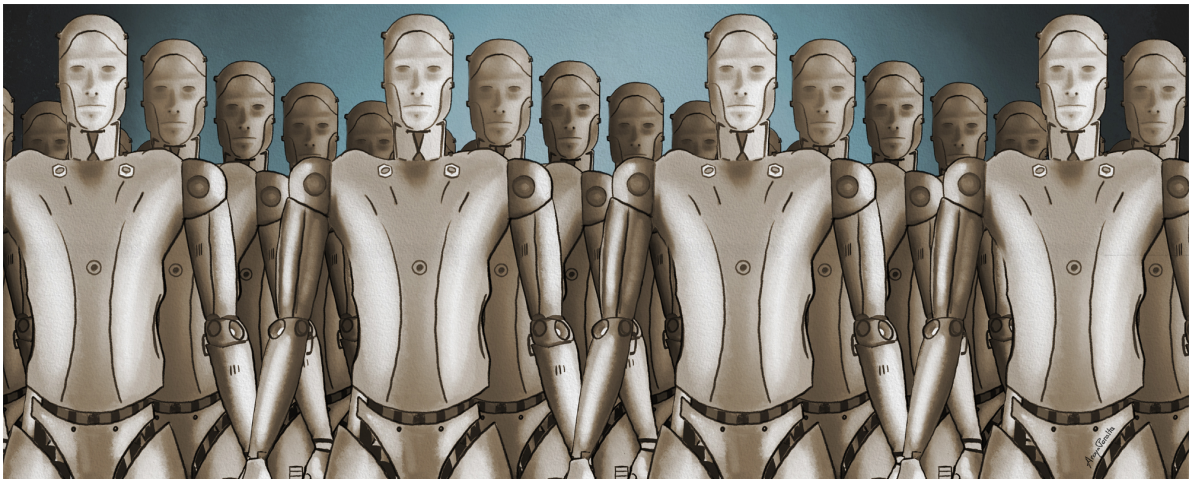
ÚS MILITAR DE LA IA

El costat fosc de la intel·ligència artificial

El cas dels sistemes d'armament letal autònom o els 'Killer Robots'

Roser Martínez Quirante, Joaquín Rodríguez



Araya Peralta

Les primeres dècades del segle XXI han vingut determinades per un procés d'acceleració tecnològica que només pot ser comparat amb altres fites com va ser la revolució industrial. Tot plegat ha comportat un seguit de canvis sistèmics difícils de valorar i comprendre en la seva totalitat.

Mites i realitats de la IA

A més, com a exemple de la magnitud d'aquesta transformació, hem de tenir en compte que, per primera vegada en la història de la nostra espècie, habitem en un context en què decisions crítiques que afecten la vida dels individus, són preses (de forma parcial o en la seva totalitat) per ens no humans, és a dir, per simulacions d'intel·ligència. Efectivament, decisions transcendentals per a la vida d'una persona, com la seva admissió en una universitat, la concessió d'un crèdit o una hipoteca són avui dia en un gran nombre de països preses per algoritmes d'aprenentatge automàtic. Intel·ligències artificials que, a més, tenen la capacitat no només d'afectar individus aïllats sinó a comunitats senceres. Cal recordar en aquest sentit el paper que van jugar determinats algoritmes financers en la

crisi econòmica de 2008, abocant al sistema a vendes massives d'accions de forma autònoma, col·laborant i accelerant la degradació econòmica. O el paper que algoritmes de "*credit scoring*" (calcificació de crèdits) estan jugant a la cristal·lització de la pobresa de les comunitats afroamericanes i hispanes als Estats Units en l'actualitat.

És a dir, ens trobem davant d'un procés de deshumanització que ve determinat per la cessió de determinades decisions a éssers sintètics sense humanitat en nom d'una automatització i estandardització de processos, la qual cosa comporta una erosió en els sistemes de responsabilitat i *accountability*. Es tracta del famós "computer says no" de la sèrie *Little Brittain* estès a la pràctica totalitat de les capes del sistema o del que Postman va descriure com la rendició de la cultura davant la tecnologia.

És evident que vivim rodejats de nous sets tecnològics que incorporen Intel·ligència Artificial: des de Siri a Cortana passant per Alexa o Google Duplex. Ens hem acostumat a compartir la nostra realitat amb simulacions d'intel·ligència, i el que és encara més important i dramàtic, ens han acostumat a regalar-los les nostres dades, en un exercici d'exhibicionisme tecnològic amb conseqüències que poden resultar dramàtiques per a la protecció de les llibertats civils i els drets humans.

> Per primera vegada en la història de la nostra espècie, decisions crítiques que afecten la vida dels individus, són preses per simulacions d'intel·ligència

Una de les principals raons que han alimentat la cessió de privacitat, individualitat i capacitat decisòria que estem vivint, el final de la qual resulta extremadament complex de preveure, són les meta-narratives que han acompanyat la nova revolució tecnològica, basades en que són la solució pels grans problemes que amenacen les nostres societats, com el canvi climàtic, el tractament i cura de malalties, la crisi per falta d'aigua, el control de fronteres com a seguretat nacional, entre d'altres.

Però oblidem que un dels múltiples riscs d'aquestes aplicacions intel·ligents és que tenen el poder no només de donar-li forma a la realitat, sinó fins i tot d'alterar la nostra percepció sobre la mateixa. Els algoritmes de personalització com *Page Rank* en són una mostra, atès que es basen en la falsa promesa de triar-nos de la xarxa i prioritzar els documents que considera que ens han d'interessar.

Sovint exagerem les capacitats i potencialitats benèfiques de la tecnologia i obviem no només les vulnerabilitats del mateix, sinó els propis riscos intrínsecs al seu desenvolupament, implementació i cristal·lització (recordem el que va passar quan Facebook va descobrir que una programa d'IA havia creat el seu propi llenguatge inintel·ligible pels seus creadors).

Així, i tal com ja es va fer amb altres sets tecnològics, com el nuclear o el transgènic, se'ns

argumenta ara que aquesta nova tecnologia resultarà clau per a la diagnosi de certes malalties, per a la distribució equitativa d'aliments o per la lluita contra el canvi climàtic. De la mateixa manera que la indústria transgènica dels 90 ens prometia acabar amb la fam al món, o la incipient indústria nuclear ens prometia una energia barata, neta i segura, ara sembla que no podem posar límits a la IA perquè el seu destí és cercar el bé de la humanitat.

La repetició de la història és clara, i en els fòrums internacionals de primer nivell es tornen a escampar promeses sobre escenaris utòpics als quals s'arriba a través d'un únic camí: cessió de dades, cessió de privacitat i en última instància, cessió d'humanitat. I és que el món de la intel·ligència Artificial està ple d'ombres que cal esvair, sobre tot, tenint en compte que moltes de les dades cedides pels usuaris corren el risc de ser utilitzades en contra seva, ja sigui per corporacions privades o per programes militars, com està ocorrent amb els programes de reconeixement facial a través d'IA.

En el món digital si alguna cosa és gratis, és simplement perquè el que està en venda ets tu. Sobretot, si tenim en compte que el comerç del petroli ha quedat superat pel de les dades ja que aquestes són la matèria primera més valorada del planeta. Això és així perquè que un algoritme sense dades, no és res, i poc a poc anem cap a una societat algorítmica, on el nostre propi comportament i llenguatge, es va adaptant de mica en mica a les necessitats dels algoritmes, i no viceversa. El cas de *Cambridge Analytica* (van mercadejar amb les dades privades de més de 50 milions de persones) resulta un exemple paradigmàtic de les nostres vulnerabilitats, socials i individuals. Però també ho són els experiments socials de Facebook i la seva FaceAPP (una aplicació que transforma la nostra cara per veure com serem de grans però que en el fons el que fem és autoritzar que trafiquin amb les nostres dades biomètriques). Tot plegat mostra cóm de relativament senzill resulta aprofitar-se'n d'una societat on el pensament crític ha estat relegat a la seva més mínima expressió per uns sistemes educatius i mediàtics que l'han devaluat.

Ara bé, el problema és molt més profund. La justificació de la teòrica necessitat d'usar tecnologies que són completament invasives de la nostra privacitat es basen en tres mites.

El primer mite és que les màquines poden adoptar comportaments ètics-morals si aquests són correctament codificats. Però és evident que una màquina no pot tenir ni ètica ni moral ni intuïció pròpia. En tot cas podrà tenir la ètica de qui l'ha codificat. Serà una simulació de la ètica del programador, una rèplica de l'enginyer o una combinació de les dades que trobi al núvol. Però, podem preguntar-nos si un cop codificada la IA, el sistema evolucionarà per sí sol? o si ens condemnarà a una societat de tipus immobilista on el bé i el mal quedin cristal·litzats en la base d'una construcció subjectivitzada en els algoritmes? I si evoluciona… quina serà la seva fita?

> La IA en cap cas pot ser considerada com un agent moral, pel simple fet que es tracta d'una simulació

En definitiva, la intel·ligència artificial, en cap cas pot ser considerada com un agent moral, pel simple fet que es tracta d'una simulació, i per tant no és capaç de comprendre, sota cap tipus de paràmetre una cosa tan senzilla i central com és el valor de una vida humana, ni sentir respecte, ni compassió.

El segon mite es basa en que la IA pot prendre decisions de forma més efectiva, més equànime i més justa que un humà. Res més lluny de la realitat, en primer lloc perquè la IA reprodueix per emulació el sistema ètic-ideològic dels seus creadors, és a dir reprodueix la nostra manca d'imparcialitat. Com ens mostra Cathy O'Neil en la seva obra *Weapons of Math Destruction*, creure en la infal·libilitat dels algoritmes pot dur a resultats dramàtics com els ocorreguts amb les avaluacions de professors en l'estat de Washington, o com ens va mostrar l'*American Civil Liberties Union* pel que fa als sistemes de reconeixement facial que tenen una alta tendència a identificar subjectes no caucàsics com criminals.

Estem davant d'una tecnologia dissenyada per homes blancs, amb el sistema mental propi dels mateixos, on les seves fílies i fòbies tendeixen a ser traslladades a les seves creacions. És més, en tractar-se d'un sistema heurístic resulta altament complex saber el procés mitjançant el qual la IA ha pres una determinada decisió. Per tant, si resulta impossible desconstruir o explicar el procés que ha portat a una determinada decisió a la IA, és una irresponsabilitat deixar-les que operin lliurement.

I, finalment, arribem al tercer mite que afirma que la intel·ligència artificial és més fiable que la intel·ligència humana, cosa que en anàlisis molt específics podria ser acceptat però mai en termes generals. Cal destacar aquí la feina feta per la ONG britànica *Big Brother is watching us* que, apel·lant a l'acte de llibertat d'informació, van aconseguir que el govern revelés la fiabilitat dels sistemes de reconeixement facial que es van fer servir durant el Carnestoltes de Candem. El resultat va ser que només un 5% de les identificacions de criminals fetes a través del sistema d'IA eres correctes, donant una errada mitjana del 95%.

Aquesta mitologia té encara resultats molt més preocupants si tenim en compte els estudis del professor Noel Sharkey, qui va elaborar la teoria de *Automation bias* on explica que els humans tenim tendència a donar per vàlids els judicis i anàlisi fets per la IA, atès que pensem que és més efectiva i fiable que nosaltres mateixos.

Però el més sorprenent, és que tot i saber que la IA no pot ser considerada com un agent moral, i tot i saber les seves limitacions a l'hora d'interpretar a la realitat a causa dels biaixos propis dels seus creadors i de la pròpia societat (especialment en sistemes que es nodreixen de llenguatge natural) la seva penetració continua augmentant, i cada vegada més processos són guiats a través d'aquests sistemes.

El control humà significatiu i els sistemes d'armament letals autònoms

És així com arribem a l'aspecte més aberrant del tema: la IA aplicada als sistemes d'armament letal autònom (*LAWS* per les sigles en anglès). Es tracta del sorgiment d'una

nova generació d'armes amb capacitat de seleccionar i eliminar objectius sense un control humà significatiu. És a dir, estem davant d'una delegació de capacitats letals a una suposada intel·ligència artificial, la qual tindrà potestat no només per decidir qui rep o no un crèdit, qui és acceptat o no en una universitat, qui accedeix o no a un determinat lloc de treball, sinó directament, qui viu i qui mor.

Estem parlant d'una tipologia d'armament que obvia la dimensió racional, cooperativa, intuïtiva, moral i ètica de les decisions humanes, contradiu el dret internacional humanitari, les lleis de la guerra i, internament, el dret administratiu atès que el monopoli de la violència legítima està en mans de l'Estat.

Tots els sistemes letals d'armes autònomes (drons o robots) desenvolupats fins ara depenen o haurien de dependre de la supervisió humana o del judici humà. És a dir, haurien de tenir un control humà significatiu previ en almenys algunes de les seves fases crítiques (selecció d'objectius o cancel·lació de l'ordre). No obstant això, s'està investigant i s'estan desenvolupant sistemes amb vocació de total autonomia, situació permesa perquè no hi ha una regulació clara al respecte. Aquesta inactivitat dels estats està portant a una espècie de carrera competitiva sense llei entre governs que pot ser molt perillosa i que a Nacions Unides, com a membres del Comitè Internacional pel control d'armes robòtiques (ICRAC) estem intentant frenar amb l'aprovació d'un tractat multilateral que prohibeixi aquest tipus d'arma genocida.

Malgrat els esforços, la majoria dels estats de pes, justifiquen la investigació d'aquesta tecnologia letal assegurant que no es farà servir en atacs sinó per a defensa nacional. Però això no sembla més que un subterfugi per ser els primers en posar en marxa sistemes absolutament letals dotats de la capacitat d'independitzar-se del seu creador i el seu responsable. Per això és essencial desenvolupar una regulació internacional que prohibeixi els usos letals de la IA, i que limiti clarament l'existència de vasos comunicants entre el desenvolupament de sistemes de defensa nacional i aquells el propòsit dels quals és l'acció letal contra les persones. En cas contrari, algú podria arribar a atribuir a una màquina sense humanitat el poder de decidir qui eliminar, és a dir, podrien crear un robot, un dron, amb llicència per matar.

Així, en els últims anys s'han començat a detectar moviments que anuncien el començament d'una nova carrera armamentista amb conseqüències que poden ser desastroses pel futur de la nostra espècie. Xina, per exemple, està modernitzant ràpidament el seu exèrcit i ha optat per armes nuclears d'última generació a través d'ogives amb IA per limitar el dany durant l'atac a objectius específics. En contrast, Estats Units segueix sent l'hereu de les armes del passat, el que fa que es moguin més lentament, en el que s'ha anomenat el "complex militar-industrial-congressional" (MICC) en referència a que el Congrés dels EEUU forma una relació tripartida anomenada Triangle de Ferro (relacions entre contractistes militars privats, el Govern i el Congrés).

D'aquesta manera, entre 2014 i 2018, la Xina va realitzar al voltant de 200 experiments de laboratori per simular una explosió nuclear, mentre que EE. UU., en el mateix període, va realitzar-ne 50. La cursa empresa per la Xina és evident. Al final, com assenyala Hartnett,

del *Bank of America*, "la guerra comercial de 2018 hauria de ser reconeguda pel que realment és: la primera etapa d'una nova carrera armamentista entre els EE. UU. i la Xina per aconseguir la superioritat en tecnologia durant a llarg termini a través de la computació quàntica, intel·ligència artificial, avions de combat hipersònics, vehicles electrònics, robòtica i ciberseguretat."

Per tant, la inversió en tecnologia està vinculada a la despesa de defensa (encara que això no sempre significa obtenir una seguretat més gran): el pronòstic de l'FMI és que la Xina superarà progressivament als EE. UU. Fins 2050, i què es convertirà en la superpotència dominant al món. Específicament, es calcula que al voltant de 2032, superarà l'economia i la força militar dels EE. UU., així com la seva influència global.

La mort a mans d'un sistema autònom amb IA va contra la dignitat humana

En els conflictes armats, el dret a la vida significa el dret a no ser assassinat de manera arbitrària o capritxosa, inexplicable o inhumana o com a dany col·lateral i no pot vulnerar el dret a la dignitat humana. Fins i tot es pot dir que la dignitat humana és un dret més important que el dret a la vida, perquè fins i tot en una societat civilitzada, pot donar-se el cas d'execucions legals, però aquestes no poden vulnerar la dignitat humana.

> Tot i que les LAWS podrien oferir millors resultats basats en un càlcul de cost-benefici, s'haurien de prohibir per raons ètiques i legals

La por a un futur distòpic sembla una raó legítima per a una prohibició total o una moratòria de les LAWS mitjançant l'aplicació del principi de precaució, però per defensar aquesta posició, la noció de dignitat humana i la clàusula de Martens s'han d'enfortir-prèviament, com així com els conceptes relacionats amb el control humà significatiu i l'autodeterminació dels sistemes letals autònoms.

A més, creiem que, tot i que les LAWS podrien oferir millors resultats basats en un càlcul de cost-benefici, s'haurien de prohibir per raons ètiques i legals. Heyns, qui té la mateixa opinió, ho basa en la concepció de Kant de la dignitat humana, segons la qual les persones tenen el dret inherent de ser tractats com a éssers humans únics i complets, especialment quan les seves vides estan en joc. Aquesta dignitat humana es negaria si les víctimes que volguessin apel·lar a la humanitat del seu botxí no poguessin fer-ho perquè es tracta d'un ésser artificial. El poder executiu ha d'oferir el degut respecte a la dignitat de la persona que considera el cas específic i realitza avaluacions i ajustos constants. A més, res d'aquesta aplicació de la llei amb les característiques de les capacitats humanes es pot garantir amb armes autònomes, ja que hi hauria una falta de judici humà adequat a les seves accions.

També cal aprofundir en noves formes de convivència considerant que la deshumanització ja provocada pels sistemes autònoms amb control humà en els conflictes de guerra deixa en el paper tot el que s'havia après a la Primera Guerra Mundial sobre la cooperació i la dignitat humana, comunicació verbal i sobre la relació humana entre combatents. El progrés en la comunicació humanitària no verbal s'atura i fins i tot retrocedeix quan es lluita amb drons autònoms letals. En paraules de Sparrow, "hem de mantenir una relació interpersonal amb altres éssers humans fins i tot durant la guerra" o no respectarem els fonaments de la llei.

Els defensors d'aquests nous sistemes d'armes intel·ligents, ignorant la necessitat d'aquest component de la humanitat, els atribueixen nombrosos avantatges com: reducció dels costos operatius, desenvolupament de certes tasques més ràpidament que els humans, alta capacitat per arribar a un objectiu fins i tot quan la comunicació dels enllaços es veuen afectats. Arkin, a més, assenyala en la seva defensa que poden dissenyar-se per acceptar els riscos més alts, que poden tenir els millors sensors, que no seran sacsejats per emocions com la por o la ira, que no patiran prejudicis cognitius i fins i tot que poden distingir legítimament i confiable els objectius legítims dels il·legítims.

Aquests avantatges podrien ser certs, però hi ha nombrosos exemples d'homes i dones de tot tipus i condicions que en un moment es van negar a pressionar el botó que hauria provocat la mort de ciutadans. Les guerres han evolucionat en humanitat perquè la comunicació no verbal des de la guerra de trinxeres va permetre moments de treva i baixa letalitat sense que els soldats haguessin rebut cap ordre en aquest sentit.

La supressió d'una vida humana per no ser considerada arbitrària, s'ha de basar en una decisió informada i un judici cognitiu humà, ja que només una decisió humana garanteix el ple reconeixement del valor de la vida individual i la importància de la seva pèrdua.

I així entren en joc tots els estàndards moderns i complexos del dret humanitari: proporcionalitat, compassió, ús de mètodes menys onerosos o menys restrictius, vigilància constant, cavallerositat… En conseqüència, les accions dels drons letals autònoms per disposar d'IA no són legítimes ni moralment justificables i s'haurien de prohibir sota el principi de dignitat humana i *ius cogens*, que com a norma obligatòria conté les normes fonamentals del dret humanitari.

D'altra banda, la intuïció és part de la nostra essència com a humans i de totes les nostres accions, i sempre ha jugat un paper fonamental en la guerra. A més, les LAWS poden dotar-se de mecanismes d'imitació i incorporar processos integradors i cognitius, però no fenomenològics. Mai poden ser intuïtius o sentir emocions, sinó només replicar. Com diu el neurocientífic G. Rizzolatti, descobridor de les neurones mirall, "els robots poden imitar, no sentir". A més, si aquest és el cas, com que els algoritmes inclosos en els sistemes autònoms letals no poden assolir les característiques humanes necessàries per prendre decisions públiques discrecionals transcendentals que es refereixen a l'exercici de la força legítima contra les persones, la transferència, la descentralització d'aquests poders als sistemes autònoms no ha de ser acceptat. El poder de decidir qui és l'enemic (dins o fora de l'Estat) i apoderar-se de vides humanes de forma discrecional és tan transcendent que no

es pot atorgar a éssers artificials sense emocions humanes.

McQuillan adverteix que la vigilància, gràcies a l'acumulació massiva i detallada de dades a través de sistemes intel·ligents, està generant canvis en la governança i danys en el nucli de la societat civil de tal nivell que l'anomena "l'estat d'excepció algorítmic."

> L'única garantia de progrés i sostenibilitat dels drets dels ciutadans davant la intel·ligència artificial en sistemes autònoms és la regulació

Fins i tot Mark Zuckerberg, CEO de Facebook, va reconèixer implícitament davant el Congrés dels Estats Units que ens enfrontem a un estat anòmic i que necessitem un regulador que no confiï tot en el lliure mercat: "la regulació federal de Facebook i altres companyies d'Internet és indispensable." Serà a través d'aquesta legislació federal quan hi haurà una projecció internacional i, en última instància, una globalització, ja que podria tenir efectes extraterritorials en altres països, com ha succeït amb altres regulacions americanes. No obstant això, fins ara no hi ha instruments internacionals legalment vinculants o fins i tot lleis nacionals que prohibeixin el desenvolupament, la producció i l'ús dels anomenats robots assassins.

L'única garantia de progrés i sostenibilitat dels drets dels ciutadans davant la intel·ligència artificial en sistemes autònoms és la regulació. La pròpia evolució de la tecnologia, que es pot veure profundament afectada per usos que van en contra del criteri de l'opinió pública, de tal manera que la totalitat de la tecnologia es vegi compromesa, com va passar amb la nuclear o la química. De la mateixa manera, una relaxació de la intervenció en aquesta tecnologia pot conduir al seu propi cap a la fi de la humanitat mateixa.

El perill més pertorbador i inquietant de les LAWS: els drons letals de butxaca

Els esforços que estem fent en les reunions d'Experts governamentals a Nacions Unides per aconseguir un tractat que prohibeixi les armes letals autònomes (CCCW) es focalitzen en grans armes per a la guerra (macro LAWS com Reaper, Tarannis, Iron Drome, etc.)

Però cal anar més enllà i reconèixer que el veritable perill són les armes petites en mans de particulars, és a dir, el que es pot arribar a anomenar microLAWS que poden passar d'estar controlats per les forces militars a estar en mans de qualsevol ciutadà per la seva seguretat privada. Seria un altre exemple de tecnologia letal de doble ús.

Una situació que pot desestabilitzar-nos i pot fer canviar els estàndards de seguretat que teníem fins ara. Si els LAWS militars ja són difícils d'atacar quan ho fan en forma d'eixams de minidrons, no és complicat d'imaginar-nos el que pot passar si aquests passen a estar

sota el poder de milers d'individus que, enlloc d'un arma convencional, escullen un dron letal per dotar-se de la seguretat que l'Estat no pot garantir-los.

El dret a portar armes que garanteix la Segona Esmena de la Constitució americana permet als seus ciutadans no només tenir una pistola o un revòlver sinó qualsevol arma que es consideri necessària per a la seva seguretat com les automàtiques o les militars. Portat a l'extrem, en la mateixa línia podrien reivindicar el dret a la possessió d'un robot autònom letal per protegir-los defensivament o ofensiva. És a dir, el dret a posseir un dron de butxaca autònom letal amb IA.

La regulació a nivell internacional i administrativa global ha de ser preventiva i ha de frenar aquesta situació atès que de l'ús militar d'aquesta tecnologia es passarà a un ús públic-civil com estem veient amb els drons utilitzats per la policia o els serveis de protecció civil, o fins i tot privats.

Això pot desencadenar una situació d'inseguretat global degut a la proliferació d'aquest tipus d'armes de forma descontrolada en mans privades i degut a la dificultat per preveure la interrelació de sistemes letals amb IA entre ells. Esperem que el dret no arribi massa tard per poder frenar aquesta pandèmia anunciada.

Roser Martínez Quirante
Roser Martínez Quirante és professora de Dret Administratiu a la Universitat Autònoma de Barcelona des del 2002. També és professora de l'Escola de Prevenció i Seguretat Integral de la UAB des de la seva fundació, l'any 2004, i ha impartit classes en diferents matèries, entre les quals destaquen la Llei de seguretat, intervenció i autoregulació i la regulació d'armes de foc als Estats Units i a Europa. Ha estat coordinadora del grup de recerca EPSI-University de Massachussetts Lowell (UMAAS) per al desenvolupament d'activitats docents i de recerca sobre seguretat. És experta en l'àmbit de les armes autònomes i defensora de la campanya *Stop Killer Robots*.

Joaquín Rodríguez
Joaquín Rodríguez és investigador a la Fundació de la Universitat Autònoma de Barcelona i coordinador local de la xarxa Leading Cities Network. És professor de l'Escola de Prevenció i Seguretat Integral, centre adscrit a la UAB, i també és un dels promotors a l'estat espanyol de la campanya *Stop Killer Robots*, que pretén prevenir la proliferació de sistemes d'armament autònom. És Doctor especialitzat en anàlisi de riscos i en les relacions entre societat i tecnologia. Té un Màster en Relacions Internacionals amb una especialització en Estudis de Pau i Seguretat per l'IBEI (Institut de Relacions Internacionals de Barcelona) i un postgrau en Gestió de projectes pel Centre d'estudis Alts Acadèmics de l'Organització d'Estats Iberoamericans.