Handhaving van de regels en fysieke verdediging - drones en killer robots
Position paper TNO voor het Rondetafelgesprek over Drones en Killer Robots op 21 januari 2019.

Inleiding
De vraagstelling 'hoe kunnen we de regels handhaven en ons fysiek verdedigen' splitsen we uit naar de bestrijding van ongewenste drones in het luchtruim en naar het uitoefenen van adequate governance over autonome wapensystemen. Tevens verwijzen we naar het TNO Position Paper over de capabilities van drones en killer robots, aangeboden voor ronde 1 van het Rondetafelgesprek.

Handhaving tegen drones
In een eerder position paper[1] voor een Rondetafelgesprek in de Tweede Kamer gehouden op 8 maart 2018 berichtten we al over de ontwikkeling van drones, de daarmee gepaard gaande veiligheidsrisico's en het gewenste beleid om een capability voor tegenmaatregelen te verwerven. Inmiddels is er een breed besef dat die capability hard nodig is, beschikken Defensie en de Nationale Politie al over tegenmaatregelen en doen TNO en NLR samen met Defensie en de Nationale Politie onderzoek om die capabilities verder uit te breiden. Maatregelen voor detectie en aangrijpen van drones zullen in de praktijk worden uitgetest, in nauwe samenwerking tussen onderzoekers en handhavers. Er bestaat namelijk niet één enkele 'silver bullet' oplossing tegen ongewenste drones. De omstandigheden en de aard en het type van de dreiging bepalen welke maatregel of combinatie van maatregelen ingezet kunnen worden, waarbij ook ongewenste neveneffecten meegenomen dienen te worden in de overweging tot inzet. Daar komt bij dat de voortschrijdende technologie zal leiden tot een kat-en-muis spel tussen kwaadwillenden en handhavers.

Tegenmaatregelen zijn onder te verdelen in zogenaamde hard-kill maatregelen (neerschieten, lasers, netten) en elektronische of soft-kill maatregelen (jamming, spoofing, hacking). Maatregelen waarbij de controle over de drone wordt overgenomen en de drone naar een veilige plaats kan worden gebracht verdienen de voorkeur; andere maatregelen hebben namelijk mogelijk ongewenste neveneffecten op de grond als een drone of zijn lading neerstort. Zo mocht er bij het drone-incident bij Gatwick rond de Kerstdagen niet op de drones geschoten worden en mochten ook niet alle elektronische tegenmaatregelen ingezet worden. De counterdrone technologie ontwikkelt zich in hoog tempo, maar is nog zeker niet volwassen. Strengere eisen voor de beveiliging van drones tegen hacken maken soft-kill tegenmaatregelen ook moeilijker uit te voeren.

Om de drones van welwillende gebruikers te onderscheiden van drones die mogelijk kwaad in de zin hebben, is een vorm van elektronische identificatie zeer wenselijk. Ook invoering van UAS Traffic Management (UTM) systemen zal leiden tot een beter overzicht van hetgeen zich in het luchtruim afspeelt, waardoor afwijkingen sneller gesignaleerd zullen worden.

'Killer robots' : TNO raamwerk voor meaningful human control
Het onderzoek naar autonome systemen heeft een grote vlucht genomen, niet in het minst door de snelle ontwikkeling van kunstmatige intelligentie (AI), de belangrijkste bouwsteen van een autonoom systeem. Als gevolg van deze ontwikkeling is ook de vraag gerezen hoe de mens controle kan blijven houden over deze AI en over de autonome systemen die het mogelijk maakt. Zelfrijdende auto's en autonome wapensystemen zijn aansprekende voorbeelden van deze roep om betekenisvolle menselijke controle (meaningful human control). Een volledig autonoom wapensysteem zonder enige vorm van meaningful human control wordt ook wel aangeduid met de term 'killer robot'.

Twee aspecten zijn hierbij van belang. Allereerst is het nodig om AI te ontwikkelen die transparant, voorspelbaar, uitlegbaar en beïnvloedbaar is. De huidige AI schiet daarin tekort. Ook vraagt de implementatie van AI aandacht voor het ontwerp van de mens-machine samenwerking, de inbedding in team en organisatie, de vormgeving van training en opleiding, het validatie- en verificatieproces en het invoeren van passende wet- en regelgeving. Kortom, goed ontworpen en goed inpasbare AI is van wezenlijk belang. Hier komt wereldwijd gelukkig meer aandacht voor; ook TNO zet zich hier actief voor in.

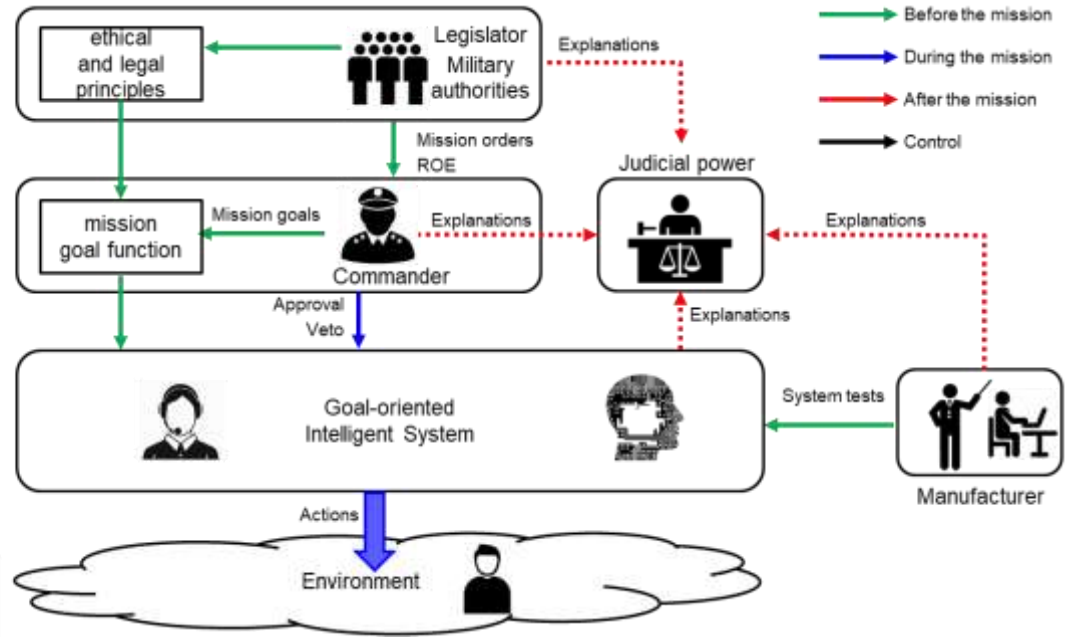
Het tweede aspect betreft het besluiten over de inzet van wapens; het is belangrijk dat mensen en autonome systemen handelen op basis van afgewogen menselijke keuzes binnen duidelijke juridische kaders, zeker als het gaat om zaken van leven en dood. Een verbod op 'killer drones' is niet waarschijnlijk, omdat binnen de VN de opvattingen verschillen over wat er dan precies verboden moet worden en omdat een aantal landen dit actief tegenhoudt. Toch lijkt er consensus over de noodzaak van meaningful human control.

[1] https://www.tweedekamer.nl/kamerstukken/detail?id=2018Z03761&did=2018D17773

Hiermee ligt de vraag op tafel hoe meaningful human control ingericht kan worden. Interessant is daarbij ook de vraag hoe de huidige militaire wapeninzet voldoet aan de eis van meaningful human control[2]. TNO onderzoekt verschillende mogelijkheden van meaningful human control, ook voor wapensystemen.

Eén van de opties voor meaningful human control is het meenemen van ethische principes in het ontwerp van een systeem (value based design). Een nadeel van deze methode is dat hij niet goed werkt in onvoorspelbare situaties met soms conflicterende waarden (‘kiezen tussen twee kwaden’).

TNO heeft een alternatieve optie ontwikkeld als raamwerk voor meaningful human control van autonome wapensystemen. Hierin worden de ethische en juridische ‘spelregels’ voor het gebruik van geweld door militairen en hun systemen vooraf vastgelegd door de wetgevende macht. Deze spelregels kunnen openbaar gemaakt worden, voor toetsing door bijvoorbeeld NGO’s. De militaire commandant formuleert voor iedere missie de doelstellingen, die niet in strijd mogen zijn met de opgelegde ethische en juridische kaders en rechtstreeks gekoppeld zijn aan de strategische doelen (‘mission orders’). Deze worden vervolgens samen met de ethische en juridische spelregels in de vorm van een mathematische functie (de mission goal functie) opgelegd aan het autonome systeem, dat deze functie niet kan veranderen. De mogelijkheid tot menselijk ingrijpen blijft open, zolang die mens zich ook verantwoordt voor dat ingrijpen. De expliciete koppeling aan de strategische doelen zorgt er voor dat het systeem steeds de juiste keuzes maakt, rekening houdend met de omstandigheden, zoals een mens dat ook zou doen, ook bij conflicterende waarden. De fabrikant is verantwoordelijk voor het goed technisch functioneren van het autonome systeem. De rechter tenslotte kan achteraf alle keuzes en beslissingen toetsen.



Het schema hierboven laat zien dat de meaningful human control in dit raamwerk zowel voorafgaand aan de missie (groene pijlen), tijdens de missie (blauwe pijlen) als na afloop van de missie (rode pijlen) plaatsvindt. Dit raamwerk is niet af: er liggen nog grote (technische) uitdagingen, maar dat doet niets af aan het principe van meaningful human control dat op deze manier bewerkstelligd kan worden. Combinatie van deze aanpak met value based design lijkt overigens ook heel goed mogelijk.

Samenvatting

Met het onderzoek naar mogelijkheden voor meaningful human control wil TNO een bijdrage leveren aan het debat over meaningful human control, om zo de voordelen van autonome systemen voor militair gebruik te benutten, op een ethisch verantwoorde, effectieve en veilige manier. Samenwerking tussen ingenieurs, ethici, juristen, militairen en wetgever is hierbij onontbeerlijk.

[2] M.A.C. Ekelhof, Autonomous Weapons – Operationalizing Meaningful Human Control, https://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/

Pieter Elands (pieter.elands@tno.nl) Rogier van Keulen (rogier.vankeulen) (contactpersoon)