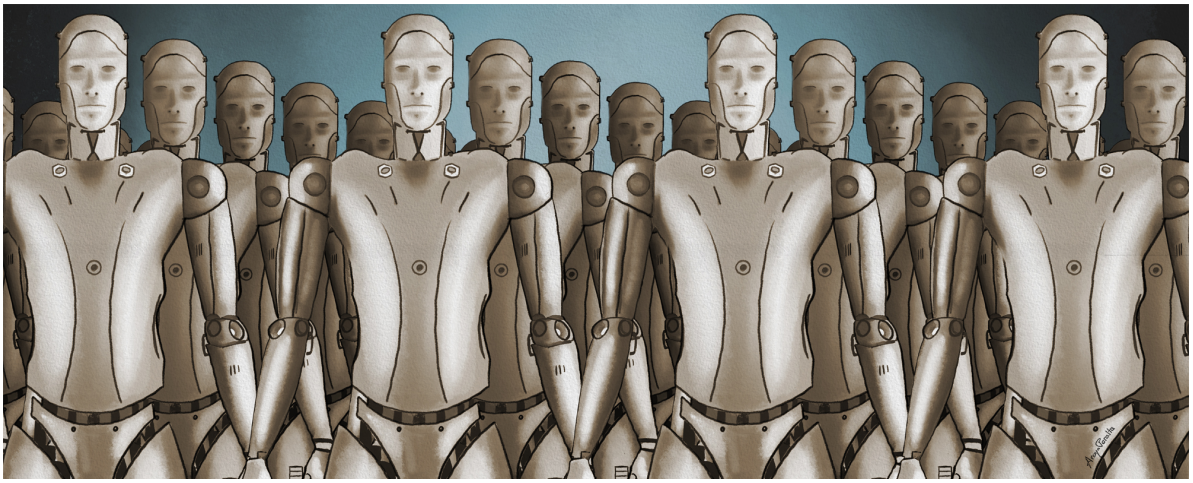
USO MILITAR DE LA IA

El lado oscuro de la Inteligencia artificial

El caso de los sistemas de armamento letal autónomo o los Killer Robots

Roser Martínez Quirante, Joaquín Rodríguez



Araya Peralta

Las primeras décadas del siglo XXI han venido determinadas por un proceso de aceleración tecnológica que sólo puede ser comparado con otros hitos como fue la revolución industrial. Al fin y al cabo ha comportado una serie cambios sistémicos difíciles de valorar y comprender en su totalidad.

Mitos y realidades de la IA

Además, como ejemplo de la magnitud de esta transformación, tenemos que tener en cuenta que, por primera vez en la historia de nuestra especie, habitamos en un contexto en el que decisiones críticas que afectan a la vida de los individuos, son tomadas (de forma parcial o en su totalidad) por entes no humanos, es decir, por simulaciones de inteligencia. Efectivamente, decisiones trascendentales para la vida de una persona, como su admisión en una universidad, la concesión de un crédito o una hipoteca son hoy día en un gran número de países tomadas por algoritmos de aprendizaje automático. Inteligencias artificiales que, además, tienen la capacidad no sólo de afectar individuos aislados sino a comunidades enteras. Hay que recordar en este sentido el papel que jugaron determinados

algoritmos financieros en la crisis económica de 2008, vertiendo al sistema a ventas masivas de acciones de forma autónoma, colaborando y acelerando la degradación económica. O el papel que algoritmos de «credit scoring» (calificación de créditos) están jugando a la cristalización de la pobreza de las comunidades afroamericanas e hispanas en Estados Unidos en la actualidad.

Es decir, nos encontramos ante un proceso de deshumanización que viene determinado por la cesión de determinadas decisiones a seres sintéticos sin humanidad en nombre de una automatización y estandarización de procesos, lo cual comporta una erosión en los sistemas de responsabilidad y *accountability*. Se trata del famoso «computer says no» de la serie *Little Brittain* extendido a la práctica totalidad de las capas del sistema o de lo que Postman describió como la rendición de la cultura ante la tecnología.

Es evidente que vivimos rodeados de nuevos sets tecnológicos que incorporan Inteligencia Artificial: desde Siri en Cortana pasando por Alexa o Google Duplex. Nos hemos acostumbrado a compartir nuestra realidad con simulaciones de inteligencia, y lo que es todavía más importante y dramático, nos han acostumbrado a regalarles nuestros datos, en un ejercicio de exhibicionismo tecnológico con consecuencias que pueden resultar dramáticas para la protección de las libertades civiles y los derechos humanos.

> Por primera vez en la historia, decisiones críticas que afectan a la vida de los individuos, son tomadas por simulaciones de inteligencia

Una de las principales razones que han alimentado la cesión de privacidad, individualidad y capacidad decisoria que estamos viviendo, el final de la cual resulta extremadamente complejo de prever, son las meta-narrativas que han acompañado a la nueva revolución tecnológica, basadas en que son la solución para los grandes problemas que amenazan nuestras sociedades, como el cambio climático, el tratamiento y cuidado de enfermedades, la crisis por falta de agua, el control de fronteras como seguridad nacional, entre otros.

Pero olvidamos que uno de los múltiples riesgos de estas aplicaciones inteligentes es que tienen el poder no sólo de darle forma a la realidad, sino incluso de alterar nuestra percepción sobre la misma. Los algoritmos de personalización como *Page Rank* son una muestra dado que se basan en la falsa promesa de escogernos de la red y priorizar los documentos que considera que nos tienen que interesar.

A menudo exageramos las capacidades y potencialidades benéficas de la tecnología y obviamos no sólo las vulnerabilidades de lo mismo, sino los propios riesgos intrínsecos a su desarrollo, implementación y cristalización (recordamos lo que pasó cuando Facebook descubrió que un programa de IA había creado su propio lenguaje ininteligible para sus creadores).

Así, y tal como ya se hizo con otros sets tecnológicos, como el nuclear o el transgénico, se nos argumenta ahora que esta nueva tecnología resultará clave para la diagnosis de ciertas enfermedades, para la distribución equitativa de alimentos o para la lucha contra el cambio climático. De la misma manera que la industria transgénica de los 90 nos prometía acabar con el hambre en el mundo, o la incipiente industria nuclear nos prometía una energía barata, limpia y segura, ahora parece que no podemos poner límites a la IA porque su destino es buscar el bien de la humanidad.

La repetición de la historia es clara, y en los foros internacionales de primer nivel se vuelven a extender promesas sobre escenarios utópicos a los cuales se llega a través de un único camino: cesión de datos, cesión de privacidad y en última instancia, cesión de humanidad. Y es que el mundo de la inteligencia Artificial está lleno de sombras que hay que invadir, sobre todo, teniendo en cuenta que muchos de los datos cedidos por los usuarios corren el riesgo de ser utilizadas en contra suya, ya sea por corporaciones privadas o por programas militares, como está ocurriendo con los programas de reconocimiento facial a través de IA.

En el mundo digital si alguna cosa es gratis, es simplemente porque el que está en venta eres tú. Sobretodo, si tenemos en cuenta que el comercio del petróleo ha quedado superado por el de los datos ya que estas son la materia prima más valorada del planeta. Eso es así porque que un algoritmo sin datos no es nada, y poco a poco vamos hacia una sociedad algorítmica, donde nuestro propio comportamiento y lenguaje, se va adaptando poco a poco a las necesidades de los algoritmos, y no viceversa. El caso de *Cambridge Analytica* (mercadearon con los datos privados de más de 50 millones de personas) resulta un ejemplo paradigmático de nuestras vulnerabilidades, sociales e individuales. Pero también lo son los experimentos sociales de Facebook y su FaceAPP (una aplicación que transforma nuestra cara para ver cómo seremos de mayores pero que en el fondo lo que hacemos es autorizar que trafiquen con nuestros datos biométricos). Todo muestra cómo de relativamente sencillo resulta aprovecharse de una sociedad donde el pensamiento crítico ha sido relegado a su más mínima expresión por unos sistemas educativos y mediáticos que lo han devaluado.

Ahora bien, el problema es mucho más profundo. La justificación de la teórica necesidad de usar tecnologías que son completamente invasivas de nuestra privacidad se basan en tres mitos.

El primer mito es que las máquinas pueden adoptar comportamientos éticos-morales si estos son correctamente codificados. Pero es evidente que una máquina no puede tener ni ética ni moral ni intuición propia. En todo caso podrá tener la ética de quien lo ha codificado. Será una simulación de la ética del programador, una réplica del ingeniero o una combinación de los datos que encuentre en la nube. ¿Sin embargo, podemos preguntarnos si una vez codificada la IA, el sistema evolucionará por sí solo? ¿o si nos condenará a una sociedad de tipo inmovilista donde el bien y el mal queden cristalizados en la base de una construcción subjetivitzada en los algoritmos? ¿Y si evoluciona…cuál será su hito?

> La IA en ningún caso puede ser considerada como un agente moral, por el simple hecho de que se trata de una simulación

En definitiva, la Inteligencia Artificial en ningún caso puede ser considerada como un agente moral, por el simple hecho de que se trata de una simulación, y por lo tanto no es capaz de comprender, bajo ningún tipo de parámetro una cosa tan sencilla y central como es el valor de una vida humana, ni sentir respeto, ni compasión.

El segundo mito se basa en que la IA puede tomar decisiones de forma más efectiva, más ecuánime y más justa que un humano. Nada más lejos de la realidad, en primer lugar, porque la IA reproduce por emulación el sistema ético-ideológico de sus creadores, es decir reproduce nuestra falta de imparcialidad. Como nos muestra Cathy O'Neil en su obra *Armas de destrucción matemática*, creer en la infalibilidad de los algoritmos puede llevar a resultados dramáticos como los ocurridos con las evaluaciones de profesores en el estado de Washington, o como nos mostró el *American Civil Liberties Union* con respecto a los sistemas de reconocimiento facial que tienen una alta tendencia a identificar sujetos no caucásicos como criminales.

Estamos ante una tecnología diseñada por hombres blancos, con el sistema mental propio de los mismos, donde sus filias y fobias tienden a ser trasladadas a sus creaciones. Es más, al tratarse de un sistema heurístico resulta altamente complejo saber el proceso mediante el cual la IA ha tomado una determinada decisión. Por lo tanto, si resulta imposible deconstruir o explicar el proceso que ha llevado a una determinada decisión en la IA, es una irresponsabilidad dejarlas que operen libremente.

Y, finalmente, llegamos al tercer mito que afirma que la inteligencia artificial es más fiable que la inteligencia humana, cosa que en análisis muy específicos podría ser aceptado, pero nunca en términos generales. Hay que destacar aquí el trabajo hecho por la ONG británica *Big Brother is watching us* que, apelando al acto de libertad de información, consiguieron que el gobierno revelara la fiabilidad de los sistemas de reconocimiento facial que se utilizaron durante el Carnaval de Candem. El resultado fue que sólo un 5% de las identificaciones de criminales hechas a través del sistema d'IA eran correctas, dando un error medio del 95%.

Esta mitología tiene todavía resultados mucho más preocupantes si tenemos en cuenta los estudios del profesor Noel Sharkey, quien elaboró la teoría del *Automation bias* donde explica que los humanos tenemos tendencia a dar por válidos los juicios y análisis hechos por la IA, dado que pensamos que es más efectiva y fiable que nosotros mismos.

Pero lo más sorprendente, es que a pesar de saber que la IA no puede ser considerada como un agente moral, y a pesar de saber sus limitaciones a la hora de interpretar a la realidad a causa de los sesgos propios de sus creadores y de la propia sociedad (especialmente en sistemas que se nutren de lenguaje natural) su penetración sigue aumentando, y cada vez más procesos son guiados a través de estos sistemas.

El control humano significativo y los sistemas de armamento letales autónomos

Es así como llegamos al aspecto más aberrante del tema: la IA aplicada a los sistemas de armamento letal autónomo (*LAWS* por las siglas en inglés). Se trata del surgimiento de una nueva generación de armas con capacidad de seleccionar y eliminar objetivos sin un control humano significativo. Es decir, estamos ante una delegación de capacidades letales a una supuesta inteligencia artificial, la cual tendrá potestad no sólo para decidir a quien recibe o no un crédito, quien es aceptado o no en una universidad, quien accede o no a un determinado puesto de trabajo, sino directamente, quién vive y quién muere.

Estamos hablando de una tipología de armamento que obvia la dimensión racional, cooperativa, intuitiva, moral y ética de las decisiones humanas, contradice el derecho internacional humanitario, las leyes de la guerra e, internamente, el derecho administrativo dado que el monopolio de la violencia legítima está en manos del Estado.

Todos los sistemas letales de armas autónomas (drones o robots) desarrollados hasta ahora dependen o tendrían que depender de la supervisión humana o del juicio humano. Es decir, tendrían que tener un control humano significativo previo en al menos algunas de sus fases críticas (selección de objetivos o cancelación del orden). No obstante, se está investigando y se están desarrollando sistemas con vocación de total autonomía, situación permitida porque no hay una regulación clara al respecto. Esta inactividad de los estados está llevando a una especie de carrera competitiva sin ley entre gobiernos que puede ser muy peligrosa y que, en Naciones Unidas, como miembros del Comité Internacional para el control de armas robóticas (ICRAC), estamos intentando frenar con la aprobación de un tratado multilateral que prohíba este tipo de arma genocida.

A pesar de los esfuerzos, la mayoría de los estados de peso, justifican la investigación de esta tecnología letal asegurando que no se utilizará en ataques sino para defensa nacional. Pero eso no parece más que un subterfugio para ser los primeros al poner en marcha sistemas absolutamente letales dotados de la capacidad de independizarse de su creador y el suyo responsable. Por eso es esencial desarrollar una regulación internacional que prohíba los usos letales de la IA, y que limite claramente la existencia de vasos comunicantes entre el desarrollo de sistemas de defensa nacional y aquellos cuyo propósito es la acción letal contra las personas. En caso contrario, alguien podría llegar a atribuir a una máquina sin humanidad el poder de decidir a quién eliminar, es decir, podrían crear un robot, un dron, con licencia para matar.

Así, en los últimos años se han empezado a detectar movimientos que anuncian el comienzo de una nueva carrera armamentista con consecuencias que pueden ser desastrosas por el futuro de nuestra especie. China, por ejemplo, está modernizando rápidamente su ejército y ha optado por armas nucleares de última generación a través de ojivas con IA para limitar el daño durante el ataque a objetivos específicos. En contraste, Estados Unidos sigue siendo el heredero de las armas del pasado, lo que hace que se muevan más lentamente, en lo que se ha denominado el «complejo militar-industrial-congresional» (MICC) en referencia

a que el Congreso de los EEUU forma una relación tripartita denominada Triángulo de Hierro (relaciones entre contratistas militares privados, el Gobierno y el Congreso).

De esta manera, entre 2014 y 2018, China realizó en torno a 200 experimentos de laboratorio por simular una explosión nuclear, mientras que EE. UU., en el mismo periodo, realizó 50. La carrera emprendida por China es evidente. Al final, como señala Hartnett, del *Bank of America*, «la guerra comercial de 2018 tendría que ser reconocida por lo que realmente es: la primera etapa de una nueva carrera armamentista entre EE. UU. y China para conseguir la superioridad en tecnología durante a largo plazo a través de la computación cuántica, inteligencia artificial, aviones de combate hipersónicos, vehículos electrónicos, robótica y ciberseguridad «.

Por lo tanto, la inversión en tecnología está vinculada al gasto de defensa (aunque eso no siempre significa obtener una seguridad mayor): el pronóstico del FMI es que China superará progresivamente en los EE. UU. Hasta 2050, y qué se convertirá en la superpotencia dominante en el mundo. Específicamente, se calcula que en torno a 2032, superará la economía y la fuerza militar de los EE. UU., así como su influencia global.

La muerte en manos de un sistema autónomo con IA va contra la dignidad humana

En los conflictos armados, el derecho a la vida significa el derecho a no ser asesinado de manera arbitraria o caprichosa, inexplicable o inhumana o como daño colateral y no puede vulnerar el derecho a la dignidad humana. Incluso se puede decir que la dignidad humana es un derecho más importante que el derecho a la vida, porque incluso en una sociedad civilizada, puede darse el caso de ejecuciones legales, pero estas no pueden vulnerar la dignidad humana.

> Aunque las LAWS podrían ofrecer mejores resultados basados en un cálculo de coste-beneficio, se tendrían que prohibir por razones éticas y legales

El miedo a un futuro distópico parece una razón legítima para una prohibición total o una moratoria de las LAWS mediante la aplicación del principio de precaución, pero para defender esta posición, la noción de dignidad humana y la cláusula de Martens se tienen que fortalecer-previamente, como así como los conceptos relacionados con el control humano significativo y la autodeterminación de los sistemas letales autónomos.

Además, creemos que, aunque las LAWS podrían ofrecer mejores resultados basados en un cálculo de coste-beneficio, se tendrían que prohibir por razones éticas y legales. Heyns, quien tiene la misma opinión, lo basa en la concepción de Kant de la dignidad humana, según la cual las personas tienen el derecho inherente de ser tratados como seres humanos

únicos y completos, especialmente cuando sus vidas están en juego. Esta dignidad humana se ahogaría si las víctimas que quisieran apelar a la humanidad de su verdugo no pudieran hacerlo porque se trata de un ser artificial. El poder ejecutivo tiene que ofrecer el debido respeto a la dignidad de la persona que considera el caso específico y realiza evaluaciones y ajustes constantes. Además, nada de esta aplicación de la ley con las características de las capacidades humanas se puede garantizar con armas autónomas, ya que habría una falta de juicio humano adecuado a sus acciones.

También debemos profundizar en nuevas formas de convivencia considerando que la deshumanización ya provocada por los sistemas autónomos con control humano en los conflictos de guerra deja en el papel todo lo que se había aprendido en la Primera Guerra Mundial sobre la cooperación y la dignidad humana, comunicación verbal y sobre la relación humana entre combatientes. El progreso en la comunicación humanitaria no verbal se detiene e incluso retrocede cuando se lucha con drones autónomos letales. En palabras de Sparrow, «tenemos que mantener una relación interpersonal con otros seres humanos incluso durante la guerra» o no respetaremos los fundamentos de la ley.

Los defensores de estos nuevos sistemas de armas inteligentes, ignorando la necesidad de este componente de la humanidad, les atribuyen numerosas ventajas como: reducción de los costes operativos, desarrollo de ciertas tareas más rápidamente que los humanos, alta capacidad para llegar a un objetivo incluso cuando la comunicación de los enlaces se ven afectados… Arkin además señala en su defensa que pueden diseñarse para aceptar los riesgos más altos, que pueden tener los mejores sensores, que no serán sacudidos por emociones como el miedo o la ira, que no sufrirán prejuicios cognitivos e incluso que pueden distinguir legítimamente y de forma confiable los objetivos legítimos de los ilegítimos.

Estas ventajas podrían ser ciertas, pero hay numerosos ejemplos de hombres y mujeres de todo tipo y condiciones que en un momento se negaron a presionar el botón que habría provocado la muerte de ciudadanos. Las guerras han evolucionado en humanidad porque la comunicación no verbal desde la guerra de trincheras permitió momentos de tregua y baja letalidad sin que los soldados hubieran recibido ningún orden en este sentido.

La supresión de una vida humana, para no ser considerada arbitraria, se tiene que basar en una decisión informada y un juicio cognitivo humano, ya que sólo una decisión humana garantiza el pleno reconocimiento del valor de la vida individual y la importancia de su pérdida.

Y así entran en juego todos los estándares modernos y complejos del derecho humanitario: proporcionalidad, compasión, uso de métodos menos onerosos o menos restrictivos, vigilancia constante, caballerosidad… En consecuencia, las acciones de los drones letales autónomos para disponer de IA no son legítimas ni moralmente justificables y se tendrían que prohibir bajo el principio de dignidad humana e *ius cogens*, que como norma obligatoria contiene las normas fundamentales del derecho humanitario.

Por otra parte, la intuición es parte de nuestra esencia como humanos y de todas nuestras

acciones, y siempre ha jugado un papel fundamental en la guerra. Además, las LAWS pueden dotarse de mecanismos de imitación e incorporar procesos integradores y cognitivos, pero no fenomenológicos. Nunca pueden ser intuitivos o sentir emociones, sino sólo replicar. Como dice al neurocientífico G. Rizzolatti, descubridor de las neuronas espejo, «los robots pueden imitar, no sentir». Además, si este es el caso, como los algoritmos incluidos en los sistemas autónomos letales no pueden alcanzar las características humanas necesarias para tomar decisiones públicas discrecionales trascendentales que se refieren al ejercicio de la fuerza legítima contra las personas, la transferencia, la descentralización de estos poderes a los sistemas autónomos no tiene que ser aceptada. El poder de decidir quién es el enemigo (dentro o fuera del Estado) y apoderarse de vidas humanas de forma discrecional es tan trascendente que no se puede otorgar a seres artificiales sin emociones humanas.

McQuillan advierte que la vigilancia, gracias a la acumulación masiva y detallada de datos a través de sistemas inteligentes, está generando cambios en la gobernanza y daños en el núcleo de la sociedad civil de tal nivel que lo llama «el estado de excepción algorítmico».

> La única garantía de progreso y sostenibilidad de los derechos de los ciudadanos ante la inteligencia artificial en sistemas autónomos es la regulación

Incluso Mark Zuckerberg, CEO de Facebook, reconoció implícitamente ante el Congreso de los Estados Unidos que nos enfrentamos a un estado anómico y que necesitamos un regulador que no confíe todo en el libre mercado: «la regulación federal de Facebook y otras compañías de Internet es indispensable». Será a través de esta legislación federal cuando haya una proyección internacional y, en última instancia, una globalización, ya que podría tener efectos extraterritoriales en otros países, como ha sucedido con otras regulaciones americanas. No obstante, hasta ahora no hay instrumentos internacionales legalmente vinculantes o incluso leyes nacionales que prohíban el desarrollo, la producción y el uso de los llamados robots asesinos.

La única garantía de progreso y sostenibilidad de los derechos de los ciudadanos ante la inteligencia artificial en sistemas autónomos es la regulación. La propia evolución de la tecnología, que se puede ver profundamente afectada por usos que van en contra del criterio de la opinión pública, de manera tal que la totalidad de la tecnología se vea comprometida, como pasó con la nuclear o la química. De la misma manera, una relajación de la intervención en esta tecnología puede conducir al suyo propio hacia el fin de la humanidad misma.

El peligro más perturbador e inquietante de las LAWS: los drones letales de bolsillo

Los esfuerzos que estamos haciendo en las reuniones de Expertos gubernamentales a Naciones Unidas para conseguir un tratado que prohíba las armas letales autónomas (CCCW) se focalizan en grandes armas para la guerra (macro LAWS como Reaper, Tarannis, Iron Drome, etc.).

Pero es necesario ir más allá y reconocer que el verdadero peligro son las armas pequeñas en manos de particulares, es decir, lo que se puede llegar a microLAWS que puede pasar de estar controlado por las fuerzas militares a estar en manos de cualquier ciudadano para su seguridad privada. Sería otro ejemplo de tecnología letal de doble uso.

Una situación que puede desestabilizarnos y puede hacer cambiar los estándares de seguridad que teníamos hasta ahora. Si los LAWS militares ya son difíciles de atacar cuando lo hacen en forma de enjambres de minidrones, no es complicado de imaginarnos lo que puede pasar si estos pasan a estar bajo el poder de miles de individuos que, en vez de un arma convencional, escogen un dron letal para dotarse de la seguridad que el Estado no puede garantizarles.

El derecho a llevar armas que garantiza la Segunda Enmienda de la Constitución americana permite a sus ciudadanos no sólo tener una pistola o un revólver sino cualquier arma que se considere necesaria para su seguridad como las automáticas o las militares. Llevado al extremo, en la misma línea podrían reivindicar el derecho a la posesión de un robot autónomo letal para protegerlos defensiva u ofensivamente. Es decir, el derecho a poseer un dron de bolsillo autónomo letal con IA.

La regulación a nivel internacional y administrativa global tiene que ser preventiva y tiene que frenar esta situación dado que del uso militar de esta tecnología se pasará a un uso público-civil como estamos viendo con los drones utilizados por la policía o los servicios de protección civil, o incluso privados.

Eso puede desencadenar una situación de inseguridad global debido a la proliferación de este tipo de armas de forma descontrolada en manos privadas y debido a la dificultad para prever la interrelación de sistemas letales con IA entre ellos. Esperamos que el derecho no llegue demasiado tarde para poder frenar esta pandemia 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.