

## ÉTICA EN EL DISEÑO PARA EL DESARROLLO DE UNA INTELIGENCIA ARTIFICIAL, ROBÓTICA Y BIG DATA CONFIABLES Y SU UTILIDAD DESDE EL DERECHO

Lorenzo Cotino Hueso\*

### Resumen

El estudio lleva a cabo una aproximación a la ética de la inteligencia artificial. En primer término, se hace una recopilación de la proclamación de la misma y de su necesidad en los diferentes documentos de referencia internacional, aunque el análisis se centra especialmente en los múltiples focos de atención en la Unión Europea. Es precisamente en la UE donde se apuesta por una ética confiable de la IA en el diseño y *made in Europe*, para posicionarse frente a Estados Unidos y especialmente China, que no conceden importancia a la cuestión. Primero, se describe el contenido de la ética de la IA y sus principios esenciales a partir de la dignidad y los derechos; segundo, se exponen los cinco principios básicos destilados de las declaraciones internacionales. Tercero, se recogen otros principios básicos y que derivan de exigencias de empatía con los humanos. Desde una perspectiva algo escéptica se argumentan las potenciales utilidades de la ética de la IA para el derecho: se considera que resulta un instrumento especialmente preventivo y que bien puede desarrollarse una gobernanza ética de la IA, siguiendo el ejemplo de las políticas y marcos de la ética pública y la integridad institucional. Se afirman las fases a seguir al respecto. Se detalla la oportunidad y el contenido básico de códigos de conducta y comités y otros sistemas de control. Se concluye con una llamada al diseño de algoritmos como guardianes del cumplimiento normativo y de la ética de la IA.

Palabras clave: inteligencia artificial; robótica; derechos fundamentales; ética en el diseño; Unión Europea.

### ETHICS IN DESIGN FOR THE DEVELOPMENT OF AN ARTIFICIAL INTELLIGENCE, TRUSTWORTHY ROBOTICS AND BIG DATA AND THEIR UTILITY FOR THE LAW

#### Abstract

*The study deals with the ethics of artificial intelligence (AI). Firstly, we explain the proclamation of the ethics and its necessity in the different international reference documents, although the analysis is focused on the actions carried out in the European Union. It is precisely in the EU where there is a special commitment to developing an ethics for a trustworthy AI in design and made in Europe, to position itself in front of United States and especially China, two countries that don't pay much attention to the issue. Firstly, we describe the content of ethics of AI and its essential principles from the point of view of the dignity and rights; second, we describe the main five principles contained in the international declarations. Third, we include other basic principles which arise from the demands of empathy with humans. From a somewhat sceptical perspective, we argue the potential utilities of ethics of AI for the law: it is considered to be an especially preventive instrument and that an ethical governance of AI can be developed, following the examples of policies and frameworks on public ethics and institutional integrity. The phases to be followed are described in this regard. We explain in detail the opportunity and the basic content of codes of conduct and committees and other control systems. Finally, we make an appeal for the design of algorithms that serve as guardians of regulatory compliance and the ethics of AI.*

*Key words: artificial intelligence; robotics; fundamental rights; ethics in design; European Union.*

\* Lorenzo Cotino Hueso, catedrático de derecho constitucional en la Universidad de Valencia y coordinador de Derecho TICs (Red de especialistas en Derecho de las Tecnologías de la Información y Comunicación). Facultad de Derecho, av. dels Tarongers, s/n. 46022 València. [cotino@uv.es](mailto:cotino@uv.es), [@cotino](https://twitter.com/cotino).

Artículo recibido el 20.03.2019. Evaluación ciega: 10.04.2019 y 12.04.2019. Fecha de aceptación de la versión final: 15.04.2019.

**Citación recomendada:** Cotino Hueso, Lorenzo. (2019). Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el derecho. *Revista Catalana de Dret Públic*, (58), 29-48. <https://doi.org/10.2436/rcdp.i58.2019.3303>

## Sumario

- 1 Introducción. La ética de la IA y su recepción y proclamación internacional
    - 1.1 Una aproximación a la ética de la IA y el internet de las cosas robóticas
    - 1.2 La recepción y proclamación internacional de la necesidad de una ética de la IA
  - 2 La ética ha pasado a ser esencial en la marca *AI made in Europe*
    - 2.1 Los diversos focos de actuación de la UE hacia una ética de la IA
    - 2.2 Una ética de la IA confiable y en el diseño *made in Europe* para competir en el mundo
  - 3 El contenido y principios de la ética de la IA
    - 3.1 Una ética de la IA que sitúa en el centro la dignidad y los derechos
    - 3.2 Los cinco principios básicos a los que se reconducen los principios éticos reconocidos internacionalmente
    - 3.3 Otros principios básicos de la ética de la IA y la empatía con los humanos como principio ético
  - 4 Las potenciales utilidades de la ética de la IA pese al natural escepticismo del jurista y el desarrollo de códigos de conducta y comités
    - 4.1 Hacia un marco, infraestructura, sistema o gobernanza ética de la IA a partir del ejemplo de la ética pública y la integridad institucional
    - 4.2 El desarrollo de códigos de conducta profesionales, empresariales, de investigación y usuarios y comités y sistemas de garantía de los mismos
  - 5 Algoritmos “guardianes” al servicio del cumplimiento normativo y de la ética de la IA
  - 6 Y para concluir: pasar “de las musas al teatro” en la ética y el derecho de la IA
- Bibliografía

## 1 Introducción. La ética de la IA y su recepción y proclamación internacional<sup>1</sup>

### 1.1 Una aproximación a la ética de la IA y el internet de las cosas robóticas

En estas páginas el objeto de atención es la llamada ética de la inteligencia artificial (IA, AI en inglés). Como se va a describir, se trata de una ética que se decanta especialmente a través de declaraciones internacionales, así como una importante acción en la UE, que se va traduciendo en una serie de principios. Quien suscribe y posiblemente el lector jurista no podrá evitar cierto escepticismo y visión crítica de la verdadera importancia y alcance de esta ética de la IA y de sus principios. No obstante, sí que se argumenta el papel y la funcionalidad que puede alcanzar esta ética de la IA para el derecho y para el desarrollo futuro de políticas y regulaciones.

Entre diversos de los estudios de este número de la *Revista* sobre el tema y de modo completo en un reciente estudio (Cotino, 2019: 2-8), el lector encontrará una amplia exposición de las nociones de *algoritmos*, *IA*, *sistemas de autoaprendizaje automático (machine learning)*, *deep learning* y *redes neuronales*, la *robótica inteligente*, así como el *big data* como alimento de la IA. Entre otras, la Universidad de Stanford brinda una excelente aproximación a los cien años de historia y futuro de la IA (Stanford University, 2016). En el presente estudio IA ha referencia en general a todas estas tecnologías disruptivas y convergentes, hacia el llamado *internet de las cosas robóticas* (Simoens *et al.*, 2018). Obviamente hay no pocos matices entre unas y otras tecnologías, si bien, al menos respecto de las principales preocupaciones éticas y jurídicas, considero que bien pueden abordarse de modo conjunto. Asimismo, en el referido estudio (Cotino, 2019: 9-15) he analizado y categorizado los diferentes riesgos, peligros e impactos por estas nuevas tecnologías, que alcanzan riesgos existenciales, errores y discriminaciones masivas y estupidez artificial. De igual modo, no hay que obviar los temores ante la “singularidad tecnológica”, una “IA fuerte” (*strong AI*), que puede dar un “giro traicionero”, una “explosión” (Kurzweil, Bolstrom, Musk, Hawking, etc.). Con estas expresiones se hace referencia al momento en el que la IA pasaría a ser lo suficientemente fuerte para atacar a los humanos e incluso considerarles una amenaza y se volvería en contra evitando los sistemas de control humano establecidos. Se trata de algo que ya ha sido objeto de atención especial en la literatura y el cine. También en aquel estudio (*id.*: 9-11) refiero la literatura de referencia respecto a los peligros de la “algorocracia” (Danaher), dictadura o fundamentalismo de los datos (Mayer-Schönberger), “dataísmo” (Harari) o la “psicopolítica” y el “enjambre” (Han). A estos autores directamente o a las estupendas aproximaciones a la materia desde el punto de vista de la ética me remito (*Ética de datos, sociedad y ciudadanía*, 2017).

En perspectiva ética no son pocas las cuestiones y consideraciones que se consideran clave (European Group on Ethics in Science and New Technologies, EGE, 2018: 8-13). Así, las relativas al control humano a la autonomía artificial, la interacción incluso emocional de seres humanos y robots, la responsabilidad, el rediseño institucional (gobernanza, regulación, diseño, desarrollo, inspección, monitoreo, pruebas y certificación), la zona gris entre el impulso o la sutil manipulación (*nudging*) hasta la manipulación, la explicabilidad y transparencia de la IA, los límites a los sistemas de puntuación social (*social scoring*). También se consideran cuestiones críticas (AI-HLEG, 2018: 10 y sig.) el perfilado humano sin consentimiento, la vigilancia masiva o los sistemas de IA encubiertos amén del uso de sistemas letales de armas autónomas. Incluso como preocupaciones a largo plazo (AI-HLEG, 2018: 12-13) se apunta la existencia de conciencias o “agentes morales artificiales” y la posibilidad de responsabilidad y derechos a los no humanos. También debe servir para responder a dilemas y ámbitos más concretos. Por ejemplo, decisiones de los vehículos autónomos ante accidentes (EGE: 12).

Dado que estas páginas se dedican a la ética de la IA, para una aproximación a los enfoques, modelos y principios de regulación de la IA el lector puede acudir a Cotino (2019: 18-30). Y ya de modo más particular en cuanto a los derechos fundamentales, que están en la base de la ética de la IA, destacan diversos estudios en reciente monográfico (Boix y Cotino, 2019), así como tuve ocasión de una primera aproximación a los derechos fundamentales en Cotino (2017).

---

<sup>1</sup> El presente estudio se realiza en el marco de los proyectos “La regulación de la transformación digital y la economía colaborativa” (PROMETEO/2017/064 Generalitat Valenciana y MINECO Retos). Asimismo, es resultado de investigación del proyecto “Derecho y *Big data*”, Grupo de Investigación en Derecho Público y TIC como investigador de la Universidad Católica de Colombia. De igual modo, el estudio se realiza en el marco de ayuda de la Generalitat Valenciana para la estancia de personal investigador en empresa (AEST/2019/013).

## 1.2 La recepción y proclamación internacional de la necesidad de una ética de la IA

Son ya diversos los hitos de la llamada ética de la IA. El Instituto de Ingenieros Eléctricos y Electrónicos (IEEE), la organización técnica profesional más grande del mundo, en diciembre de 2016 lanzó la primera versión del documento de políticas del IEEE sobre el diseño éticamente alineado, actualizadas en diciembre de 2017 (IEEE, 2016-2017). Más de 250 líderes de opinión mundiales fueron consultados en las primeras ediciones. Desde la International Telecommunication Union (ITU), desde 2017 se lideran las cumbres anuales *AI for Good*, que culminan con documentos y declaraciones relevantes (ITU, 2017). Asimismo, la ACM (Association for Computing Machinery) también desde 2017 realiza conferencias sobre IA, ética y sociedad.<sup>2</sup> De particular relevancia, la Alianza sobre la IA (Partnership on AI, 2018), que en septiembre de 2018 proclamó siete principios, reúne a más de 80 entidades públicas y privadas entre las que se encuentran las grandes compañías tecnológicas. También, especialmente dedicada a la investigación de la IA para el bien común, destaca Open AI<sup>3</sup> con empresas e investigadores muy destacados. Una actividad significativa desde 2015 la ha realizado el Future of Life Institute,<sup>4</sup> que hizo una movilización para evitar una escalada de la IA en el armamento. La Conferencia de enero de 2017 (Asilomar, 2017) concluyó afirmando 23 principios sobre cuestiones de investigación (5), ética y valores (6-18) y problemas a largo plazo. La Universidad de Montreal también lideró un proceso participativo que llevó al primer borrador de la Declaración para un desarrollo responsable de la inteligencia artificial (Montreal Declaration, 2017), que proclamó diez principios. Desde Naciones Unidas cabe destacar el centro UNICRI de Inteligencia Artificial y Robótica en La Haya,<sup>5</sup> del que han surgido diversos estudios, y la UNESCO ha adoptado un papel activo,<sup>6</sup> al tiempo de cerrar estas páginas aboga por un enfoque humanístico y una gobernanza de la IA que respete los derechos.<sup>7</sup>

En el ámbito internacional de la protección de datos destaca inicialmente el International Working Group on Data Protection in Telecommunications. En este foro se adoptó en 2014 el documento de trabajo *Working Paper on Big Data and Privacy. Privacy principles under pressure in the age of Big Data analytics* (International Working Group on Data Protection in Telecommunications, 2014). Destaca en cualquier caso la actividad de la Conferencia Internacional de Autoridades de Protección de Datos (ICDPPC). Las amenazas del *big data* fueron objeto de una breve resolución de dos páginas en la 36 Conferencia, celebrada en 2014 en Mauricio (ICDPPC, 2014). Y de modo concreto, para la 38 Conferencia Internacional de Marrakech se elaboró un documento de trabajo, Room Document, sobre el tema (ICDPPC, 2016). Sobre esta base, destaca finalmente la Declaración sobre ética y protección de datos en el sector de la inteligencia artificial en el marco de la 40 Conferencia Internacional de Bruselas, de octubre de 2018. En ella se afirma la necesidad del diseño bajo el principio de justicia, la necesidad del uso justo, la vigilancia continua y la rendición de cuentas, la transparencia e inteligibilidad, la privacidad por defecto y en el diseño, el empoderamiento de cada individuo y reconocimiento efectivo de sus derechos, así como la lucha frente a los sesgos ilegales o las discriminaciones. De cara al futuro se “pide que se establezcan principios de gobernanza común sobre inteligencia artificial” (ICDPPC, 2018:5).

Entre la actividad de referencia de algunos países, en EE. UU. hay que destacar sobremanera el profundo análisis de la Comisión Federal del Comercio (FTC, 2016) con especial énfasis en los peligros de discriminación. Asimismo en mayo de 2016 el Comisionado de Información Australiano (Australian Information Commissioner, 2016) afirmó once principios de privacidad australianos con relación al *big data* (*Australian Privacy Principles*, APPs). Son destacables en el Reino Unido los trabajos parlamentarios. Inicialmente la Cámara de los Comunes (2016) en su Comité de Ciencia y Tecnología y, en particular, el 29 de junio de 2017 la Cámara de los Lores eligió un comité de inteligencia “para considerar las implicaciones económicas, éticas y sociales de los avances en inteligencia artificial”. De dicho comité resultó en abril de 2018 el amplio informe de casi doscientas páginas (Cámara de los Lores, 2018). En el ámbito de la ética incluye cinco principios para un código de IA (parágrafo 417). Singapur ha hecho de la IA toda la

---

2 [www.aies-conference.com](http://www.aies-conference.com)

3 <https://openai.com>

4 <https://futureoflife.org/team>

5 [http://www.unicri.it/in\\_focus/on/UNICRI\\_Centre\\_Artificial\\_Robotics](http://www.unicri.it/in_focus/on/UNICRI_Centre_Artificial_Robotics)

6 <https://en.unesco.org/artificial-intelligence>

7 Conferencia Mundial UNESCO “Principios para la Inteligencia Artificial: ¿hacia un enfoque humanístico?”, celebrada el 4 de marzo de 2019 en la sede de la UNESCO, <https://en.unesco.org/artificial-intelligence/principles-ai-towards-humanistic-approach>

estrategia de país<sup>8</sup> y en enero de 2019 lanza la necesidad de un modelo de marco de gobernanza de IA vinculada a principios. No obstante, no se centra tanto en la ética y los derechos —como la UE— sino más específicamente en la confianza del consumidor y el cumplimiento de estándares de tratamiento de datos (PDPC, 2019: 5 y sig.).

## 2 La ética ha pasado a ser esencial en la marca *AI made in Europe*

### 2.1 Los diversos focos de actuación de la UE hacia una ética de la IA

En Europa, la actividad sobre IA del Consejo de Europa<sup>9</sup> ha sido inicialmente más discreta que la de la UE. No obstante, cabe destacar el muy sugerente informe del parlamentario Jean-Yves Le Déaut (Le Déaut, 2017). A partir de este cabe seguir la no muy ambiciosa Recomendación 2102 (2017)<sup>1</sup>, de 28 de abril, de la Asamblea Parlamentaria (2017b), y su respuesta de 17 de octubre por el Consejo de Ministros (Committee of Ministers, 17 de octubre de 2017). En el momento de cerrar estas páginas se ha dado una breve Declaración del Comité de Ministros sobre las capacidades de manipulación de los procesos algorítmicos (Consejo de Europa, 2019a). Destaca, no obstante, la adopción el 3-4 de diciembre de 2018 de una amplia Carta ética europea sobre el uso de la IA en los sistemas judiciales y su entorno (CEPEJ, 2018). También hay que seguir las conclusiones finales de la Conferencia de Helsinki el 26-27 de febrero de 2019 (Consejo de Europa, 2019b).

En el ámbito de la UE son múltiples y muy activos los diversos focos de actuación a partir de diversos grupos técnicos y de expertos que apoyan o asesoran más o menos directamente a la Comisión o al Parlamento Europeo; lo que va unido a actuaciones más formalizadas a través de importantes resoluciones del Parlamento Europeo o comunicaciones de la Comisión. Cuando menos desde el punto de vista declarativo, esta actividad ha situado a la UE en la vanguardia mundial de la atención ética y normativa en la materia.

Ya el 16 de septiembre de 2014, las autoridades de protección de datos de los Estados de la UE en el antiguo Grupo de Trabajo del Artículo 29 (2014) adoptaron la Declaración sobre el impacto del desarrollo de *big data* en la protección de las personas con respecto al procesamiento de sus datos personales en la UE. A esta actuación siguió el Dictamen 4/2015 del Supervisor Europeo de Protección de Datos (2015a), en el que recordó que “no basta con respetar la ley sino que es preciso tener en cuenta la dimensión ética del tratamiento de datos”. También se dedicó a la cuestión el Dictamen 7/2015 y el Dictamen 8/2016 (Supervisor, 2015b; 2016) y la decisión de constituir en diciembre de 2015 un grupo consultivo sobre ética.<sup>10</sup> El Parlamento Europeo cuenta con el Panel for the Future of Science and Technology, un grupo de expertos del organismo llamado Science and Technology Options Assessment (STOA), que proporciona estudios e información a los parlamentarios y organiza eventos.<sup>11</sup> El 28 de junio de 2016 lanzaron los amplios documentos de referencia (STOA, 2016a; 2016b) sobre aspectos e impactos éticos, legales y políticos de la IA y robótica en diversas áreas (discapacidad, salud, agricultura, fabricación, energía, logística y transporte, seguridad y protección). De estos trabajos derivaría la importante Resolución del Parlamento Europeo (2017a), de 16 de febrero de 2017, sobre normas de derecho civil sobre robótica y un importante anexo de principios y códigos de conducta. Y al mes siguiente se adoptó la no menos importante Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales (Parlamento Europeo, 2017b).

La Comisión Europea cuenta con la asesoría del Grupo Europeo de Ética en Ciencia y Nuevas Tecnologías (EGE).<sup>12</sup> El 9 de marzo de 2018 el EGE (2018) publicó su *Declaración sobre inteligencia artificial, robótica y sistemas autónomos*, donde se realizan una serie de reflexiones morales, preguntas y consideraciones clave para afirmar unos principios éticos y requisitos democráticos. Y la Comisión Europea en abril de 2018 publicó su comunicación *Inteligencia artificial para Europa*, que afirma una hoja de ruta para elaborar

---

<sup>8</sup> [www.aisingapore.org](http://www.aisingapore.org)

<sup>9</sup> La información al respecto se concentra en <https://www.coe.int/en/web/artificial-intelligence>

<sup>10</sup> [https://edps.europa.eu/data-protection/our-work/publications/ethical-framework/ethics-advisory-group\\_en](https://edps.europa.eu/data-protection/our-work/publications/ethical-framework/ethics-advisory-group_en)

<sup>11</sup> <http://www.europarl.europa.eu/stoa/en/home/highlights>

<sup>12</sup> European Commission's European Group on Ethics in Science and New Technologies. [https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/scientific-support-eu-policies/european-group-ethics-science-and-new-technologies-ege\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/scientific-support-eu-policies/european-group-ethics-science-and-new-technologies-ege_en)

directrices *éticas* en relación con la IA con un proyecto para fines de 2018 teniendo en cuenta la Carta de los Derechos Fundamentales y contando con todas las partes interesadas pertinentes (Comisión Europea 2018a: 17). Precisamente tales directrices habían de partir del trabajo del EGE, a su vez inspirado en las iniciativas básicas en la materia. También de la Comisión Europea (2018b) destaca el Plan coordinado sobre la inteligencia artificial de 7 de diciembre de 2018, así como su importante anexo (Comisión Europea, 2018c) con el especial énfasis puesto en una IA *made in Europe* y confiable. También en abril de 2018 la Comisión Europea anunció la creación de una alianza europea de la IA<sup>13</sup> como “foro para intercambiar buenas prácticas y para estimular la inversión privada y las actividades relacionadas con el desarrollo de la IA” (Comisión Europea 2018a: 20-21).

Asimismo, también en abril de 2018, la Comisión nombró a 52 miembros del Grupo de Expertos de Alto Nivel en IA (AI-HLEG, 2018).<sup>14</sup> Y el 18 de diciembre de 2018 este grupo de expertos publicó el Proyecto de directrices de ética para una IA confiable (AI-HLEG, 2018). Se trata del inicio de un proceso de consulta —a través de la Alianza de la IA— que acabó en febrero de 2019 con más de 500 comentarios. En abril de 2019 el grupo de expertos difunde la versión revisada, de la cual puede afirmarse quizá una mayor relevancia del papel del derecho (“Lawful AI” pasa a ser primer componente: AI-HLEG, 2019: 2, 5 y sig., 9). El Proyecto de directrices parte de los derechos fundamentales hacia la búsqueda de una inteligencia confiable. Además de afirmar los principios éticos resulta especialmente destacable la visión más práctica que afirma los “requisitos de la IA de confianza” y los “métodos técnicos y no técnicos para lograr una IA confiable” (AI-HLEG, 2018: 14-24). Precisamente en la versión definitiva el *check list* para comprobar una IA confiable que incluye se ha duplicado respecto del inicial, incluyendo 147 preguntas (2019: 28 y sig.) frente a las 82 iniciales.

Paralela e interrelacionada es la actividad de AI4People,<sup>15</sup> en el marco de la plataforma Atomium-EISMD, inicialmente desde el Parlamento Europeo. Sus integrantes, en noviembre de 2017, lanzaron una hoja de ruta para tres años a fin de establecer las directrices éticas para diseñar políticas y las prácticas fundamentales sobre las cuales construir una “buena sociedad de AI”. Un comité científico de doce expertos presididos por Floridi inició sus trabajos en febrero de 2018. Estos trabajos científicos de la AI4People fueron difundidos en noviembre de 2018 (Floridi *et al.*, 2018) y sus resultados se pueden apreciar con claridad en el primer borrador de directrices del grupo de expertos —en el que está también Floridi— de diciembre de 2018 (AI-HLEG, 2018). En particular se aprecia la influencia en la síntesis de los principios éticos en cinco —que se analiza más tarde— así como en una visión más práctica que se traduce en 20 recomendaciones concretas (para evaluar, desarrollar, incentivar y respaldar una buena IA).

Por su parte, la Agencia de Derechos Fundamentales (FRA, 2018) ya ha lanzado un estudio específico sobre sesgos y discriminación algorítmica y está en marcha una evaluación de los desafíos actuales que enfrentan los productores y los usuarios de nuevas tecnologías con respecto al cumplimiento de los derechos fundamentales.<sup>16</sup>

## 2.2 Una ética de la IA confiable y en el diseño *made in Europe* para competir en el mundo

Las llamadas a la ética y a subrayar su papel fundamental en el desarrollo de la IA y complementar la regulación han sido muchas en el ámbito europeo. En este sentido la Resolución del Parlamento de la UE sobre robótica señala que se precisa un “marco ético claro, estricto y eficiente que oriente el desarrollo, diseño, producción, uso y modificación de los robots a fin de complementar [...] las recomendaciones jurídicas” (n.º 11) (Parlamento Europeo, 2017a). Tras los ya referidos estudios de STOA en 2016, esta Resolución del Parlamento sobre robótica va más allá e incluye un completo anexo, carta integrada, con código de conducta para los ingenieros en robótica, un código deontológico destinado a los comités de ética de la investigación, y licencias tipo para los diseñadores y los usuarios que luego se detalla. También la Resolución del Parlamento

---

<sup>13</sup> <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

<sup>14</sup> <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

<sup>15</sup> <http://www.eismd.eu/ai4people>

<sup>16</sup> Se trata del proyecto Big Data and Fundamental Rights <http://fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights>

de la UE sobre *big data* afirma que “son fundamentales normas científicas y éticas estrictas” (n.º 2), “un marco ético común sólido” (n.º 20), “las normas éticas más elevadas” (n.º 32) (Parlamento Europeo, 2017b).

El Supervisor Europeo de Protección de Datos (2015a: 4 y sig.) señala que en el entorno digital actual no basta con respetar la ley sino que es preciso tener en cuenta la dimensión ética del tratamiento de datos. En esta dirección, la Comisión Europea (2018a: 2-3) recuerda que hay que “centrar esfuerzos en garantizar que las nuevas tecnologías estén basadas en valores”, que “la IA se desarrolle y aplique en un marco adecuado, que promueva la innovación y respete los valores y derechos fundamentales de la Unión, así como principios éticos tales como la obligación de rendir cuentas y la transparencia”. Se debe “garantizar el establecimiento de un marco ético y jurídico apropiado, basado en los valores de la Unión y en consonancia con la Carta de los Derechos Fundamentales de la UE” (Comisión Europea, 2018a: 4, 14 y sig.).

En marzo de 2018 el EGE (2018: 14) afirmó que lanzar unos principios fundamentales éticos “es un asunto de gran urgencia”. La referida plataforma AI4People, en el marco Atomium-EISMD (Floridi *et al.*, 2018b: 701 y sig.), entre sus veinte recomendaciones generales éticas, subraya el papel de la ética y que es clave evaluar “qué tareas y funciones de toma de decisiones no deben delegarse a los sistemas de inteligencia artificial” (2ª rec.). Igualmente, hay que evaluar “si las regulaciones actuales están suficientemente fundamentadas en la ética” (3ª rec.). La ética debe servir también para señalar qué tecnologías de inteligencia artificial subvencionar por ser “socialmente preferibles” (12ª rec.). Castellanos (2019: 26-27) recientemente ha planteado en términos maquiavélicos la necesaria mitigación de la acción de la IA, pese a su posible mayor eficacia y eficiencia frente a los humanos.

En la nueva economía digital y en particular de la IA hay una fuerte competencia entre las grandes regiones —y países— del mundo. Y no parece que la UE esté nada bien situada en esta competición en patentes, grupos de investigación, inversión, subvenciones y tantos otros indicadores (Sandri, 18 de febrero de 2019). El EGE (2018: 18) llama la atención del peligro de que haya un mercadeo a la baja respecto de la ética y regulación a respetar por la IA. Se alerta de un peligro de reubicación del desarrollo y el uso de la IA en regiones con estándares éticos más bajos —o inexistentes. Y la UE apuesta sin duda por altos estándares de ética, derecho y derechos fundamentales. Esta estrategia podría considerarse de cierta candidez de la UE. Puede entenderse como *poner unos palos éticos y jurídicos en las ruedas al desarrollo de la IA*; unas barreras que no se introducen para las multinacionales de EE. UU. y especialmente para las empresas, universidades y el sector público chino, entre otros. No obstante, la UE no puede renunciar a sus “valores humanistas intrínsecamente europeos y universales que caracterizan la contribución de Europa a la sociedad” (Parlamento Europeo, 2017a: letra U), valores humanistas y principios basados en la dignidad y los derechos fundamentales. El grupo de expertos (AI-HLEG, 2018: 4) apunta que no se trata de frenar la innovación, “sino para usar la ética como inspiración en el desarrollo de una marca propia de IA. Europa podrá posicionarse como líder en una IA de vanguardia, segura y ética”. Así pues, afirma el objetivo de acordar estándares éticos para el diseño, tipo etiqueta *made in Europe* (AI-HLEG, 2018: 21). De ahí que se afirme que esta marca es “nuestra estrella del norte” y “nuestro objetivo es crear una cultura de ‘IA confiable en Europa’” (AI-HLEG, 2018: 29). Esta idea se refuerza en el Plan coordinado sobre la IA (Comisión Europea, 2018b) de diciembre de 2018: la IA confiable es la pieza angular de la *AI made in Europe*, que es el propio título del Plan de IA en su importante anexo.

A mi juicio, parece inteligente la estrategia de la UE de situarse a la vanguardia del mundo en una IA ética y confiable como marca. Bien puede entenderse como *hacer de la virtud necesidad*: dado que son irrenunciables unos altos estándares de derechos fundamentales de la UE, aprovecharlo para *venderlo* en el mundo. O, si se quiere, *hacer de la necesidad virtud*: dadas nuestras dificultades para competir en IA con otras regiones, ofrecer al mundo un producto de IA ética específico.

Y el mayor distintivo de esta marca Europa es el *Ethics & Rule of law by design (X-by design)* (AI-HLEG, 2018: 19). Se trata de la inserción del cumplimiento ético y legal en el mismo código, en términos de Lessig (2001), en nuestro caso, en el *código* de las tecnologías disruptivas. Y este es precisamente el tercer modelo regulatorio de la IA que menciona Petit para la IA (2017: 3 y 11), uno de los miembros del grupo de expertos: “Normas [que] se pueden incorporar directamente a la tecnología en el sentido de que un comando y su cumplimiento está imbuido en la tecnología misma”. En otros lugares (Cotino, 2017; Cotino, 2019: 27 y sig.)

he subrayado la importancia de adoptar técnicas relativas a los principios de precaución, así como técnicas de responsabilidad activa. Como es sabido, la privacidad debe integrarse desde el inicio en la gestión y ciclo de vida del tratamiento de datos. La privacidad en el diseño (art. 25) es un principio esencial en la nueva filosofía de acción de la responsabilidad proactiva del RGPD de la UE. Ahora se va incluso más allá de la protección de datos: “Para una IA ‘made in Europe’ un principio clave será el de ‘ética por diseño’ según el cual los principios éticos y legales, conforme al Reglamento general de protección de datos, el cumplimiento de la ley de competencia, la ausencia de sesgo de datos se implementan desde el inicio del proceso de diseño” (Comisión Europea, 2018b: 5-9 y 19-21). Y a ello debe añadirse la exigencia de la transparencia y explicabilidad también en el diseño mismo. También, la Conferencia Internacional de Protección de Datos afirma un “enfoque general de ‘ética por diseño’” (ICDPPC, 2018: ap. 4º).

### 3 El contenido y principios de la ética de la IA

#### 3.1 Una ética de la IA que sitúa en el centro la dignidad y los derechos

El punto de partida y premisa ética de la IA en las numerosas declaraciones y los diferentes documentos no es otro que la dignidad y los derechos fundamentales. A partir de ello, se derivan diversos principios. Es posible que el lector jurista vea diversos de los fundamentos y principios éticos que ahora se exponen como una serie de generalidades y lugares comunes para un jurista, especialmente porque la dogmática jurídica y el ordenamiento internacional y constitucional español parten de la dignidad como presupuesto de los derechos fundamentales y los derechos sean derechos subjetivos sean principios objetivos del ordenamiento jurídico.

La referida marca de la IA confiable y en el diseño *made in Europe* parte de la dignidad humana y los derechos fundamentales. Especial acierto tiene el Supervisor Europeo de Protección de Datos (2015: 14 y sig.) cuando considera que hay que situar “la dignidad en el centro de una nueva ética digital”, “un mayor respeto de la dignidad humana y una mayor salvaguardia de la misma podrían servir de contrapeso a la vigilancia generalizada y la asimetría de poder a la que se enfrentan las personas”. También el EGE (2018: 16) sitúa como primer principio la dignidad humana. Y del mismo derivan limitaciones en las “determinaciones y clasificaciones” que se realizan con IA, especialmente cuando los afectados no están informados. También la dignidad humana genera la obligación de dar a conocer si se da una interacción artificial con humanos. Asimismo de la dignidad deriva la necesidad de reservar diversas tareas a los humanos y no a las máquinas. El grupo de expertos de la UE, en su elaboración de directrices de ética que se le ha encomendado, parte de una IA centrada en el hombre (*human-centric AI*): además de garantizar los valores humanos, “el uso de la IA no debe verse como un medio en sí mismo, sino con el objetivo de aumentar el bienestar a mayor número de ciudadanos”.

Y una ética que parte de la dignidad implica que sus principios se nutren de los derechos fundamentales. Así lo expresa con claridad el Parlamento de la UE en su declaración de robótica (Parlamento Europeo, 2017a: por ejemplo, letra O, principios éticos 10 y 13). El grupo de expertos del Parlamento de la UE (STOA, 2016a: 36-38) ya afirmaba que el futuro código de conducta que luego adoptó el Parlamento “debería basarse en los principios consagrados en la Carta de los Derechos Fundamentales de la UE [...] marco normativo en el que se puede construir un entendimiento común de los riesgos éticos asociados con el funcionamiento de los robots”. También el EGE (2018: 18) afirma que los principios éticos de la IA están basados en la Carta de Derechos Fundamentales. Y la Comisión Europea (2018: 17) encomienda al grupo de expertos que las directrices éticas han de “tener debidamente en cuenta” la Carta. Así, para el grupo de expertos, los derechos fundamentales son la base de concepto de IA confiable y del fin ético de la IA, que es precisamente garantizar su cumplimiento. Los derechos son “la base para la formulación”, “el trampolín [*stepping stone*] para identificar principios y valores éticos abstractos y concretos y operativos” (AI-HLEG, 2018: IV, 5-6). En la misma línea, están la 40 Conferencia Internacional de Protección de Datos de Bruselas, de octubre de 2018 (ICDPPC, 2018: punto 1), o la Cumbre de la UNESCO de 4 de marzo de 2019 (*supra* nota 4).

#### 3.2 Los cinco principios básicos a los que se reconducen los principios éticos reconocidos internacionalmente



La declaración del Parlamento de la UE (2017a) sobre robótica es una buena expresión de lo que constituyen los principios éticos esenciales. Así, además del principio de transparencia (n.º 12) señala que “este marco de orientaciones éticas debe basarse en los principios de beneficencia, no maleficencia, autonomía y justicia, así como en los principios consagrados en la Carta de los Derechos Fundamentales de la Unión Europea [...] así como en otros principios [...] como la no estigmatización, la transparencia, la autonomía, la responsabilidad individual, y la responsabilidad social” (n.º 13). Esto es, ya condensó en buena medida todos los principios en estos cinco.

En 2018, el proyecto AI4People ha llegado a contabilizar 47 principios éticos proclamados internacionalmente. Y considera que hay cinco principios que sintetizan o captan el significado de los 47 (Floridi *et al.*, 2018: 696): beneficencia (hacer el bien), no maleficencia (no hacer daño), autonomía o acción humana (*human agency*, respeto por la autodeterminación y elección de los individuos) y justicia (trato justo y equitativo para todos). A estos cuatro añade el también mencionado por el Parlamento, principio de explicabilidad o transparencia (operar de modo transparente o como inteligibilidad y responsabilidad). El Grupo de Expertos de Alto Nivel de la UE los asume directamente por su carácter inclusivo y, sobre todo, porque “estos principios generales proporcionan una guía hacia la operacionalización de los valores fundamentales” (AI-HLEG, 2018: 8). Los cinco principios pasan a ser hoy día los pilares de la IA confiable *made in Europe*. Con todo hay que decir que los cuatro principios básicos traen causa del ámbito de la biomedicina desde los años 2000 (Beauchamp y Childress, 2001). “Los cuatro principios bioéticos se adaptan sorprendentemente bien a los nuevos desafíos éticos planteados por la inteligencia artificial”, pues la bioética es la que más se parece a la ética digital (Floridi *et al.*, 2018: 696).

#### *Haz el bien y no hagas daño (principios de beneficencia y de no maleficencia). Principio de justicia*

Toda declaración internacional recoge el principio de beneficencia (haz el bien). Así, la IA debe desarrollarse para el bien común y el beneficio de la humanidad, mejorar el bienestar individual y colectivo, generar prosperidad, valor y maximizar la riqueza y sostenibilidad. Asimismo, debe buscar una sociedad justa, inclusiva y pacífica, ayudando a aumentar la autonomía de los ciudadanos, con una distribución equitativa de oportunidades económicas, sociales y políticas. También debe tener objetivos como la protección del proceso democrático y el Estado de derecho; la provisión de bienes y servicios comunes a bajo costo y de alta calidad; alfabetización y representatividad de los datos; mitigación de daños y optimización de la confianza hacia los usuarios. De igual modo ha de servir de modo amplio para el logro de los objetivos de desarrollo sostenible de la ONU o la sostenibilidad (entre otros, AI-HLEG, 2018: 8; Floridi *et al.*, 2018: 696; *Cámara de los Lores*: n.º 417).

No siempre fácil de distinguir del anterior, el principio de no maleficencia tiene una funcionalidad esencialmente preventiva: no hacer daños físicos, psicológicos, financieros o sociales. Se trata de subrayar una cautela en torno a los límites superiores de las capacidades futuras de IA, así como de evitar el uso indebido (IEEE, 2016-2017). Ello vale tanto para quienes desarrollan y configuran la IA, como para sus resultados. Incluye evitar los daños accidentales (uso excesivo), así como los daños deliberados (“uso indebido”). Como he analizado recientemente con detalle (Cotino, 2019), es posible categorizar los riesgos de la IA para también adoptar modelos de respuesta y regulación (*id.*: 9-15 y 18-21). Y precisamente frente a los daños, el principio de precaución y las técnicas de responsabilidad activa son idóneos (*id.*: 27-31). En particular con este principio de no maleficencia se apuntan los variados bienes en riesgo: todo derecho fundamental, la “privacidad, seguridad y precaución de capacidad” (*capability caution*, AI-HLEG, 2018: 9; Floridi *et al.*, 2018: 697). Especialmente en el ámbito de los principios de Asilomar (2017), se trata de evitar una carrera de armamentos. También debe protegerse el proceso democrático, evitar la polarización ideológica y el determinismo algorítmico. Para asegurar la inclusión y la diversidad debe haber especial cautela con datos sobre minorías, especialmente en el diseño. Otro daño a evitar es sobre el medio ambiente, teniendo en cuenta el consumo que requieren los cómputos masivos, el almacenamiento de datos, así como los minerales para las baterías necesarias.

El principio de justicia es otro de los cinco principios éticos básicos. Su contenido es también muy amplio e inclusivo de muchos otros principios (AI-HLEG, 2018: 10; Floridi *et al.*, 2018: 698). Alcanza al uso de la IA para la distribución de recursos y acceso equitativo; para eliminar todos los tipos de discriminación

(Montreal, 2017). Incluye eliminar las discriminaciones históricas, evitar sesgos, estigmatización y nuevas discriminaciones sociales. También se trata de que se generen beneficios compartidos y prosperidad compartida (Asilomar, 2017). El principio de justicia también implica altos estándares de responsabilidad y la reparación efectiva o remedio si se produce un daño. Igualmente se hace referencia a que se evalúe el rendimiento de la IA y que se asegure que en salud se faciliten tratamientos nuevos y la disponibilidad de la atención médica.

### *La libertad y autonomía o acción humana frente a la “autonomía” artificial*

Otro principio ético básico es el de autonomía o acción humana (*human agency*). Se parte de que con la IA “cedemos voluntariamente parte de nuestro poder de decisión a las máquinas. Así, se debe promover la autonomía humana y restringir la autonomía de las máquinas y hacerla intrínsecamente reversible”. “Los humanos siempre deben conservar el poder de *decidir qué decisiones tomar*” (Floridi *et al.*, 2018: 697-698). El humano ha de decidir si se decide (AI-HLEG, 2018: 9-10). Se apuesta así por el también llamado principio de supervisión humana de la IA (gobernanza de la autonomía) (AI-HLEG, 2018: 14-15): más control humano a mayor autonomía e impacto social de la IA, así como más registros auditables y una gobernanza más estricta. También el humano ha de tener la capacidad de desviarse de la decisión elegida por la máquina.

El EGE (2018: 11-12) recuerda que la autonomía como derecho a ser libre solo puede atribuirse a los seres humanos, por lo que no cabe atribuir autonomía a los sistemas por inteligentes que sean. La autonomía humana queda bajo el principio de la libertad del ser humano (letra b: 16), que se traduce en responsabilidad humana y, por lo tanto, control y conocimiento sobre los sistemas autónomos y su mayor transparencia y previsibilidad. En la misma dirección (Floridi, 2018: 18 y sig.) insiste en que las mejoras sociales de la IA no pueden ser a costa de reducir el control humano ni de limitar la prevención de daños. Frente al peligro de manipulación algorítmica, se apuesta por el complejo concepto de *nudging* (Thaler y Sunstein, 2009), como autoempujón, refuerzo positivo indirecto para influir en el comportamiento. No obstante, alerta del riesgo de que las tecnologías erosionen la autodeterminación humana por el cambio de comportamientos y rutinas que puede generar la IA al facilitar el trabajo y la vida de las personas.

### *El principio de explicabilidad y transparencia*

El principio de rendición de cuentas, explicabilidad y transparencia (AI-HLEG, 2018: 10; Floridi *et al.*, 2018: 699-700) se añadiría a los otros cuatro principios básicos. Al respecto y jurídicamente, cabe remitir a estudios como los de Cerrillo, Ponce o Arellano en Boix y Cotino (2019). Como tal principio ético implica la necesidad de entender y pedir cuentas. Debe poder responder a la pregunta “¿cómo funciona?”. Y también en el sentido ético de responsabilidad (“¿quién es responsable de la forma en que funciona?”). Se trata de la pieza crucial que falta en el rompecabezas pues “complementa los otros cuatro principios” (AI-HLEG, 2018: 10). Así, sin transparencia no se entiende el bien o el daño de la IA, sin transparencia no se sabe cómo actuaría la IA en lugar de nosotros (autonomía). Tampoco se puede determinar la justicia y la responsabilidad. La explicabilidad, auditabilidad y transparencia varía mucho en razón de los sujetos, pues va desde la persona de la calle hasta los máximos expertos. El círculo de personas que deban acceder al conocimiento del algoritmo puede ser también restringido por muchos motivos, como seguridad o propiedad intelectual.

La explicabilidad ha generado no pocas recomendaciones (Floridi *et al.*, 2018: 700 y sig.): hay que implicar a industrias y profesionales (rec. 4). En justicia, hay que “desarrollar procedimientos legales apropiados y mejorar la infraestructura de TI del sistema de justicia para permitir el escrutinio de las decisiones algorítmicas en los tribunales” (rec. 5). De igual modo, hay que “desarrollar mecanismos de auditoría [...] para identificar consecuencias no deseadas, como un sesgo injusto”. El grupo de expertos subraya la necesaria trazabilidad y auditabilidad (AI-HLEG, 2018: 20) siempre que la decisión algorítmica tenga un *impacto significativo* en la vida de las personas. En estos casos los legos deben poder entender y debe configurarse el sistema para que los organismos reguladores puedan verificar y auditar. Ya en el ámbito de la investigación se aboga por la *Explanation (XAI research)*, es necesario poder entender por qué tuvo un

comportamiento determinado y por qué ha proporcionado una interpretación determinada. Y es que pequeños cambios pueden generar efectos dramáticos (como confundir un autobús escolar con un avestruz).

### 3.3 Otros principios básicos de la ética de la IA y la empatía con los humanos como principio ético

Los cuatro principios esenciales y el de transparencia van acompañados de otros principios básicos. Según se vio, las premisas básicas de la dignidad y la propia libertad del ser humano como principio de autonomía. Además, se configuran como principios los derechos fundamentales y con especial significación la no discriminación —y evitación del sesgo— y la privacidad. Igualmente se afirman principios autónomos como el de responsabilidad, Estado de derecho y la rendición de cuentas o la no estigmatización (Parlamento UE, 2017a). De la democracia se afirma que “las decisiones clave sobre la regulación del desarrollo y la aplicación de la IA deben ser el resultado del debate democrático y el compromiso público” (EGE, 2018: 16). Asimismo la necesidad de “educación o acceso a información sobre nuevas tecnologías y sus implicaciones éticas”. Se afirman también los derechos de las “futuras generaciones” a las precondiciones de vida y el medio ambiente.

El Parlamento UE (2017a), en su anexo al código de conducta para ingenieros de robótica, afirma otros principios como los de precaución, participación, rendición de cuentas, seguridad, reversibilidad e inversión de acciones y volver a la fase “buena” de su trabajo, privacidad y maximización de beneficios y reducir al mínimo los daños. Asimismo, en el código de comités de ética incluye principios como evitar conflictos de intereses; acreditada competencia y experiencia de sus miembros, independencia, transparencia y obligación de rendir cuentas, carácter multidisciplinar y como mínimo un miembro con formación filosófica, ética o jurídica. En el Reino Unido la Cámara de los Lores (2018: n.º 417) parte de cinco principios generales para un código de IA además de beneficencia (1), transparencia o inteligibilidad (2) y no maleficencia—no disminuir derechos (3), incluye la limitación del poder autónomo de los sistemas para herir, destruir o engañar a los seres humanos (5) y el derecho a ser educados para permitirles prosperar mental, emocional y económicamente junto con la inteligencia artificial (4).

El grupo de expertos de la UE (AI-HLEG, 2018: 14-18) afirma diez requisitos y principios de la IA, entre los que cabe subrayar el de responsabilidad (1.º), como posible compensación monetaria en la detección de fallos e incluso simple reconciliación sin compensaciones. El de gobernanza de los datos (2.º) pone el énfasis en asegurar la calidad de los datos, la prevención de sesgos y que la anonimización permita la división de los datos en conjuntos. Se insiste en la evitación de datos maliciosos especialmente para que sistemas de autoaprendizaje no entrenen con ellos. Asimismo, la garantía de que los datos recopilados no se utilizarán contra las personas que los proporcionaron. Se afirma también (3.º) el principio de diseño para todos, pero no con un enfoque único, sino centrado en el usuario y teniendo en cuenta toda gama de particularidades.

Resulta de interés el principio de robustez (*robustness*, 8.º). Con él se reclaman sistemas seguros, confiables y suficientemente sólidos frente a errores o inconsistencias durante la fase de diseño, desarrollo, ejecución, implementación y uso del sistema IA, y para enfrentar adecuadamente los resultados erróneos. Se incluyen las exigencias de fiabilidad (*reliability*) por cuanto los resultados deben poder ser avalados por evaluación independiente y de reproducibilidad (*reproducibility*): que los resultados sean consistentes en diferentes situaciones, marcos computacionales y datos de entrada. Este principio incluye también la exactitud (*accuracy*) para clasificar correctamente la información en las categorías correctas, o su capacidad para hacer predicciones, recomendaciones o decisiones correctas basadas en datos o modelos. Igualmente, la resiliencia frente a las vulnerabilidades y ataques que pueden afectar al funcionamiento mismo del sistema, a las decisiones, su sesgo o generar daños. Asimismo, se exige un plan de retroceso (*fall back plan*) en caso de problemas con el sistema de IA, esto es, que el sistema de IA cambie de procedimiento o directamente que actúe un operador humano.

Más allá de los principios contenidos en las importantes declaraciones, Bentani (2019) afirma principios que me atrevo a calificar de obligada *empatía artificial con el humano*, esencialmente para la “urgente” necesidad de “adaptar nuestras estructuras sociales a esta nueva sociedad formada por agentes autónomos mixtos” y “garantizar la convivencia entre el hombre y la IA, entre el hombre y el robot”. Así, por el principio de intuición (1) la IA debe poder adivinar lo que los hombres quieren hacer y adaptarse a sus intenciones cooperando. La máquina debe ser adaptada al hombre y su singularidad y no viceversa. Su enfoque del

principio de inteligibilidad (2) también es particular: que la persona que comparte el espacio de trabajo con la máquina siempre pueda entender lo que la máquina va a hacer. También se señala la adaptabilidad (3), en el sentido de que el robot también debe adaptarse a la personalidad humana con la que coopera y, en este sentido, ser emocional. Igualmente se afirma el principio de adecuación de los objetivos (4), en la dirección de cierta humildad artificial, esto es, la eficacia no es la primera prioridad operacional, sino adaptar su acción y objetivos al humano con el que coopera.

#### **4 Las potenciales utilidades de la ética de la IA pese al natural escepticismo del jurista y el desarrollo de códigos de conducta y comités**

Para la cultura continental europea más legalista o al menos para quien suscribe puede verse la ética de la IA con cierto escepticismo. Si bien también hay que tener en cuenta los límites del derecho. Y en el ámbito de la IA puede haber especial lugar para la ética pues complementa al derecho y tiene un carácter preventivo que puede evitar que el derecho reactivo tenga que aplicarse. Asimismo, la eficacia misma del derecho puede reforzarse con políticas y estrategias de ética de la IA. Aun a pesar de cierto escepticismo del jurista, pueden sostenerse diversas ventajas y aportaciones del desarrollo de una ética de la IA, a saber:

- Con la regulación, autorregulación y políticas que se adopten, es una pieza más para construir un marco, una estrategia, un sistema y una gobernanza de la IA.
- Es el marco propio para la reflexión de las cuestiones clave y profundas que se plantean y la guía inicial de las soluciones jurídicas que deben plantearse adoptarse.
- Es una fuente inspiradora de las normas y las políticas concretas que han de desarrollarse, al tiempo de marcar límites que deben concretarse.
- Complementa la interpretación de las normas que existan o que vayan apareciendo y facilita su aplicación en supuestos concretos.
- Se adapta mejor a escenarios privados, empresariales, organizaciones corporativas y sectores técnicos y profesionales, especialmente por cuanto se vehicula a través de autorregulación, normas éticas y códigos de conducta a los que adherirse. Asimismo, puede contarse con instrumentos de garantía como comités éticos u otros órganos que tienen, a la postre, mayor efectividad en tales escenarios.
- Es un mecanismo especialmente preventivo, e incluso puede verse como un sistema de alerta temprana frente a impactos y riesgos, a diferencia del derecho que ofrece especialmente respuestas reactivas una vez producidos incumplimientos.
- Permite detectar a tiempo líneas o caminos que aunque sean lícitos se consideran socialmente inaceptables.
- Tanto para los colectivos implicados como para la sociedad general, es un mecanismo eficaz para generar una adhesión personal y colectiva, para facilitar también la pedagogía y una cultura de la futura IA confiable y humana.

En la línea de lo expuesto, lejos de desdeñar la importancia y papel de la ética de la IA, no solo hay que afirmar su complementariedad, sino la necesidad de la convergencia de la misma con las distintas fórmulas de regulación y la política y las concretas políticas a adoptar en la materia, así como una cultura y clima ético en el sector privado y público llamado a desarrollar la transformación digital.

Así, la Conferencia Internacional de Protección de Datos en 2016 señaló con acierto que “en el ambiente de hoy, la adhesión a la ley no es suficiente; tenemos que considerar la dimensión ética de las tecnologías [...]”. Independientemente del marco regulatorio, existen preguntas profundas sobre el impacto de estas tecnologías en la dignidad y la libertad individual” (ICDPPC, 2016 núm. 3: 16). Floridi (2018: 4) subraya esta dimensión inspiradora y positiva de la ética: “La ética digital moldea la regulación digital y el gobierno digital”, “en resumen, la ética es lo que hace o da forma a la ley”. También, con AI4People, apunta que “el cumplimiento de la ley es meramente necesario (es lo mínimo que se requiere), pero es significativamente insuficiente” (Floridi *et al.*, 2018: 694). De hecho afirma que un enfoque ético tiene doble ventaja: apreciar el valor

social, identificar y aprovechar nuevas oportunidades que son socialmente aceptables o preferibles. Y, dos, un valor preventivo “permite a las organizaciones anticipar y evitar o al menos minimizar errores costosos”, detectar y evitar lo “socialmente inaceptable”, “incluso cuando son legalmente incuestionables” (Floridi *et al.*, 2018: 695). El eficaz papel preventivo de la ética lo subrayó también en la Cámara de los Lores del Reino Unido (2018: ap. 378) la firma Baker McKenzie: el Gobierno debería “facilitar marcos éticos (en lugar de legales) para el desarrollo de tecnologías de IA para apoyar la autorregulación en la industria”, esto es, una “intervención proactiva y basada”, “en lugar de una regulación reactiva, implementada después de que algo salga mal”. En esta línea, Floridi *et al.* (2018: 695) señalan que “un enfoque ético de la IA también puede verse como un sistema de alerta temprana (*early warning*) contra riesgos que podrían poner en peligro a organizaciones enteras”.

#### 4.1 Hacia un marco, infraestructura, sistema o gobernanza ética de la IA a partir del ejemplo de la ética pública y la integridad institucional

Como se ha afirmado, un marco o enfoque ético es cuando menos complementario y enriquecedor y potencia la eficacia de la regulación, así como de las decisiones políticas que se adopten en este terreno. Una gobernanza de la IA o gobernanza responsable (AI-HLEG, 2018: 22) parece integrar sistemáticamente estas acciones, más o menos en la línea que se está construyendo en la UE o que se acaba de proclamar en Singapur aunque con menor papel de la ética (PDPC, 2019). Considero de utilidad seguir los avances que se han dado en el terreno de la llamada ética o integridad pública y el buen gobierno, donde la variada literatura y *soft law* nacional e internacional afirman conceptualmente una infraestructura ética un sistema ético o de integridad institucional. Especialmente en los últimos años, la crisis, la desconfianza y la deslegitimación de la política han llevado a reforzar estas estrategias de mayor tradición anglosajona, superando el enfoque solo legalista propio del continente europeo. Este sistema incorpora estrategias, políticas, marcos de integridad institucional (*integrity framework*), marcos normativos y gestión de la integridad. Así, por ejemplo, la reciente Recomendación OCDE (2017: 4) sobre integridad pública afirma una estrategia para la integridad pública y recomienda a los países que desarrollen un sistema coherente y global en materia de integridad pública. La ordenación de todo ello queda englobada de infraestructura ética y las combinaciones de estos conceptos llevan a hablar, asimismo, de un sistema de gobernanza ética (Londo, 2015), marcos que en España especialmente estudian Jiménez Asensio (2016 y 2017) y Villoria e Izquierdo (2016: 202-204).

Asimismo, hay que tener en cuenta que el sistema y marco ético de IA ha de implantarse especialmente en el sector privado y profesional. Es por ello que hay que aprovechar la nueva cultura del *compliance* —también anglosajona— que también se está insertando paulatinamente en los países de tradición continental europea, particularmente en el ámbito de las sociedades mercantiles y de los delitos societarios (entre otros, Matallín, 2018). Así, es bien posible partir de tales esquemas para emularlos y proyectarlos al sector público y privado y los colectivos profesionales de desarrollo de la IA, como ya ha sucedido en terrenos como la biomedicina, en el que están bien asentados. Sin excluir el papel de la ley, en aras de la eficacia especialmente en el sector privado, cabe seguir la experiencia de países anglosajones, que suelen partir de una premisa conceptual de autorregulación de esta materia apoyada con algunas previsiones excepcionales o genéricas en las leyes.

Y en estos ámbitos, la ética pública se ha subrayado por su especial carácter preventivo, frente a un papel más reactivo propio del derecho. También en el desarrollo de la IA se trata más de prevenir que curar. Más allá de la reacción penal o disciplinaria que pueda existir, se requiere el desarrollo de mecanismos de alerta y vigilancia eficaces, coherentes y generalizados. Se han subrayado en esta dirección ejes de actuación que considero proyectables al ámbito de la IA. Así, Sauv  (2014: 103-104) señala la necesidad de fijar las reglas y principios b sicos a seguir para que queden claros a todos los actores; introducir medidas basadas en una cultura de prevenci n; dar respuestas o herramientas m s concretas para los riesgos m s importantes generados por la IA; se puede encomendar el control a una autoridad. Tambi n, siguiendo conjuntamente la Recomendaci n OCDE (2017) y a Jim nez Asensio (2016: 18) y proyect ndolo para nuestro  mbito, el logro de la gobernanza de la  tica de la IA requiere:

– Afirmar un compromiso, con un dise o institucional fijando responsabilidades, estrategia, as  como un marco normativo y de espacios de autorregulaci n.

- Definir una política de ética de la inteligencia artificial, que incluye los códigos éticos o de conducta que conviven con los marcos normativos y legales.
- Una cultura de ética de la IA con liderazgos, formación tanto a la sociedad en general como a los sectores públicos, privados y profesionales implicados. Igualmente debe generarse un debate abierto en la sociedad y en los distintos sectores a través tanto de la discusión como de las denuncias de los peligros o malas prácticas.
- Una rendición de cuentas eficaz que incluya órganos de supervisión y participación para que todas las partes desarrollen e implementen las políticas y también se den responsabilidades e incluso sanciones por la quiebra de los principios éticos. Como señala Sauvé (2014: 104), pese a que la prevención es esencial, hay que implementar mecanismos de control incluso con garantías jurídicas concretas y no puede quedar todo en una soledad ética.

#### 4.2 El desarrollo de códigos de conducta profesionales, empresariales, de investigación y usuarios y comités y sistemas de garantía de los mismos

La ética en el diseño incluye que los principios éticos —y el derecho— se inserten en el código tecnológico en el sentido de Lessig (2001). Pero también en los códigos y la deontología profesional y, por ende, en las reglas y deberes que rigen las profesiones y la conducta de aquellos vinculadas con el desarrollo de la IA y sus relaciones con sus clientes y el público. Además, debe haber una infraestructura ética para la garantía con comisiones, comisionados, autoridades o responsables de ética algorítmica, así como una evaluación y seguimiento. STOA (2016a: 36-38) propuso al Parlamento UE la inclusión del anexo con diversos códigos y principios y que no tuviera forma jurídica. Y el Parlamento Europeo (2017a) decidió incluir como anexo una carta integrada, que “sentará las bases para la identificación, la supervisión y el cumplimiento de los principios éticos fundamentales desde la fase de diseño y desarrollo” y con el objetivo “también de introducir un procedimiento para la resolución de los dilemas éticos y permitir que estos sistemas puedan funcionar de una manera éticamente responsable”.

El Código de conducta ética para los ingenieros en robótica dirigido a investigadores y diseñadores afirma que los organismos públicos y privados de financiación de la investigación deberían exigir una evaluación del riesgo para cada propuesta de financiación. Asimismo, asumir el compromiso de “adoptar una conducta estricta en materia de ética y de deontología” y, en concreto, respetar una serie de principios a los que se ha hecho concreta referencia *supra*. El anexo también incluye el Código deontológico para los comités de ética de la investigación, en el que destaca que el proceso de revisión ética ha de ser independiente de la propia investigación, así como evitar conflictos de intereses. Asimismo, debe garantizarse la competencia y la experiencia adecuada y la transparencia y obligación de rendir cuentas. Se recomienda que sean multidisciplinarios, con especialistas técnicos, usuarios de servicios especializados y que se incluya al menos un miembro con formación filosófica, ética o jurídica.

También el anexo cuenta con una licencia para los diseñadores que deben tener en cuenta los principios básicos “antes, durante y después del proceso de concepción, desarrollo y de aplicación de esas tecnologías”, incluida la necesidad de no perjudicar, herir, engañar o explorar a los usuarios (vulnerables). Se insiste en que deberán introducir principios de diseño tanto para la concepción del material y de programas informáticos, como para el tratamiento de datos. Asimismo, deberán asegurar que las informaciones privadas se conserven con total seguridad. También se exige a los diseñadores que integren “teclas de interrupción de urgencia”. Igualmente deberán asegurarse de que las etapas de toma de decisión del robot puedan ser objeto de reconstrucción y trazabilidad bajo la transparencia máxima y previsibilidad del sistema. Han de existir instrumentos de rastreo para explicar los comportamientos de los robots, al igual que protocolos específicos de seguridad. Asimismo, “los diseñadores deberán obtener el dictamen favorable de un comité de ética de la investigación antes de probar un robot en un entorno real o implicando a seres humanos en los procedimientos de concepción y desarrollo”.

Finalmente el anexo incluye una licencia para los usuarios con garantías como la posibilidad de “hacer uso de un robot sin miedo de perjuicio físico o psicológico, el derecho a esperar que un robot efectúe las tareas para las que haya sido expresamente concebido”. También, que los usuarios “no están autorizados a recoger, utilizar o divulgar información personal”, ni para “utilizar un robot de modo contrario a los principios y normas éticas o jurídicas”, y tampoco pueden “modificar los robots para utilizarlos como armas”.

Tras el anexo, en 2018 AI4People (Floridi *et al.*, 2018b: 701 y sig.) recomienda el desarrollo de códigos de conducta autorregulados “para los datos y las profesiones relacionadas con la inteligencia artificial”, con la certificación correspondiente de IA ética a través de etiquetas de confianza. En el ámbito empresarial se llama a que los consejos de administración asuman la responsabilidad de las implicaciones éticas y que cuenten con comité de ética con poderes de auditoría interna. De igual modo se afirma que las organizaciones que desarrollan o utilizan sistemas de inteligencia artificial cuenten de una forma obligatoria con una junta de revisión ética corporativa (19ª rec.). El grupo de expertos de la UE postula una gobernanza responsable (*accountability governance*) y para ello que se nombre a una persona a cargo de los problemas de ética en relación con la IA, un panel o junta de ética tanto interna como externa a la organización. Esta persona debe supervisar problemas y asesorar en los procesos. Ello sin perjuicio del cumplimiento legal o de las competencias de las autoridades (AI-HLEG, 2018: 22).

Paralelamente, la Cámara de los Lores (2018: párrafos 419-420 y conclusiones 73-74) apuesta por que el gobierno coordine las iniciativas de muchas organizaciones públicas y privadas de elaborar códigos de conducta. Podría resultar útil un código IA “que las empresas y organizaciones que implementan IA podrían firmar” (párrafo 419; concl. 73). Para ello, recomienda que el Centro de Ética e Innovación de los Datos de Reino Unido<sup>17</sup> lo elabore, con aportes del Consejo de IA y del Instituto Alan Turing, con un grado de urgencia. Entre los posibles contenidos señala el establecimiento de consejos asesores éticos y que el futuro código IA podría ser la base para la regulación legal (concl. 74). El Centro ha iniciado el plan de actividades desde noviembre de 2018.<sup>18</sup>

## 5 Algoritmos “guardianes” al servicio del cumplimiento normativo y de la ética de la IA

Señalan los Etzioni que, dadas las características únicas de la IA (Etzioni y Etzioni 2018: 136-138), los humanos (auditores, contadores, inspectores, policías, jueces, etc.) no pueden llegar a controlar y garantizar que los instrumentos guiados por IA cumplan con la ley. Para determinar la intención, la responsabilidad por los actos de los instrumentos inteligentes, se necesitan “AI guardians” (*id.*: 138), esto es, programas de IA para examinar los programas de IA. “La ley necesita instrumentos inteligentes para tratar con instrumentos inteligentes” (*id.*: 139). Siguiendo el principio de transparencia, también apuestan por que los sistemas inteligentes estén también diseñados de modo que permitan ser supervisados y monitoreados ya por humanos, ya por otros sistemas inteligentes. También, en la línea del principio de reversibilidad y resiliencia ya enunciados, los autores en todo caso concluyen la necesidad de interruptores de apagado como forma de responder a la eterna pregunta de “¿quién vigilará a los guardianes?” (*id.*: 146).

Desde el STOA del Parlamento Europeo, Kritikos (2018b) se ha preguntado: “¿Qué pasaría si los algoritmos pudieran cumplir los principios éticos?”. Y señala la emergencia de soluciones de mercado, herramientas y sistemas para auditoría de la IA. Así, se mencionan soluciones de IA como AI Fairness Toolkit, Audit-AI por Pymetrics u ORCAA o la herramienta de Facebook Fairness Flow para encontrar sesgos en sus algoritmos. En el ámbito del vehículo autónomo aspira a que los autos vayan equipados con un mando ético, que cuente con tecnología de IA que lleve a cabo juicios morales. También se hace referencia a *software* abierto de IBM para que los desarrolladores puedan lidiar con algoritmos de caja negra y entender cómo se adoptan decisiones. En otro aporte, Kritikos (2018a) se pregunta: “¿Y si las tecnologías tuvieran sus estándares éticos propios?”. Se trataría de equipar los sistemas con principios o procedimientos éticos, dotarles de potencial para funcionar de una manera éticamente responsable. Bentani (2019) también apuesta por algoritmos de verificación independientes que puedan evaluar inteligentemente la idoneidad de las inteligencias artificiales destinadas a coexistir y cooperar con el trabajador humano. Se trataría de tecnologías para cuantificar y certificar de alguna manera el cumplimiento de los principios éticos básicos de intuición, inteligibilidad, adaptabilidad y adecuación de los objetivos del robot. Tales algoritmos de evaluación habrían de ser independientes y confiarse a organismos de certificación de terceros.

Sin duda que hay que apostar por herramientas de IA para controlar a la propia IA. Obviamente la cuestión última es la de cómo controlar al controlador. Habrá que aplicar los principios éticos y la normativa y poder

---

17 <https://www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei>

18 <https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation>

monitorear a las propias tecnologías de control. A partir de las evocaciones del cine y la literatura, no faltará quien se pregunte si la singularidad o giro de la IA hacia la superinteligencia lleve a que la IA autónoma intente eludir tanto los controles humanos como los controles a través de IA. Es más, hasta podría pensarse en un complot o connivencia entre sistemas de IA para eludir el control humano. Al menos, por lo que orientan los expertos, quedan algunas décadas para que al parecer sea posible.

## **6 Y para concluir: pasar “de las musas al teatro” en la ética y el derecho de la IA**

El derecho y la ética han prestado escasa y, sobre todo, tardía atención a nuevas tecnologías emergentes y disruptivas. Sin embargo, la IA ya ha concentrado un muy importante interés social, institucional y doctrinal (al menos en otros países). Tal atracción no creo que sea excesiva, pero sí preventiva frente a los efectos, impactos y variados desafíos de todo tipo que ya provoca la IA. Esta vez se quiere llegar a tiempo. La IA está cautivando la atención social y de especialistas y, al tiempo, está generando una demanda de una ética de la IA respecto de las amenazas que se anuncian como más inminentes e incluso frente a los peores vaticinios sobre el futuro de la humanidad. Estamos en un periodo temprano para las respuestas regulatorias más concretas y globales, pero en el momento oportuno para la proclamación de la ética de la IA. Se trata de una noción inclusiva nutrida de ya consolidados principios jurídicos de nuestros sistemas políticos democráticos liberales (dignidad, democracia, derechos fundamentales, etc.). Además, con la proclamación de la ética se subraya también la pretensión del cumplimiento normativo en el diseño y por defecto. Según se ha visto, ello pasa a ser emblema de la ética confiable de la UE con la que se quiere competir en el mundo.

Incluso para un jurista escéptico se ha sostenido que el desarrollo ético de la IA es un elemento más para una más amplia política y gobernanza de la IA, que puede inspirar futuras regulaciones; también puede hacer más eficaz el cumplimiento de las variadas fórmulas de regulación pública y privada. Es más, se ha argumentado que puede desarrollar escenarios de prevención y cumplimiento normativo mejores que las herramientas de heterorregulación más propias del derecho. Como me he ocupado en otros estudios, es necesaria también una respuesta del derecho, tanto por vías más clásicas como con nuevas o actualizadas fórmulas. Y precisamente se ha de contar con la complementariedad de la ética de la IA.

Además, en el momento actual, la ética de la IA puede ser útil para socializar y visibilizar las mayores preocupaciones jurídicas tanto en la sociedad en general como en los sectores y colectivos más vinculados al desarrollo de la IA. Así, la ética de la IA ha de implementarse en los sectores empresariales, profesionales, tecnológicos e incluso en promotores de políticas públicas y de futuras regulaciones. La inserción de una ética en el diseño de la IA puede servir para impulsar u obligar a que las organizaciones se estructuren y se adecúen al cumplimiento normativo, a generar una gobernanza sensible a los problemas e insertar variados sistemas de control interno y externo.

Es más, ya se debe pasar —en expresión de Lope— “de las musas al teatro”. Ya se debe superar una etérea ética de la IA con principios éticos genéricos y por lo general ya asentados (la dignidad y los derechos, principios de beneficencia y de no maleficencia, justicia, libertad y autonomía humana frente a la autonomía artificial y explicabilidad y transparencia). De tales *musas*, ya se puede pasar al *teatro* de más concretas reglas de contenido, organización y procedimiento y la implantación práctica de procesos, organismos, comités, sistemas de cumplimiento. Como muestra, valgan las más de cien preguntas del *check list* que poder contestar al inicio de cualquier proyecto y desarrollo de IA, lanzadas desde la UE en abril de 2019 (AI-HLEG, 2019: 28 y sig.).

Juristas y otros especialistas de ciencias sociales más sensibles a estas cuestiones han de ponerse desde ya a trabajar codo con codo con informáticos, técnicos, matemáticos y tantos especialistas que desarrollan IA. Ello en buena medida puede ser una de las claves de éxito del desarrollo de la IA en la Unión Europea y, esperamos, en otras regiones del mundo.



## Bibliografía

- AI-HLEG, Grupo de Expertos de Alto Nivel en Inteligencia Artificial. (2018). [\*Draft ethics guidelines for trustworthy AI: Working document for stakeholders' consultation\*](#).
- AI-HLEG, Grupo de Expertos de Alto Nivel en Inteligencia Artificial. (2019). [\*Ethics Guidelines for Trustworthy AI. Working Document for stakeholders' consultation\*](#).
- [\*Asilomar AI Principles\*](#). (2017). Future of Life Institute.
- Australian Information Commissioner. (2016). [\*Consultation draft: Guide to big data and the Australian privacy principles\*](#).
- Beauchamp, Tom L. y Childress, James F. (2001). [\*Principles of biomedical ethics\*](#) (5.<sup>a</sup> edición). Nueva York: Oxford University Press.
- Boix Palop, Andrés y Cotino Hueso, Lorenzo (coords.). (2019). [\*Monográfico Derecho Público, derechos y transparencia ante el uso de algoritmos, inteligencia artificial y big data\*](#). *Revista General de Derecho Administrativo*, (50).
- Cámara de los Comunes, Science and Technology Committee. (2016). [\*Robotics and artificial intelligence\*](#) [Quinto informe de la sesión 2016-17].
- Cámara de los Loes. (2018). [\*AI in the UK: ready, willing and able?\*](#) [Informe de la sesión 2017-19].
- Castellanos Claramunt, Jorge. (2019). [\*La democracia algorítmica: inteligencia artificial, democracia y participación política\*](#). En: Andrés Boix Palop y Lorenzo Cotino Hueso (coords.), *Monográfico Derecho público, derechos y transparencia ante el uso de algoritmos, inteligencia artificial y big data*. *Revista General de Derecho Administrativo*, (50).
- CEPEJ, Comisión Europea para la Eficiencia de la Justicia. (2018) [\*European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment\*](#). Consejo de Europa, 79 pp.
- Comisión Europea. (2018a). [\*Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia artificial para Europa\*](#) [COM(2018) 237 final].
- Comisión Europea. (2018b). [\*Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Plan coordinado sobre la inteligencia artificial\*](#) [COM(2018) 795 final].
- Comisión Europea. (2018c). [\*Anexo de la Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Plan coordinado sobre la inteligencia artificial\*](#) [COM(2018) 795 final].
- Committee of Ministers. (17 de octubre de 2017). [\*Technological convergence, artificial intelligence and human rights\*](#) [Reply to Recommendation]. Consejo de Europa.
- Consejo de Europa. (2017). [\*Technological convergence, artificial intelligence and human rights\*](#) [Recomendación 2102 (2017)]. Asamblea Parlamentaria.
- Consejo de Europa. (2019a). [\*Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes\*](#) [Adoptada por el Comité de Ministros el 13 de febrero de 2019 en la 1337<sup>a</sup> reunión].
- Consejo de Europa. (2019b). [\*Governing the game changer – Impacts of artificial intelligence development on human rights, democracy and the rule of law\*](#) [conclusiones de la Conferencia, Presidencia de Finlandia, 26-27 de febrero].

- Cotino Hueso, Lorenzo. (2015). [Algunas cuestiones clave de protección de datos en la nube. Hacia una “regulación nebulosa”](#). *Revista Catalana de Dret Públic*, (51), 85-103.
- Cotino Hueso, Lorenzo. (2017). [Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales](#). *Dilemata. Revista Internacional de Éticas Aplicadas*, (24), 2017, 131-150.
- Cotino Hueso, Lorenzo. (2019). Riesgos e impactos del *big data*, la inteligencia artificial y la robótica. Enfoques, modelos y principios de la respuesta del Derecho. *Revista General de Derecho Administrativo*, (50) (número monográfico de derecho público, derechos y transparencia ante el uso de algoritmos, inteligencia artificial y big data).
- EGE, European Group on Ethics in Science and New Technologies. (2018). [Statement on Artificial Intelligence, Robotics and “Autonomous” Systems](#). Comisión Europea, Dirección General de Investigación e Innovación.
- [Ética de datos, sociedad y ciudadanía](#). (2017). *Dilemata, Revista Internacional de Éticas Aplicadas*, (número monográfico, 24).
- Floridi, Luciano, *et al.* (2018). [AI4People —An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations](#). *Minds and Machines*, 28(4), 689-707.
- Floridi, Luciano. (2018). [Soft ethics and the governance of the digital](#). *Philosophy & Technology*, 31(1), 1-8.
- FRA, Agencia de Derechos Fundamentales UE. (2018). [#BigData: Discrimination in data-supported decision making](#).
- FTC, Comisión Federal del Comercio. (2016). [Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues \(FTC Report\)](#).
- Grupo del Artículo 29 (WP29). (2014). [Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU](#).
- ICDPPC, International Conference of Data Protection and Privacy Commissioners. (2014). [Resolución sobre Big Data](#) [36th International Conference of Data Protection and Privacy Commissioners].
- ICDPPC, International Conference of Data Protection and Privacy Commissioners. (2016). [Artificial Intelligence, Robotics, privacy and data protection: Room document for the 38<sup>th</sup> International Conference of Data Protection and Privacy Commissioners](#) [conferencia celebrada en Marrakech en octubre de 2016].
- ICDPPC, International Conference of Data Protection and Privacy Commissioners. (2018). [Declaration on ethics and data protection in artificial intelligence](#) [40 Conferencia Internacional de Protección de Datos de Bruselas, de octubre de 2018].
- IEEE. (2016-2017). [Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems](#). Actualización del documento de 2016 en diciembre de 2017. Consultado en <https://ethicsinaction.ieee.org>
- International Telecommunication Union (ITU). (2017). [AI for Good](#). Declaraciones, Informe de 2017.
- International Working Group on Data Protection in Telecommunications. (2014). [Working paper on big data and privacy. Privacy principles under pressure in the age of big data analytics](#) [55th Meeting, 5-6 de mayo, Skopje].
- Jiménez Asensio, Rafael. (2016). [Informe final. Asistencia Técnica: Formulación de herramientas para la construcción de Códigos de Integridad y Buen Gobierno en las Entidades del Estado Colombiano](#).

- Jiménez Asensio, Rafael. (2017). *Cómo prevenir la corrupción. Integridad y transparencia*. Madrid: Los Libros de la Catarata.
- Kritikos, Michael. (2018a). [What if technologies had their own ethical standards?](#) *Scientific Foresight (STOA): What if?*
- Kritikos, Michael. (2018b). [What if algorithms could abide by ethical principles?](#) *Scientific Foresight (STOA): What if?*
- Le Déaut, Jean-Yves. (2017). [Technological convergence, artificial intelligence and human rights](#) [Informe doc. 14288]. Consejo de Europa.
- Lessig, Lawrence. (2001). *El código y otras leyes del ciberespacio*. Madrid: Taurus.
- Londo, Francisco. (2015). *Administración pública con valores. Instrumentos para una gobernanza ética*. Madrid: INAP.
- Matallín, Ángela (dir.). (2018). *Compliance y prevención de delitos de corrupción*. Valencia: Tirant lo Blanch, Universidad de Valencia.
- [Montreal Declaration for a Responsible Development of Artificial Intelligence](#). (2017). [Pronunciada en la clausura del *Forum on the Socially Responsible Development of AI*].
- OCDE. (2017). [Integridad pública: Recomendación de la OCDE sobre integridad pública](#).
- Parlamento Europeo. (2017a). [Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica \(2015/2103\(INL\)\)](#).
- Parlamento Europeo. (2017b). [Resolución de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley \(2016/2225\(INI\)\)](#).
- Partnership on AI. (2018). [Tenets](#).
- PDPC, Personal Data Protection Commission Singapore. (2019). [A Proposed Model AI Governance Framework](#).
- Petit, Nicolas. (2017). [Law and regulation of artificial intelligence and robots - conceptual framework and normative implications](#).
- Sandri, Piergiorgio. (18 de febrero de 2019). [China y Estados Unidos dejan atrás a Europa en inteligencia artificial](#). *La Vanguardia*.
- Sauvé, Jean-M. (2011). [Pour une nouvelle déontologie de la vie publique: Rapport de la Commission de réflexion pour la prévention des conflits d'intérêts dans la vie publique, remis au Président de la République le 26 janvier 2011](#).
- Simoens, Pieter; Dragone, Mauro y Saffiotti, Alessandro. (2018). The Internet of Robotic Things. A review of concept, added value and applications. *International Journal of Advanced Robotic Systems*, 15(1).
- STOA, Science and Technology Options Assessment. (2016a). [Ethical Aspects of Cyber-Physical Systems](#). European Parliamentary Research Service.
- STOA, Science and Technology Options Assessment. (2016b). [Legal and ethical reflections concerning robotics](#). European Parliamentary Research Service.
- Stone, Peter (dir.). (2016) [Artificial Intelligence and life in 2030. One hundred year study on artificial intelligence](#). Stanford University.

- 
- Supervisor Europeo de Protección de Datos. (2015a). [Dictamen 4/2015. Hacia una nueva ética digital. Datos, dignidad y tecnología.](#)
- Supervisor Europeo de Protección de Datos. (2015b). [Opinion 7/2015. Meeting the challenges of big data. A call for transparency, user control, data protection by design and accountability.](#)
- Supervisor Europeo de Protección de Datos. (2016). [Opinion 8/2016. EDPS Opinion on coherent enforcement of fundamental rights in the age of big data.](#)
- Thaler, Richard H. y Sunstein, Cass R. (2009). *Un pequeño empujón. El impulso que necesitas para tomar mejores decisiones sobre salud, dinero y felicidad.* Madrid: Taurus.
- Villoria, Manuel e Izquierdo, Agustín. (2016). *Ética pública y buen gobierno. Regenerando la democracia y luchando contra la corrupción desde el servicio público.* Madrid: Tecnos/INAP.