWS används av såväl förespråkare som motståndare. I vissa debattinlägg har

FOI Memo 6953
Forskningsområde: Övrigt
Godkänd av: Anders Lindström

LAWS förespråkats eftersom de inte har (negativa) mänskliga egenskaper. De blir inte rädda, stressade eller arga och kommer därför inte bryta mot lagar och regler. Men samma argument framförs också mot LAWS, att de saknar mänskliga (positiva) egenskaper och inte kan ha medkänsla för sina motståndare och till exempel inte bryta mot felaktig order och skona människoliv. Skönlitteraturen är rik på såväl löftesrika framtidsscenarier som dystopiska beskrivningar om tekniska system som utgör ett hot mot människan. De psykologiska aspekterna påverkar debatten, men kan ibland vara svåra att urskilja.

Gör autonoma vapen världen mer osäker?

LAWS är inte nödvändigtvis dödligare än eller har andra verkansdelar än andra vapensystem. Vissa anser att stater lättare kommer att ta till våld i internationella relationer om de, på grund av autonoma system, inte behöver riskera människoliv i den egna staten. Enligt detta argument skulle tröskeln för staters våldsanvändning alltså kunna sänkas. Vissa är också rädda att väpnade konflikter, på grund av autonoma system, skulle kunna starta helt oavsiktligt. Andra ger uttryck åt risken att vapensystemen sprids till fel personer, även om det är oklart huruvida LAWS i händerna på fel personer skulle vara värre än andra vapensystem. Medan vissa uttrycker en oro för en destabiliserande kapprustning argumenterar andra att LAWS i händerna på stormakter snarare skulle verka avskräckande på viljan att ta till våld.

Varför behövs automation i militära tillämpningar?

Det finns flera skäl att införa automation i militära tillämpningar. Minskad risk för egen personal förs ofta fram som en drivande kraft, men lika viktigt kan vara att öka prestanda i olika avseenden. Tidsprestanda är ett sådant exempel, att snabbt kunna bearbeta mycket information och agera när tiden mellan signaler från sensorer till aktivering av skyddssystem måste vara mycket kort, som vid ett robotanfall. Uthållighet är ett annat exempel, där automation kan lämpa sig för utdragna spaningsuppdrag vars upprepade och enformiga arbetsmoment är uttröttande för människor.

Vad är ett autonomt system?

En återkommande fråga i debatten kring LAWS är vad som definierar ett autonomt system. Vad skiljer egentligen autonoma system från de som finns idag och är det någon skillnad mellan automatiska och autonoma system? Ofta hamnar ordet autonom i fokus som en beskrivning på en egenskap som skiljer ett autonomt system från ett automatiskt. Ordet autonom har ingen entydig mening, betydelsen varierar beroende på sammanhang och i vilket område det används. Det är en avgörande skillnad mellan automatisk och autonom inom filosofin. Autonom beskriver en människas oberoende och moraliska rätt att fatta beslut. Inom statsvetenskapen betyder autonomi (en stats) självständighet från yttre kontroll. Den beskrivningen är mer relevant för tekniska system än den filosofiska definitionen.

Skillnaden mellan ett autonomt och ett automatiskt system är mer en semantisk fråga än en skillnad i teknik. Autonom är ett ord som ofta används för komplex automation medan automatisk är ett begrepp som används för mer välkända automatiska funktioner och vad som uppfattas som enkla system. Det är bättre att använda uttrycket *system med autonoma funktioner* istället för autonoma system. Att ett tekniskt system har en automatisk, eller autonom, funktion innebär att funktionen agerar utan påverkan av en operatör. Ett system kan vara autonomt i alla sina funktioner eller i en eller flera delfunktioner.

Det är inte bara benämningen autonom om tekniska system som kan skapa oklarhet i debatten. Att beskriva komplicerad teknik i termer av mänskliga egenskaper, så kallad *antropomorfisering*, är praktiskt. Det finns många exempel på detta inom AI-området, exempelvis ”lärande system” och ”maskiner som fattar beslut”, och andra liknande beskrivningar av tekniska system som leder tankarna till mänskliga egenskaper där sådana inte finns. Även om sådana formuleringar förenklar sättet att beskriva tekniken skapar det lätt föreställningar om, och förväntningar på systems förmåga som inte alltid stämmer överens med verkligheten. Ett antropomorft språkbruk kan därmed skapa förvirring och otydlighet kring systemens faktiska egenskaper. Debatten om LAWS skulle bli tydligare om människoliknande beskrivningar övergavs till förmån för mer relevanta tekniska sådana av de system som är i fokus. Sådana beskrivningar måste samtidigt vara begripliga för de utan tillgång till den tekniska begreppsapparaten.

FOI Memo 6953
Forskningsområde: Övrigt
Godkänd av: Anders Lindström

Vad innebär tekniken?

Teknik för automation har funnits länge och med datorer kan idag många komplexa funktioner automatiseras. Ett system med automatiska, autonoma, funktioner kan beskrivas på olika sätt. Gemensamt för alla dessa är att de har någon form av sensor som hämtar information från omgivningen, en matematisk modell av omgivningen och en metod, en algoritm, för att beräkna om den inhämtade informationen innebär att någon form av åtgärd behöver vidtas samt en metod för att utföra åtgärden. En matematisk modell av omgivningen kan till exempel bestå av kända fysikaliska samband, till exempel mellan acceleration, hastighet och position, men för komplexa system där sådana samband är svåra att beskriva på ett explicit och direkt sätt kan andra metoder vara nödvändiga.

En teknik som fått ökad uppmärksamhet och som ligger till grund för många av den senaste tidens framsteg inom artificiell intelligens är så kallad *maskininlärning*. Detta är ett exempel på en antropomorf term där ordet ”inlärning” i maskininlärning egentligen inte har något att göra med hur människor lär sig. Maskininlärning är ett samlingsnamn för ett antal tekniker som i grunden handlar om att identifiera strukturer i stora datamängder. Om det inte finns något enkelt sätt att beskriva den omgivning ett system ska verka i, i form av fysikaliska samband eller med matematiska begrepp, kan det ibland vara möjligt att ha en stor datamängd som beskriver sambanden på ett indirekt sätt. Exempel på när det används är system för att tolka bilder eller talat språk. Datamängden måste vara representativ för den miljö systemet ska verka i. Om datamängden (ofta kallad träningsdata eftersom systemet ”tränas” baserat på datamängden) inte beskriver miljön på ett relevant sätt kommer systemet inte fungera som avsett, utan spegla samma skevhet jämfört med verkligheten som använd träningsdata. Det finns flera exempel på sådana effekter, till exempel system för att tolka tal som är baserade på ett visst uttal och därför inte förstår personer med dialekter eller ett annat röstläge än det som användes för att utveckla systemet.

I debatten om LAWS har exempel på system som uppvisat oväntade brister väckt farhågor att vapensystem byggda med teknik baserad på maskininlärning inte kommer kunna uppfylla de krav som måste ställas på militära system. Ett sätt att beskriva maskininlärningsmetoder är att de fungerar som en ”svart låda”, vilket gör det svårt för en människa att förstå hela processen och kunna förutse hur tekniken kommer fungera i alla olika situationer. Det är emellertid inte en unik egenskap för system som bygger på maskininlärning, utan något som kan sägas karakterisera de flesta komplexa datorsystem. Stora forskningsansträngningar görs för att öka transparensen och förståelsen för komplexa system, där AI-tekniker är ett särskilt utpekat område.

Vad innebär meningsfull mänsklig kontroll?

Svårigheten att komma överens om en definition av LAWS och hur man ska kunna skilja LAWS från dagens system med autonoma funktioner har lett till att begreppet *meningsfull mänsklig kontroll* växt fram i debatten. Många aktörer är överens om att begreppet är användbart och att vapenanvändning helt utan mänsklig kontroll varken vore önskvärd eller tillåten. Begreppet används för att diskutera vilka krav som ska ställas på ett vapensystem, utan att behöva definiera det utifrån ett tekniskt perspektiv eller begreppet autonomi. Termer som ”meningsfull” och ”kontroll” är dock relativa, och trots bred konsensus kring begreppets användbarhet är det svårare att precisera dess exakta innebörd. I debatten återkommer vissa gemensamma krav som kan brytas ner till möjliga huvudkomponenter av begreppet meningsfull mänsklig kontroll. Exempel på dessa är bland annat krav på förutsägbarhet och tillförlitlighet hos tekniken och konsekvenserna av dess användning, förståelse för systemet och sammanhanget det används i, möjligheten att övervaka vapenanvändningen samt att ingripa genom att exempelvis avbryta ett anfall vid förändrade omständigheter.

Dessa huvudkomponenter av meningsfull mänsklig kontroll är användbara som utgångspunkt för fortsatta diskussioner om vilken teknik och användning som ska anses vara tillåten i fråga om autonom våldsanvändning. Men många frågor kvarstår vad gäller hur varje komponent ska tolkas i specifika situationer. Begrepp som förutsägbarhet, tillförlitlighet och förståelse är ord som saknar en bestämd betydelse i sammanhanget och måste tolkas vid varje konkret tillämpning. Termen ”förstå” har exempelvis olika betydelser för en programmerare som implementerat algoritmerna, en officer som planerar striden och en soldat som använder vapnet.

Kan meningsfull mänsklig kontroll utövas innan ett vapen aktiveras? Måste operatörer alltid ha möjlighet att avbryta en operation? En utmaning i debatten är att precisera vad meningsfull mänsklig kontroll innebär i förhållande till specifika frågor som dessa och tillse att begreppet förankras i den

FOI Memo 6953
Forskningsområde: Övrigt
Godkänd av: Anders Lindström

operativa verkligheten. Det får inte heller bli motsägelsefullt i relation till existerande icke-kontroversiella vapensystem, såsom exempelvis indirekt eld, där det inte är möjligt att avbryta ett anfall efter avfyring.

Vilka krav ställer folkrätten?

Det är okontroversiellt att fastslå att ny teknologi som vapensystem med autonoma funktioner måste kunna användas i enlighet med folkrätten. Frågan som debatteras är dock om existerande folkrätt är tillräcklig eller om kompletterande reglering, i form av specifika begränsningar eller ett förbud, är nödvändig. Det är viktigt att ha i åtanke att folkrätten ställer krav som människor och/eller stater ska uppfylla, inte vapensystemet i sig. Debatten om autonom våldsanvändning fokuserar till stor del på systemanvändningens förenlighet med krigets lagar som gäller i väpnad konflikt. All vapenanvändning måste uppfylla de grundläggande principerna om *distinktion, proportionalitet* och *försiktighet*. Enligt distinktionsprincipen ska åtskillnad göras mellan militära mål å ena sidan och civila personer och egendom å andra sidan och anfall får bara riktas mot militära mål. Proportionalitetsprincipen säger att oavsiktliga förluster, i form av civila personer som skadas eller dödas och civil egendom som tar skada eller förstörs, inte får vara överdrivna i jämförelse med den konkreta och direkta militära fördelen som förväntas med ett anfall. Försiktighetsprincipen fastslår bland annat att parterna i väpnad konflikt är skyldiga att göra allt praktiskt möjligt för att kontrollera att anfall riktas mot militära mål och att välja stridsmedel i syfte att undvika, eller i möjligaste mån minska, oavsiktliga förluster.

Relevant för debatten om autonom våldsanvändning är även de mänskliga rättigheterna, som är tillämpliga i både fred och krig. Under väpnad konflikt har krigets lagar företräde om regelverken motsäger varandra, men i praktiken kompletterar de snarare varandra. De mänskliga rättigheterna blir gränssättande regelverk om vapensystem med autonoma funktioner används i fredstida situationer. Exempel på mänskliga rättigheter som kan bli aktuella vid autonom våldsutövning är rätten till liv och förbudet mot omänsklig eller förnedrande behandling. Enligt Europakonventionen måste all våldsanvändning vara absolut nödvändig för att till exempel försvara någon mot en olaglig våldshandling eller för att verkställa en laglig arrestering. Europadomstolen har i sin tur fastställt att våldsanvändningen måste vara proportionerlig i relation till det legitima ändamålet med den och ställer krav på att planeringen av operationer som innebär våldsanvändning måste hålla en viss standard för att uppfylla kravet på absolut nödvändighet. Rätten till liv innebär också en processuell skyldighet för stater att effektivt utreda potentiella kränkningar till följd av våldsanvändning, vilket kan tänkas ställa krav på transparent teknologi vars konsekvenser kan förutses och förklaras.

Kan juridiska beslut automatiseras?

Tillämpning av folkrätten på användningen av vapensystem med autonoma funktioner handlar i viss mån om möjligheten att automatisera juridiska beslut, om rättsautomation. Juridiska beslut fattas med stöd av rättsregler och kan bestå av allt från okomplicerade delbeslut till avancerade rättsavgöranden. Huvuddelen av befintliga tekniska system stödjer enstaka delmoment i juridiska beslutsprocesser. Ett avancerat rättsavgörande skulle kräva att varje moment i en beslutsprocess identifieras och att system utvecklas som kombinerar flera tekniska lösningar. Vilka juridiska beslut som är möjliga att automatisera är otydligt, men ett utgångsläge är att något som ska automatiseras måste kunna beskrivas. Ju mer komplext ett beslutsunderlag är, desto svårare blir det att automatisera beslutsprocessen. Idag fattas många juridiska beslut om saker som går att räkna, helt utan mänsklig inblandning. System med autonoma funktioner räknar och fattar beslut som till exempel skattelagstiftningen föreskriver.

Att exempelvis särskilja militära mål från civila och väga oavsiktliga förluster mot militär fördel kräver dock kvalitativa bedömningar som många menar att ett tekniskt system omöjligtvis kan utföra med tillfredsställande resultat. Till exempel kan militära mål under vissa omständigheter övergå till att vara skyddade och på motsatt vis kan civila övergå till att vara legitima mål. Militär fördel kan i sin tur omfatta flera olika hänsynstaganden, variera under tidens gång och bero på såväl utvecklingen av konflikten som befälhavarens planer. Att automatisera juridiska beslut i dynamiska miljöer är alltså ytterligare en utmaning, vilket tyder på att oföränderliga eller åtminstone stabila miljöer lämpar sig bättre för rättsautomation. Debatten kring rättsliga förutsättningar för autonom våldsanvändning måste därför utgå bland annat från huruvida det är ett delmoment eller en hel beslutsprocess som ska automatiseras, hur beslutsunderlaget kan beskrivas och miljön beslutet ska fattas i.

Vem bär ansvaret när något går fel?

En del av debatten kring vapensystem med autonoma funktioner handlar om ansvarsutpekande och den återkommande

FOI Memo 6953
Forskningsområde: Övrigt
Godkänd av: Anders Lindström

frågan om vem som bär ansvaret för eventuella oönskade följder av användningen. Vissa uttrycker en rädsla för ett ansvarsvakuum. Detta beror inte nödvändigtvis på bristen på straffrättsliga regler, utan kanske snarare på utmaningen i att tillämpa reglerna på användningen av ett system med autonoma funktioner. Enligt krigets lagar kan varje individ hållas straffrättsligt ansvarig för eventuella krigsförbrytelser och militära befälhavare kan i vissa fall också hållas ansvariga för krigsförbrytelser som individer under deras befäl begår. Vad gäller de mänskliga rättigheterna kan bara stater hållas ansvariga för kränkningar. Innan Europadomstolen för de mänskliga rättigheterna tar upp ett mål om kränkning till prövning måste alla nationella rättsmedel ha uttömts. För debatten kring användningen av vapensystem med autonoma funktioner är det därför av intresse att titta på hur olika staters rättssystem skulle hantera ansvarsutpekandet i dessa fall. I debatten diskuteras om och hur ansvarsfrågan påverkas och hur komplicerat ansvarsutpekandet kan bli med tanke på komplexiteten hos ett system med autonoma funktioner.

Folkrättslig granskning av vapen

Alla stater är skyldiga att granska om användningen av ett nytt vapen, under vissa eller alla omständigheter, skulle vara förbjuden enligt statens folkrättsliga förpliktelser, en så kallad *artikel 36-granskning*. Med ”nytt vapen” menas till exempel att staten utvecklar eller köper ett nytt vapen, modifierar ett befintligt vapen eller börjar använda det på ett sätt som inte tidigare förväntats. Granskningen ska baseras på normal användning av vapnet, som förväntas vid bedömningstillfället. Stater är inte skyldiga att förutse eller analysera all tänkbar användning av ett vapen, eftersom i princip alla vapen kan missbrukas på ett eller annat sätt. Granskningen kräver en bedömning av all relevant information som rör vapnet och en analys av det sammanhang där det är tänkt att användas. Granskningen måste alltså svara på huruvida vapnet med dess egenskaper, såsom effekter, vid förväntad användning i vissa förutsägbara situationer och under vissa specifika omständigheter kan användas i enlighet med gällande folkrätt. Komplexiteten i vapensystem med autonoma funktioner kan göra denna folkrättsliga granskning svår och en särskild vägledning för hur detta ska gå till kommer förmodligen behövas.

Vägen framåt

Debatten kring autonom våldsanvändning är mångsidig och komplex, såväl perspektiv på ställningstaganden som argument för och emot användningen är åtskilliga. Det är av stor vikt att förstå vilka begränsningar rättsliga ramverk sätter på användningen av vapensystem med autonoma funktioner. Innebörden av meningsfull mänsklig kontroll och hur det kan införas eller beskrivas för olika typer av högautomatiserade system är en fortsatt viktig fråga. Detta då det är svårt att ur ett tekniskt perspektiv definiera vad som är eller bör vara förbjudet respektive tillåtet. En annan viktig fråga är hur en artikel 36-granskning av nya vapen bör vara utformad. Den folkrättsliga granskningen kan förväntas bli mer komplex för framtida vapensystem med komplex automation då kraven på systemen, organisation och metoder för användning kommer vara mer omfattande än för dagens vapen. Ett systematiskt arbete bör därför inledas för att fördjupa kunskapen om och förståelsen för en utvecklad artikel 36-granskning av komplexa vapensystem.

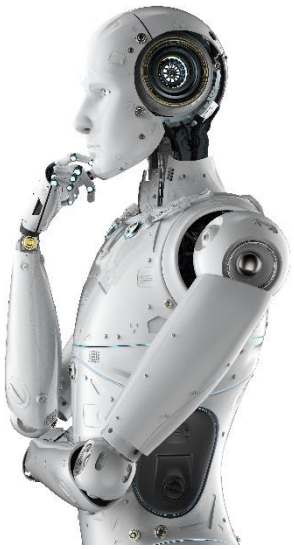

Foto: Shutterstock

Författare:
Jessica Appelgren
Tâm Beran
Amanda Musco Eklund
Martin Hagström

FOI Memo 6953
Forskningsområde: Övrigt
Godkänd av: Anders Lindström