IA confiable del Instituto de Ingeniería Eléctrica y Electrónica | Reliable AI of the Institute of Electrical and Electronic Engineering

4.5.5º.- IA confiable de [IEEE](#) | [IEEE Reliable AI](#)

<https://ethicsinaction.ieee.org/>

↩ Ir al índice principal del libro	↩ Go to the main index of the book
Contenidos	Contents
4.5.1º.- Introducción a la IA empresarial	4.5.1º.- Introduction to business AI
4.5.2º.- IA confiable de Google	4.5.2º.- Google Reliable AI
4.5.3º.- IA confiable de Microsoft	4.5.3º.- Microsoft Reliable AI
4.5.4º.- IA confiable de IBM	4.5.4º.- IBM Reliable AI
4.5.5º.- IA confiable de IEEE	4.5.5º.- IEEE Reliable AI

Si cabe es más notable denotar la existencia de los estándares desde el P7000 al P7014. (El protocolo P7000 - Proceso modelo para abordar inquietudes éticas durante el diseño del sistema). | [If possible, it is more remarkable to denote the existence of standards from P7000 to P7014. \(The P7000 protocol - Model process to address ethical concerns during system design\).](#)

Principios generales	General principles
1. Derechos humanos: A/ IS se creará y operará para respetar, promover y proteger los derechos humanos internacionalmente reconocidos.	1. Human Rights-A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.
2. Bienestar: los creadores de A/IS adoptarán un mayor bienestar humano como criterio principal de éxito para el desarrollo.	2. Well-being-A/IS creators shall adopt increased human well-being as a primary success criterion for development.
3. Agencia de datos: los creadores de A / IS deben capacitar a las personas con la capacidad de acceder y compartir sus datos de manera segura, para mantener la capacidad de las personas de tener control sobre su identidad.	3. Data Agency-A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity.
4. Eficacia: los creadores y operadores de A/IS deberán proporcionar evidencia de la efectividad y aptitud para el propósito de A / IS.	4. Effectiveness-A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.
5. Transparencia: la base de una decisión particular de A/IS siempre debe ser reconocible.	5. Transparency-The basis of a particular A/IS decision should always be discoverable.
6. Responsabilidad: A/IS se creará y operará para proporcionar una justificación inequívoca de todas las decisiones tomadas.	6. Accountability-A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.
7. Conciencia del mal uso: los creadores de A/IS deben protegerse contra todo mal uso potencial y los riesgos de A/IS en funcionamiento.	7. Awareness of Misuse-A/IS creators shall guard against all potential misuses and risks of A/IS in operation.

8. Competencia: los creadores de A/IS deben especificar y los operadores deben cumplir con el conocimiento y la habilidad necesarios para una operación segura y efectiva.

8. Competence-A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

* Autonomous and Intelligent systems: (A/IS)

4.5.5.1.- Introducción | Introduction

IEEE es la organización profesional técnica más grande del mundo dedicada al avance de la tecnología en beneficio de la humanidad. IEEE y sus miembros inspiran a una comunidad global a innovar para un mejor mañana a través de sus más de 423,000 miembros en más de 160 países, y sus publicaciones, conferencias, estándares tecnológicos y actividades profesionales y educativas altamente citadas. Es la 'voz' de confianza para la información de ingeniería, informática y tecnología en todo el mundo^{C4.5.5-1}. Su publicación La biblioteca **digital IEEE Xplore** es un poderoso recurso para el descubrimiento de contenido científico y técnico publicado por el IEEE (Instituto de Ingenieros Eléctricos y Electrónicos) y sus socios editoriales. IEEE Xplore proporciona acceso web a más de cuatro millones de documentos de texto completo de algunas de las publicaciones más citadas del mundo en ingeniería eléctrica, informática y electrónica.

[English]

IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity. IEEE and its members inspire a global community to innovate for a better tomorrow through its more than 423,000 members in over 160 countries, and its highly cited publications, conferences, technology standards, and professional and educational activities. IEEE is the trusted "voice" for engineering, computing, and technology information around the globe^{C4.5.5-1}. The **IEEE Xplore digital library** is a powerful resource for discovery of scientific and technical content published by the IEEE (Institute of Electrical and Electronics Engineers) and its publishing partners. IEEE Xplore provides web access to more than four-million full-text documents from some of the world's most highly-cited publications in electrical engineering, computer science, and electronics.

Uno de los elementos notables del que podemos hacer uso es de su glosario técnico (en inglés). Cuya finalidad es: "El propósito de este glosario es proporcionar a los equipos interdisciplinarios un recurso compartido para referenciar términos que pueden tener significados específicos de la disciplina. Dentro de este documento hay 6 definiciones dadas para la mayoría de los términos. En aquellos casos en los que no se dan 6 definiciones completas para cada término, no se pudo encontrar una definición adecuada dentro de la disciplina examinada. Las 6 categorías de disciplinas de las cuales se extrajeron las definiciones incluyen: lenguaje ordinario; disciplinas computacionales (por ejemplo, matemática y estadística); economía y ciencias sociales; disciplinas de ingeniería; filosofía y ética y derecho y política internacional".

Del que vamos a aprovechar qué definición nos entrega de '**Inteligencia Artificial**'.

[English]

The purpose of this glossary is to give interdisciplinary teams a shared resource for reference to terms which may have meanings that are discipline specific. Within this document there are 6 definitions given for most terms. In those cases where 6 full definitions are not given for each term a suitable definition within the discipline examined could not be found. The 6 categories of disciplines from which definition were draws include: ordinary language; computational disciplines (e.g., mathematics and statistics); economics and social sciences; engineering disciplines; philosophy and ethics; and international law and policy.

From which we will take advantage of what definition gives us '**Artificial Intelligence**'.

Fig. I A. C4.5.5-2. Glosario IEEE, Ética de Sistemas Autónomos e Inteligentes | [IEEE Glossary, Ethics of Autonomous and Intelligent Systems](#). Crédito imág (IEEE). URL:

https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e_glossary.pdf

-
- **TÉRMINO** | **TERM**: INTELIGENCIA ARTIFICIAL | **ARTIFICIAL INTELLIGENCE**
 - **LENGUAJE ORDINARIO** | **ORDINARY LANGUAGE**: "La capacidad de las computadoras u otras máquinas para exhibir o simular comportamientos inteligentes" (OED) | ["The capacity of computers or other machines to exhibit or simulate intelligent behaviour"](#) (OED)
 - **DISCIPLINAS COMPUTACIONALES** | **COMPUTATIONAL DISCIPLINES**:

"La IA será un programa que en un mundo arbitrario no será peor que un humano" (Dobrev 2004, 2); "La inteligencia artificial es la empresa de construir un sistema de símbolos que pueda pasar la prueba de Turing de manera confiable" (Ginsberg 2012, 9); Ver Figura 1.1 Russell y Norvig (1995

página 5). “La inteligencia artificial es un campo de la ciencia de la computación relacionada con la comprensión computacional de lo que comúnmente se llama comportamiento inteligente y con la creación de artefactos que exhiben tal comportamiento. Esta definición puede examinarse más de cerca al considerar el campo desde tres puntos de vista: psicología computacional (cuyo objetivo es comprender el comportamiento humano inteligente creando programas informáticos que se comporten de la misma manera que las personas), computacional filosofía (cuyo objetivo es formar una comprensión computacional del comportamiento inteligente a nivel humano, sin limitarse a los algoritmos y estructuras de datos que la mente humana realmente usa), y la inteligencia de la máquina (cuyo objetivo es expandir el frente de lo que sabemos programar ”(Reilly 2004, 40-41).

[English]

“AI will be such a program which in an arbitrary world will cope not worse than a human” (Dobrev 2004, 2); “Artificial intelligence is the enterprise of constructing a symbol system that can reliably pass the Turing test” (Ginsberg 2012, 9); See Figure 1.1 Russell and Norvig (1995 page 5). “Artificial intelligence is a field of computer science concerned with the computational understanding of what is commonly called intelligent behavior and with the creation of artifacts that exhibit such behavior. This definition may be examined more closely by considering the field from three points of view: computational psychology (the goal of which is to understand human intelligent behavior by creating computer programs that behave in the same way that people do), computational philosophy (the goal of which is to form a computational understanding of human-level intelligent behavior, without being restricted to the algorithms and data structures that the human mind actually does use), and machine intelligence (the goal of which is to expand the frontier of what we know how to program” (Reilly 2004, 40-41).

- **INGENIERIA | ENGINEERING:**

La ingeniería de inteligencia artificial se ha comparado con la ingeniería del conocimiento. Un "diseño de sistema basado en el conocimiento" de IA abarca 3 niveles: la vista del 'nivel de conocimiento' de un sistema basado en el conocimiento describe el conocimiento que utiliza e integra ese sistema. La vista de 'nivel de algoritmo' describe el sistema como un algoritmo de búsqueda, configurado a partir de tipos de componentes estándar (por ejemplo, generadores, probadores, parches, propagadores de restricciones, revisores de creencias, etc.). La vista de 'nivel de programa' expresa el sistema en términos de los elementos de los paradigmas de programación existentes (reglas, objetos, procedimientos, etc.) (Tong y Sriram 2012, 8-9)

[English]

Artificial intelligence engineering has been compared to knowledge engineering. A “knowledge based system design” of AI encompasses 3 levels: “the ‘knowledge level’ view of a knowledge-based system describes the knowledge that is used and embedded in that system. The ‘algorithm level’ view describes the system as a search algorithm configured out of component types (e.g., generators, testers, patchers, constraint propagators, belief revisers, etc). The ‘program level’ expresses the system in terms of the elements of existing programming paradigms (rules, objects, procedure calls)” (Tong and Sriram 2012, 8-9)

• ECONOMIA Y CIENCIAS SOCIALES | ECONOMICS & SOCIAL SCIENCES:

"Los enfoques de IA se pueden dividir en "IA estrecha" y "IA general". Los sistemas de IA estrecha realizan tareas individuales en dominios especializados y bien definidos, como el reconocimiento de voz, el reconocimiento de imágenes y la traducción. En contraste, el objetivo a largo plazo de la IA general es crear sistemas que exhiban la flexibilidad y versatilidad de la inteligencia humana en una amplia gama de dominios cognitivos, que incluyen el aprendizaje de lenguaje, la percepción, el razonamiento, la creatividad y la planificación" (NITRD 2016, 19).

[English]

"AI approaches can be divided into "narrow AI" and "general AI." Narrow AI systems perform individual tasks in specialized domains, such as speech recognition, image recognition and translation. In contrast, the long-term goal of general AI is to create systems that exhibit the flexibility and versatility of human intelligence in a broad range of cognitive tasks, including learning, language, perception, reasoning, creativity, and planning" (NITRD 2016, 19)

• ÉTICA Y FILOSOFÍA | ETHICS & PHILOSOPHY:

"Diremos que una entidad es inteligente si tiene un modelo adecuado del mundo (incluido el mundo intelectual y simbólico, la comprensión de sus propios objetivos y otros procesos mentales), si es lo suficientemente inteligente como para responder a una amplia variedad de preguntas sobre la base de este modelo, si puede obtener información adicional del mundo externo cuando sea necesario, y puede realizar tareas en el mundo externo según lo exijan sus objetivos y sus capacidades físicas" (McCarthy y Hayes 1969, 4)

[English]

"We shall say that an entity is intelligent if it has an adequate model of the world (including the intellectual world of man, understanding of its own goals and other mental processes), if it is clever enough to answer a wide variety of questions on the basis of this model, if it can get additional information from the external world when required, and can perform tasks in the external world as its goals demand and its physical abilities permit" (McCarthy and Hayes 1969, 4)

• DERECHO INTERNACIONAL Y POLÍTICA | INTERNATIONAL LAW & POLICY:

"Un subcampo de la informática relacionado con los conceptos y métodos de inferencia simbólica de una computadora y la representación simbólica del conocimiento que se utilizará para hacer inferencias" (INIS); "La rama de la informática relacionada con el desarrollo de máquinas capaces de realizar actividades que normalmente se cree que requieren un tipo de inteligencia humana. Y, adicionalmente, la capacidad de una computadora u otra máquina para realizar actividades que normalmente se considera que requieren un tipo de inteligencia humana." (UNHQ: A / S-15/12 Anexo II, pr. 12 de E, párrafo (re)).

[English]

"A subfield of computer science concerned with the concepts and methods of symbolic inference by a computer and the symbolic representation of the knowledge to be used in making inferences" (INIS);

"The branch of computer science concerned with the development of machines capable of performing activities that are normally thought to require a human type of intelligence. And, additionally, the ability of a computer or other machine to perform activities normally thought to require a human type of intelligence." (UNHQ: A/S-15/12 Annex II, pr 12 of E, paragraph (d))

4.5.5.2.- *DISEÑO ÉTICAMENTE ALINEADO* (Una visión para priorizar el bienestar humano con sistemas autónomos e inteligentes) | *ETHICALLY ALIGNED DESIGN* (A vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems). IEEE.

Este trabajo se publica bajo los auspicios de la Iniciativa Global IEEE sobre Ética de los Sistemas Autónomos e Inteligentes con el propósito de promover la comprensión pública de la importancia de abordar las consideraciones éticas en el diseño de sistemas autónomos e inteligentes.

"Supone un conjunto de principios que ubican a la IA dentro de un marco de derechos humanos con referencias al bienestar, la responsabilidad, la responsabilidad corporativa, el valor por diseño y la IA ética" C4.5.5-2.

[English]

This work is published under the auspices of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems for the purposes of furthering public understanding of the importance of addressing ethical considerations in the design of autonomous and intelligent systems.

"It means a set of principles that place AI within a human rights framework with references to wellbeing, accountability, corporate responsibility, value by design, and ethical AI" C4.5.5-2.

Fig. I A. C4.5.5-3. Web Iniciativa Global IEEE sobre Ética de Sistemas Autónomos e Inteligentes | Web IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Crédito imagen (IEEE). URL: <https://ethicsinaction.ieee.org/>

4.5.5.3. El estándar P7000 | [The IEEE P7000 standard](#)

El estándar P7000 establece un modelo de proceso mediante el cual los ingenieros y tecnólogos pueden abordar la consideración ética a lo largo de las diversas etapas de inicio, análisis y diseño del sistema. Los requisitos de proceso esperados incluyen la visión de gestión e ingeniería del desarrollo de nuevos productos de IT, la ética informática y el diseño del sistema de IT, el diseño sensible al valor y la participación de los interesados en el diseño ético del sistema de IT.

La Asociación de Estándares de IEEE actualmente tiene catorce proyectos de estándares, la serie de estándares **IEEE P7000**, en desarrollo para permitir un nivel elevado de confianza entre las personas y la tecnología que se necesita para su uso fructífero y generalizado en nuestra vida diaria.

A medida que el uso y el impacto de los sistemas autónomos e inteligentes se generalizan, es necesario establecer pautas sociales y políticas para que dichos sistemas permanezcan centrados en el ser humano y sirvan a los valores y principios éticos de la humanidad.

[English]

The P7000 standard establishes a process model by which engineers and technologists can address ethical consideration throughout the various stages of system initiation, analysis and design. Expected process requirements include management and engineering view of new IT product development, computer ethics and IT system design, value-sensitive design, and, stakeholder involvement in ethical IT system design.

The IEEE Standards Association currently has fourteen standards projects, the **IEEE P7000** series of standards, under development to allow for an elevated level of trust between people and technology that is needed for its fruitful, pervasive use in our daily lives.

As the use and impact of autonomous and intelligent systems become pervasive, there is a need to establish societal and policy guidelines in order for such systems to remain human-centric, serving humanity's values and ethical principles.

P70XX: (01,02... 14)

1. Transparencia de sistemas autónomos | [Transparency of Autonomous Systems](#)
2. Proceso de privacidad de datos | [Data Privacy Process](#)
3. Consideraciones de sesgo algorítmico | [Algorithmic Bias Considerations](#)
4. Norma sobre Gobierno de Datos de Niños y Estudiantes | [Standard on Child and Student Data Governance](#)
5. Norma sobre gobierno de datos del empleador | [Standard on Employer Data Governance](#)
6. Grupo de trabajo sobre el agente de AI de datos personales estándar | [Standard on Personal Data AI Agent Working Group](#)
7. Norma ontológica para sistemas de automatización y robótica con ética | [Ontological Standard for Ethically driven Robotics and Automation Systems](#)
8. Estándar para empujar éticamente para sistemas robóticos, inteligentes y autónomos | [Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems](#)
9. Norma para el diseño a prueba de fallos de sistemas autónomos y semiautónomos | [Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems](#)
10. Estándar de métricas de bienestar para inteligencia artificial ética y sistemas autónomos | [Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems](#)
11. Proceso de identificación y calificación de la confiabilidad de las fuentes de noticias | [Process of Identifying and Rating the Trustworthiness of News Sources](#)
12. Términos de privacidad personal legibles por máquina | [Machine Readable Personal Privacy Terms](#)
13. Estándares para la tecnología de análisis facial automatizado | [Standards for Automated Facial Analysis Technology](#)
14. Norma para consideraciones éticas en la empatía emulada en sistemas autónomos e inteligentes

Bibliografía | [Bibliografy](#)

[C4.5.5-1] Web [ieee.org](https://www.ieee.org). [Recuperado (05/09/2019) de: <https://www.ieee.org/about/index.html>]

[C4.5.5-2] Rossi, F. (6.2.2019). *Building Trust In Artificial Intelligence*. Columbia SIPA | *Journal of International Affairs*. [Recuperado (05/09/2019) de: <https://jia.sipa.columbia.edu/building-trust-artificial-intelligence>]

© 2019. Licencia de uso y distribución / [License for use and distribution](#): [[Los estados de la inteligencia artificial \(IA\)](#) | [The states of artificial intelligence \(AI\)](#)] [creative commons CC BY-NC-ND](#) | [ISSN 2695-3803](#) |

- [Notas legales](#) / [Legal notes](#)
 - [Página web de Formaempleo](#) / [Formaempleo website](#)
 - [Formulario de contacto](#) / [Contact Form](#)
-

[Revisión #2](#)

Creado el Sat, Nov 23, 2019 10:05 AM por [Juan Antonio Lloret Egea](#)

Actualizado el Sat, Nov 23, 2019 10:52 AM por [Juan Antonio Lloret Egea](#)