

ARTICLES CENTRALS

## Quan perdem el control de la nostra humanitat

**Noel Sharkey / Maya Brehm**

Noel Sharkey és president del Comitè Internacional per al Control de les Armes Robòtiques i catedràtic emèrit d'Intel·ligència Artificial i Robòtica a la Universitat de Sheffield, al Regne Unit. / Consultora sobre Ginebra per a l'ONG britànica Article 36.

En ple segle XXI, una de les principals amenaces emergents contra la seguretat mundial és el desenvolupament de les armes autònomes, també denominades “armes totalment autònomes”, “robots assassins”, “robots autònoms letals” o “sistemes autònoms d'armament letals”. Són sistemes d'armament que, un cop activats, seleccionen objectius i els ataquen sense cap més intervenció humana.

Cada vegada és més difícil trobar una nou artefacte tecnològic que no estigui controlat per circuits informàtics. Les tecnologies de la violència no en són una excepció: els aparells informàtics esdevenen ubics en la majoria de sistemes actuals d'armament i sistemes de control. Avui dia, gairebé totes aquestes armes són sota “control de supervisió”, en què el control humà senzillament està mediat per un programa d'ordinador.

Alguns estats ja utilitzen diversos sistemes automatitzats d'armament que intercepten per si sols objectes inanimats d'alta velocitat, com míssils, projectils d'artilleria, granades de morter o atacs de saturació; per exemple, els sistemes C-RAM, Phalanx, NBS Mantis i Iron Dome. Aquests sistemes completen el seu procés de detecció, avaluació i resposta en qüestió de segons i, per tant, fan extremament difícil per als operadors humans d'exercir un control de supervisió significatiu més enllà d'activar-los i desactivar-los. Fins ara, molts sistemes s'han desplegat en entorns relativament lliures d'obstacles, despoblats de civils.

Tanmateix, hi ha alguns estats que impulsen cada vegada amb més intensitat el desenvolupament d'armament a distància que podria sobrepassar el control de supervisió humana: els Estats Units han fet proves avançades de diverses plataformes d'armament autònom, com l'X-47b -un reactor subsònic d'alta velocitat que ja pot enlairar-se i aterrar a portaavions-, el Crusher -un robot autònom terrestre de 7 tones- i un submarí d'atac autònom; els xinesos treballen en l'Anjain, un vehicle de combat autònom de combat aire-aire; els russos desenvolupen el reactor de caça autònom Skat; Israel disposa del robot terrestre autònom Guardian; i al Regne Unit la Mantis, una aeronau intercontinental de combat totalment autònoma, és en fase avançada de proves.

## **Manca de precisió del reconeixement d'objectius**

Un dels grans problemes de les armes autònomes és que identificar i seleccionar objectius necessita un programari ben definit. Tanmateix, els actuals mètodes de reconeixement d'objectius emprats pels militars no s'adeqüen al seu propòsit, excepte en entorns molt lliures d'obstacles. Actualment, hi ha tres mètodes principals:

1 - La detecció de siluetes permet de reconèixer un tanc en un entorn sense obstacles, com un desert de sorra d'orografia plana. Malgrat les dècades de recerca, s'ha demostrat que és extremament difícil distingir entre un tanc i un camió, o qualsevol altre vehicle, si es troben entre altres figures, com arbres.

2 - La formació d'imatges tèrmiques detecta el calor irradiat per un objecte i en mostra el moviment, però amb aquest mètode seria difícil distingir un tanc d'un autobús escolar.

3 - Per a la detecció per radar es fa servir munició passiva (o en espera) per detectar la senyal dels radars enemics i bombardejar-los. Cal donar per fet que l'objectiu és una instal·lació antiaèria, perquè en cas contrari la detecció d'un radar no n'estableix la legitimitat com a objectiu.

En bona part a causa de les greus limitacions d'aquests mètodes, la idea mateixa de desenvolupar robots assassins per utilitzar-los sense unes limitacions molt estrictes però respectant els requisits de la utilització de força letal establerts en el dret

internacional dels drets humans i el dret internacional humanitari és, en el millor dels casos, pura especulació i, en el pitjor, pura fantasia. Pot ser que les descobertes futures resolguin els problemes tècnics. Tanmateix, és important tenir en compte que, fins i tot si en millora la capacitat de reconèixer objectius, això no permet a les màquines valorar si atacar un determinat objectiu és legítim o si l'atac és, en conjunt, permisible.

L'adequació i la legalitat d'un atac depenen del context i tendeixen a avaluar-se cas per cas. Si es deixa que els estats segueixin desenvolupament els robots assassins amb l'esperança que podran funcionar, se'ns planteja el risc greu que aquestes armes siguin desplegades amb independència que compleixin la llei o no. L'única solució és una prohibició integral i preventiva del desenvolupament, la producció i l'ús d'aquests sistemes.

**“ No hi ha cap garantia que ens permeti preveure una utilització dels sistemes autònoms d'armament letal que compleixi amb el dret internacional. ”**

## **Actuar contra els robots assassins**

L'abril de 2013 es va posar en marxa la Campanya per Aturar els Robots Assassins, una campanya internacional de la societat civil que exigia la prohibició preventiva del desenvolupament, la producció i l'ús dels sistemes d'armament completament autònoms. La campanya no intenta prohibir els vehicles o robots autònoms de tot tipus, sinó que el seu enfocament se centra clarament en evitar l'automatització de la decisió de matar.

Un mes més tard, Christof Heyns, relator especial de les Nacions Unides sobre les execucions extrajudicials, sumàries o arbitràries, va demanar una moratòria mundial sobre la utilització i el desenvolupament dels robots autònoms letals, la qual cosa representava una pausa per permetre als països de reflexionar sobre les implicacions

del desenvolupament d'aquestes armes. Heyns conclouïa que: “[Adoptar-les] podria tenir efectes importants i duradors sobre els valors socials, especialment la protecció i el valor de la vida i de l'estabilitat i la seguretat internacionals.” Com molts d'altres, Heyns considera que delegar a les màquines la decisió de matar pot vulnerar un criteri moral fonamental.

Gràcies a la pressió creixent de la societat civil per abordar els reptes plantejats pels sistemes autònoms d'armament, el novembre de 2013 França presentà als 117 estats parts de la Convenció sobre Certes Armes Convencionals (CCAC) la proposta d'una reunió informal d'experts. El mandat s'aprovà i la qüestió es debatrà el maig de 2014.

Aquests debats d'experts empenyen els estats a expressar la postura de cada país sobre la qüestió, quelcom d'urgentment necessari. Fins ara, només els Estats Units han publicat una declaració política oficial. La Directiva 3000.09 del Departament de Defensa estableix unes directrius sobre les armes autònomes. Alhora que n'impulsa un major desenvolupament, la Directiva exigeix que, de moment, un humà participi en el procés de decisió relatiu a l'ús d'una força letal. Així mateix, el govern del Regne Unit ha afirmat que d'acord amb la seva normativa actual totes les armes romanen “sota control humà”.

Malgrat aquestes afirmacions, encara no és clar què vol dir “control humà” o “participació en el procés de presa de decisions”. Aquests termes podrien significar quelcom de tan senzill com prémer un botó per iniciar un atac o, fins i tot, programar una arma per a una missió. Clarament, això seria molt diferent del tipus de control humà que es considera apropiat en relació amb els sistemes d'armament existents.

**“ Encara que millori la precisió, les màquines no poden valorar si atacar un determinat objectiu és legítim o si l'atac és, en conjunt, permisible. ”**

**Control humà efectiu**

Una anàlisi de la recerca científica sobre el control humà de supervisió ens permet desenvolupar una classificació consistent en cinc tipus de control:

1. L'operador humà entra en contacte amb un objectiu, el selecciona com a tal i és qui inicia qualsevol atac.
2. Un programa suggereix objectius alternatius i l'operador humà decideix quin atacar.
3. Un programa selecciona un objectiu i l'operador humà n'ha d'aprovar la tria abans de l'atac.
4. Un programa selecciona un objectiu i l'operador humà disposa d'un temps limitat per vetar-ho.
5. El programa selecciona l'objectiu i inicia l'atac sense cap participació humana.

Per a un nivell de control 1, és crucial entendre que el *control humà efectiu* està sotmès a requisits molt estrictes: un comandant o operador humà té un coneixement complet del context i la situació de l'àrea objectiu en el moment d'un atac específic i pot percebre qualsevol canvi o situació imprevista que hagi pogut sorgir des de la planificació de l'atac i respondre-hi. Cal que hi hagi una participació cognitiva en l'atac i el temps suficient per reflexionar sobre la naturalesa de l'objectiu, la seva rellevància en termes de necessitat i adequació de l'atac, i els efectes indirectes probables i possibles de l'atac. També cal que es disposi dels mitjans per a suspendre o avortar l'atac amb rapidesa.

El nivell de control 2 podria ser acceptable si es demostra que compleix els requisits d'un control humà efectiu. Un ésser humà que controlés l'atac hauria de poder avaluar si un atac és necessari i apropiat, si totes les alternatives suggerides (o, fins i tot, alguna de les alternatives) constitueixen un objecte d'atac permisible, i seleccionar l'objectiu per causar el mínim dany a persones i béns civils.

El nivell 3 és inacceptable. La recerca experimental ha mostrat que aquest tipus de control causa el denominat "biaix de l'automatització", d'acord amb el qual els operadors humans acaben acceptant com a correctes les decisions generades per ordinador i menystenen la informació que les contradueix (o bé no en cerquen).

El nivell 4 també és inacceptable. No promou la identificació d'objectius, mentre que el poc temps disponible per exercir el veto reforçaria el biaix de l'automatització i no deixaria cap possibilitat per dubtar o reflexionar. El fet que s'executaria l'atac tret que un humà l'interrompés faria que es vulneressin supòsits consolidats del dret internacional humanitari que protegeixen la condició i la posició dels civils.

En el cas del nivell de control 5, no hi ha participació humana en la selecció de l'objectiu ni en l'atac, de manera que, com ja hem explicat anteriorment, aquests sistemes d'armament no podrien complir el dret internacional.

## **Conclusió**

No hi ha cap garantia que ens permeti preveure una utilització de les armes autònomes que compleixi amb el dret internacional. És impossible predir com un complex programari reaccionarà en totes les circumstàncies. És ben sabut que a les regions en conflicte hi abunden els esdeveniments imprevistos i canviants i que la tecnologia pot ser manipulada, bloquejada, piratejada o eludida per un enemic adaptatiu. En conseqüència, és clar que hi ha raons sòlides de caire moral, legal i de seguretat per a la prohibició preventiva, d'acord amb el dret internacional, del desenvolupament, la producció i la utilització dels robots assassins.

Quan els estats declaren que sempre hi haurà "control humà" o que "un humà participarà en la presa de decisions" d'un sistema d'armament automatitzat, cal que ens plantejem si aquest tipus de control compleix els requisits d'un control humà efectiu. També cal esperar que els estats expliquin com s'asseguraran que es compleixen aquests requisits en els seus processos de revisió de l'armament.

Fotografia: Official U.S. Navy flickr Page. Modificada. [Link a la llicència.](#)

© Generalitat de Catalunya