



JURIDISKA FAKULTETEN
vid Lunds universitet

Jakob Rydin-Malmqvist

Autonoma Vapensystem och folk- rätten

Problem, definition och överblick

LAGF03 Rättsvetenskaplig uppsats

Kandidatuppsats på juristprogrammet

15 högskolepoäng

Handledare: Aurelija Lukoseviciene

Termin: HT 23

Innehåll

SUMMARY	1
SAMMANFATTNING.....	2
FÖRKORTNINGAR	3
1 INLEDNING.....	4
1.1 Bakgrund.....	4
1.2 Definition av AMS.....	5
1.3 Syfte och frågeställning	9
1.4 Avgränsningar	11
2 AUTONOMA VAPENSYSTEM OCH IHR	13
2.1 En introduktion till IHR och jus in bello	13
2.2 Frågan om "responsibility gap" och det s.k. "black box" – problemet.....	15
2.3 Den Internationella rätten och användning av AMS	18
KÄLLFÖRTECKNING	24

Summary

This paper examines the issue of autonomous military systems and their unique challenges for international law, with a focus on the use and development of these systems in the framework of international regulation regarding “jus bello”. It also examines the issue of which definition of “autonomous military systems” should be used. The paper examines the question if there is a so called “responsibility gap” which is unique to AMS and how this relates to the question of individual criminal liability as well as the issue of state-responsibility.

The paper gives the reader a short introduction to international law in regards to this issue, and what suggestions have been brought forward regarding its regulation. It asserts that it should be possible to demand some form of liability for violations of international law, either on an individual level or as a last resort, in the form of state-responsibility. Furthermore, it concludes that this issue is in no way unique for AMS, but that a so called “responsibility gap” can happen even with the use of other weapons and due to other circumstances. Finally, it concludes that the issue is one that is under development, and presents some proposals that have been brought forward in academic literature and in the public debate on the subject.

Sammanfattning

Den här uppsatsen går igenom Autonoma militära system och dess unika utmaningar för den internationella rätten. Med ett fokus på användning, och utveckling av dessa inom ramen för krigets lagar angående ”jus bello”. Uppsatsen utreder också frågan om vilken definition som bör användas, om det åligger ett s.k. ”responsibility gap” som är unikt för AMS och hur detta förhåller sig till frågan om individuellt straffrättsligt ansvar och frågan om statsansvar.

Inom ramen för uppsatsen ges läsaren en kort introduktion till allmän folkrätt och hur det förhåller sig till användningen samt utveckling av AMS och vad det finns för tankar angående reglering inom ämnet. I uppsatsen konstateras att det med nuvarande internationella regelverk borde gå att utkräva ansvar i åtminstone någon form på antingen ett individuellt plan eller som sista utväg, statsansvar och att denna fråga inte är unik för just AMS. Utan att ett s.k. ”responsibility gap” kan föreligga av fler anledningar och inom andra situationer. Slutligen så konstateras det också att frågan är under utveckling och uppsatsen presenterar vissa förslag på reglering som framkommit inom facklig litteratur och den allmänna debatten.

Förkortningar

AMS	autonoma militära system
ARSIWA	Responsibility of States for Internationally Wrongful Acts (FN 2001: resolution 56/83)
ASS	Autonoma Spaningssystem (Autonomous Scouting Systems)
AVS	autonoma spaningssystem (Autonomous Scouting Systems)
AVS	autonoma vapensystem
CCW	FN:s konvention av 1980 om vissa särskilt inhumana vapen. (SÖ 1982:27)
ICJ	International Court of Justice (Internationella domstolen i Hague)
IHR	Internationell humanitär rätt
NGO	Non-governmental organization (icke-statlig organisation)

1 Inledning

1.1 Bakgrund

Autonoma vapensystem (AVS) innebär enligt många en revolution inom krigsföringen. Systemens förmåga att på egen hand välja mål, sortera data, spaningsinformation, och sammanfatta olika typer av underrättelseinformation kan innebära har inneburit och kan komma att innebära ett stort steg framåt för militärens förmåga att välja mål och sortera den oerhörda mängd underrättelseinformation som ofta står att finna på ett kaotiskt slagfält.¹ I en nyskriven rapport ifrån försvarsberedningen för år 2023 konstateras det att AI-teknologi förväntas innebära att militära operationer kommer kunna ske i ett snabbare tempo och leda till en ökad samordning mellan olika förband.² Dess fram tills nyligen inom militärens outnyttjade förmåga att kunna operera helt autonomt utan mänsklig styrning eller kontroll innebär också en för militären helt ny typ av förmåga som kan komma att innebära ett stort steg framåt för att kunna undgå t.ex. elektronisk krigföring som stör ut drönaroperatörens kommunikation med drönaren och en förmåga att agera som en s.k. drönarsvärm eller för ett luftvärnsbatteri att på egen hand välja och avfira motmedel mot en inkommande robot eller dylikt militärt mål.³ Men även inom informationskrigsföringen har AI kommit att spela roll genom att både hjälpa till med att samla in, lägga ut information, underrättelser och genom att skapa samt sprida propaganda på internet.⁴

Men denna nya teknologi ställer också till ett nytt problem för den internationella humanitära rätten, vem är ansvarig för händelse av brott mot folkrätten och den IHR? Och hur ska AI-teknologi kunna utvecklas på ett säkert sätt utan att bryta mot internationell rätt?

¹ Se t.ex. Aided Detection on the Future Battlefield (army.mil) och Boutin, Berenice (2022) s. 137

² Försvarsberedningens säkerhetspolitiska rapport 2023 Ds 2023:19 s. 38

³ Ukraine AI Drones Seek, Attack Russian Forces Without Humans Oversight (forbes.com)

⁴ Ukraine's Secret Weapon - Artificial Intelligence (cepa.org)

Det finns idag redan ett stort antal åsikter i frågan som står att finna i t.ex. policydokument framtagna av amerikanska regeringen,⁵ det nyligen antagna ”Call to Action” från REAIM 2023, antaget av över 60 länder,⁶ inom juridiska journaler,⁷ blogginlägg från akademiker,⁸ uttalanden från NGO:s⁹ och akademisk litteratur¹⁰.

Men åsikterna går isär om hur man skall gå tillväga och i vilken utsträckning, och om det behövs reglering i frågan. En intressant fråga som väcks är frågan om statsansvar i frågan om användande av AVS på slagfältet och dess utveckling/upphandlande av staten innan. Medan vissa menar att det existerar ett s.k. ”responsibility gap” när det kommer till ansvar för brott begångna av robotar och andra autonoma AI-system,¹¹ menar andra att det går att utkräva statsansvar på båda dessa områden redan nu med redan befintliga internationella konventioner såsom ARSIWA.¹²

Frågan är omstridd och väcker många intressanta juridiska och filosofiska aspekter, om vem som är ansvarig för brott eller skador som uppkommit till följd av autonomt handlande av ett AI-system. Även frågan om det går tillskriva juridisk handlingskraft och ansvar från AI-systemet självt. Men denna fråga väljer jag att lämna utanför omfattningen av denna uppsatsen då det är en alltför teoretisk diskussion som skulle ta alldeles för lång tid att gå igenom.

1.2 Definition av AMS

⁵ Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy - United States Department of State, Global summit calls for ‘responsible use’ of AI in the military (theiet.org)

⁶ REAIM Call to Action, Global summit calls for ‘responsible use’ of AI in the military (theiet.org)

⁷ Boutin, Berenice (2022) s. 137

⁸ The law and the U.S.’ new declaration on military uses of AI: some observations – Lawfire (duke.edu)

⁹ Key elements of a treaty on fully Autonomous Weapons – (stopkillerrobots.org)

¹⁰ Seixas-Nunes, Afonso (2022)

¹¹ Danaher, J (2016) och Matthias (2004)

¹² Boutin, Berenice (2022) s. 144-145

Men frågan om vad ett AVS är, är viktigt att redan nu klarlägga för att bringa klarhet i vilka system som avses med AVS i denna uppsats och vilka frågor dessa väcker inom den Internationella humanitära rätten (IHR). Enligt amerikanska försvarsdepartementets direktiv 3000.09¹³ så är AVS ett system som:

När det aktiverats på egen hand kan välja och anfälla mål utan vidare instruktioner från en mänsklig användare. Generellt sätt så kan man säga, att ett AVS är främst ett system som har kapacitet att identifiera och anfälla ett militärt mål utan att det krävs mänsklig övervakning eller instruktioner.

Det är en intressant definition som vid första anblick kan kännas lockande men den har också sina problem som jag kommer gå igenom i större detalj senare i detta avsnitt. Men för frågan om statsansvar vid användande av AVS avser jag inte att enbart stanna vid system som har både kapacitet att identifiera och anfälla ett mål utan mänsklig övervakning eller instruktioner. Utan att även gå igenom frågan om statsansvar vid användning av vad jag från och med nu väljer att definiera som ”autonoma spaningssystem”(ASS) och därmed särskiljer från AVS-termen i dess generella helhet. Dessa system använder AI för att samla in samt gå igenom underrättelseinformation på slagfältet som sedan används av en mänsklig aktör för att välja mål på slagfältet. Jag anser att det finns en avgörande skillnad i användningen av dessa två system och att det därför är viktigt att klarlägga skillnaden mellan dessa och gå igenom frågan om individuellt ansvar vid operativt användande av dessa, frågan om ansvar vid design och uppköp samt frågan om statsansvar vid användande av dessa. Skillnaden mellan användningen av dessa typer av system samt dess potentiella implikationer för både statsansvar samt individuellt ansvar vid användande av dessa har tidigare nämnts i den vidare debatten, men särskiljande termer såsom jag kommer att använda i denna uppsatsen, har inte vad jag kunnat se, tidigare uppkommit.¹⁴ Jag kommer därmed hädanefter att använda en ny samlingsterm för båda systemen, nämligen autonoma militära system eller AMS förkortat.

¹³ Se: USA Department of Defense Directive 3000.09, citerad och av uppsatsförfattaren översatt sammanfattning ifrån: Seixas-Nunes, Afonso (2022) s. 13

¹⁴ Boutin, Berenice (2022) s. 142-143

Definitionsproblemet för AMS är också genomgående när man får en översyn av litteraturen inom ämnet och exemplifieras bäst av FN organet GGE:s misslyckande att få till en sådan under sitt arbete med frågan under åren 2014–2019.¹⁵

Det finns många olika sätt att definiera autonomi för ett vapen eller spanings-system, t.ex. så menar amerikanska regeringen och röda korset att alla militära system som kan ta beslut utan mänsklig kontroll eller inverkan i någon mån är autonoma. Denna definition är mer vidsträckt än t.ex. Storbritanniens definition som menar att ett autonomt vapensystem som menar att ett sådant system måste kunna förstå ”higher level intent and direction” eller kinesiska regeringen definition som är att ett sådant system måste kunna ”learn autonomously, expand it’s functions and capabilities in a way exceeding human expectations”.¹⁶

Det bör noteras att den definition som röda korset och amerikanska regeringen har valt att använda även inkluderar olika former av militära system som varit i bruk under en längre tid, bl.a. luftvärnssystem som det amerikanska ”patriot”, israeliska ”iron dome”, missilsystem såsom AMRAAM och brimstone samt s.k. ”loitering munitions” (på svenska ibland kallat patrullrobot) såsom harop och harpy. Men även mer primitiva system såsom landminor skulle kunna falla in under denna definitionen, vilket gör att den i mina ögon blir problematisk, då det finns en stor och avgörande skillnad mellan ett system som bara saknar mänsklig kontroll, system som jag snarare skulle vilja kalla automatiska än de facto autonoma på ett för IHR meningsfullt sätt.

Detta eftersom dessa system istället för en djupare form av självständigt tänkande och ordentlig autonomi snarare kan sägas ha en reaktiv form av intelligens. Det bör i detta sammanhang självklart noteras att det finns en gradskillnad i den kognitiva förmågan mellan en primitiv landmina och mer avancerade system såsom brimstone, där brimstone kan anpassa sig till situationen

¹⁵ Galliot, Jai, MacIntosh, Duncan och Ohlin (2021) s. 107

¹⁶ Ibid s. 107

och omgivningen på ett betydligt djupare plan med hjälp av de förprogrammerade processer som lagts in i dess mjukvara och med hjälp av dessa välja vilken typ av sensor den använder sig av, vilket mål den väljer och vilket sätt den väljer att ta sig an situationen. ¹⁷

Dessa system som visserligen kan ta in information från omgivningen och ändra sitt beteende beroende på omgivningens signaler men bara på ett sådant vis som är förprogrammerat och systemen kan därmed inte anpassa sig till oväntade situationer eller självt komma på nya sätt att ta sig an ett problem.

I detta sammanhang kan det vara värt att notera att autonomi som begrepp är svårdefinierbart och inte lätteligen kan anses ha en binär innebörd som ett system som är antingen autonomt eller inte utan snarare bör ses som ett spektrum, för att citera ifrån boken: ”Lethal Autonomous Weapons”:

the range of technological capabilities suggests that ‘autonomy’ cannot be conceptualized as a binary concept— in other words, a system being either autonomous or not. Rather, autonomy is a spectrum and any specific system, or more specifically its various functions, may sit at different points of that spectrum. ¹⁸

Det är därmed ingen sak som man kan ta lätt på, vilken definition man väljer att ha för ett autonomt militärt system kan ha starka juridiska, samhälleliga och tekniska konsekvenser längre fram.

För att förtydliga skillnaderna så skulle jag också vilja att på ett någorlunda kort och koncist vis förklara skillnaderna mellan ett system som styrs av en mer autonom, maskinlärande AI och en mer traditionellt deterministisk ”autonom” maskin som styrs av deterministiska algoritmer.

Maskininlärning går ut på att man ”matar” en mjukvara med data som mjukvaran sedan använder för att lära sig att tänka på ett mer autonomt och självständigt plan än traditionellt kodad mjukvara som bara kan reagera med hjälp av de förutbestämda mänskligt kodade algoritmer som har kodats in i den.

¹⁷ Seixas-Nunes, Afonso (2022) s.112

¹⁸ Galliot, Jai, MacIntosh, Duncan och Ohlin (2021) s. 107

Detta innebär en stor skillnad mellan de olika systemen på flera plan, för det första så innebär det att en mjukvara med maskininlärning kan reagera och agera i alla fall någon mån självständigt och anpassa sig till nya oförutsedda situationer som skaparen av algoritmen inte har kodat in i den. Medan en mer traditionell utformad mjukvara är beroende av instruktioner i form av kod som programmeraren har lagt in i den. Självklart är denna i någon mån en förenkling av en mer komplicerad process, men räcker gott och väl för denna uppsatsens syfte.¹⁹

Med andra ord så står vi nu inför en ny era med system som kan förändra sitt eget beteende utan förprogrammerade instruktioner, för att kunna utföra sitt uppdrag och nå sitt mål. Dessa system kan sägas vara autonoma på ett djupare plan än de system som vi har sett tidigare men når inte upp till den nivå av autonomi som vi hittills bara kan spekulera om eller synas inom science fiction, där den artificiella intelligensen kan ifrågasätta sitt mål och sitt syfte och därmed avgöra själv vad den vill göra och på ett än djupare plan hur detta ska gå till än de nuvarande system.²⁰

Det går därmed att ifrågasätta om det i dagens läge går att prata om en autonomi där den artificiella intelligensen kan anses ha ansvar för sitt eget handlande och behöva hållas ansvarig för detta inom IHR. Frågan om det kan ge upphov till ett s.k. ”responsibility gap” är något jag kommer att gå in i mer detalj senare.

1.3 Syfte och frågeställning

Uppsatsen syftar till att klargöra relationen mellan Autonoma militära system och IHR. Specifikt de viktiga principerna inom IHR som reglerar ”jus in

¹⁹ Seixas-Nunes, Afonso (2022) s. 23-24

²⁰ Ibid s.112-115

bello”, dvs. användningen av militär styrka i konfliktsituationer. Min ambition är att försöka presentera en nulägesbild av den rättsliga reglering som vi kan komma att vänta inom internationell rätt av dessa relativt nya militära system. Jag kommer också försöka besvara frågan om det finns ett s.k. ”responsibility gap” och om systemen kan anses vara autonoma till den grad att de antingen själva kan anses ansvariga för sitt handlande och hur detta autonoma handlande kan påverka möjligheten att hålla operatören, ingenjören eller staten som använder systemet ansvariga. Frågeställningen som jag har är därmed följande:

I vilken mån är autonoma militära system kompatibla med nuvarande regler inom IHR angående jus in bello, finns det ett s.k ”responsibility gap” problem inneboende i de nya AI-systemen, och kan de anses vara autonoma så till den grad att systemen själva kan behöva hållas ansvariga?

Metod och material

Eftersom uppsatsens syfte är att försöka bringa klarhet i vad som är gällande rättslig relation mellan IHR och AMS så kommer jag att använda mig av rättsdogmatisk metod. Rättsdogmatik innebär att man beskriver gällande rätt genom tolkning och systematisering av denna. Eftersom rättsdogmatiken har till syfte att beskriva gällande rätt så innebär det att man som grund ska kolla på de auktoritativa källor som i enlighet med rättskällevärdet är relevanta för att besvara frågeställningen.²¹

Eftersom internationell rätt inte har sin grund i nationella lagar utan bygger på internationella källor såsom praxis från internationella rättsliga organ såsom ICJ, och uttalanden samt konventioner/fördrag ifrån internationella organ såsom FN och EU samt olika mellanstatliga förbindelser emellan stater, så är det dessa som jag kommer kolla på och försöka analysera. Jag kommer också vända mig till rättsvetenskapliga källor såsom artiklar i akademiska journaler och akademisk litteratur inom ämnet. Eftersom AMS är nytt både som fenomen och speciellt inom internationell rättslig reglering så finns det

²¹ Se Sandgren (2018) s. 49-51

inte mycket av stora bindande konventioner/fördrag²² att kolla på med egentlig rättslig verkan eller någon för AMS specifik rättspraxis. Därför kommer rättsvetenskaplig litteratur, artiklar, uttalanden från nationer, nationella organ samt NGO:s få än större betydelse i den här uppsatsen. Jag kommer även använda mig av vanlig nyhetsmedia för att beskriva omvärldsläget och den geopolitiska samt rättsliga relevansen för AMS. Även en inom den för debatten relevanta artikel inom filosofi kommer att användas.

Det finns de som menar att den rättsvetenskapliga litteraturen plats i rättskälleliteraturen är omdiskuterad.²³ Men eftersom AMS som faktiskt användbart system är relativt nytt och det inte finns många av de mer tunga rättskällorna att vända sig till, och eftersom litteraturen fyller sin plats genom tyngden och kraften av sina argument och kan anses ha en potentiellt framtida rättsskapande kraft genom dessa så anser jag att dessa har sin plats, och är därmed i denna uppsats högst relevanta. Den fyller också funktionen att lyfta fram olika tolkningar och perspektiv av det rättsliga problem som användning av AMS inom krigsföring kan anses ha och angående frågan om ett s.k. ”responsibility gap” föreligger.

1.4 Avgränsningar

I denna uppsats har jag valt att förhålla mig till AMS inom ramen för framförallt några av de mer relevanta artiklarna inom Genevekonventionen för användningen och utvecklingen av AMS. Frågan om statsansvar inom ramen för ARSIWA. Frågan om individuellt ansvarskrävande inom ramen för ICC och den praxis som där har utvecklats t.ex. med sin grund inom genevekonventionen. Frågan om det om det föreligger ett s.k. ”responsibility gap” med anledning av de unika omständigheterna som föreligger vid användning av AMS. Samt att ge läsaren en kort introduktion både till krigets lagar och hur dessa förhåller sig till AMS och vilka definitioner som kan tänkas för AMS

²² Med visa undantag se resolution 51/22 från HRC FN och resolution L.56 från FN:s generalförsamling

²³ Se Sandgren (2018) s. 45

som helhet. Detta har jag gjort för att ge läsaren en enkel överblick över frågan samt de viktigaste och av akademiker mest uppmärksammade frågorna inom ämnet och för att kunna analysera de unika utmaningar som frågan kräver av det internationella samfundet och den internationella rättspraxis som har utvecklats av denna.

2 Autonoma vapensystem och IHR

2.1 En introduktion till IHR och jus in bello

Sedan 1800-talets mitt och den inhumana slakt som uppkom på slagfältet i Krimeahalvön så har det internationella samfundet på ett allt starkare och mer robust sätt försökt att reglera krigets inverkan på civila och soldater. Detta syfte och skyddsbehov som står att finna både i genevakonventionen och i den av FN upprättade Internationella Domstolen och den praxis som den har upprättat med hjälp av olika vägledande domslut bl.a. gällande nödvändighet och proportionalitet²⁴ syftar till att upprätthålla humanitära principer även under krigsföring. ”jus in bello” som det kallas eller ”krigets lagar” är de regler som det internationella samfundet har kommit överens om skall gälla vid krigsföring och skall ses som vägledande för militärens sätt att strida som helhet.

Reglerna skall ses som en balans mellan ”militär nödvändighet” och ”humanitära principer”.²⁵ De regler som får ses som mest relevanta för den här uppsatsens del och även AMS i sin helhet är de som reglerar vapenanvändning och principen om ett förbud mot att använda vapen eller stridsmetod som orsakar onödigt lidande. Genevekonventionen som är den främsta rättskällan inom IHR när det gäller till både reglering av jus in bello och kontroll över vapenframställning i allmänhet. Ett av de mer för denna uppsatsen intressanta artiklarna är artikel 36. Artikel 36 i Genevekonventionen syftar till att reglera framställningen av nya vapen genom att:

In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.²⁶

²⁴ Se t.ex. Nicaragua v. United States (1986) ICJ

²⁵ Seixas-Nunes, Afonso (2022) s. 141

²⁶ Artikel 36 TPI, Genevekonventionen infört i svensk lag genom: SÖ 1979:22

Med andra ord så syftar artikeln till att förekomma nya vapen och stridsmetoder som riskerar att strida mot folkrätten genom att kräva en genomgång av om vapnet/stridssättet redan under framtagningsprocessen.

Frågan som uppkommer då är vilka internationella rättsprinciper och konventionsartiklar som blir aktuella vid en sådan genomgång. Självklart så finns det ett stort antal sådana rättsprinciper som blir viktiga att gå igenom för denna artikels potentiella rättsverkan, och jag skulle vilja anse att dessa framförallt är: distinktionsprincipen, proportionalitetsprincipen, försiktighetsprincipen, förbudet mot onödigt lidande, principen om icke-diskriminering och Martensklausulen.

Distinktionsprincipen innebär att krigförande parter alltid måste skilja mellan militära mål å ena sidan och civila personer/objekt å andra sidan.

Proportionalitetsprincipen innebär att ett anfall mot ett militärt mål måste stå i proportion till de risker som civilbefolkningen och civil egendom utsätts för.

Försiktighetsprincipen innebär ett krav på de stridande parterna att vidta alla rimliga försiktighetsåtgärder för att undvika skador på civila och civila objekt, detta blir speciellt relevant i detta fall då det innebär att stridsmedel ska väljas så att risken för civila förluster undviks eller minskas.

Förbudet mot onödigt lidande innebär att det är förbjudet att åsamka motpartens soldater onödig skada eller lidande vid krigsföring, även denna blir högst relevant för denna uppsatsen samt tillämpningen av artikel 36 då det inom den traktatbaserade rätten återfinns som både ett generellt samt specifikt förbud av stridsmedel samt stridsmetoder.

Principen om icke-diskriminering innebär att IHR ska tillämpas utan att någon diskrimineras på grund av ras, hudfärg, religion, tro, kön, börd, förmögenhet eller dylik omständighet. Martensklausulen innebär att stater även i de situationer som saknar direkta folkrättsliga regler, ska beakta humanitetens

krav inom sin krigföring.²⁷ En något mer detaljerad genomgång av dessa kommer ske löpande i texten när jag analyserar tillämpningen av dessa regler för de aktuella militära systemen.

2.2 Frågan om ”responsibility gap” och det s.k. ”black box” – problemet.

En i allra högsta grad för folkrätten intressant fråga är frågan om det finns ett s.k. ”responsibility gap”, redan 2004 så tillkom begreppet då filosofen och akademikern Andreas Matthias myntade det i en artikel för tidskriften ”Ethics and information technology”.²⁸ Andreas menar i den artikeln att möjligheten för mjukvara att ta självständiga beslut öppnar upp för ett moralisk och i slutändan även juridiskt problem som vi tidigare inte stått för. Nämligen att det finns potential för att en skada för egendom, människor eller finansiell skada uppkommer p.g.a. programmet helt utan att någon kan hållas ansvarig för det, då operatören eller designern av systemet inte kunnat förutse det på förhand och därmed inte kan hållas moraliskt eller juridiskt ansvariga. Artikeln har fått stor spridning och har skapat en debatt som fortfarande pågår angående om autonoma maskiner skapar ett s.k. ”responsibility gap” som stället till problem både på ett juridiskt och rent moraliskt plan.

Om vi för stunden bortser från det faktum att det faktiskt visst går att ibland hålla personer strikt skadeståndsrättsligt ansvariga i vissa fall för skador som uppkommer utan något uppsåt (dolus) eller oaktsamhet och vårdslöshet (culpa).²⁹ Så får det ändå för folkrätten anses vara en högt intressant fråga. Om en autonom maskin tar ett beslut att använda militär styrka på ett för folkrätten särskilt skyddat mål såsom ett civilt objekt, och designern samt operatören menar att de använt och designat maskinen på ett sätt som gjort att de inte kunnat förutse handlandet, kan någon då hållas ansvarig?

²⁷ Denna korta genomgång av dessa principer är baserad på följande text från röda korsets hemsida, se denna för en något mer genomgående genomgång av dessa principer samt internationella rättsregler angående krigföring: Viktiga regler och principer i krigets lagar | Röda Korset (rodakorset.se)

²⁸ Matthias (2004) s.175-183

²⁹ Se t.ex. NJA 1989 s. 389

Uppsåt eller vårdslöshet har länge varit ett viktigt fundament för folkrätten.³⁰ Det är en del av sedan länge utarbetad sedvanerätt och står att finna som grund för de straffrättsliga sanktioner som Internationella domstolen i Haag använder för att hålla den enskilda operatören av ett vapen eller befälhavaren över en sådan, ansvariga för ett begånget krigsbrott. Frågan som därmed blir aktuell är följande: Blir någon ansvarig och i sådana fall, vem?

För att kunna besvara den frågan gäller det att gå igenom flera olika folkrättsliga samt allmänrättsliga principer, konventioner och hur den rättspraxis samt sedvanerätt som utvecklats internationellt skulle kunna tillämpas på sådana fall och i vilka olika situationer de blir aktuella. Samt hur man skulle kunna reglera skapandet av sådana system för att de i högsta möjliga mån skall vara förenliga med folkrätten. Jag går igenom detta i större detalj i avsnitt 2.3, 2.4 samt 2.5. Men det går redan nu att konstatera, att i varje fall av begånget krigsbrott oavsett natur, så krävs det culpa eller dolus för individuellt straffrättsligt ansvar för operatör eller befälhavare.

Den folkrättsliga princip som torde bli allra mest aktuell med anledning av black box problemet eller oförutsägbarhet vid användning är försiktighetsprincipen som får sitt främsta uttryck i artikel 57 första tilläggsprotokollet till Genevekonventionen. De delar av artikeln som torde bli främst tillämpliga är följande:

Ifråga om anfall, skall följande försiktig hetsmått vidtagas:

a) de som planerar eller beslutar om ett anfall skall

1) göra allt som är praktiskt möjligt för att kontrollera att de mål mot vilka anfallen skall riktas varken är civilpersoner eller civil egendom och icke föremål för särskilt skydd utan att de är militära mål enligt punkt 2 i artikel 52 samt att det enligt bestämmelserna i detta protokoll icke är förbjudet att anfälla dem;

2) vidtaga alla praktiska möjliga försiktighetsmått vid valet av stridsmedel och anfallsmetoder i syfte att undvika och i varje fall i möjligaste mån minska oavsiktliga förluster i människoliv bland civilbefolkningen samt skador på civilpersoner och civil egendom;

³⁰ Se bland annat: Artikel 25(3) (a) Artikel 28 Romstadgan (ICC) (SÖ 2002: 59)

3) avstå från beslut att inleda anfall som kan förväntas medföra oavsiktliga förluster i människoliv bland civilbefolkningen, skador på civilpersoner och civil egendom eller en kombination därav, vilka skulle vara överdrivna vid jämförelse med den avsedda, konkreta och direkta militära fördelen med anfallet.

b) ett anfall skall inställas eller avbrytas om det framgår att målet icke är militärt eller att det är föremål för särskilt skydd eller att anfallet kan förväntas medföra oavsiktliga förluster i människoliv bland civilbefolkningen, skador på civilpersoner och civil egendom eller en kombination därav, vilka skulle vara överdrivna vid jämförelse med den avsedda, konkreta och direkta militära fördelen.³¹

Som vi kan se så ställer artikeln upp ett antal starka krav innan användning av militär styrka kan anses godtagbart. Dessa artiklar är besvärande för användningen av AMS, då vi har att göra med en autonom agent som i varje fall till viss del kan ta självständiga beslut om t.ex. vilket mål som väljs och/eller vilken typ av våld som mot detta mål skall användas.

Frågan blir därmed, kan ett AMS anses ha en tillräckligt hög tillförlitlighetsgrad för att kunna användas utan att bryta mot artikel 57?

Detta torde bli en svår fråga att reda ut framöver, t.ex. kan man fråga sig om det skulle behövas någon typ av mekanism som möjliggöra mänsklig övervakning av AMS och t.ex. de-aktivering eller självförstörande om det kan antas att allvarliga brott mot folkrätten är nära förestående. Hur detta ska göras rent praktiskt och om det ens är möjligt utan att den militära fördelen med användande av AMS försvinner är svårt att säga i förhand och kommer kräva en närmare efterforskning. Men det är definitivt en fråga som är värd att ställa, och är något som t.ex. Alfonso Seixas-Nunes frågar sig i sin bok ”The Legality and Accountability of Autonomous Weapon Systems”.³²

Statsansvar borde däremot oavsett situation kunna utgå för användning av AMS som bryter mot folkrätten oavsett om mens rea går att fastställa vid användningen av sådana vapen.³³

³¹ Artikel 57 TPI, Genevekonventionen införd i svensk lag genom: SÖ 1979:22

³² Seixas-Nunes, Afonso (2022) s. 166

³³ Se t.ex. Artikel 25(4) ICC:s stadgar och fjärde Haagkonventionen (1949) artikel 3

I enlighet med de regler som har fastställs i ARSIWA angående statsansvar för händelser som bryter mot folkrätten så borde det gå att komplettera oväntade händelser där individuellt ansvar ej utgår med dessa regler om statsansvar.

2.3 Den Internationella rätten och användning av AMS

Efter den 7:e Oktober, har ett krig startat mellan Hamas-anknutna aktörer å ena sidan och den israeliska staten å den andra sidan startat. Det har kallats för det första AI-ledda kriget av vissa medier, då fortfarande är ett relativt nytt system används för att välja mål på slagfältet. ”The Gospel” (Evangeliet) heter systemet, och lockelsen i användningen av det är uppenbar, enligt en officer i den israeliska armén så är systemet kapabelt att 50 gånger snabbare än en grupp av 20 underrättelseofficerare välja mål på slagfältet, en kapacitetsutveckling som ger varje användare av ett sådant system ett enormt övertag över en analog motpart i de flesta situationer.³⁴ Systemet som egentligen inte är ett nytt system utan användes redan 2021 i en dåtida konflikt mellan Israel och Hamas, samt gruppen ”Islamic Jihad”, har dock utvärderats i en rapport ifrån den amerikanska NGO:n JINSA³⁵ som menar att systemet visserligen har varit effektivt på att välja mål och har inneburit en minskning av civila dödsoffer jämfört med mer analoga system för underrättelseanalys och spaning men att det varit ineffektivt på att sälla bort civila mål. Det bör i sammanhanget förövrigt noteras att rapporten kommer ifrån en NGO som i allra högsta grad kan anses ha en bias till förmån för Israel som aktör och att dess resultat därmed även om de kan anses vara intressanta bör tas med en nypa salt.

Frågan om mänsklig kontroll och statsansvar vid användande av ASS är därmed mer relevanta än någonsin och allting tyder på att användningen av dessa system enbart kommer att bli vanligare ju mer tid som går då det ger ett övertag mot mer analoga sätt att sälla underrättelseinformation. Frågan är också i

³⁴ Here's how Israel is using artificial intelligence to find targets in Gaza (npr.org)

³⁵ Gaza Conflict 2021 Assessment: Observations and Lesson – rapport (jinsa.org) s. 31

vilken mån mänsklig kontroll kan utövas när en så stor mängd information presenteras till denna på så kort tid och i vilken utsträckning mänsklig psykologi samt fysiska limitationer påverkar denna förmåga. Det kan t.ex. antas att det kommer finnas en bias till förmån för att välja mål snabbt och för att systemet funkar som tänkt. Lägg till detta det faktum att många mål kommer att visas under mycket kort tid och det finns många som menar att den mänskliga kontrollen i detta fall, även i bästa fall kommer att vara ytlig och inte i någon mening vara tillräcklig för att kunna säkerställa en faktiskt operativ kontroll och översyn av systemet.³⁶

Det finns såklart därmed redan nu tydliga problem med användande av ett sådant problem och därmed kan man konstatera att det bör ställas starka krav på designen, inmatningen (av inlärningsdata) och användandet av sådana här system.

Men detsamma kan såklart sägas om vanliga konventionella sätt att välja mål som redan nu finns. Människor begår misstag, människor har fördomar och bryter även medvetet mot IHR, ibland utan och ibland med sina överordnades vetskap. Frågan som blir intressant i sammanhanget är därmed inte om det finns problem med denna nya spanings och analysteknik utan handlar snarare om att ställa det i relation till de system som för nuvarande finns tillgängligt för detta syfte.

Jag kan själv tycka i sammanhanget att det för nuvarande är svårt att uttala sig på ett definitivt sätt som sådana system kan anses vara eller inte vara kompatibla med IHR, det kommer behövas en större användning och analys av systemen för framtiden för att kunna avgöra detta. Men man kan redan i nuläget konstatera vissa saker.

Det går att utforma dessa system på olika sätt, beroende på vilken data man matar algoritmen med så kommer den välja målen på olika sätt. Här finns helt klart en potential för mänsklig påverkan som kan anses problematisk för det

³⁶ Boutin, Berenice (2022) s. 138

nuvarande regelverket inom IHR både angående jus in bello och mer allmänna folkrättsliga principer. T.ex. så skulle man kunna mata systemet med data som leder till att den artificiella intelligensen väljer civila mål, mål som genom sitt urval kan anses vara diskriminerande såsom en viss etnisk, religiös eller kulturell grupp, eller att systemet används i ett sammanhang som kan anses vara problematiskt i förhållande till IHR.

Om detta gör att det medföljer ett folkrättsligt ansvar för brott mot IHR för operatören, ingenjören eller staten är svårt att uttala sig om allmänt, men jag anser att det bör kunna föras ett resonemang i varje enskilt fall om detta och att systemet egentligen inte innebär ett stort avsteg ifrån gällande Internationell rätt i den mån att det skulle krävas en översyn av gällande principer och konventioner i nuläget. Men det bör såklart krävas att systemen både skapas och används på ett sätt som är förenligt med gällande principer inom IHR.

Med gällande praxis inom IHR angående t.ex. försiktighetsprincipen och förbudet mot onödigt lidande och proportionalitetsprincipen så bör det redan nu kunna ställas krav på detta. Men man bör också ha i beaktande att dessa principer ställs mot det som inom IHR ofta kallas ”militär nödvändighet”, dvs. hur en militär styrka ska kunna besegra sin motståndare på snabbaste och mest effektiva sätt, med beaktande av sina egna militära resurser.

CCW har sedan tidigare antagits för att förbjuda vissa typer av vapen, för att reglera stridssätt och vilket ansvar stater har för att komma tillrätta med dessa problem efter att kriget är över. Konventionen förbjuder bland annat lasrar som gör kombattanter blinda, användningen av landminor, förståtsminering och liknande lönnfällor.

Förbudet mot förblindande lasrar gjordes förövrigt innan ett sådant vapen ens för första gången använts på slagfältet, och är därmed unikt genom att det innebar ett förebyggande förbud mot ett vapensystem. Det går därmed att fråga sig om det inte skulle kunna vara i sin ordning att med det i åtanke förbjuda iaf. vissa användningssätt och design av ASS i förebyggande syfte. Jag tänker då framförallt på t.ex. matande av data till algoritmen som leder

till att mål väljs på basis av speciellt skyddade grupper såsom civila, civila objekt, etnicitet, kultur, religion, kön osv.

Men något förebyggande förbud av autonoma militära system tror jag inte kommer bli aktuellt. Det finns också en stor motvillighet bland många stora och viktiga stater mot ett sådant förbud och nya internationella rättsregler angående användningen av AMS, se t.ex. USA:s, Rysslands och Israels avvikande meningsskiljaktigheter under CCW 2018.³⁷ Framförallt eftersom utvecklingen går framåt i rask takt, det finns uppenbara och stora fördelar enligt tidiga användare av systemen.³⁸

Det kan såklart finnas vissa problematiska aspekter av dessa system, t.ex. det möjliga black box eller ”responsibility gap” problemet, men som tidigare visats i avsnitt 2.2 så finns det redan nu en möjlighet att åligga ansvar på någon för användning av AMS som leder till brott mot folkrätten eller iaf. möjlighet att kräva t.ex. skadestånd via statsansvar. Jag ser inte heller den tilltänkta situationen som helt unik, t.ex. kan man ställa sig en hypotetisk fråga, om ett stridsflyg under krigsoperationer åker över ett tätbebyggt område och p.g.a. ett svår-förutsätt tekniskt problem råkar tappa hela eller delar av sin last på detta område och detta leder till dödsfall och skada på egendom, går det alltid att hålla någon straffrättsligt ansvarig? Möjligtvis skulle man kunna postulera att den ansvarige är t.ex. designern av systemet eller underhållsingenjören för farkosten, men det är inte enligt min mening säkert. För designen kanske är sund, möjligtvis så beror problemet på något som inte går att upptäcka vid en underhållskontroll och det finns inte t.ex. produktionsfel eller liknande som är orsaken. Det är t.ex. möjligt att farkosten utsatts för helt unika omständigheter under sin användning som innan denna olycka aldrig tidigare givit upphov till en olycka av detta slag. Går det då att hålla någon straffrättsligt ansvarig? Såsom vi tidigare konstaterade så finns det i alla fall inom folkrätten fortfarande ett krav på dolus eller culpa för att straffrättsligt ansvar ska kunna

³⁷ Seixas-Nunes, Afonso (2022) s. 201

³⁸ Se t.ex. Here's how Israel is using artificial intelligence to find targets in Gaza: (npr.org) men även Gaza Conflict 2021 Assessment: Observations and Lesson – (jinsa.org)

utgå. Men i denna hypotetiska situation så finns det varken eller. Detta scenario kanske verkar osannolikt eller överkligt men speciellt inom krig och med åtanke på den vikt som inom "jus bello" lagts vid t.ex. militär nödvändighet som alltså ska ställas i proportion till de folkrättsligt skyddande principer som finns till för skydd av individer och samhället såklart alltid en risk för att situationer kan uppkomma utan att någon kan hållas straffrättsligt ansvarig. Jag ställer mig därmed frågande om det verkligen är ett unikt problem för autonoma vapensystem och kanske snarare ett problem som förstoras upp p.g.a. det för det mänskliga psyket skrämmande scenario där en främmande typ av intelligens tar beslut över liv och död.

Det är såklart fortfarande system vars användning i längden kommer behöva utvärderas för att se potentiella risker med, men generellt sett så tror jag att det blir både svårt att helt förbjuda systemen och att det inte finns något som i nuläget tyder på en risk med användningen som gör att det kommer bli nödvändigt med ett generellt förbud, speciell med tanke på den tydliga militära fördelen med att använda autonoma system.

Det som jag tror är troligare, är att vi kommer se någon form av reglering av dessa system t.ex. genom att det byggs in någon form av "failsafe" i systemen, att mänsklig övervakning på något sätt kommer krävas eller att utvecklingen av dessa system på något sätt kommer att regleras för att se till att folkrättsliga principer efterlevs av systemen vid praktisk användning.

Det finns t.ex. ett resonemang av roboticisten Ronald C. Arkin i hans bok "*Governing Lethal Behavior in Autonomous Robots*" angående potentialen för en mjukvara som skulle kunna installeras i AMS vid namn "Ethical Governor". Denna mjukvara skulle enligt denna tanke kunna agera som en slags "failsafe" eller i alla fall ytterligare en komponent som av den autonoma maskinen aktiveras innan dödligt våld används och kunna "rådfrågas" angående användningen innan auktorisering sker.³⁹

³⁹ Resonemanget återges av: Seixas-Nunes, Afonso (2022) i t.ex. s. 170

Redan idag finns det ett krav inom folkrätten på att det vid utvecklingen av nya vapensystem skall krävas att vapnet t.ex. inte strider mot någon av de humanitära principer som ställs upp inom folkrätten till skydd för människoliv, mänskliga rättigheter och civil egendom. Hur detta kommer att se ut i praktiken är en fråga som är högst relevant och är en fråga som kommer att aktualiseras alltmer efterhand som utvecklingen och användningen av AMS blir allt vanligare.

Men jag kan konstatera att även om det i nuläget inte finns någon form av bindande traktat så fortgår utvecklingen alltjämt på området, och det är mycket möjligt kanske till och med troligt att vi snart kommer att se detta, kanske som ett resultat av det tidigare nämnda arbetet inom REAIM 2023, eller det många olika nationella samt internationella arbetsgrupper, kampanjer och föreslagna traktat som finns inom ämnet.⁴⁰

Men fram till dess så är det min åsikt i frågan att det i nuläget borde gå att med nuvarande folkrättsliga principer, gå att kräva av stater att vid användningen och utvecklingen av AMS se till att det autonoma systemet förhåller sig till allmänt hållna folkrättsliga principer inom ramen för t.ex. krigets lagar på ett fullgott sätt och att det inte finns något unikt s.k. ”responsibility gap” med anledning av systemens utformning. Men att det alltjämt finns unika omständigheter med systemen som torde kunna leda till någon form av unik reglering av dessa inom ramen för t.ex. Förenta Nationerna.

⁴⁰ Se t.ex. REAIM 2023, kampanjen: ”Stop Killer Robots”, och resolutionsförslag ”A/C.1/78/L.56” från generalförsamlingens första utskott under UNGA 78

Källförteckning

Litteratur:

Artiklar:

Boutin, Berenice, *State Responsibility in Relation to Military Applications of Artificial Intelligence*. (August 31, 2022). T.M.C. Asser Institute for International & European Law, Asser Research 09, 2022, Forthcoming in: *Leiden Journal of International Law*, Cambridge University Press (2022), Forthcoming, Available at SSRN: <https://ssrn.com/abstract=4214292>

Danaher, J. Robots, law and the retribution gap. *Ethics and Information Technology*, volym 18, 299–309 (2016). <https://doi.org/10.1007/s10676-016-9403-3>

Matthias, A. *The responsibility gap: Ascribing responsibility for the actions of learning automata*. *Ethics and Information Technology* 6, 175–183 (2004). <https://doi.org/10.1007/s10676-004-3422-1>

Böcker:

1. Seixas-Nunes, Afonso, *The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective*. Cambridge University Press; 2022.
2. Galliot, Jai, MacIntosh, Duncan och Ohlin, Jens David, *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare*. Oxford University Press; 2021
3. Claes Sandgren. *Rättsvetenskap för uppsatsförfattare upplaga 4*. Norstedts Juridik AB; 2018

Offentligt tryck:

Departementsserier:

Wallmark, Hans, *Försvarsberedningens säkerhetspolitiska rapport 2023*
Ds 2023:19

Direktiv:

Department of Defense Directive 3000.09 *Autonomy In Weapon Systems*
(November 21, 2012)

Traktater:

Genevekonventionen 1864, 1906, 1929 och 1949

Romstadgan för Internationella brottmålsdomstolen 1998

IV:e Haagkonventionen 1949 (del av Genevekonventionen)

FN-resolutioner:

resolution 51/22: FN:s råd för mänskliga rättigheter

resolution L.56: FN:s generalförsamling

Rättsfall:

Avgöranden inom Internationell rätt:

Nicaragua v. United States (1986) ICJ – Internationella domstolen

Nytt juridiskt arkiv:

[NJA 1989 s. 389](#)

Övrigt:

Internetkällor:

[Aided Detection on the Future Battlefield | Article](#) | (army.mil), Besökt:
2024-02-03

[Ukraine AI Drones Seek, Attack Russian Forces Without Humans Oversight](#) (forbes.com), Besökt: 2024-02-03

[Ukraine's Secret Weapon - Artificial Intelligence](#) (cepa.org)

[Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy](#) – (state.gov) besökt 2024-02-05

[Global summit calls for 'responsible use' of AI in the military | Engineering and Technology Magazine](#) (theiet.org) besökt 2024-02-05

[The law and the U.S.' new declaration on military uses of AI: some observations](#) – (duke.edu) besökt 2024-02-05

[Key elements of a treaty on fully Autonomous Weapons](#) – (stopkillerrobots.org) besökt 2024-02-05

[Viktiga regler och principer i krigets lagar](#) - (rodakorset.se) besökt 2024-02-05

[Here's how Israel is using artificial intelligence to find targets in Gaza](#) – (npr.org) besökt 2024-02-05

[Gaza Conflict 2021 Assessment: Observations and Lesson](#) – (jinsa.org) besökt 2024-02-05