



HACIA UNA ÉTICA DEL ECOSISTEMA HÍBRIDO DEL ESPACIO FÍSICO Y EL CIBERESPACIO

TOWARDS AN ETHICS OF THE HYBRID ECOSYSTEM OF PHYSICAL AND CYBER SPACES

Ángel Gómez de Ágreda
Ministerio de Defensa / Universidad Politécnica de Madrid
agomde@ea.mde.es

Claudio Feijóo
Universidad Politécnica de Madrid
claudio.feijoo@upm.es

Fecha recepción artículo: 03/11/2020 • Fecha aprobación artículo: 09/12/2020

RESUMEN

Las tecnologías emergentes relacionadas con el entorno cognitivo posibilitan la creación de un ecosistema híbrido entre el espacio físico y el digital en el que crecientemente se desarrolla la actividad humana. Su construcción ética y su posterior regulación jurídica han de considerar esta realidad híbrida y el uso dual de las tecnologías que sustentan el mundo digital. La elaboración de ambos códigos -ético y legal- puede extraer conclusiones interesantes de los principios que se están diseñando para los casos más extremos como los sistemas de armas autónomos letales, o la neurociencia o la biotecnología con su capacidad de suplantación de los sentidos o de alteración del sistema cognitivo. De estas experiencias, el artículo concluye que se debe mantener el foco en los valores humanos a preservar en este entorno -en especial, la libertad y la dignidad humanas- y no únicamente en las características tecnológicas del sustrato sobre el que se desarrolla la actividad.

Palabras clave: Inteligencia artificial, Uso dual, Sistemas de armas autónomos, Neurociencia, Biotecnología, Influencia, Ética, Autonomía, Percepción.

Ángel Gómez de Ágreda es coronel del Ejército del Aire, Diplomado de Estado Mayor, Máster en Terrorismo por la UNIR y doctorando en Ingeniería de Organización por la Universidad Politécnica de Madrid. Ha sido jefe de cooperación en el Mando Conjunto de Ciberdefensa y ejerce actualmente como analista geopolítico en el Ministerio de Defensa. Es miembro fundador y director del Área de Seguridad del Observatorio del Impacto Social y Ético de la Inteligencia Artificial (OdiseIA).



Claudio Feijóo es Doctor Ingeniero de Telecomunicaciones y Licenciado en Economía Cuantitativa, Catedrático en la Universidad Politécnica de Madrid y su director para Asia. Es codirector del Campus Hispano-chino en la Universidad de Tongji, en Shanghai, China. Es también fundador y director de la incubadora Xiji en China. Ha trabajado también en la Comisión Europea y el Ministerio de Industria, y es autor de más de 300 artículos, libros y contribuciones en obras científicas.

ABSTRACT

Emerging technologies related to the cognitive environment make it possible to create a hybrid ecosystem between the physical and digital spaces in which human activity increasingly takes place. Their ethical structuring and subsequent legal regulation must be carried out bearing in mind their hybrid nature and the dual use of the technologies that support the digital world. The development of both type of codes -ethical and legal- can draw interesting conclusions from the principles that are being designed for the most extreme cases such as lethal autonomous weapons systems and neuroscience or biotechnology with their capacity to supersede the senses or alter the cognitive system. From these experiences, the paper concludes that the focus must be maintained on the human values to be preserved in this environment -especially human freedom and dignity- and not only on the technological characteristics of the substratum on which the activity is carried out.

Keywords: Artificial intelligence, Dual use, Autonomous weapon systems, Neuroscience, Biotechnology, Influence, Ethics, Autonomy, Perception.

1. INTRODUCCIÓN

Las tres primeras revoluciones industriales y la recién comenzada cuarta –basada en tecnologías tales como la conectividad ubicua, la realidad virtual y aumentada, la inteligencia artificial (IA), el procesado masivo de datos, la interconexión de múltiples dispositivos entre ellos y con nosotros, y las incipientes biotecnologías– no son comparables a pesar de las continuas referencias cruzadas que recoge gran parte de la literatura. El error de pretender asimilarlas en función de aspectos meramente coyunturales –como el impacto en el mercado laboral, en la industria o en la propia productividad– puede resultar fatal a la hora de abordar el tratamiento de un modelo futuro de convivencia social en un marco donde las fronteras de lo físico y de lo virtual comienzan a borrarse.

La primera revolución Industrial que tuvo lugar a caballo de los siglos XVIII y XIX y los numerosos descubrimientos que la acompañaron alteraron únicamente las herramientas con las que los humanos interactuamos con el entorno. Nuestras capacidades para movernos o manipular el mundo físico se vieron incrementadas en varios órdenes de magnitud. Sin embargo, los cambios se produjeron solamente en los instrumentos sin afectar a la naturaleza de los sujetos ni a la de los objetos.¹ Prácticamente lo mismo se puede decir de la segunda revolución industrial de finales del siglo XIX y comienzos del XX, una aceleración de las tendencias de la primera revolución junto con nuevas fuentes de energía, materiales y modos

¹ Si bien es cierto que, como recoge (Aguado, 2020), “las herramientas son solo huellas de la tecnología y ésta no se limita a un modo de hacer, sino más bien a un modo de hacer en el mundo” (Munford, 1971). El cambio se refiere al “guion” que siguen los usuarios, no a alteraciones en el “escenario” sobre el que se desarrolla la acción.



de comunicación. Al igual que en el caso de la primera, el trabajo se modificó, las empresas se hicieron verdaderamente globales por vez primera y con ellas, los modos de consumo.

La tercera revolución, que podríamos llamar digital, comenzó en una línea similar. Se trataba de utilizar la tecnología para dar un paso más en la evolución que se había seguido hasta el momento. El correo pasaba a ser electrónico, los procesadores de texto sustituían a las máquinas de escribir, y el software se hizo ubicuo. Sin embargo, el proceso de digitalización ha supuesto mucho más que convertir a formato PDF los viejos manuscritos (Vial, 2019). La revolución digital ha terminado creando un nuevo ecosistema –el ciberespacio– al cual se está trasladando, al menos en parte, la convivencia humana. Las redes digitales no dejan de ser una poderosa herramienta, pero su verdadero valor está en su papel de nuevo entorno social (Gómez-De-Ágreda, 2012). Este es precisamente el portal que nos lleva a esta cuarta revolución. Ya no es simplemente industrial, es profundamente social y, por tanto, afecta a las relaciones que gobiernan la sociedad y a la ética y, de aquí, a las leyes y regulaciones que la sostienen.

De hecho, como entorno social en el que se produce la interacción de miles de millones de personas, el ciberespacio requiere la elaboración de unos criterios de convivencia éticos y morales, así como de unas reglas jurídicas, que idealmente debieran ser universalmente aceptadas, para ordenar las relaciones. Tanto unos como otras tendrán la complicación añadida de tener que compatibilizar las características de los entornos físico y digital que empleamos de forma simultánea. El ciberespacio altera los principios, derechos y deberes tradicionales, y requiere de la adopción de otros adaptados a su idiosincrasia. La ética y el derecho, aplicables solamente a los seres humanos, deberán contemplar la dualidad de ecosistemas en que ya habitamos, los nuevos agentes que aparecen en este ciberespacio y evitar principios o reglas incompatibles para ambas esferas.

Precisamente este breve artículo revisa algunos de los principios éticos que nos pueden servir para manejar este doble ecosistema interconectado del ciberespacio y el espacio físico. Para ello, después de esta breve introducción, el siguiente apartado ahonda en las implicaciones éticas de lo que podemos llamar singularidades tecnológicas –por su capacidad de disrupción. A partir de aquí, se utilizan dos casos de ética aplicada a situaciones extremas para extraer algunas lecciones de interés general. Un apartado de conclusiones cierra el artículo.

2. SINGULARIDADES TECNOLÓGICAS Y ÉTICA

La aplicación de soluciones tecnológicas como la biotecnología, la neurociencia o la inteligencia artificial está creando las nuevas interfaces entre el mundo físico y el digital. Es en este lugar híbrido, por lo tanto, donde se produce el conflicto de valores derivado de las diferentes características de ambos entornos.

Sería un error contemplar la inteligencia artificial o cualquiera otra de las “tecnologías singulares” de una forma aislada entre ellas o respecto de la sociedad en la que van a estar inmersas. Las implicaciones y los principios éticos y jurídicos que habrá que desarrollar seguirán afectando a la naturaleza y a la convivencia humanas, que deben permanecer como los objetos de estudio ético.

El desarrollo de soluciones tecnológicas asociadas a disciplinas como la neurociencia, la biotecnología o la inteligencia artificial debería traer asociado un auge en los estudios éticos (Goering & Yuste, 2016) y humanísticos, y resucitar los debates filosóficos sobre la esencia del ser humano en tanto individuo y como miembro de la sociedad (Gómez-de-Ágreda, 2019).



El verdadero carácter disruptivo de estas tecnologías no reside en su poder de actuar sobre el entorno, sino en su capacidad para (re)definirlo. Su esencia está en la alteración de las funciones propias de los sentidos humanos, bien para incrementar su capacidad, bien para redirigirla, y todo el procesado posterior a la adquisición de información por los sentidos. En este aspecto, la tecnología está cambiando definitivamente el modo en el que interactuamos con procesos económicos como el trabajo (Santos, Kissamitaki, & Chiesa, 2020), con la misma realidad y con la forma en la que aprehendemos información sobre esta para definir nuestra visión exterior y nuestra actitud hacia ella.

Como ejemplo concreto, la inteligencia artificial no solamente mejora nuestras capacidades de interacción con el mundo, sino que tiene el potencial de alterar la forma en la que lo percibimos y lo sentimos (Macdonald, 2006). Es la mente, la cognición, lo que conforma las sociedades (Boyer, 2018). Por primera vez en la Historia, el ser humano no solo conforma, sino que “crea” el mundo en el que habita. Un mundo nuevo requerirá de un manual de instrucciones. Elaborar el manual durante el diseño permitiría concebir los planos del mundo que se desea crear. Hacerlo a posteriori supondría limitarse a describir el resultado del proceso creativo, al que habríamos llegado sin tener una idea clara y global de lo que se pretendía hacer.

Como otro ejemplo, nos encontramos todavía en una etapa en la que el medio condiciona el mensaje y la tradición sigue imponiendo el medio a emplear. Se mantienen los formatos visuales clásicos (pantallas) con algunas variantes y evoluciones (resolución, tamaño, carácter táctil, flexibilidad del soporte). Su capacidad de influencia se ve, sin embargo, reducida por la necesidad de adaptación y por el origen extrínseco del mensaje. Un mensaje que hay que decodificar y que, consiguientemente, no se acepta como propio. El abandono de este tipo de soportes para enlazar directamente con los sentidos a través de la realidad virtual ubicua o más directamente todavía vía la interacción con las redes neuronales del cerebro supondrá una interiorización absoluta del mensaje (Hopenhayn, 2020) y un paso más en la conexión entre mundo físico y ciberespacio.

3. LAS LECCIONES DE LOS CASOS EXTREMOS (I): ARMAS AUTÓNOMAS

Para esta fase de transición en la que nos encontramos entre la interpretación de la inteligencia artificial y tecnologías asociadas como herramientas y su concepción como ecosistema en el cual se desarrollará la actividad humana podemos acudir a la consideración de ámbitos que ya han debido tener en cuenta – independientemente de su aplicación real– los principios éticos. Particularmente si los principios éticos son una de las cuestiones clave del desarrollo de tales ámbitos.

Este es justamente el caso del diseño y desarrollo de los “sistemas de armas autónomos letales” (SALAS) – los conocidos popularmente como “robots asesinos” (Stop Killer Robots, 2018)–, que pueden considerarse como un referente extremo para la generación de códigos generales para el resto de los sistemas.

A este respecto, hay que notar que los SALAS pertenecen fundamentalmente todavía al ámbito de las herramientas y son, por lo tanto, más cercanos a los conceptos que manejamos habitualmente, aunque algunas de sus implementaciones forman ya parte tanto del espacio físico como del nuevo ciberespacio. Por otro lado, su carácter potencialmente letal los sitúa como un asunto cuyo abordaje se sugiere urgente e importante simultáneamente (Future of Life Institute, 2016).

A los efectos de este estudio se sugiere una división de los sistemas de inteligencia artificial basada en el sustrato sobre el que actúan. De este modo, podríamos decir que los SALAS en su acepción más tradicional



serían una subcategoría de los sistemas robóticos en general a los que hemos denominado “sistemas de inteligencia artificial dura” (Hard-AI). Se trata de forma casi invariable de máquinas herramientas con efectos equivalentes a los de instrumentos convencionales derivados de las dos primeras revoluciones industriales. Otros sistemas de IA, también con potencial letal, actuarían sobre un sustrato intangible. En un caso se trataría de software (Soft-AI) y en un tercero sobre sistemas biológicos en relación con las percepciones y las redes neuronales animales (Bio-AI).

La elaboración de códigos éticos relacionados con el diseño, desarrollo y empleo de sistemas autónomos dotados de inteligencia artificial en el ámbito de la defensa se ha centrado, hasta ahora, en los sistemas de armas autónomos letales. Es decir, solo contempla *Hard-AI* y, dentro de ella, se limita a aquella susceptible de generar efectos letales. Los trabajos tienen lugar en el seno del Convenio sobre Prohibiciones o Restricciones en el Empleo de Ciertas Armas Convencionales que Pueden Considerarse Excesivamente Nocivas o de Efectos Indiscriminados (CCW)² que elabora un grupo de trabajo que se reúne semestralmente en la sede de Naciones Unidas en Ginebra. Desde el ámbito civil apenas si se mencionan estas aplicaciones, centrando los riesgos en aspectos más generales como los sesgos, la justicia o la equidad (Á. Gómez-de-Ágreda, 2020).

Los principios que deben guiar el diseño, desarrollo y uso de los sistemas de armas autónomos letales, de acuerdo con este comité de Naciones Unidas, deben ser (CCW, 2018):

1. El Derecho Internacional Humanitario sigue siendo de completa aplicación a todos los sistemas de armas, incluyendo el potencial desarrollo y utilización de sistemas de armas autónomos letales.
2. La responsabilidad humana sobre las decisiones relacionadas con el uso de sistemas de armas tiene que mantenerse ya que la responsabilidad jurídica no puede transferirse a las máquinas. Este principio debería ser de aplicación a lo largo del ciclo de vida completo del sistema de armas.
3. Se debe asegurar la responsabilidad jurídica por el desarrollo, despliegue y uso de cualquier sistema de armas emergente en el marco de la CCW de conformidad con el Derecho Internacional aplicable, incluyendo por la operación de dicho sistema de armas dentro de una cadena humana de mando y control responsable.
4. De conformidad con las obligaciones estatales bajo el Derecho Internacional, durante el estudio del desarrollo, adquisición o adopción de una nueva arma, medio o método de guerra se debe determinar si su empleo podría, en algunas o todas las circunstancias, estar prohibido por el Derecho Internacional.
5. Durante el desarrollo o adquisición de nuevos sistemas de armas basados en tecnologías emergentes en el área de las SALAS se deberían considerar las salvaguardas físicas y no físicas (incluyendo la ciberseguridad contra hackeo o suplantación de datos), el riesgo de adquisición por grupos terroristas y el riesgo de proliferación.
6. La evaluación de riesgos y de medidas de mitigación deberían ser parte del ciclo de diseño, desarrollo, pruebas y despliegue de tecnologías emergentes en cualquier sistema de armas.
7. Se debería prestar consideración al uso de tecnologías emergentes en el área de las SALAS en el aseguramiento del cumplimiento de las normas de Derecho Internacional Humanitario y otras obligaciones legales internacionales.

² <https://meetings.unoda.org/meeting/ccw-gge-2020/>



8. En el desarrollo de potenciales medidas normativas, las tecnologías emergentes en el área de los SALAS no deberían ser antropomorfizadas.
9. Las deliberaciones y cualquier medida normativa potencial que se tome en el contexto de los SALAS no deberían evitar el progreso de o el acceso a usos pacíficos de tecnologías autónomas inteligentes.
10. El CCW (...) busca conseguir un equilibrio entre la necesidad militar y las consideraciones humanitarias.

La relación introduce ciertas posibles redundancias y reiteraciones de conceptos, pero recoge de forma inequívoca una serie de preceptos que se perciben como de necesaria aplicación más allá de los sistemas letales o de las armas en general: (1) la retención de la responsabilidad en el elemento humano con independencia del grado de autonomía concedido a las máquinas (Floridi, 2016); (2) la obligatoriedad de supervisión humana en todas las fases de desarrollo y utilización de las tecnologías y no solo en la adopción de la decisión final (Lewis, 2018); (3) la responsabilidad asociada a ese control (en lo que podría entenderse como un principio de responsabilidad compartida en el que cada integrante de la cadena asume la responsabilidad de un uso ilícito del sistema); (4) la adopción del principio de precaución frente a posibles usos ilícitos (especialmente en sistema supervisados) (Amoroso, Sauer, Sharkey, Suchman, & Tamburrini, 2018); (5) la necesidad de adopción de medidas frente a posibles efectos indeseados derivados de fallas de seguridad física, lógica o política del sistema; (6) el riesgo emocional derivado de la antropomorfización de los robots; y (7) la necesidad de minimizar el impacto sobre el desarrollo pacífico y beneficioso de estas mismas tecnologías.

4. LAS LECCIONES DE LOS CASOS EXTREMOS (II): SUPLANTACIÓN DE LOS SENTIDOS Y ALTERACIÓN DEL SISTEMA COGNITIVO

El ciberespacio va mucho más allá de las redes. Estas son solo la base tecnológica sobre la que se construye una parte cada vez más importante del ecosistema cognitivo. Por ellas discurren los datos que nos definen como usuarios y aquellos en los que basamos un creciente porcentaje de la información sobre la que elaboramos nuestro conocimiento del mundo.

Se puede argumentar que, en el entorno digital, los datos son nosotros. Más que una relación de propiedad con respecto a ellos, se debe establecer una relación de identidad con ellos. Somos nuestros datos en la misma medida en que somos nuestro ADN en el mundo físico. El acceso a nuestros datos comporta riesgos asimilables al que tendría el conocimiento de nuestras peculiaridades biológicas, aunque también posibilidades equiparables.

Por otro lado, los datos y la información –o la desinformación– procedente de esas redes configura en gran medida el conjunto de percepciones que tenemos del mundo. El confinamiento que casi todo el planeta ha sufrido en mayor o menor medida con ocasión de la pandemia de COVID-19 ha evidenciado la dependencia informativa y comunicacional que tenemos de los medios digitales. Un porcentaje muy elevado de las percepciones que recibimos proceden de las pantallas.³ Esas percepciones repercuten en la forma en que el receptor entiende la realidad (su “verdad”) a través de los datos que le proporcionan y de las emociones que generan en él.

³ En la franja de edad de los menores de 30 años, los *social media* representaban ya en 2018 el medio más utilizado para informarse con un 36% del total. El estudio de *Pew Research* (Shearer, 2018) se refiere, no obstante, a destinatarios finales de la información. La autora no tiene en cuenta la incidencia de lo digital en la generación de noticias en los demás medios. Cabe inferir que las preferencias de la franja más joven de edad se irán desplazando hacia la derecha según vayan envejeciendo.



Las tecnologías ubicuas combinan tres aspectos determinantes (Aguado, 2020): “la fusión entre representación y comunicación, la vinculación al cuerpo y la confluencia entre ubicuidad e inmediatez”. De este modo, la tecnología facilita una capa de “sobrerrealidad” (Rettberg Jill, 2014) que tiene el potencial de sustituir a los sentidos naturales del mismo modo que ha relativizado la importancia de la memoria o del sentido de la orientación.

Una sobrerrealidad permanente se superpone inicialmente al mundo físico para posteriormente sustituirlo. Altera las percepciones sin necesidad de afectar a la materialidad. Las aplicaciones en publicidad o, incluso, en diseño y urbanismo, resultan evidentes. Pero también lo es, como veremos, su capacidad manipulativa al ofrecer una imagen distorsionada de la realidad subyacente (Bose & Aarabi, 2019).

Las interfaces hombre-máquina (la Bio-AI) pretenden suplir o complementar las características de los sentidos humanos. Resulta evidente que, en muchos casos, este auxilio será beneficioso y fundamentalmente neutral (Mudgal, Sharma, Chaturvedi, & Sharma, 2020). Su aplicación a personas con mermas en sus capacidades sensoriales permitirá una vuelta a la paridad de oportunidades, su uso por parte de personas cuya edad o historial clínico haya degradado su agudeza sensorial habilitará vidas satisfactorias más largas y productivas (Yuste et al., 2017).

Sin embargo, más allá de las desigualdades evidentes que puede generar un acceso censitario a estas tecnologías, es preciso considerar el impacto que una percepción cualitativamente distinta de la realidad tendrá en la construcción de la verdad personal y, por lo tanto, en el ejercicio de la libertad. La inserción a nivel sensorial o cognitivo de simulaciones indistinguibles de la realidad permite modificar la raíz del conocimiento sin afectar a la naturaleza de lo observado (Rashkov, Bobe, Fastovets, & Komarova, 2019).

Tanto el ser humano en su faceta íntima como la sociedad en su conjunto son la consecuencia de las capacidades y limitaciones de los sentidos naturales. El universo humano está diseñado atendiendo a las necesidades y oportunidades que se derivan de estos. Una alteración de sus características debería llevar a un rediseño del entorno basado en las nuevas circunstancias. Si la personalidad se construye en base a la experiencia, una alteración en la capacidad perceptiva de unos sentidos mejorados o modificados debe suponer una base distinta sobre la que construir esa personalidad. La ampliación, por poner un ejemplo simple, de la paleta de colores o del abanico de frecuencias perceptibles por una persona ofrece una imagen diferente de la misma realidad, incluso aunque no se manipulen dichos inputs. Las “imágenes” compuestas generadas por sensores en frecuencias de radar o infrarrojas combinadas con las frecuencias en el espectro visible que se utilizan en sistemas tanto civiles como militares permiten, por ejemplo, un grado de comprensión más profundo de la realidad.

Sobre la base de estos elementos, cada usuario construye su propia verdad. La libertad se ejerce partiendo de esa realidad percibida y los sesgos inducidos que llegan por la misma vía. Por lo tanto, la libertad se encuentra mediatizada grandemente por los datos y la información que procede de las redes.⁴ Esta información –que está, en principio, disponible sin restricciones en las redes– sufre un filtrado interesado por parte de las plataformas –que (Pariser, 2012) denomina “burbuja de filtros”– para encerrar a la audiencia en cámaras de eco que replican y amplifican sus propios prejuicios (Jamieson & Capella, 2010).

El uso ubicuo de interfaces tecnológicas que modifiquen las percepciones o la interpretación que hacemos de ellas supone el mayor grado de manipulación imaginable. En estas circunstancias estaremos generando

.....

⁴ Se trata de prácticas como la ingeniería social o las *fake-news* que llevan a una fractura social como la desinformación o las distintas modalidades de acoso en la redes que han proliferado en los últimos años (Aguado, 2020; von Foerster, 2003).



realidades personalizadas, bien por el propio usuario, bien por las plataformas sobre las que se ubique esa realidad paralela,⁵ para acabar modelando una nueva sociedad.

La defensa de la libertad es la función compuesta de la protección de la veracidad de los datos sobre los que cada cual construye su verdad y de aquellos que introducen sesgos o prejuicios en la evaluación de esta. Por lo tanto, la alteración de las percepciones o la manipulación de las emociones es un arma más que una herramienta, siguiendo la concepción clásica de guerra como una confrontación de voluntades –esto es, de libertades de acción (Von Clausewitz, 2014).

El uso de la inteligencia artificial blanda (Soft-AI) como arma –su “weaponization”, en versión anglosajona– se muestra claramente en la creciente importancia que han adoptado las llamadas “operaciones de influencia”, en las que el campo de batalla es el entorno cognitivo y emocional, y en el que se pretende alterar la voluntad del adversario sin cambiar necesariamente la realidad sobre el terreno (Calvo Albero et al., 2020; Lin, 2018; Pomerleau, 2020).

Aunque en este ámbito no existen intentos de códigos éticos más allá de los genéricos propuestos para la inteligencia artificial, podemos observar como nuevamente la libertad –la autonomía como la propia capacidad de pensar, decidir y actuar independiente sin influencia de otros– es el componente ético que se ve inicialmente afectado y, a través de él, el resto de los aspectos éticos individuales –dignidad humana, privacidad– o sociales –equidad, derecho a una vida digna, justicia (Vesnic-Alujevic, Nascimento, & Pólvora, 2020).

5. CONCLUSIONES

Para comenzar, es necesario considerar el balance entre los beneficios obtenibles por una aplicación temprana de tecnologías emergentes frente a los potenciales perjuicios derivados de un uso no adecuado de las mismas. Es preciso recordar que las soluciones algorítmicas que forman sistemas de inteligencia artificial se encuentran presentes –de forma idéntica en algunos casos– tanto en los avances más beneficiosos para la humanidad como en los sistemas de armas letales, de suplantación de los sentidos o de alteración del sistema cognitivo. Las tecnologías son de uso dual, lo que dificulta su desarrollo pacífico sin habilitar el ilícito o el bélico (Aicardi, 2018; Gómez-de-Ágreda, 2020; Jonathon-Penney, 2018)

A partir de aquí y en general, más que a las similitudes o diferencias tecnológicas, los principios éticos aplicables al nuevo ecosistema híbrido, mezcla de realidad y ciberespacio, tienen que estar basados en los efectos que provocan sobre los seres humanos de forma individual o colectiva. La ética y la moral están vinculadas a la esencia y la actividad humana con independencia del escenario en que se desarrollen. Por consiguiente, los principios deben regir igualmente en el caso de que la IA y tecnologías similares se apliquen al uso de una herramienta en un entorno físico o a una manipulación de un entorno virtual o cognitivo.

⁵ Las plataformas son agrupaciones de hardware, software, servicios, infraestructuras y usuarios (Feijoo, Gómez-Barroso, Aguado, & Ramos, 2012). Su centralidad en el panorama digital comienza en los entornos móviles (Ballon, 2009) sobre los sistemas operativos de Android e IOS para trasladarse después también al resto del entorno en lo que (Castellet & Feijóo, 2013) denominan “paradigma del software”.



Desde ese punto de vista, como se ha mencionado, el principal derecho afectado es la libertad humana –la autonomía que constituye su base–, se vea ésta condicionada por una amenaza cinética (física) o por una alteración de las percepciones a nivel sensorial o cognitivo. La dignidad no tiene ninguna relación con la letalidad, sino con la capacidad para ejercer la autonomía de criterio en tanto se mantiene la condición humana.

Lo que resulta evidente cuando se utiliza el contraste que proporciona la lente de la letalidad no es menos cierto cuando desaparece la lente.⁶ El foco, como recogen las conclusiones del Grupo de Expertos Gubernamentales de Naciones Unidas, no debe situarse en la letalidad de los sistemas, sino en la autonomía que pueden proporcionar las tecnologías a las máquinas en detrimento de la de los humanos (Group of Governmental Experts, 2018) o en la capacidad de los algoritmos para hacer prevalecer la voluntad de unos humanos sobre otros, como en el caso de las suplantaciones sensoriales y alteraciones cognitivas. Es para esta tarea para la que puede resultar útil explorar la aplicabilidad de los principios éticos de los sistemas de armas autónomos letales al resto de los sistemas dotados de inteligencia artificial.

La inteligencia artificial y las tecnologías afines son más que unas herramientas con las que se puede actuar sobre el mundo natural. Su desarrollo reciente coincide con el de un entorno de movilidad en el que los dispositivos permanecen fusionados al usuario y, en muchas ocasiones, actúan como mediadores entre la realidad y sus percepciones (Feijoo, Pascu, Misuraca, & Lusoli, 2009). Los algoritmos se encuentran, en ese caso, en disposición de alterar dichas percepciones y condicionar, por lo tanto, la libertad del usuario.

Es esta interacción entre las diferentes tecnologías y disciplinas singulares como la biotecnología o la neurociencia la que crea un ecosistema cognitivo paralelo. Como todo ecosistema susceptible de habitación humana, éste requiere de una regulación, y unos principios éticos y morales de comportamiento y convivencia. Sus características, como se ha visto, son diferentes de las del mundo físico natural. Sin embargo, la actividad humana no se va a desarrollar de forma exclusiva en el entorno digital por lo que los principios y normas que se desarrollen tendrán que ser compatibles con ambos mundos para seguir manteniendo al ser humano en el centro de su actividad.

“Cuando los ojos ven lo que nunca vieron, el corazón siente lo que nunca sintió”.

Baltasar Gracián.

.....
⁶ Resulta patente que videos como https://www.youtube.com/watch?v=HipTO_7mUOw, en los que se muestran los posibles efectos de las armas autónomas tienen un mayor impacto emocional que se debería extrapolar a actuaciones menos llamativas.



6. REFERENCIAS BIBLIOGRÁFICAS

- Aguado, J. M. (2020). *Mediaciones ubicuas*. (GEDISA, Ed.).
- Aicardi, C. et al. (2018). Opinion on ' Responsible Dual Use .' *Human Brain Project*, 1–21.
- Amoroso, D., Sauer, F., Sharkey, N., Suchman, L., & Tamburrini, G. (2018). *Autonomy in Weapon Systems The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy* (Vol. 49).
- Ballon, P. (2009). The platformisation of the European mobile industry. *Communications & Strategies*, (75), 15–34.
- Bose, A. J., & Aarabi, P. (2019). Virtual Fakes: DeepFakes for Virtual Reality. In *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)* (p. 1).
- Boyer, P. (2018). *Minds Make Societies: How Cognition Explains the World Humans Create*. Yale University Press.
- Calvo Albero, J. L., Andrés Menárguez, D. F., Peirano, M., Moret Millás, V., Peco Yeste, M., & Donoso Rodríguez, D. (2020). *El entorno cognitivo: implicaciones para las operaciones militares*.
- Castellet, A., & Feijóo, C. (2013). Los actores en el ecosistema móvil. In J. M. Aguado, C. Feijóo, & I. J. Martínez (Eds.), *La comunicación móvil. Hacia un nuevo ecosistema digital* (pp. 27–56). Barcelona: Gedisa.
- CCW. (2018). Report of the 2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems.
- Feijoo, C., Gómez-Barroso, J. L., Aguado, J. M., & Ramos, S. (2012). Mobile gaming: Industry challenges and policy implications. *Telecommunications Policy*, 36, 212–221. <https://doi.org/10.1016/j.telpol.2011.12.004>
- Feijoo, C., Pascu, C., Misuraca, G., & Lusoli, W. (2009). The Next Paradigm Shift in the Mobile Ecosystem: Mobile Social Computing and the Increasing Relevance of Users. *Communications & Strategies*, 75(3rd quarter), 57–77.
- Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083). <https://doi.org/10.1098/rsta.2016.0112>
- Future of Life Institute. (2016). Lethal Autonomous Weapons Systems.
- Goering, S., & Yuste, R. (2016, November). On the Necessity of Ethical Guidelines for Novel Neurotechnologies. *Cell*. Cell Press. <https://doi.org/10.1016/j.cell.2016.10.029>
- Gómez de Ágreda, Á. (2020). Ethics of autonomous weapons systems and its applicability to any AI systems. *Telecommunications Policy*, 44(6). <https://doi.org/10.1016/j.telpol.2020.101953>
- Gómez De Ágreda, Á. (2012). El ciberespacio como entorno social y de conflicto. *Instituto Español de Estudios Estratégicos*.
- Gómez de Ágreda, Ángel. (2019). *Mundo Orwell. Manual de supervivencia para un mundo hiperconectado* (1st ed.). Barcelona: Ariel.
- Group of Governmental Experts. (2018). Emerging Commonalities, Conclusions and Recommendations CCW 2018, (August), 1–5.
- Hopenhayn, D. (2020). Rafael Yuste, neurobiólogo: "Creo que vamos en camino hacia un nuevo Renacimiento." *La Tercera*.
- Jamieson, K. H., & Capella, J. M. (2010). *Echo Chamber*. Oxford University Press.
- Jonathon-Penney, J. P. A. (2018). Advancing Human-Rights-By-Design In The Dual-Use Technology Industry. *Journal of International Affairs*, (April).



- Lewis, L. (2018). Redefining Human Control: Lessons from the Battlefield for Autonomous Weapons. *Center for Autonomy and AI*, 1–23.
- Lin, H. (2018). Developing Responses to Cyber-Enabled Information Warfare and Influence Operations. *Lawfare*.
- Macdonald, S. (2006). *Propaganda and Information Warfare in the Twenty-First Century*. *Propaganda and Information Warfare in the Twenty-First Century*. <https://doi.org/10.4324/9780203967393>
- Mudgal, S. K., Sharma, S. K., Chaturvedi, J., & Sharma, A. (2020). Brain computer interface advancement in neurosciences: Applications and issues. *Interdisciplinary Neurosurgery: Advanced Techniques and Case Management*, 20. <https://doi.org/10.1016/j.inat.2020.100694>
- Munford, L. (1971). *Técnica y civilización*. Alianza Editorial.
- Pariser, E. (2012). *The filter bubble*. Penguin Books.
- Pomerleau, M. (2020). Why the Pentagon needs to fully embrace influence operations. *C4ISRNET*.
- Rashkov, G., Bobe, A., Fastovets, D., & Komarova, M. (2019). Natural image reconstruction from brain waves: a novel visual BCI system with native feedback. *BioRxiv*, 787101. <https://doi.org/10.1101/787101>
- Rettberg Jill, W. (2014). *Seeing Ourselves Through Technology How We Use Selfies, Blogs and Wearable Devices to See and Shape Ourselves Please respect intellectual property rights*. Palgrave MacMillan. <https://doi.org/10.1057/9781137476661>
- Robots, S. K. (2018). CAMPAÑA PARA DETENER ROBOTS ASESINOS Retener el control humano sobre los sistemas de armas, 1–5.
- Santos, S., Kissamitaki, M., & Chiesa, M. (2020). Should humans work? *Telecommunications Policy*, 44(6), 101910. <https://doi.org/10.1016/j.telpol.2020.101910>
- Shearer, E. (2018). *Social media outpaces print newspapers in the U.S. as news source*.
- Vesnic-Alujevic, L., Nascimento, S., & Pólvara, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, 44(6). <https://doi.org/10.1016/j.telpol.2020.101961>
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *JOURNAL OF STRATEGIC INFORMATION SYSTEMS*.
- Von Clausewitz, K. (2014). *De la guerra*. La esfera de los libros.
- von Foerster, H. (2003). *Understanding Understanding: Essays on Cybernetics and Cognition*. New York, NY: Springer New York. <https://doi.org/10.1007/b97451>
- Yuste, R., Goering, S., Agüeray Arcas, B., Bi, G., Carmena, J. M., Carter, A., ... Wolpaw, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature*, 551(7679), 159–163. <https://doi.org/10.1038/551159a>

