



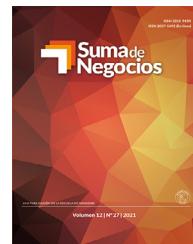
Suma de Negocios
ISSN: 2215-910X
Fundación Universitaria Konrad Lorenz

Verdegay, José Luis; Lamata, M.^a Teresa; Pelta, David; Cruz, Carlos
Inteligencia artificial y problemas de decisión: la necesidad de un contexto ético
Suma de Negocios, vol. 12, núm. 27, 2021, Julio-Diciembre, pp. 104-114
Fundación Universitaria Konrad Lorenz

DOI: <https://doi.org/10.14349/sumneg/2021.V12.N27.A2>

Disponible en: <https://www.redalyc.org/articulo.oa?id=609970431002>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org



Artículo de investigación

Inteligencia artificial y problemas de decisión: la necesidad de un contexto ético

José Luis Verdegay¹, M.^a Teresa Lamata², David Pelta³ y Carlos Cruz⁴

¹Doctor en Ciencias. Profesor del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada. Granada, España. Email: verdegay@ugr.es ORCID: 0000-0003-2487-942X

²Doctora en Ciencias. Profesora del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada. Granada, España. Email: mtl@decsai.ugr.es (autor de correspondencia). ORCID: 0000-0002-9433-644X

³Doctor en Informática. Profesor del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada. Granada, España. Email: dpelta@decsai.ugr.es ORCID: 0000-0002-7653-1452

⁴Doctor en Informática. Profesora del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada. Granada, España. Email: carloscruz@decsai.ugr.es ORCID: 0000-0003-2072-4949

INFORMACIÓN DEL ARTÍCULO

Recibido el 06 de abril de 2021

Aceptado el 23 de abril de 2021

Online el 28 de mayo de 2021

Códigos JEL:

A13, A14, Y80, Y90, Z19

Palabras clave

Inteligencia artificial,
sistemas autónomos de decisión,
procesos de decisión,
problemas de decisión,
ética, concurrencia.

R E S U M E N

Los computadores procesan información y toman decisiones. Hasta hace poco, las decisiones que tomaban no eran complejas, pero debido a los incesantes avances tecnológicos que se producen, los sistemas basados en inteligencia artificial están logrando niveles de competencia en la toma de decisiones que en muchos contextos igualan o superan a las personas. Se trata de sistemas autónomos de decisión que, si bien pueden aumentar la capacidad y eficiencia de las personas en sus ámbitos de actuación, también podrían sustituirlas, algo que preocupa al conjunto de la sociedad. Evitar disfunciones en estos sistemas es un objetivo social, científico y tecnológico prioritario, que exige contar con modelos teóricos, que recojan toda la riqueza y variedad de los problemas de decisión, que definen precisamente los elementos que los caracterizan y que atiendan los planteamientos éticos que deben guiar su funcionamiento. Este artículo describe en sendas secciones cada uno de estos aspectos.

Artificial intelligence and decision making problems: The need for an ethical context

A B S T R A C T

Keywords:

Artificial Intelligence,
autonomous decision systems,
decision processes,
decision problems,
ethics, concurrency.

Computers process information and make decisions. Until recently, the decisions they made were not complex, but due to the incessant technological advances that occur, systems based on Artificial Intelligence are achieving levels of competence in decision-making that in many contexts equal or exceed people. These are Autonomous Decision Systems that, although they can increase the capacity and efficiency of people in their fields of action, could also replace them, something that worries the whole of Society. Avoiding missfunctions in these systems is a priority social, scientific and technological objective, which requires having theoretical models that collect all the richness and variety of decision problems, that define precisely the elements that characterize them and that address the ethical approaches which should guide its operation. This paper describes each of these aspects in a separate section.

Introducción

Parece que nuestro día a día lo controla la inteligencia artificial (IA), los algoritmos, y que nuestro futuro va a depender de ellos. Están presentes a diario en los medios, en nuestras conversaciones o en las redes sociales. Y la verdad es que las noticias que se refieren a ellos nos preocupan y nos atemorizan hasta provocar nuestra desconfianza. Solemos hablar de ellos y de sus consecuencias con rotundidad, pero muchas veces con gran desconocimiento, habilitando escenarios irrealizables o de difícil justificación científica (Verdegay, 2019).

Un algoritmo es una secuencia ordenada de pasos, exentos de ambigüedad, de tal modo que, al llevarse a cabo con fidelidad, en un tiempo finito dará como resultado la solución del problema planteado, habiendo realizado por tanto la tarea para la que se diseñó. Los algoritmos no son por tanto como las recetas de cocina, que pueden tener reglas imprecisas. Por el contrario, son procesos iterativos que generan una sucesión de puntos, conforme a un conjunto dado de instrucciones y un criterio de parada, y como tales no están sujetos a restricciones tecnológicas de tipo alguno, es decir, son absolutamente independientes del equipamiento tecnológico disponible para resolver el problema que afronten (Bazaraa, Sherali & Shetty, 2013). Es el programa en el que se escriba el algoritmo el que depende de la tecnología disponible, es decir, un algoritmo que pueda resolver un problema complicadísimo no sirve si no puede ejecutarse, si no puede desarrollar sus cálculos en una máquina apropiada; lo que, como se ha dicho, es misión del programa en el que se codifique el algoritmo y del computador que se use.

En muchos casos, esto ha sido motivo de una importante ralentización del progreso científico en diferentes ámbitos del saber, como puede ser el caso de las heurísticas, de la programación dinámica, de la prospectiva meteorológica o del análisis exploratorio de datos. Conocíamos los algoritmos que podían dar respuesta a problemas trascendentales para la sociedad (reconocimiento de secuencias de ADN, itinerarios óptimos, alarmas de catástrofes, reglas de conducta, predicciones meteorológicas, etc.), pero no podíamos programarlos y ejecutarlos porque la tecnología disponible no lo permitía, es decir, los computadores disponibles no podían calcular las soluciones buscadas en tiempos razonables.

Sin embargo, el abaratamiento de los computadores, el aumento de su velocidad y el tremendo incremento de sus capacidades, que se produce de forma incesante desde finales del siglo XX, ha tenido dos consecuencias importantes. La primera, el hecho de que hoy en día no quede prácticamente ninguna parcela que tenga un problema que no se pueda abordar con un computador. Este es el caso precisamente de la IA, que en la actualidad es omnipresente en cualquier actividad real de nuestra cotidianidad. La segunda consecuencia es que estemos siendo testigos de la cuarta revolución industrial. En efecto, si la tercera revolución se caracterizó por la democratización de la información, dando lugar a la conocida sociedad de la información, amparada por las nuevas tecnologías de la información y la comunica-

ción (TIC) y las energías renovables, esta cuarta revolución que estamos viviendo está marcada por avances tecnológicos emergentes en una serie de campos (robótica, computación cuántica, biotecnología, internet de las cosas, etc.), entre los que sobresale la IA, los algoritmos, y justifica que se la denomine la *revolución digital*, describiendo una también llamada sociedad 5.0.

La IA desempeña un papel clave en esa cuarta revolución industrial debido principalmente a cinco características, que de manera singular le confieren una naturaleza especial: la transparencia, porque no solemos detectarla cuando interactuamos con algún sistema dotado de IA; su dificultad, porque la referencia esencial para su trabajo la constituyen ni más ni menos que las personas humanas; su adaptabilidad, puesto que es contexto dependiente y por tanto puede resultar polimórfica; la transversalidad, porque en la actualidad no hay ámbito que esté al margen de sus aplicaciones; y, por último, su necesaria y permanente renovación y mejora, como es inherente al área de las TIC, que conducen a una necesidad de inmediatez en las respuestas como en ningún otro campo.

Las sospechas y prevenciones que muestra la sociedad ante los sistemas basados en IA están fundamentadas porque está comprobado que cuando estos sistemas se gestionan con técnicas de automatización inteligente pueden aumentar y, en algunos casos, sustituir, por medio de programas completamente autónomos, la capacidad de actuar y de tomar decisiones de las personas. En lo que sigue, para abreviar, a estos sistemas los denominaremos sistemas autónomos de decisión (SAuD) (Verdegay, 2020), si bien es muy frecuente que también se les llame sistemas automatizados de decisión.

El hecho crucial como hemos dicho es que esa sustitución de funciones podría producir, más pronto que tarde, cambios estructurales en la sociedad en su conjunto; pérdidas masivas de puestos de trabajo y por tanto la minusvaloración y descalificación de las personas que los desempeñaban; efectos desconocidos en los sistemas que puedan gestionar o situaciones de ingobernabilidad no deseadas. Por todo ello resulta patente que:

- a) las cuestiones éticas relativas al comportamiento de los SAuD deben incluirse en su diseño tecnológico, para facilitar que, más que como amenazas, se entiendan como promotores de la innovación, y
- b) en caso de producirse esa sustitución de funciones, que no produzca disfunciones, es decir, que el correspondiente sistema actúe exactamente igual como el decisivo humano de turno, reproduciendo y mejorando su conducta y procurando evitar los ineludibles e imprevisibles fallos que las personas podemos tener a la hora de tomar decisiones, sobre todo cuando estas han de tomarse en ambientes desconocidos (Lamata, Pelta & Verdegay, 2018; 2021).

Conscientes de estas dos exigencias sobre el carácter ético y el correcto funcionamiento de los SAuD, múltiples organizaciones de todo el mundo (European Group on Ethics in Science and New Technologies: Statement on AI 2018; Futureoflife, 2017; Informatics Europe, 2018) entre las que

no faltan las españolas (ENIA, 2020; Ministerio de Asuntos Económicos y Transformación Digital, 2020a; 2020b; Ministerio de Ciencia, Innovación y Universidades, 2019) han comenzado a debatir las condiciones bajo las que deberían desempeñarse dichos sistemas, así como las premisas que deberían guiar su diseño, construcción y ubicación material, a las que habría que añadir la más importante de todas: que esos sistemas estén construidos y diseñados sobre sólidas bases teóricas que minimicen los errores que un mal diseño pudiera producir, ya que las disfunciones en su actividad pueden tener consecuencias extraordinariamente importantes (Foot, 1967).

El objetivo de este artículo es reseñar algunos aspectos teóricos indispensables para la construcción y diseño de los SAuD, así como comentar algunas de las condiciones que múltiples organizaciones han planteado para el desempeño de dichos sistemas. Para ello, el artículo se estructura de la siguiente manera. En la segunda sección se describe lo que es un proceso de decisión y en la tercera se definen los elementos esenciales de un problema general de decisión (y tácitamente su contrapartida de problema de optimización) destacando el papel que desempeñan dos elementos fundamentales: el tipo de información disponible y el contexto en el que se desempeñará el SAuD. La cuarta sección la dedicamos a abundar en los diferentes tipos de información que podemos considerar y a describir una serie de contextos posibles, muchos de ellos propuestos de manera original. Finalmente, se muestran algunas líneas de actuación referentes a la inclusión en el diseño de los SAuD de aspectos éticos o de responsabilidad legal.

Procesos de decisión

La teoría de la decisión ha estado tradicionalmente asociada a los ámbitos de la economía y de la estadística y la investigación operativa. En cualquiera de ellos el principal problema que se ha abordado, más que cómo decidir, ha sido cómo tomar la mejor decisión, de manera que, de forma implícita, el principal problema que subyace en un problema de decisión es el de optimización, es decir, seleccionar la mejor alternativa ante una determinada situación.

La toma de decisiones es consustancial con los humanos, es decir, con las personas. Todo el día no hacemos más que tomar decisiones, intentando que sean las mejores posibles, en cualquier ámbito de la actividad humana. No hay campo profesional que escape a la necesaria y continua toma de decisiones. Ya sea en el terreno laboral, familiar, comercial, etc., tenemos que tomar decisiones. Y tenemos además que tomar las mejores decisiones, de modo que hemos de considerar un doble aspecto: el normativo y el descriptivo, o, dicho de otro modo, hemos de considerar cómo se ha de decidir, lo que es la vertiente normativa del proceso de decisión, y cuál es la decisión que hay que tomar, lo que se conoce como la versión descriptiva del proceso.

Durante mucho tiempo los progresos en teoría de la decisión han discursado por rutas paralelas, dependiendo de si el estudio se realizara desde el punto de vista normativo o descriptivo, de tal forma que parecía que los teóricos de la teoría de la decisión trabajaban de espaldas a los decisores

del mundo real y, como si fueran mundos distintos, ambos al margen de lo que ocurría en el mundo de la economía. De hecho, siendo los modelos de base idénticos, y por tanto los resultados, lo que se traducía en cada caso como solución del problema podía tener versiones diferentes, y por consiguiente parecer soluciones distintas.

Las sucesivas transformaciones sociales de los últimos años, han traído computadores, robots, teléfonos móviles y, en definitiva, todo tipo de artefactos inteligentes en nuestra vida diaria. Con ello, ha surgido la necesidad de enseñar a esos aparatos, a esos SAuD, a que puedan decidir, y a hacerlo de manera óptima. Por eso, conocer tan específica como detalladamente todos los elementos que hacen parte de un proceso de decisión es tan importante como tener garantía de que el resultado de las acciones de un SAuD será tan correcto como el propio modelo permita.

En este contexto, la teoría de la decisión surge como un elemento esencial que juega un destacadísimo papel en esta sociedad digital en la que vivimos y en la que los sistemas basados en IA, y muy particularmente los SAuD, modificarán nuestros hábitos de vida, provocarán cambios estructurales en la sociedad en su conjunto, efectos desconocidos en los sistemas que puedan gestionar, situaciones de ingobernabilidad no deseadas o pérdidas masivas de puestos de trabajo, con el consiguiente riesgo de la descalificación de las personas que los desempeñaban. Por eso, que los SAuD operen correctamente, es decir, que tomen las decisiones adecuadas, requiere conocer tanto como sea posible la forma de desarrollar un proceso de toma de decisiones, algo que Herbert Simon (1960) identificó en tres etapas: concretar las opciones disponibles para tomar alguna decisión; determinar los cursos de acción, las metodologías de solución a nuestro alcance y por último decidir. Con un poco más de detalle estas tres fases se denominan y describen del siguiente modo:

- Etapa de inteligencia, en la que se trata de determinar cómo podemos decidir, es decir, concretar cuánto se pueda cuáles son los datos disponibles, el tipo de información que tendremos que manejar o los mecanismos de razonamiento lógico que podrán usar los decisores.
- Etapa de diseño, en la que se aborda todo lo relativo a la modelización del problema para poder definir claramente las opciones, sus consecuencias o los mecanismos de comparación a emplear. Es una etapa de marcado carácter teórico en la que el conocimiento de otras situaciones similares es fundamental a fin de distinguir similitudes y diferencias.
- Etapa de elección, que incluye la aplicación de los métodos necesarios para optimizar nuestra decisión, pero atendiendo a posibles revisiones de nuestra elección a la vista de lo que el modelo nos proponga como primera opción. La elección, aunque guiada por un proceso de optimización, antes de su adopción definitiva puede estar sometida a negociaciones y modificaciones, sugeridas o impuestas por condicionantes exteriores.

Como se muestra en la figura 1, la superación de esas tres etapas supone su realización secuencial y circular ya que, al culminar la tercera, la reconsideración de nuevo desde

la primera etapa del modelo que tenemos que resolver es obligada.



Figura 1. Desarrollo de un proceso de decisión

Fuente: elaboración propia.

En otras palabras, en su conjunto, a lo largo de la realización de estas tres etapas hay que abordar diferentes tareas: la identificación del problema, la obtención de la información necesaria para su planteamiento, la determinación de posibles soluciones, la evaluación de dichas soluciones y la selección de una estrategia de ejecución, es decir, de puesta en práctica de la que eventualmente se podría reconsiderar la solución final.

Todo ello supone el manejo de conceptos como decisor, consecuencias, orden, contexto de la elección, entre otros, que, aunque conocidos, en el mundo actual pueden adoptar nuevas versiones, que podrían conllevar mecanismos de decisión que merecen ser analizados a la luz de esta nueva situación, a fin de poder exportar a los SAuD con la mayor fidelidad los mecanismos que empleamos las personas para la toma de decisiones, asegurando así su seguridad, eficacia y eficiencia.

Trataremos por tanto de formalizar en lo que sigue un marco que sirva para identificar los elementos necesarios para poder abordar un problema de decisión con la consiguiente optimización de los intereses del decisor. Para ello, a continuación, se presentan los componentes que describen un Problema de Decisión y se presentan algunos conceptos nuevos.

Elementos de un problema general de decisión

Clásicamente, el establecimiento de un problema de decisión pasa por conocer los siguientes elementos esenciales (Ríos, 1976):

- Un decisor, que puede ser una única persona, ya sea física o jurídica, o un grupo de decisores.
- Un conjunto de acciones sobre las que el decisor puede elegir. Como es obvio, para que haya problema, deberá haber dos elementos al menos.
- Un conjunto, denominado *ambiente*, constituido por las situaciones (habitualmente llamadas *estados de la naturaleza*) que puede encontrarse el decisor a la hora de elegir, y que no puede controlar.
- Un conjunto de consecuencias asociadas a cada acción y estado.
- Un criterio que ordena las consecuencias. Este criterio no tiene por qué ser único en el caso de que haya más de un decisor.
- El tipo de información disponible, que no tiene porqué ser necesariamente de naturaleza probabilística.

- La duración del proceso, que podría ser unietápico o multietápico, independientemente del número de etapas, como las descritas en la figura 1, que se lleven a cabo para la resolución del problema.
- El contexto, no aleatorio, en el que finalmente se determinará la acción elegida, que puede influir en la metodología de elección.

Así, si suponemos en todo lo que sigue: un decisor individual, es decir, no consideramos los modelos que involucran múltiples decisores, que dan lugar a los problemas de toma de decisión en grupo (GDM por su sigla en inglés), un único criterio de ordenación, evitando por tanto los problemas de toma de decisión con múltiples criterios (MCDM por su sigla en inglés), y que la duración del proceso se limita a una etapa, para no entrar en problemas más propios de la teoría de control, sin perjuicio de que esa etapa se lleve a cabo en varias fases, conforme a lo descrito en la figura 1, resultará que un problema de decisión queda descrito por un sexteto (X, I, E, C, \leq, K) que incluye el conjunto X de acciones posibles para el decisor, la información disponible I , el ambiente E , el conjunto C de las consecuencias de las acciones, el criterio \leq que ordena las consecuencias, y el contexto K en el que el decisor decide.

Sobre estos dos elementos, información disponible y contexto, hasta ahora no incluidos en la definición estándar de los problemas generales de decisión, tenemos que concretar determinados aspectos que nos ayudarán a modelizar mejor el problema que haya que resolver en cada caso.

En los problemas del mundo real, en los problemas que nos planteamos y resolvemos a diario las personas y, por tanto, en los problemas que queremos modelizar para su implementación en los SAuD de nuestro interés, es prácticamente imposible disponer de toda la información necesaria. Entonces, en lo que queda de artículo supondremos que para la resolución del problema en cuestión tenemos una información incompleta. Aunque tradicionalmente se supone que esa incompletitud es de tipo probabilístico, esa incertidumbre sobre la información a lo que verdaderamente se refiere es a que no conocemos con exactitud qué pasará cuando elijamos un curso de acción concreto, es decir, cuál será la consecuencia concreta de nuestra decisión si algún dato de partida no es exacto (Lindley, 1971). Como es evidente, esa incompletitud puede tener características distintas de las probabilísticas. En este sentido, particularmente útil y reveladora puede resultar la taxonomía de la ignorancia de Smithson (1989) que se ilustra en la figura 2.

La descripción completa y detallada de esta taxonomía excede los límites de este trabajo, pero nótese que, de acuerdo con ella, el hecho de tener una información incompleta sobre los elementos que intervienen en nuestro problema puede deberse a una de las dos siguientes razones:

- Que haya cierta incertidumbre, es decir, falta de certeza sobre los verdaderos valores de los datos y parámetros que intervienen en el modelo.
- Que se dé una auténtica ausencia de información (caso que no consideraremos aquí), es decir, que haya valores del modelo considerado que se desconocen por completo.

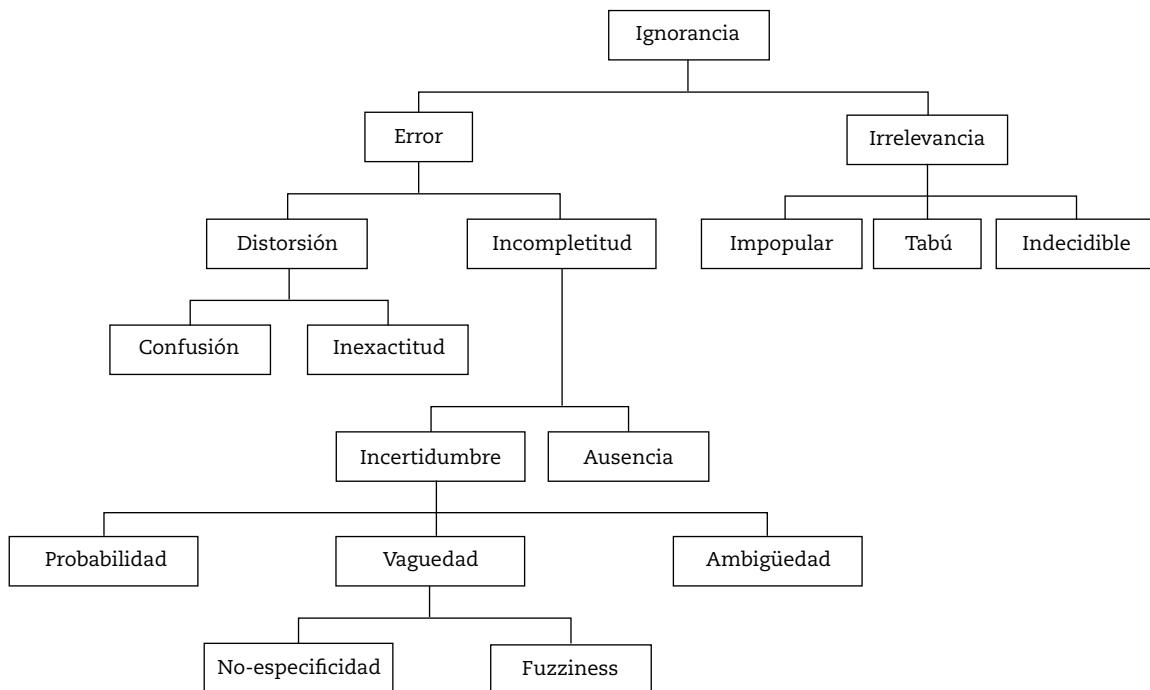


Figura 2. Taxonomía de la ignorancia de Smithson

Fuente: elaboración propia.

De hecho, se desconocería la existencia de las variables que darían esos valores.

A su vez podemos considerar tres tipos distintos de incertidumbre, dependiendo de si su origen es ambiguo, vago o probabilista, es decir, según que:

- haya un número finito de opciones para cada valor (ambigüedad): los hoteles están en el área metropolitana, o
- los datos vengan dados mediante rangos de valores (vaguedad): viajaremos de día, es decir, entre las 7 y las 17 o que,
- intervenga el azar (probabilista): la duración del viaje depende del medio de transporte que nos toque.

En el caso particular, menos tratado desde el punto de vista de desarrollo de un SAuD que aquí estamos considerando, de que la información que tengamos sea de naturaleza vaga o ambigua, que por otro lado es la más habitual en las aplicaciones prácticas reales, dicha información podrá ser gestionada con metodologías y técnicas propias del ámbito de los conjuntos y sistemas difusos (Bellman & Zadeh, 1970; Zadeh, 1965) y así, de forma paralela a cuando la información tiene naturaleza aleatoria, podemos identificar tres ambientes:

- Ambiente de certidumbre: se caracteriza porque se conoce el estado de la naturaleza que se dará, pero ese estado está definido mediante un conjunto fuzzy o una variable lingüística. Por ejemplo, los estados de la naturaleza para el tiempo meteorológico que habrá cierto día pueden ser: fresco, agradable o calor. En ambiente de certidumbre podemos saber que mañana hará calor, pero eso no significa que conozcamos con exactitud

(con certeza, como sería en el caso aleatorio) la temperatura que hará.

- Ambiente de posibilidad: se da cuando sobre los estados de la naturaleza existe una distribución de posibilidad. Se trata de una situación similar a la que se da en el ambiente de riesgo cuando la información es probabilística, pero ahora más asociada al concepto de factibilidad que al de aleatoriedad y, por tanto, sin tener que verificar dicha información axiomática alguna.
- Ambiente de indeterminación: surge cuando se ignora por completo cualquier información sobre cuál de los estados de la naturaleza se presenta, es decir, sabemos que hay una distribución de posibilidad sobre los estados de la naturaleza, pero desconocemos el resto de la información que serviría para concretar dicha posibilidad.

En segundo lugar, respecto al contexto, como hemos adelantado arriba al definir la etapa de elección que describe Simon, la acción que finalmente escoge el decisor como óptima puede estar condicionada por el contexto, K, en el que se desarrolle el problema.

Un contexto, independientemente de la naturaleza que tenga la información disponible, se define como un conjunto de reglas, a menudo establecidas en forma de predicados lógicos, que establecen las características cualitativas que han de tener las decisiones que elegimos para dar solución a nuestro problema.

Los contextos más conocidos son los clásicos de ética y concurrencia (Ríos, 1976):

- Contexto ético, que puede aparecer en procesos de decisión que se desarrollan en ámbitos muy específicos

y profesionales, como puede ser el jurídico, el militar, el médico o, en general, cualquier otro en donde la decisión final esté sometida al cumplimiento de cierto "código". En este contexto no se trata solo de tomar decisiones ajustadas a ciertos comportamientos morales, sino más bien de que las decisiones se ciñan a intereses que se ajusten a códigos éticos. Es el caso típico que se puede presentar cuando en un gabinete jurídico se decide una determinada línea de acción por ser la óptima, sin embargo, hay que abandonarla por cuestiones de ética profesional. El contexto ético suele venir definido por un conjunto de "buenas prácticas" a las que han de ajustarse los decisores.

- Contexto de concurrencia o de competitividad, idóneo para procesos de decisión en los que varios decisores compiten por conseguir un resultado que sea el mejor posible para cada uno a costa del perjuicio que su decisión pueda causarles a los demás. Se presenta en situaciones principalmente asociadas a juegos en los que lo que un jugador gana es lo que pierde su oponente. Es importante comentar que el hecho de que haya más de un decisor no supone que se trate de un GDM problema, ya que no se persigue alcanzar una decisión para el grupo, sino que cada cual actúa por su cuenta.

Con estos contextos también podemos considerar como clásicos, debido a su reconocimiento inicial en el campo de la economía, aunque posteriormente en todas las áreas, estos otros:

- Contexto inducido, directamente inspirado en la teoría del empujón del Nobel Richard Thaler y Sunstein (2008), que recomienda usar refuerzos positivos y sugerencias indirectas para lograr que los decisores, las personas en general, hagan determinadas cosas. Este enfoque, y por tanto este contexto, se fundamenta en la creencia de que con este tipo de refuerzos y sugerencias se logran mejores resultados que con instrucciones precisas, legislación estricta o controles autoritarios
- Contexto disuasorio, de manera contraria al anterior, se considera cuando el planteamiento del problema está exento de peculiaridades que pueden influir en las decisiones que se han de tomar, interesando exclusivamente encontrar la mejor solución con respecto al criterio de comparación predefinido. En general, este contexto, en el que el decisor elige sus cursos de acción con criterios racionales, sin incidencias externas y para variables que toman valores reales positivos, será el que asumamos aquí a efectos de estudio teórico de soluciones del problema.

Pero, además de estos contextos clásicos, hay otros que han surgido recientemente que merecen ser mencionados. En concreto, son nuevos contextos que pueden condicionar nuestras decisiones los siguientes:

- Contexto de presencia de adversarios no visibles (ADM por su sigla en inglés), que se da cuando las decisiones que tomamos son conocidas por nuestros adversarios. La toma de decisiones en presencia de adversarios plantea dificultades propias de una situación que

a veces exige recurrir a decisiones subóptimas solo con el objeto de confundir a los adversarios. Este tipo de situaciones surgen claramente en el ámbito militar, pero también en áreas como la vigilancia de perímetros, el desarrollo de juegos de ordenador, el diseño de sistemas inteligentes para el entrenamiento de personal, el cibercrimen, etc. En términos generales, un adversario es una entidad cuyos beneficios (en algún sentido) son inversamente proporcionales a los nuestros. Este adversario es capaz de alterar nuestros beneficios tomando ciertas acciones y, además, puede observar nuestras acciones/decisiones teniendo la oportunidad de aprender nuestro patrón de comportamiento. Este aprendizaje le llevará a ser más efectivo en su intento de maximizar sus beneficios y minimizar los nuestros. Este contexto es diferente del contexto de concurrencia, ya que en el primero no sabemos si existe o no adversario y, si existe, no conocemos nada acerca de él, mientras que en el segundo se sabe que hay un oponente que compite contra nosotros por disminuir nuestras ganancias.

- Contexto de crisis, que se presenta cuando se dan circunstancias excepcionales (por ejemplo, catástrofes, accidentes...) y en el que hay que tomar la mejor decisión posible entre las disponibles, que no suelen ser todas las posibles. En ocasiones, esa mejor decisión entre las disponibles puede coincidir con la solución óptima del problema. Sin embargo, en la mayoría de los casos no será así, debido a diversos factores como puede ser la falta de recursos para explorar todo el espacio de alternativas, la posible desaparición de alternativas, la infactibilidad súbita de algunas otras, etc. En estos casos, una buena estrategia de solución, inspirada en el diseño de algoritmos de precondicionamiento, puede ser la protocolización del problema, de manera que cuando surja la emergencia se pueda consultar un protocolo de actuación que minimice los riesgos de una mala actuación al máximo, pudiendo aumentar la posibilidad de que coincidan la solución al problema en casos de emergencia y la solución óptima del problema sin emergencias.
- Contexto de sostenibilidad, asociado a lo que se entiende por decisiones sostenibles en un ecosistema concreto. De forma paralela a como se define "desarrollo sostenible" (United Nations, 1987) para que una decisión sea sostenible ha de satisfacer las expectativas del momento en que se toma, es decir, ser óptima en algún sentido que establece el decisor, y a la vez no comprometer las elecciones que sobre el problema en cuestión se puedan tomar en el futuro. Por tanto, tiene perfecto sentido que consideremos los contextos para tomar decisiones que, ajustándose a las necesidades del problema en cuestión, permitan poder volver a resolver el mismo problema cuando se vuelva a presentar, sin estar condicionados por las decisiones precedentes. Este contexto, aunque generalmente se asocia a cuestiones relativas al medio ambiente, no está limitado a ese entorno. La adquisición "ocasional" de equipamientos sin un mínimo análisis de su sostenibilidad,

aun siendo el resultado de un proceso de toma de decisiones perfectamente desarrollado, con más frecuencia de la deseada produce resultados indeseados que, al final, demuestran que la toma de decisiones se efectuó deficientemente. Por otro lado, y aunque una conducta sostenible es siempre éticamente plausible, este contexto no se parece al anterior contexto ético, puesto que este último está más orientado a cuestiones de conflictos de intereses y de carácter moral.

- Contexto dinámico, en el que en el momento en que se toma la mejor decisión posible, las condiciones que han llevado a esa decisión cambian y pueden provocar que aquella primera decisión óptima ya no lo sea. Un ejemplo sencillo explica estas situaciones en las que la mejor solución puede cambiar conforme se va desarrollando el proceso de toma de decisiones. Queremos comprar cierto complemento por internet. Encontramos un modelo que nos satisface y que es nuestra mejor opción (eventualmente puede ser una oferta temporal). Ejecutamos todos los trámites que nos solicitan y abonamos el importe. Sin embargo, con más frecuencia de la que cabría esperar, al poco tiempo se nos informa desde el e-proveedor que el complemento solicitado no se nos puede enviar porque “se agotó” el inventario (fruto de una avalancha masiva de compradores que de forma independientemente, pero simultánea, optaron por el complemento en cuestión). Este contexto es propio de problemas de transporte, de gestión o inversiones (Yankelevich, 2017), y se da cada vez más en las redes sociales.
- Contexto de responsabilidad social corporativa (RSC), entendida como una forma de dirigir las empresas, basada en la gestión de los impactos que su actividad genera sobre sus clientes, empleados, accionistas, comunidades locales, medio ambiente y sobre la sociedad en general (Observatorio de RSC, s. f.). En definitiva, la RSC es un concepto a través del cual la empresa integra de forma voluntaria las dimensiones social y ambiental en sus operaciones de negocio y en sus relaciones con los grupos de interés, que es de plena actualidad y que puede y debe modificar los cursos de acción de los procesos de decisión. El contexto que define la RSC no es el mismo que el ético, aunque puntualmente puedan tener alguna similitud, ya que este último se enfoca más en las consecuencias morales de las decisiones. La toma de decisiones en un contexto de RSC, sobre todo en el caso de corporaciones públicas, está condicionada por lo que se denomina la “rendición de cuentas” que, en todos los casos, puede cambiar la selección de la mejor acción que se escogería de no contemplar este contexto de RSC.
- Contexto de estrés, que surge como consecuencia de que hoy en día encontrar un punto de equilibrio entre el trabajo y la familia, que concilie ambos mundos, cada vez se está volviendo más complicado. Encontrar tiempo de “desconexión” en el que estemos tranquilos es difícil y las exigencias en el trabajo son abrumadoras. Así, la vida se está volviendo más estresante y con sucesivamente el estrés se ha convertido en uno de los

factores que los decisores deben tener en cuenta en la mayoría de las situaciones que suponen cierta criticidad. En los negocios, el estrés puede ser perjudicial para el éxito de los gerentes al tomar decisiones clave. Tomar decisiones estratégicas es el componente más crítico del trabajo de un ejecutivo y aunque las decisiones ejecutivas generalmente tienen consecuencias muy importantes, los ejecutivos deben tomar decisiones de gran repercusión independientemente de las situaciones y condiciones que asumen, sabiendo que los ejecutivos que toman decisiones bajo estrés en un tiempo limitado y recursos o incertidumbre pueden verse obligados a reducir sus alternativas (Hejase, Hamdar, Hashem, & Bou, 2017). La toma de decisiones bajo estrés puede tener consecuencias desastrosas.

Como es obvio, cada uno de estos contextos dependerá de cada situación concreta, por lo que no es fácil profundizar mucho más en cada uno de ellos. Pero en todos los casos, e independientemente de que dispongamos de información completa o no, estos contextos están definidos por reglas, i. e. por predicados lógicos, que describen cada situación analizada y permiten su modelización con relativa facilidad. Por tanto, la relevancia que pueden tener en el diseño, construcción y desarrollo de los SAuD es más que patente y, debido a las consecuencias indeseadas que puede tener tomar una decisión en un contexto inadecuado, especialmente cuando estemos trabajando en un contexto ético, su análisis y modelización se han convertido en una línea prioritaria de investigación en IA a la que se le está dedicando mucho esfuerzo y atención en todo el mundo. La siguiente sección la dedicamos a describir como se está abordando en el ámbito europeo la implementación de comportamientos éticos en los SAuD.

Ética y SAuD

Como ya se ha apuntado, el desempeño práctico de los SAuD, es decir, su operatividad en los diferentes entornos del mundo laboral real, ha de ser muy cuidadoso ya que cuando estos sistemas se gestionan con técnicas de automatización inteligente, pueden llegar a sustituir la capacidad de actuar y de tomar decisiones de las personas, lo que no es precisamente fácil de asumir según los campos de aplicación en los que se produzca, sobre todo, si se hace de forma masiva o puramente comercial. El tema es trascendente y acapara gran atención en todo el mundo, tanto desde el punto de vista, podríamos decir, académico (Baum, 2020; Goria, 2020; Hickok, 2021; Nath & Sahu, 2020), como práctico, que es en el que nos vamos a concentrar en lo que sigue.

Informatics Europe es una organización privada que representa ante la Unión Europea a la comunidad investigadora y académica en informática de Europa. Esta institución, junto con el ACM Europe Council, que persigue incrementar en Europa el nivel y visibilidad de las actividades de la Association for Computing Machinery (ACM), y el ACM Europe Policy Committee (EUACM), publicaron en el 2018 un informe que recoge las medidas que se habrán de tener en cuenta para lograr un desarrollo equilibrado y eficaz de los

SAuD en la sociedad (Informatics Europe, 2018), cuyas principales líneas recogemos a continuación.

El punto de partida de dicho informe es que como en la práctica sería peligroso responsabilizar de las cuestiones éticas relacionadas con los SAuD a comités de expertos o a la industria implicada en cada caso de uso, ya que principalmente lo que se requiere es una profunda comprensión e incorporación de la ética en todo el diseño de la tecnología, los valores sociales y morales no deberían ser vistos como simples "factores de riesgo" o restricciones, sino como los principales impulsores y moldeadores de la innovación y, por tanto, deben estar incorporados en los SAuD desde el primer momento de su concepción. Para ello se hacen una serie de recomendaciones de tipo técnico, ético, legal, social, económico y educativo, que de forma resumida exponemos a continuación.

a) Recomendaciones de tipo técnico

1. Establecer medios, medidas y estándares para asegurar que los SAuD sean objetivos. Todos los actores clave (instituciones gubernamentales, academia, industria, instituciones internacionales, ONG y ciudadanos) deben estar involucrados en la formulación de normas y prácticas que aseguren el bien público como primer criterio que ha de conducir el diseño y la construcción de los SAuD. Estas normas deben estar formuladas de manera flexible para que perduren ante la rápida evolución de la tecnología y de las aplicaciones industriales de los SAuD. Para facilitar este objetivo, es necesario incentivar la investigación en IA de cara a desarrollar una base teórica sólida sobre la toma automatizada de decisiones.

b) Recomendaciones de tipo ético

2. Asegurar que la ética se mantenga a la vanguardia del desarrollo y la implementación de SAuD y que sea parte integral de esta. Al igual que con la salud y la biología, los países miembros y la Unión Europea deberían desarrollar comités de ética para asesorar a las organizaciones sociales, políticas, académicas y legales sobre las consecuencias positivas y negativas de las iniciativas, herramientas y sistemas relativos a los SAuD. También, como garante del interés público, debería crearse una nueva agencia (europea) que supervise el desarrollo y despliegue de los SAuD en toda Europa.
3. Promover el diseño de SAuD sensibles a valores. En la educación superior deberían diseñarse programas especiales sobre técnicas sensibles a los valores, resaltando que los valores sociales y las prioridades éticas de los usuarios deben tenerse en cuenta en todos los aspectos y elementos asociados a un SAuD.

c) Recomendaciones de tipo legal

4. Definir claramente la responsabilidad legal del uso e impacto de los SAuD. Los principios básicos que gobiernan

actualmente el desarrollo de SAuD desde el punto de vista informático-profesional deben ser la base de un amplio debate entre expertos legales y técnicos, medios de comunicación y sociedad en busca de nuevas normas legales para gobernar la implementación masiva de SAuD. En particular, hay que volver a considerar el descargo general de responsabilidades que tiene casi todo el software actual, y revisarse o rechazarse si, como parece, no es aplicable a muchos de los actuales o futuros usos de los SAuD. La agencia (europea) propuesta en la segunda recomendación debería fomentar y facilitar este debate, así como proponer una legislación adecuada.

d) Recomendaciones de tipo económico

5. Garantizar que las consecuencias económicas de la adopción de SAuD sean siempre consideradas en su totalidad. Entre sus primeras iniciativas oficiales, con el fin de emitir directrices y reglamentos apropiados, la nueva agencia propuesta debería comenzar emitiendo informes sobre una serie de problemas concretos de índole económica y socioeconómica, a los que el desarrollo y la acelerada aplicación de los SAuD probablemente den lugar.

Se debe reconocer explícitamente que la misión de la referida agencia siempre se orientará a dos objetivos inherentemente interrelacionados: fomentar la evolución y el uso responsable de los SAuD y minimizar sus posibles disruptores de tipo personal, social y económico sobre los individuos y las naciones. En este sentido, un aspecto fundamental que se debe tener en cuenta es el de la sostenibilidad de los SAuD.

Como se sabe, la industria de la IA a menudo se compara con la industria petrolera, ya que una vez extraídos y refinados los datos, como pasa con el petróleo, pueden ser un producto altamente lucrativo. Así, igual que ocurre con los combustibles fósiles, el proceso de entrenamiento de los algoritmos que subyacen en los SAuD tiene un impacto ambiental descomunal.

En efecto, Strubell, Ganesh y McCallum (2019) muestran una evaluación del consumo energético necesario para entrenar varios modelos de gran dimensión habituales en IA descubriendo que el proceso puede contaminar el equivalente a más de cinco veces el CO₂ que emite un coche de gama media a lo largo de toda su vida útil (incluyendo la fabricación en sí), por no comparar con las emisiones que se derivan del viaje de ida y vuelta en avión de un pasajero entre dos aeropuertos estadounidenses (NYC y SFO), lo que sin duda demuestra un impacto ambiental muy sustancial (figura 3):

A este dato también habría que sumar los perjuicios causados por los residuos que generamos, la denominada basura digital, es decir, los correos electrónicos, los CD, los videos, los audios, etc., que abandonamos de cualquier forma todos los días y que exigen cada vez mayores servidores para su almacenamiento, con el consiguiente consumo energético. De acuerdo

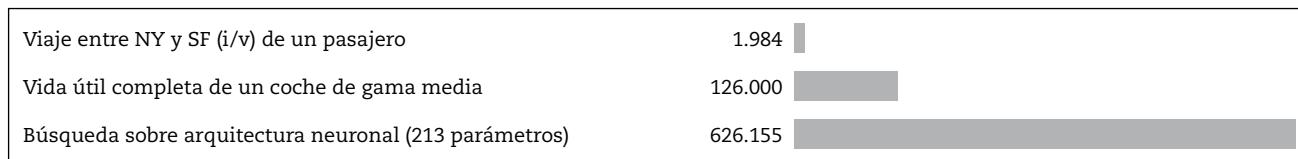


Figura 3. Equivalente a consumo en libras de CO₂

Fuente: elaboración propia a partir de Strubell, Ganesh y McCallum (2019).

con Schwartz, Dodge, Smith y Etzioni (2019) y con informes realizados por Greenpeace, disponibles en su página web, la electricidad consumida globalmente por el almacenamiento de este tipo de material digital aumenta sin cesar, pasando desde los 632 billones de kilovatios del 2007 hasta los 1963 billones previstos para el 2020, lo que, equivalente al CO₂ emitido, supone 1034 megatonnes (1034 millones de toneladas).

e) Recomendaciones de tipo social

6. Imponer legalmente que se informe con claridad a los usuarios de los SAuD sobre todas las prácticas de privacidad y adquisición de datos de sus implementadores. El aprendizaje automático funciona a partir de datos y por tanto cuando y donde se recopile la información, lo que se recopile y los usos que se le dará, deben ser descritos obligatoriamente por el proveedor de los datos de manera concisa y clara.

Este es un aspecto fundamental ya que, en una época de captura masiva y generalizada de datos, el derecho a la protección de la información personal y el derecho al respeto a la privacidad son un desafío crucial. Tanto los SAuD como los programas con IA que operan a través de la web, deben cumplir con las leyes sobre protección de datos y no recopilarlos, ni difundirlos, ni ejecutarlos sobre conjuntos de datos para cuyo uso y difusión no se haya dado consentimiento. Los sistemas con IA no deben interferir con el derecho a la vida privada, que comprende el derecho a estar libre de tecnologías que influyan en el desarrollo y las opiniones personales, el derecho a establecer y desarrollar relaciones con otros seres humanos y el derecho a estar libres de vigilancia. En este sentido, podría ser oportuno considerar nuevos derechos, como el derecho a un contacto humano significativo y a no usar los perfiles personales ni a ser medido, analizado, entrenado o persuadido.

7. Aumentar significativamente la financiación pública para la investigación no comercial relacionada con SAuD. Es necesario incentivar la investigación dirigida a comprender mejor el aprendizaje automático y su uso en sistemas que puedan influir en el comportamiento humano. Quedan por investigar muchas cuestiones fundamentales, pero el conocimiento público y riguroso de los resultados logrados por estas técnicas, no exclusivamente dependientes de la industria, ha de ser un requisito previo para un posterior debate sobre su aceptabilidad y adopción efectiva por parte de las empresas europeas.

f) Recomendaciones de tipo educativo

8. Fomentar la formación técnica universitaria relacionada con la IA. Todos los estudiantes universitarios deberían recibir formación sobre los aspectos prácticos y el potencial de la IA. Los estudiantes de todas las disciplinas deben ser conscientes del impacto que esta tecnología tendrá en su campo y en su futuro trabajo.
9. Complementar la formación técnica con formación sociohumanista del mismo nivel. Debido al impacto cada vez mayor que la tecnología tendrá en la sociedad, los currículos técnicos también deberían formar a los estudiantes para enfrentar escenarios complejos al complementar las habilidades técnicas con el desarrollo del pensamiento crítico, la formación digital y el juicio ético. Los planes de estudio de educación superior deberían fomentar los estudios interdisciplinares, basados en el patrimonio cultural europeo, tanto en las disciplinas científicas como en las artes liberales. También debería incluirse en los planes de estudio de secundaria una introducción a la IA y a los problemas que plantean.
10. Aumentar la conciencia y comprensión del público sobre los SAuD y sus impactos. Existe una clara necesidad de formar a la sociedad sobre esta tecnología, ya que se está introduciendo rápidamente y nos afectará prácticamente a todos en nuestras vidas profesionales y privadas. Como la mayoría de las personas no siguen cursos adicionales después de completar su formación, los medios de comunicación públicos representan los medios *de facto* más adecuados para formar a la población en general. En consecuencia, los profesionales informáticos y los responsables de las políticas tecnológicas deben actuar coordinadamente con la prensa para transmitir la información a la que se refieren las recomendaciones recogidas aquí. Se debe prestar la debida atención al uso preocupante de las técnicas de IA para influir en la opinión pública.

Estas recomendaciones que se recogen en Informatics Europe (2018) no son las únicas que hay al respecto; se destacan también las publicadas en European Group on Ethics in Science and New Technologies: Statement on A. I. (2018) y en Futureoflife (2017), en cuya descripción y detalle no entramos porque sobrepasaríamos los límites razonables de una contribución como esta. En cualquier caso, una panorámica muy completa sobre cómo se está trabajando este tema en todo el mundo puede verse en Dutton (2019).

Conclusiones

Este trabajo se dedicó principalmente a profundizar en algunos aspectos importantes relacionados con los sistemas basados en IA, a los que nos hemos referido por sistemas autónomos de decisión (SAuD). Entre ellos se destacan:

a) La necesidad de incorporar a los elementos que definen un problema de decisión, y más generalmente a un proceso de decisión, la naturaleza de la información disponible, que no necesariamente ha de ser de tipo probabilístico, y el contexto en el que se ha de desarrollar el proceso de decisión considerado. De hecho, se han introducido nuevos contextos, no contemplados hasta el momento, que habría que tener en cuenta para poder elegir el mejor curso de acción en cualquier proceso de decisión.

A partir de ese nuevo concepto de proceso de decisión son innumerables las posibilidades que se abren para abordar, con seguridad para las personas y confianza y rigor sobre los resultados, el desempeño de los SAuD. Por ejemplo, uno de los ámbitos en los que los SAuD están llamados a desempeñar un papel protagonista es en el de la optimización de recursos asociados a los nuevos modelos de transporte, a la sostenibilidad de diferentes y variados ecosistemas, al cuidado del medio ambiente, etc., ya que cada vez tienen más parámetros, mayores dimensiones y necesitan algoritmos de resolución de eficiencia creciente, que actúen y reproduzcan la forma que los expertos emplean para resolver los problemas que se les plantean en la práctica, es decir, con información insegura, considerando diferentes objetivos, con varios tipos de datos, etc.

b) La importancia de contemplar un marco ético y legal común y, por tanto reconocido internacionalmente, para el diseño, producción, uso y gobierno de los SAuD, que garantice el derecho a la protección de la información personal y el derecho al respeto a la privacidad, así como que concrete dónde recae la responsabilidad legal del uso e impacto de los SAuD, mediante nuevas normas legales que gobiernen la implementación masiva de estos sistemas y su aplicación. Especial atención habrá que poner en los sistemas diseñados para automejorarse o autorreplicarse recursivamente, pudiendo producir un aumento rápido de su calidad o cantidad, ya que inexcusablemente tendrán que estar sometidos a estrictas medidas de seguridad y control.

Financiación

Este artículo fue parcialmente financiado por el proyecto TIN2017-86647-P, del Ministerio de Economía y Competitividad del Gobierno de España, incluyendo fondos Feder de la Unión Europea.

REFERENCIAS

- Baum, S. D. (2020). Social choice ethics in artificial intelligence. *Artificial Intelligence and Society*, 35(1), 165-176.
- Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (2013). *Nonlinear programming: Theory and algorithms*. John Wiley & Sons.
- Bellman, R. E., & Zadeh, L. A. (1970). Decision-making in a fuzzy environment. *Management Science*, 1, B-141-B-164.
- Dutton, T. (2019). An overview of national AI strategies. <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>
- ENIA. (2020). Estrategia nacional de inteligencia artificial. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIAResumen2B.pdf>
- European Group on Ethics in Science and New Technologies: Statement on A. I. (2018). *Robotics and Autonomous Systems*. European Commission. Directorate-General for Research and Innovation Unit RTD.01
- Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, 5, 5-15 (1967)
- Futureoflife. (2017). Asilomar principles. <https://futureoflife.org/ai-principles>
- Govia, L. (2020). Coproduction, ethics and artificial intelligence: A perspective from cultural anthropology. *Journal of Digital Social Research*, 2(3), 42-64.
- Hejase, H. J., Hamdar, B., Hashem, F., & Bou, S. R. (2017). Decision making under stress: An exploratory study in Lebanon. *Journal of Middle East and North Africa Sciences (MENA Sci.)*, 3(12), 1-16.
- Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *Artificial Intelligence and Ethics*, 1(1), 41-47.
- Informatics Europe. (2018). When computers decide: European recommendations on machine-learned automated decision making. Informatics Europe & EUACM. <https://www.informatics-europe.org/publications.html>
- Lamata, M. T., Pelta, D. A., & Verdegay, J. L. (2021). The role of the context in decision and optimization problems. En M.-J. Lesot & C. Marsala (eds.), *Fuzzy Approaches for soft computing and approximate reasoning: Theories and applications. studies in fuzziness and soft computing* (pp. 75-84). Springer Cham. https://doi.org/10.1007/978-3-030-54341-9_7
- Lamata, M. T., Pelta, D. A., & Verdegay, J. L. (2018). Fuzzy information and contexts for designing automatic decision-making systems. En *Advances in artificial intelligence. Lecture notes in artificial intelligence* (pp. 174-183) 11160. Springer Cham 174-183. https://doi.org/10.1007/978-3-030-00374-6_17
- Lindley, D. V. (1971). *Making decisions*. John Wiley & Sons Ltd.
- Ministerio de Asuntos Económicos y Transformación Digital. (2020a). Plan España Digital 2025. https://portal.mineco.gob.es/RecursosArticulo/mineco/prensa/ficheros/noticias/2018/Agenda_Digital_2025.pdf
- Ministerio de Asuntos Económicos y Transformación Digital. (2020b). Plan Nacional de Competencias Digitales. https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/210127_plan_nacional_de_competencias_digitales.pdf
- Ministerio de Ciencia, Innovación y Universidades. (2019). Estrategia Española de I+D+I en Inteligencia Artificial. https://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf
- Nath, R., & Sahu, V. (2020). The problem of machine ethics in artificial intelligence. *Artificial Intelligence and Society*, 35(1), 103-111.
- Observatorio de RSC. (s. f.). *La responsabilidad social corporativa (RSC)*. <http://observatoriorsc.org/la-rsc-que-es/>
- Ríos, S. (1976). *Ánalisis de decisiones*. ICE Ediciones.
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2019). Allen Institute for AI, Seattle, Washington, USA. Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. University of Washington, Seattle, Washington, USA July 2019. arXiv preprint arXiv:1907.10597

- Simon, H. (1960). *The new science of management decision*. Harper & Brothers.
- Smithson, M. (1989). *Ignorance and uncertainty: Emerging paradigms*. Springer Verlag.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and policy considerations for deep learning in NLP*. En 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence, Italy. (arXiv:1906.02243)
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- United Nations. (1987). *Our common future: Brundtland Report*. 20 March 1987.
- Verdegay, J. L. (2019). ¿En qué piensan los algoritmos? Discurso de Apertura del Curso Académico 2019-2020 de la Universidad de Granada. Granada. <https://digibug.ugr.es/handle/10481/57034>
- Verdegay, J. L. (2020). Consideraciones éticas sobre el diseño, desarrollo y aplicación de los sistemas autónomos de decisión. *Revista Cubana de Transformación Digital y Cibersociedad*, 1(1), 75-95. <https://rctd.uic.cu/rctd/article/view/6>
- Yankelevich, D. (2017). *Prediciendo un futuro predicho*. <https://www.linkedin.com/pulse/prediciendo-un-futuro-predicho-daniel-yankelevich/?originalSubdomain=es>
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.