

SELECCIÓN DE ESCRITOS SOBRE INTELIGENCIA ARTIFICIAL

"INTELIGENCIA ARTIFICIAL: ALGUNOS ASPECTOS DE SU IMPACTO"

Centro de Estudios en Tecnologías Inteligentes (CETI)

SERIE CONTRIBUCIONES COMPILADAS N° 8



ANCBA 2023

Ferreri, Juan Carlos

Inteligencia artificial : algunos aspectos de su impacto : serie contribuciones compiladas N° 8 año 2023 / Juan Carlos Ferreri ; compilación de Juan Carlos Ferreri. - 1a ed. - Ciudad Autónoma de Buenos Aires : Academia Nacional de Ciencias de Buenos Aires, 2023.

Libro digital, PDF/A

Archivo Digital: descarga y online

ISBN 978-987-537-180-4

1. Inteligencia Artificial. I. Título.

CDD 006.301

Fecha de catalogación: 12/2023

Academia Nacional de Ciencias de Buenos Aires

Av. Alvear 1711, 3° piso – 1014 Ciudad de Buenos Aires – Argentina

El presente trabajo se encuentra disponible sólo en versión electrónica

www.ciencias.org.ar

correo-e: info@ciencias.org.ar

ISBN 978-987-537-180-4

La publicación de los trabajos de los Académicos y disertantes invitados se realiza bajo el principio de libertad académica y no implica ningún grado de adhesión por parte de otros miembros de la Academia, ni de ésta como entidad colectiva, a las ideas o puntos de vista de los autores.

INDICE

<i>Presentación</i> Juan Carlos Ferreri	1
<i>Aprendizaje Automático en Situaciones Ambiguas</i> Daniela López De Luise	2
<i>Inteligencia Artificial y Ciencia de Datos</i> <i>Panorámica de tecnologías emergentes y disruptivas</i> Luis Joyanes Aguilar	9
<i>Pronóstico de Series Temporales Aplicada A La Predicción De Eventos</i> <i>Climáticos Severos Para La Región Del Valle De Catamarca</i> Julián Antonio Pucheta	59
<i>Migración Inteligente de Procesos para balancear la carga de trabajo en</i> <i>Sistemas Distribuidos</i> David Luis La Red Martínez	67
<i>Ética e Inteligencia Artificial: Encuentro Ineludible</i> Antonio A. Martino	112
<i>Sobre los Autores</i>	128

INTELIGENCIA ARTIFICIAL: ALGUNOS ASPECTOS DE SU IMPACTO

PRESENTACIÓN

Juan Carlos Ferreri

El presente volumen es la compilación de trabajos relacionados con diversos aspectos de la Inteligencia Artificial y sus aplicaciones y que han sido presentados en diversas reuniones en la Academia. En lo posible se ha tratado de evitar redundancias, pero algunas son inevitables por el desarrollo de los razonamientos. El énfasis de este volumen es mayormente técnico. La cuestión ética asociada a las aplicaciones y a los desarrollos está otra vez analizada.

Es importante agradecer a los autores por sus contribuciones y a mis colaboradores, la Lic. Karina Líbano por su trabajo de edición y su permanente colaboración y al Prof. Juan Cejas por la coordinación y colaboración en las presentaciones remotas de algunas de las conferencias.

Resumen

Las producciones más sofisticadas en los sistemas actuales suelen encontrar una limitación severa para resolver problemas con el empleo de la lógica clásica. Es frecuente que estos sistemas caigan en lo que se suele denominar inteligencia computacional, donde las técnicas de aprendizaje automático liberan al humano de la complejidad de establecer categorías con solapamientos parciales. Con estas estrategias son posibles de ser aplicadas en sectores ligados al procesamiento de imágenes, tomas de decisiones y procesamiento de lenguaje natural.

Abstract

The most sophisticated productions in current systems find a hard limitation to solve problems by using classical logic. Frequently, those systems fall in what is known as computational intelligence, where machine learning approaches release human from the complex activity consisting in determining categories with partial overlapping. They are able to be applied in fields related to image processing, decision-making, and natural language processing.

Palabras Clave

Sistemas inteligentes, aprendizaje automático, lógica difusa

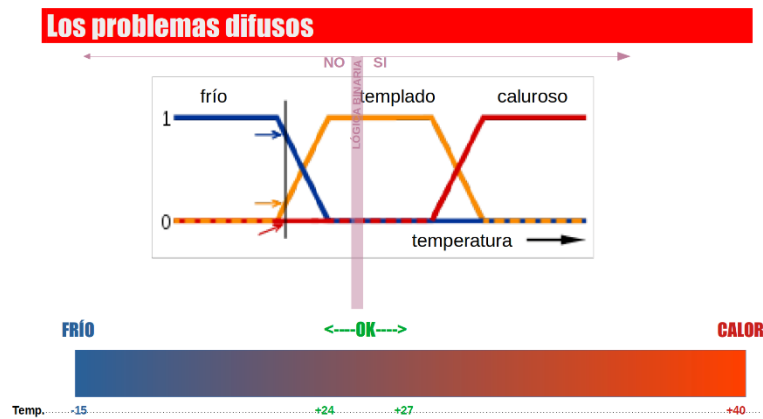
1. El mundo de los problemas difusos

La vida cotidiana enfrenta al individuo con situaciones de las que puede afirmarse que son verdaderas o falsas, y frecuentemente se puede evaluar y “predicar” de ellas uno de dichos estados. Pero la naturaleza humana es compleja y un exquisito caleidoscopio de casos que no encajan en una mera evaluación binaria. Por ejemplo, cuando se trata de sensaciones, opiniones, o evaluaciones subjetivas, las divisiones taxativas se vuelven impracticables o artificios poco útiles para trabajar correctamente. Una de las áreas más afectadas por este problema es el procesamiento de lenguaje.

Palabras como lindo, feo, bueno, malo, etc. tienen una cierta connotación general, pero es imposible determinar exactamente la intensidad de lindo, bueno, feo o malo de una situación. Por el contrario, la sobre simplificación que exige el declarar que es verdadero o falso la afirmación “X es bueno” deja de lado cualquier sutil interpretación, ignorando la amplia gama de connotaciones que puede implicar en cada individuo el considerar que X es bueno. Es por ello que por los años 60 el matemático Lofti Zadeh crea la denominada lógica difusa.

Su característica más importante es que trabaja en un continuo de posibilidades aplicables a X, sin imponer restricciones que no sean las de los hechos a representar. La siguiente figura muestra visualmente la diferencia, entre la lógica clásica y la difusa. Se basa en cómo diferentes personas consideran a la temperatura. Establece tres posibilidades o niveles de verdad de la temperatura: frío, templado y caluroso.

Etiquetas y pertenencias de un sistema difuso.



Nótese que de acuerdo con la gráfica es posible que algo sea frío y templado a la vez pero con distinto nivel de posibilidad. La barra gris muestra el caso de una temperatura que tiene mayor posibilidad de ser catalogada como frío que como templado. Como contraste, la barra rosa indica una frontera rígida entre alternativas de verdad que es propia de la lógica binaria aristotélica. Es fácil apreciar que la sutil diferencia entre la dominancia de templado en cierto rango, y la combinada con alternativas como frío y caluroso, se pierden por completo, queda en su lugar una frontera rígida y abrupta que no representa la blanda frontera de opiniones.

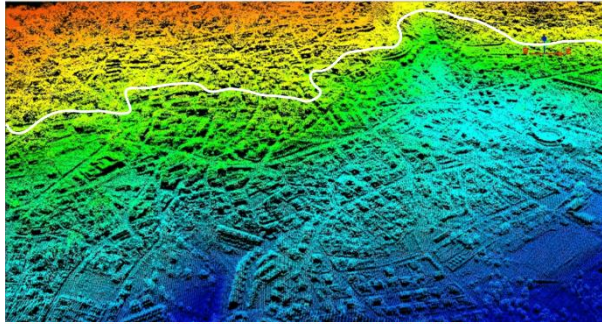
2. El Aprendizaje Automático (AA) como herramienta

Entre las técnicas basadas en sistemas inteligentes se encuentra el denominado Aprendizaje automático (AA), que pretende generar modelos o maquetas prácticas que permitan resolver situaciones específicas. Muchas veces ese aprendizaje se obtiene con algorítmicas que parecen un tanto oscuras pero que indefectiblemente pueden validarse de algún modo.

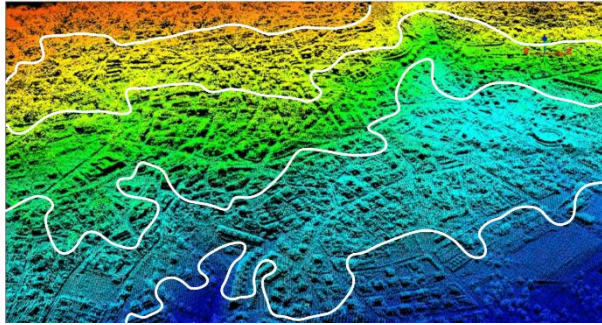
La lógica difusa es una de las muchas herramientas empleadas en este campo, y su aplicación es variada; típicamente en cuestiones relacionadas con encuestas, opiniones, y desempeño del lenguaje natural. En el caso de procesamiento de imágenes puede aplicarse para detectar regiones de interés.

La figura siguiente muestra cómo separa regiones la lógica binaria (basada en verdad o falsedad) y la denominada lógica multivaluada (una extensión con una cantidad finita y numerable de estados de verdad).

AA: uso de lógica binaria para clasificar

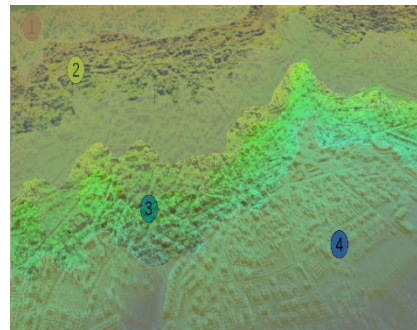
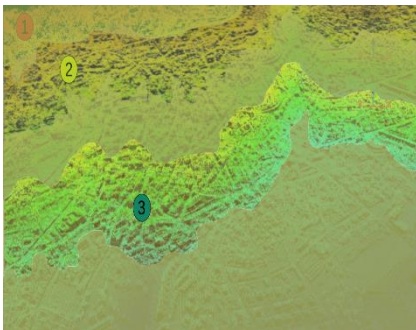
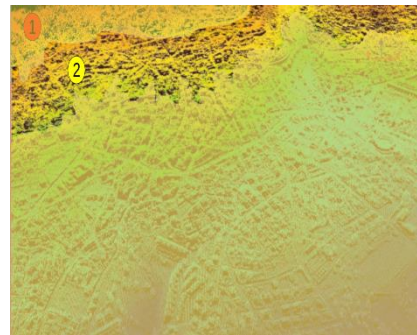
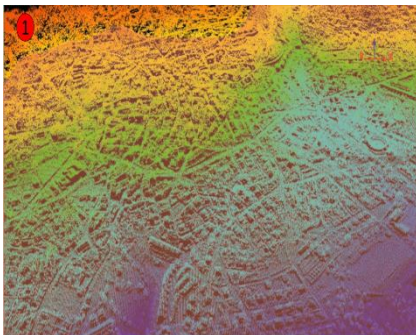


AA: uso de lógica binaria y multivaluada para clasificar



Créditos: Nube de puntos LIDAR clasificada por su altura (Núcleo urbano de Miraflores de la Sierra). ©Instituto Geográfico Nacional de España.

Para el caso de la lógica difusa, la selección de sectores o segmentos se transforma en la disposición solapas con un tope en la zona central de la región en cuestión, decreciendo el nivel hasta cero en el centro de la zona adyacente. En la figura que sigue se ven las solapas progresivamente desde la zona 1 (en rojo) hasta la zona 4 (en azul).



En la última versión de la imagen segmentada (la que tiene cuatro etiquetas), se puede imaginar que cada punto contiene la superposición de todas las posibilidades. Cada zona tiene en su centro el valor máximo de posibilidad y una pendiente hasta cero al acabar toda posibilidad de pertenencia. Al comportamiento mencionado se lo representa con una función de pertenencia $\mu_i(x)$, con $i \in [1,2,3,4]$ una de las cuatro etiquetas.

3. El Aprendizaje Automático (AA) como herramienta

En la lógica clásica se supone que dado un predicado P, por ejemplo P = es bueno y una variable x, es posible aplicar P(x). Por caso P(Juan) se interpreta como el predicado “Juan es bueno”. Según la lógica aristotélica luego es posible afirmar si P(Juan) es verdad o falso.

Lofti Zadeh fue quien en 1965 introdujo un manejo más flexible de la noción de verdad y la cambia por posibilidad representada por una función $\mu_L(x)$ (o funciones de pertenencia). Un ejemplo sería considerar las posibilidades (o etiquetas):

Etiquetas: {Bueno, Malo, Indiferente}

Considerando:

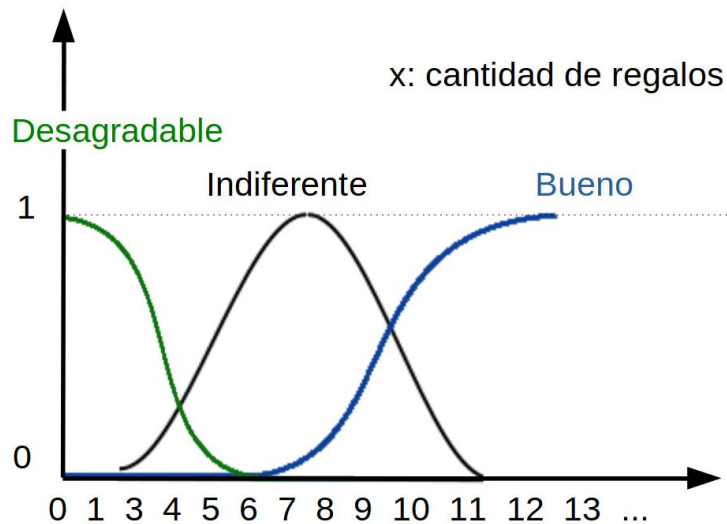
$\mu_B =$ posibilidad de que es cierto que es bueno

$\mu_M =$ posibilidad de que es cierto que es malo

$\mu_I =$ posibilidad de que es cierto que es indiferente

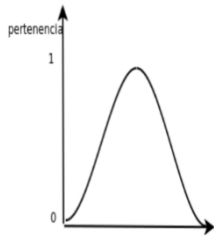
Queda por definir la medición sobre la que se evaluarán las etiquetas, denominada variable métrica de referencia. Sea por caso $x =$ cantidad de regalos que Juan realiza

La siguiente figura da una idea de cómo podrían ser las distribuciones de bueno, malo e indiferente de acuerdo con los posibles valores de x.

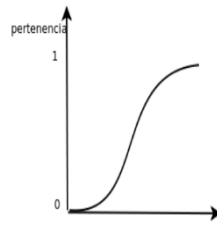


Una consideración importante es que las funciones de pertenencia pueden tener distintas formas.

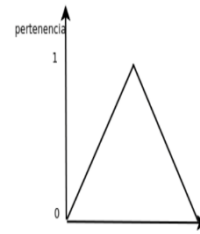
Funciones de pertenencia:



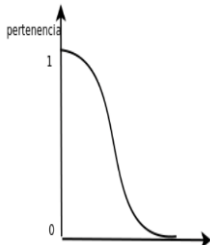
Campana (curva π)



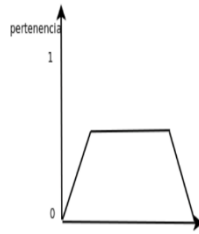
Forma S (curva s)



Triangular



S invertida (curva Z)

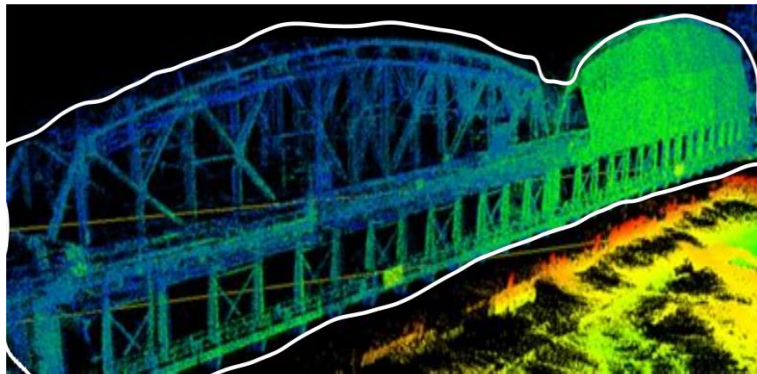


Trapezoidal

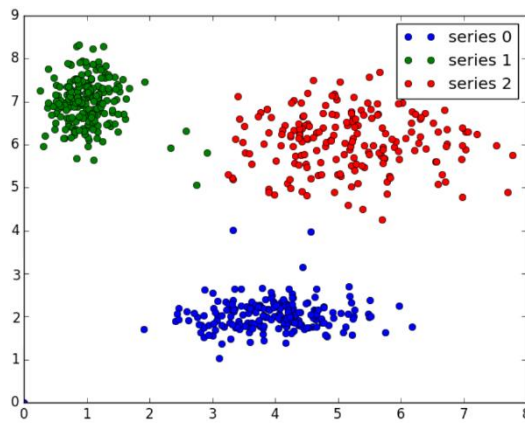
4. Agrupaciones difusas como herramienta de AA

Es posible aplicar el aprendizaje automático con una clasificación difusa de regiones. Esto permite detectar objetos respetando de manera más correcta sus límites en situaciones de iluminación confusa o de solapamiento con otras imágenes.

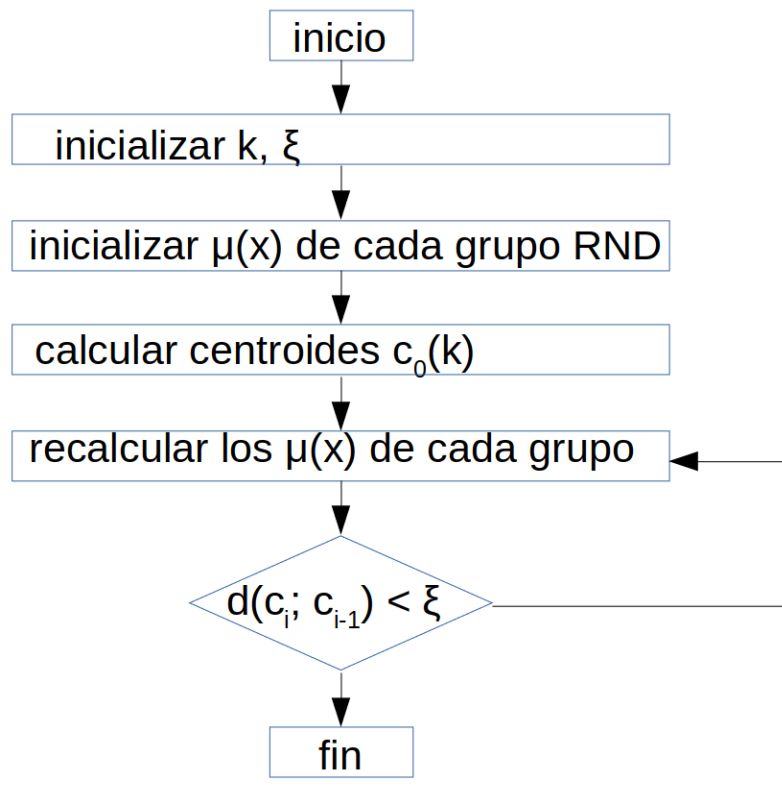
La figura que sigue fue segmentada aplicando la clasificación difusa y luego un umbral a la pertenencia.



Detrás de este resultado existe un proceso donde se han declarado regiones de acuerdo a la temperatura del color y de acuerdo a la cantidad de puntos con la misma temperatura que se encuentran en el entorno inmediato. Luego cada punto es identificado con sus coordenadas y estas cantidades como variables métricas de base a las etiquetas. La figura siguiente supone tres condiciones (etiquetadas como series), correspondientes a la zona de puente, zona exterior y fondo.



La clasificación luego permitirá determinar a qué zona naturalmente pertenece cada punto. La figura siguiente esquematiza la lógica del algoritmo denominado fuzzy c-means, que implementa una variante simple y conocida de este tipo de clasificaciones.



5. Comentarios finales

La lógica difusa fue presentada muy brevemente a fin de mostrar algunas de sus características esenciales. Es posible observar aún en estas condiciones básicas las bondades de contar con un mecanismo que permita diferenciar situaciones donde no es posible determinar taxativamente alternativas por segmentación (es decir por regiones sin solapamientos). Generalmente los procesos que caen en estas condiciones son aquellos más estrechamente ligados a los procesos subjetivos de los seres humanos.

Bibliografia

- Dernoncourt, Franck, "Introduction to fuzzy logic", MIT, 2013.
- VaniSri, ClusteringKhadri et al., "A Modern Optimized Fuzzy C-Means Clustering Using Machine Learning", International Journal of Advanced Science and Technology. Vol. 29, No. 5, (2020), pp. 9417-9428
- Karasulu, Bahadir, "Image Segmentation Using Fuzzy Logic", (2010) Machine Graphics and Vision 19(4):367-409

INTELIGENCIA ARTIFICIAL Y CIENCIA DE DATOS

Panorámica de tecnologías emergentes y disruptivas

Luis Joyanes Aguilar¹

Resumen

Los términos *tecnologías emergentes* y *tecnologías disruptivas*, y *transformación digital* son palabras muy populares o “de moda” (*buzzwords*, en inglés) con decenas de millones de referencia en buscadores como Google, y muy resaltadas en todo tipo de medios de comunicación, así como en la vida diaria de organizaciones, empresas y usuarios. Estas tecnologías de impacto en la sociedad nacen, se desarrollan, implantan, despliegan y se mantienen por periodos de tiempo cortos, medios o largos, y constituyen el soporte de la inteligencia artificial (IA) y la ciencia de datos (CD). Pilares tecnológicos fundamentales de la inteligencia artificial y la ciencia de datos son: *computación en la nube*, *big data*, *conectividad (5G-6G/Wifi 6-7)*, *internet de las cosas*, *ciberseguridad*, *blockchain* y tecnologías asociadas como NFT; otras tecnologías disruptivas son: *gemelos digitales (digital twins)*, *vehículos autónomos*, *drones*, *robots de software (RPA)*, *experiencias inmersivas (RV, RA y RM)*, *asistentes virtuales inteligentes...* y la computación cuántica, el metaverso y la Web 3.0, Industria 5.0 y la Quinta Revolución Industrial. La inteligencia artificial conversacional y la inteligencia artificial generativa junto con las numerosas disciplinas de soporte de la ciencia de datos constituyen los grandes beneficios de la sociedad actual y futura, así como de organizaciones y empresas de todos los sectores.

Abstract

The terms emerging technologies and disruptive technologies, and digital transformation are very popular or "fashionable" words with tens of millions of references in search engines like Google, and highly highlighted in all kinds of media, as well as in the daily life of organizations, companies, and users. These technologies with an impact on society are born, developed, implanted, deployed, and maintained for short, medium, or prolonged periods of time, and constitute the support of artificial intelligence (AI) and data science (CD). Fundamental technological pillars of artificial intelligence and data science are: cloud computing, big data, connectivity (5G-6G/Wi-Fi 6-7), internet of things, cybersecurity, blockchain and associated technologies such as NFT; Other disruptive technologies are: digital twins (digital twins), autonomous vehicles, drones, software robots (RPA), immersive experiences (VR, AR and MR), intelligent virtual assistants... and quantum computing, the metaverse and Web 3.0, Industry 5.0 and Fifth Industrial Revolution. Conversational AI and Generative AI along with the many supporting disciplines of data science are great benefits for today's and tomorrow's society, as well as for organizations and businesses across all industries.

¹ Dr. Ingeniero en Informática y Dr. en Sociología

Presidente de la Fundación I+D del Software Libre (Fidesol), Granada (España); ljoyanes@fidesol.org

Palabras clave

Inteligencia artificial, ciencia de datos, redes neuronales artificiales, procesamiento de lenguaje natural, aprendizaje automático, aprendizaje profundo, científico de datos, modelo de lenguaje, chatbot, asistente virtual inteligente, ética, privacidad de datos, industria 5.0.

Keywords

Artificial intelligence, data science, natural language processing, machine learning, deep learning, data scientist, language model, chatbot, virtual intelligent assistant, ethics, data privacy, industry 5.0

1. Tecnologías emergentes y disruptivas

La RAE (Real Academia Española) define **disruptivo** como “Que produce disrupción” y el término **disrupción** como “Rotura o interrupción brusca” y la Fundación FundéuRAE, amplía los conceptos: “*El sustantivo **disrupción**, el adjetivo **disruptivo** y el menos frecuente verbo **disrumpir** son adecuados para aludir a un proceso o un modo de hacer las cosas que supone una ‘rotura o interrupción brusca’ y que se impone y desbanca a los que venían empleándose*”, *‘innovación disruptiva’ es aquella que supone una ruptura en relación con productos y/o procesos existentes hasta ese momento, a los cuales reemplaza.*

1.1 Tecnología Disruptiva

El profesor Clayton M. Christensen de la Harvard Business School en su libro "*The Innovator's Dilemma*"², publicado en 1997, clasificaba las nuevas tecnologías en dos categorías: *mantenimiento (sostenible)* y *disruptivas*. “*La tecnología de mantenimiento se basa en mejoras incrementales a una tecnología ya establecida*”. “*La tecnología disruptiva carece de refinamiento, a menudo tiene problemas de rendimiento porque es nueva, atrae a una audiencia limitada y es posible que aún no tenga una aplicación práctica comprobada. (Un caso de éxito fue la máquina eléctrica de habla que hoy llamamos teléfono)*”. Ejemplos típicos de tecnologías disruptivas ya asentadas y desplegadas en la sociedad y en las industrias y en organizaciones y empresas son: el computador personal (PC); el sistema operativo Windows; el correo electrónico; los teléfonos celulares e inteligentes; la computación móvil; la computación en la nube; redes de comunicaciones 5G y Wifi 6, redes sociales...

Otra fuente importante para considerar es el prestigioso portal tecnológico, *Techtarget* que define Tecnología Disruptiva como “*La tecnología que afecta al funcionamiento normal de un mercado o industria. Desplaza un producto o tecnología bien establecido, creando una nueva industria o mercado*”³.

¿Cuáles son las características de la tecnología disruptiva? Una tecnología disruptiva reemplaza un proceso, producto o hábito anterior. Por lo general, tiene atributos superiores

² *Professor en Harvard Business School. The Innovator's Dilemma. 1997.*

www.claytonchristensen.com

Nueva edición: The Innovator's Dilemma: The Revolutionary Book that Will Change the Way You Do Business (Collins Business Essentials), 1 de enero 2003. Edición en *Harvard Business Review Press*, “The Innovator’s Dilemma: When New Technologies Cause Great Firms to Fail (Management of Innovation and Change)”, 5 de enero, 2016.

³ *whatis.techtarget.com/definition/disruptive-technology.*

que son inmediatamente obvios, al menos para los primeros usuarios. Las nuevas empresas en lugar de las empresas establecidas son la fuente habitual de tecnologías disruptivas.

Así, se puede considerar que las tecnologías disruptivas son innovaciones que llegan para sustituir un proceso, un producto o una tecnología que ya está establecida, originando una nueva forma de operación, ya sea para los consumidores, las organizaciones o para ambos. La reconocida consultora internacional Gartner en su prestigioso informe sobre tecnologías emergentes⁴ de 2022, plantea que *“Las tecnologías disruptivas permiten llegar hasta los mercados que eran inaccesibles hasta hace poco, por lo que lo disruptivo se convierte en una necesidad para las empresas, siendo oportunidad de continuar mejorando y mantenerse en la competencia. ¿Por qué es importante la tecnología disruptiva?”*

El concepto de «**Innovación Disruptiva**» también fue introducido, definido y analizado por Clayton Christensen. En sus libros y conferencias nos ofrece una definición de la innovación disruptiva: «es aquella innovación que transforma un producto o servicio que sólo era accesible para muy pocas personas con muchos recursos, y la hace accesible a una población mucho mayor que dispone de menos recursos». Amplía su definición, en el citado libro, refiriéndose “a cómo puede un producto o servicio que en sus orígenes nace como algo residual o como una simple aplicación sin muchos seguidores o usuarios convertirse en poco tiempo en el producto o servicio líder del mercado”.

La innovación disruptiva es una innovación que crea un nuevo mercado y una red de valor y eventualmente disrumpe (interrumpe) un mercado y una red de valor existentes, desplazando a los líderes y alianzas de mercado establecidos. Uno de los ejemplos más claros de la innovación disruptiva es el computador, que al principio llegaba a suponer una inversión de varios millones de dólares y mucho tiempo en desarrollar y aprender a utilizar, y ahora está accesible para prácticamente cualquiera hasta en dispositivos tan básicos y accesibles como un teléfono móvil o celular. Otro caso de éxito notable fue la citada máquina eléctrica de habla inventada por Alexander Graham Bell y que hoy se llama teléfono.

Las tecnologías digitales como los dispositivos móviles, las redes sociales, los teléfonos inteligentes, los macrodatos, el análisis predictivo y la nube nacieron como tecnologías emergentes y fueron evolucionando para convertirse en tecnologías disruptivas de impacto en la sociedad actual.

1.2 Tecnología emergente

En 1995 la prestigiosa consultora Gartner comenzó a publicar lo que consideraba las principales tecnologías emergentes y desde entonces, sus predicciones son una institución entre la comunidad tecnológica. Desde 1995, este modelo supone una guía, para organizaciones y empresas de impacto global y para la adopción de nuevas tecnologías. *¿Qué significa tecnología emergente?* Gartner en su Informe *Hype Cycle for Emerging Tech 2022*, considera que las tecnologías emergentes se caracterizan en general por ser tecnologías innovadoras que aportan mejoras frente a otras más tradicionales ya consolidadas pero que aún no han alcanzado su nivel máximo de madurez, por lo que se encuentran aún en vías de desarrollo. *“Las tecnologías emergentes son disruptivas por naturaleza, pero la ventaja competitiva que brindan aún no es bien conocida o probada en el mercado; la mayoría de las tecnologías emergentes tardará más de cinco años, y algunos más de 10, en alcanzar la Meseta de la Productividad y otras caen en el abismo de la desilusión”*.

⁴ Informe de Gartner: *Hype Cycle for Emerging Tech 2022 (Agosto 2022)*

La gráfica *Hype Cycle* representa sus predicciones de tecnologías emergentes (Figuras 1 y 2) y está dividida según su madurez, a saber, emergente o desencadenante tecnológico, excesivo entusiasmo o cómo se inflan las expectativas sobre una tecnología; la decepción y la adopción gradual de dicha tecnología. En la Figura 1 se representan las tecnologías emergentes que predicen despegaran en los siguientes años, desde el lanzamiento de la tecnología y se sitúan según el tiempo esperado de su llegada al despliegue comercial o de impacto en la sociedad en las fases específicas siguientes: *Lanzamiento de la tecnología* (despierta un gran interés entre los medios de comunicación y los sectores industriales), *pico de expectativas sobredimensionadas* (fase de entusiasmo excesivo y proyecciones no realistas), *abismo de desilusión* (la tecnología no cumple con las expectativas desmesuradas y rápidamente se queda fuera de moda), *rampa de consolidación* (los métodos y herramientas comerciales ya disponibles facilitan el proceso de desarrollo) y *meseta de productividad* (fase final, en la que se demuestran y se aceptan en el mundo real los beneficios de la tecnología; existe un menor riesgo de adopción en esta fase, por lo que cada vez más organizaciones tienden a implementar la solución).



Figura 1. Ciclo de expectativas de tecnologías emergentes. Fuente: Gartner

En el informe *Hype Cycle* de Agosto 2022 (Figura 2), Gartner destaca que la evolución de las tecnologías emergentes a observar en los siguientes años (2 a 10 años y más) y más destacadas son: *metaverso*, *tokens no fungibles (NFTs)*, *super aplicaciones*, *Web 3*, *identidad descentralizada*, *humanos digitales*, *gemelos digital del cliente* y *mercados de talento interno (marketplaces)*, *inteligencia artificial generativa de diseño*, *ecosistemas de datos en la nube* y *las super aplicaciones (Superapps)*.

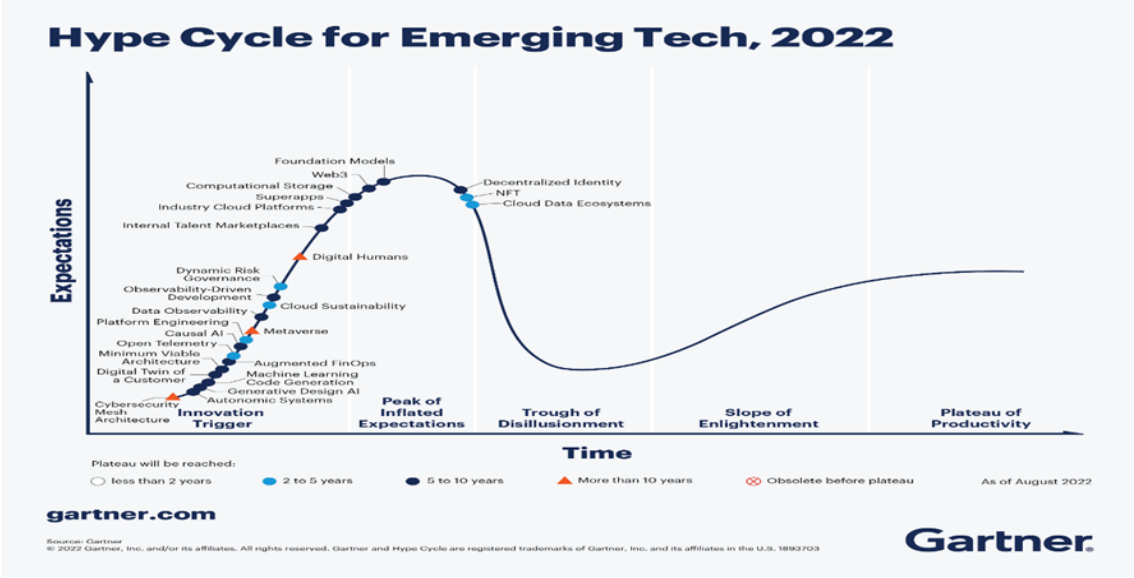


Figura 2. Ciclo de sobre expectativa (*Hype Cycle*) de tecnologías emergentes de Gartner 2022.

1.3 Caso de estudio de tecnologías emergentes y disruptivas en la NATO (OTAN)

La gran importancia de la fusión de las tecnologías emergentes y disruptivas llevó a la NATO (OTAN, Organización del Tratado del Atlántico Norte, en las siglas en español) - organización militar y política con numerosos países del mundo como miembros activos- a crear en julio de 2020 el *The NATO Advisory Group on Emerging and Disruptive Technologies*⁵ (*EDT*), un grupo independiente que proporciona asesoramiento externo a la NATO sobre cómo puede optimizar sus esfuerzos de innovación. El Grupo consta de 12 expertos del sector privado y académico de toda la Alianza que han liderado investigaciones de vanguardia, desarrollado políticas de EDT y gestionado iniciativas de innovación.

En el portal oficial del Grupo de la NATO, se destacan que tecnologías como *big data*, inteligencia artificial (IA), sistemas autónomos y tecnologías cuánticas están cambiando el mundo y la forma en que opera la OTAN. Estas y otras tecnologías emergentes y disruptivas (EDT) presentan tanto riesgos como oportunidades para la OTAN y los aliados. Es por eso por lo que la Alianza está trabajando con socios del sector público y privado, la academia y la sociedad civil para desarrollar y adoptar nuevas tecnologías, establecer principios internacionales de uso responsable y mantener la ventaja tecnológica de la OTAN.

“Las tecnologías emergentes y disruptivas están tocando cada vez más todos los aspectos de la vida, desde dispositivos electrónicos como teléfonos y computadoras, hasta actividades cotidianas como comprar alimentos en el supermercado y administrar el dinero en el banco. Estas tecnologías también están teniendo un profundo impacto en la seguridad. Las tecnologías innovadoras están brindando nuevas oportunidades para los ejércitos de la OTAN, ayudándolos a ser más efectivos, resistentes, rentables y sostenibles. Estas tecnologías, sin embargo, también representan nuevas amenazas por parte de actores estatales y no estatales, tanto militares como para la sociedad civil”.

2. Panorámica de tecnologías disruptivas en el horizonte 2025

Brian Burke, vicepresidente de investigación de Gartner, en la presentación del informe *Hype Cycle* de tecnologías emergentes 2022⁶ (agosto 2022) señalaba y reiteraba otras presentaciones de años anteriores: ***“las tecnologías emergentes son disruptivas por naturaleza, pero la ventaja competitiva que brindan aún no es bien conocida o probada en el mercado”***; por esta razón Burke considera que ***“la mayoría tardará más de cinco años, y algunos más de 10, en alcanzar la Meseta de la Productividad (Figura 2), pero algunas tecnologías en el ciclo Hype madurarán en el corto plazo (dos a cinco años) y los líderes de innovación tecnológica deben comprender las oportunidades para estas tecnologías, particularmente aquellas con alto impacto o con capacidad transformadora”***.

¿Cuáles son las tecnologías disruptivas que están desplegando y consolidándose en estos años de la actual década y continuaran en los siguientes?

Una síntesis de tecnologías disruptivas publicadas en informes de 2022 y principios de 2023 por las grandes consultoras tecnológicas como Gartner, Forrester, McKinsey, Accenture,

⁵ OTAN (NATO). *Emerging and disruptive technologies*. 08 Dec. 2022. https://www.nato.int/cps/en/natohq/topics_184303.htm

⁶ IT Trends. *Tendencias emergentes que impulsan la innovación*.

<https://www.itrends.es/transformacion-digital/2020/08/cinco-tendencias-emergentes-que-impulsan-la-innovacion>

Deloitte, IDC, ... destacan las más sobresalientes y de mayor impacto para la toma de decisiones de organizaciones y empresas:

- Inteligencia Artificial
- Computación en la nube (*Cloud Computing*)
- Big Data
- Conectividad (5G/Wi-Fi 6)
- Internet de las Cosas (IoT)
- Robótica
- Experiencias Inmersivas
- *Blockchain* (Cadena de bloques)
- Gemelos digitales
- Ciberseguridad

2.1 Inteligencia Artificial

Existen un gran número de definiciones de inteligencia artificial, pero tal vez la que más se ajusta al título de este apartado, es la dada por el prof. Mira⁷: «la inteligencia artificial se debe entender como **ciencia** (entonces la tarea de la IA es una tarea de análisis) y se debe entender como **ingeniería** (ingeniería del conocimiento, con un nuevo objeto formal, el *conocimiento* que como la información es pura forma)». Y ya de un modo más pragmático, Mira define los cuatro grandes objetivos de la IA: *modelar, formalizar, programar e implementar máquinas soporte capaces de interactuar de forma no trivial con el medio*. Hace ya una década que se publicó la obra de referencia donde definió Mira su concepto de inteligencia artificial y desde entonces, la inteligencia artificial ha llegado a la sociedad, a las organizaciones y empresas, así como a la industria, de un modo práctico, como comentaremos a continuación y, por eso, creemos que el mejor concepto actual de tan decisiva disciplina es *inteligencia artificial aplicada*.

La Inteligencia Artificial (Figura 3) tiene un componente fundamental como eje vertebral de la tecnología: aprendizaje automático (*Machine Learning*). De igual forma el aprendizaje automático tiene a su vez otro componente importante: aprendizaje profundo (*Deep Learning*). La IA evoluciona de la mano de los algoritmos de aprendizaje automático y aprendizaje profundo junto con los avances en lenguaje de procesamiento natural y redes neuronales artificiales, embebidas en la actualidad en todas las tendencias tecnológicas disruptivas que estamos considerando en el capítulo y en las que se inserta y de las que se nutre en la actualidad y en el futuro, tales como *Big Data*, Internet de las Cosas, Realidad Virtual y Realidad Aumentada, Ciberseguridad, etc. embebidas, a su vez en la computación en la nube (*cloud computing*).

⁷ José Mira Mira. “Aspectos conceptuales de la Inteligencia Artificial y la Ingeniería del Conocimiento” en *Inteligencia Artificial, Técnicas, métodos y aplicaciones*. José T. Palma y Roque Marín (eds). Madrid: McGraw-Hill, 2008.

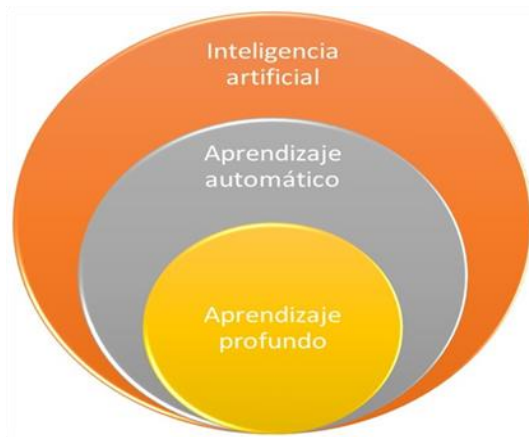


Figura 3. Inteligencia Artificial, Aprendizaje Automático (*Machine Learning*) y Aprendizaje Profundo (*Deep Learning*)

2.2 Computación en la nube

Cloud computing (computación en la nube, o la *nube*)⁸ constituye la evolución natural de la adopción de las diferentes tecnologías de la información y las comunicaciones junto con la expansión de los centros de datos y ha permitido la posibilidad de acceder a Internet a través de una serie de dispositivos, básicamente desde cualquier lugar del mundo y de forma más segura y rápida que en el pasado. *Cloud computing* se ha constituido en una plataforma soportada con una gran cantidad de tecnologías que permiten almacenar, gestionar, compartir y poner a disposición de cualquier usuario -con conexión a Internet- software, aplicaciones y todo tipo de servicios. Cuando compartimos y recibimos un correo electrónico por Gmail, escuchamos una canción en *streaming* (*Spotify*) o vemos un video en *YouTube* o una película en *Netflix*, estamos haciendo uso de la computación en nube. Esta tecnología ya forma parte de nuestra vida diaria y la mayoría de las empresas utilizan y residen en la nube.

Las compañías ya usan software basado en la nube para algunas aplicaciones de empresa y de análisis, pero con la Industria 4.0 un mayor número de tareas relacionadas con la producción requerirán mayor intercambio de datos entre lugares y compañías. Al mismo tiempo, el rendimiento de las tecnologías en la nube mejorará, alcanzando tiempos de reacción de sólo unos milisegundos. Como resultado, los datos y la funcionalidad de las máquinas irán poco a poco haciendo uso cada vez más de la computación en la nube, permitiendo más servicios basados en datos para los sistemas de producción. Incluso los sistemas de monitorización y control de procesos podrán estar basados en la nube

Edge Computing

La computación en el borde (*edge computing*) es, gracias a la Internet de las Cosas, el modelo de nube más utilizado en la actualidad y previsiblemente en el futuro. Una primera definición de *Edge Computing* es la siguiente: Acercar el poder de procesamiento lo más

⁸ JOYANES, Luis (2021). *Computación en la nube. El impacto del cloud computing en las empresas*. Segunda edición. Ciudad de México: Alfaomega; Barcelona: Marcombo. En el libro se describen los orígenes, característica, modelos y fundamentos de la computación en la nube, y todos los aspectos fundamentales para su desarrollo e implementación en organizaciones y empresas, desde la perspectiva de principio de la segunda década del siglo XXI.

cerca posible de donde los datos están siendo generados. Es decir, consiste en acercarse a la nube hasta el usuario, hasta el borde mismo (*edge*, en inglés) de la red. Este modelo tiene las grandes ventajas de que los datos se procesan cerca de donde se generan por lo cual no es preciso llevarlos a la nube, y se reduce la latencia en grandes porcentajes

El Internet de las Cosas (IoT), es el sistema que conforman millones de dispositivos, máquinas y objetos interconectados entre sí y a internet. Numerosos objetos se conectan hoy en día con asiduidad al internet de las cosas, por ejemplo, bombillas, termostatos, sensores industriales en fábricas para controlar la producción, enchufes inteligentes, altavoces virtuales con asistentes de voz como Movistar Home, Alexa o Google Home, o coches (carros) autónomos, como los de Tesla. La consultora Gartner estima que para 2025, el 75% de los datos se procesarán fuera del centro de datos tradicional o *cloud*.

La transición del procesamiento de datos a los bordes (*edge computing*) de la red se espera que suceda en las etapas más tempranas del ciclo de vida de desarrollo de internet de las cosas. El cambio a la computación en el borde tendrá un impacto significativo en los sistemas TIC de las organizaciones.

2.3 Big Data

El análisis de grandes cantidades de datos ha surgido recientemente en el mundo industrial, permitiendo optimizar la calidad de la producción, ahorrar energía y mejorar el equipamiento. En la industria 4.0, la obtención y exhaustiva evaluación de datos procedente de numerosas fuentes distintas (equipos y sistemas de producción, sistemas de gestión de clientes...) se ha convertido en norma para el apoyo de toma de decisiones en tiempo real. El tratamiento y análisis de grandes volúmenes de datos requiere de una gran potencia analítica. El análisis de *Big Data* debe ayudar a tomar mejores decisiones y evaluar las medidas que se han de tomar del modo más eficiente y rentable posible.

2.4. Conectividad

Redes 5G

La quinta generación 5G se anticipó en pruebas al año 2018 y su despliegue comercial y global comenzó el año 2019 aunque de un modo generalizado se produjo en 2021 y en el año 2022 se ha consolidado en la mayoría de las regiones del mundo. La Unión Internacional de Telecomunicaciones UIT (ITU) ha definido los marcos de trabajo y las especificaciones correspondientes con la publicación de las normas correspondientes. En la UIT, la red 5G se conoce con el término *International Mobile Telecommunications (IMT) for 2020 and beyond*, y técnicamente IMT-2020.

¿Qué son tecnologías 5G? Houlin Zhao⁹, Secretario General de la UIT anticipaba la llegada de la nueva generación de redes móviles 5G: «Las tecnologías IMT-2020 (5G) de quinta generación se aproximan rápidamente y tendrán el enorme potencial de transformar nuestras vidas para mejor. Mejor atención de la salud, ciudades más inteligentes, procesos de fabricación mucho más eficientes, todo ello se está haciendo posible al ir avanzando en la era de la Internet de las cosas». Zhao destaca como características importantes de 5G que aportará velocidades de transmisión mucho más rápidas, una conectividad fiable y baja latencia para las telecomunicaciones móviles internacionales (IMT) y crearán un nuevo

⁹ Houlin Zhao. Prólogo de la revista no. 2 de 2017, *ITU News Magazine*. Dossier especial de redes 5G. Abrir sendas hacia 5G. itunews.itu.int (6 números al año).

ecosistema de comunicaciones mundiales de dispositivos conectados que envían grandes volúmenes de datos a través de la banda ancha ultrarrápida. Las redes 5G por sus características especiales serán más rápidas y flexibles e idónea para su uso con el Internet de las Cosas.

Redes Wi-Fi 6.0

La Wi-Fi Alliance¹⁰ -responsable de la regulación y estandarización de las redes y protocolos Wi-Fi- ha lanzado oficialmente en septiembre de 2019 su programa de certificación de **Wi-Fi 6** (antiguo nombre 802.11ax)¹¹ o **wifi 6**¹². Wifi 6 mejora de modo notable las características de Wi-Fi 5 y la Alliance Wi-Fi considera que los beneficios clave¹³ de la tecnología Wi-Fi CERTIFIED 6 incluyen:

- *Mayores tasas de datos.*
- *Mayor capacidad.*
- *Rendimiento en entornos con muchos dispositivos conectados.*
- *Eficiencia energética mejorada.*

Según la organización, Wi-Fi 6 proporciona: *“las bases para una gran cantidad de usos actuales y emergentes desde la transmisión de películas (en streaming) de alta definición hasta aplicaciones comerciales de misión crítica que requieren gran ancho de banda y baja latencia para mantenerse conectados y productivos mientras atraviesa redes grandes y congestionadas en aeropuertos y estaciones de tren”.*

2.5 Internet de las cosas

No existe una definición única de Internet de las Cosas -Internet of Things, IoT-, muy al contrario, son numerosas las definiciones del concepto, algunas de las más relevantes las iremos viendo a continuación. La idea original y desarrollo de IoT fue introducida por Kevin Ashton, en 1999, un ingeniero que trabajaba por aquel entonces en P&G (Procter&Gamble) en Estados Unidos, con ocasión de una conferencia que impartió sobre la idea del uso de las etiquetas RFID (identificación por radiofrecuencia) en la optimización del proceso de la cadena de suministro y como mejora eficiente de los códigos de barras. Posteriormente Kevin Ashton¹⁴ en un artículo publicado en el *RFID Journal* el 12 de julio de 2009 ya introdujo el concepto formal de Internet de las Cosas para referirse al hecho de conectar todas las cosas que nos rodean con la finalidad de poder contarlas, saber su posición en cualquier momento,

¹⁰ En la página oficial de Wi-Fi Alliance están publicadas las características principales de las redes Wi-Fi 6: <https://www.wi-fi.org/discover-wi-fi/wi-fi-certified-6>

¹¹ Desde octubre de 2018, la Wi-Fi Alliance decidió cambiar los nombres tradicionales de las denominaciones de las redes Wi-Fi por unas más fáciles de entender. Los nuevos nombres son:

802.11 ax pasa a ser Wi-Fi 6:
802.11 ac ... Wi-Fi 5
802.11 n ... Wi-Fi 4
802.11 g ... Wi-Fi 3
802.11 a ... Wi-Fi 2
802.11 b ... Wi-Fi 1

¹² El Diccionario de la Real Academia Española (DRAE) ha incorporado en una de sus últimas revisiones, el término **wifi** (con letras minúsculas y sin cursivas, también **wi fi**), con la siguiente definición: *“Sistema de conexión inalámbrica, dentro de un área determinada, entre dispositivos electrónicos, y frecuentemente para acceso a internet”.*

¹³ <https://www.wi-fi.org/discover-wi-fi/wi-fi-certified-6>

¹⁴ Kevin Ashton. Cofundador de Auto-ID-Center del MIT. *RFID Journal*, *IoT Journal*, 1999. www.rfidjournal.com/articles/view?4986

así como aportarnos información sobre el entorno que las rodea. La definición de Ashton destacaba¹⁵

*“Si tuviésemos computadores que fuesen capaces de saber todo lo que pudiese saberse de cualquier cosa (‘las cosas’) –usando datos recolectados sin intervención humana– seríamos capaces de hacer seguimiento detallado de todo, y poder reducir de forma importante los costes y malos usos. Sabríamos cuando las cosas necesitan ser reparadas, cambiadas o recuperadas, incluso si están frescas o pasadas de fecha. El **Internet de las Cosas** tiene el potencial de cambiar el mundo como ya lo hizo Internet. O incluso más.”*

En esencia, Internet de las Cosas, es una nueva dimensión de Internet y una nueva generación de servicios que significa que cualquier cosa u objeto se pueden comunicar entre sí y a Internet, en cualquier lugar y en cualquier momento; es decir, de un modo *ubicuo* y, al realizar las comunicaciones mediante numerosos sistemas y tecnologías, el mundo estará *hiperconectado*.

Definición de Internet de las Cosas de ITU (UIT)

La definición¹⁶ de la Unión Internacional de Telecomunicaciones, UIT (ITU) se publicó el 4 de diciembre de 2012 y destaca que la IoT es una visión global que no sólo tendrá tecnologías sino numerosas consecuencias tecnológicas y sociales:

“Infraestructura mundial para la sociedad de la información que propicia la prestación de servicios avanzados mediante la interconexión de cosas (físicas y virtuales) en base a la información existente y evolución interoperable de tecnologías de la información y la comunicación (TIC) presentes y futuras.”

Una definición más simple que subyace en la ITU es: *“una sociedad definida por cosas inteligentes que pueden comunicarse entre sí directamente o a través de una red. Internet de las cosas comprende la posibilidad de conectar prácticamente cualquier dispositivo a la Red, incrementando sus posibilidades de interacción con otros elementos gracias al nivel de interacción que permite Internet”¹⁷.*

La OCDE define el internet de las cosas en sentido muy amplio, en concreto, como *“todos los dispositivos y objetos cuyo estado puede ser alterado a través de internet, con o sin la implicación directa de las personas”*. Incluye, por tanto, computadores portátiles, servidores, tabletas y teléfonos inteligentes, que pertenecen tradicionalmente al *“internet de las personas”*, por llamarlo de alguna manera. Por otra parte, el verdadero internet de las cosas es lo que ellos denominan M2M (*Machine To Machine*) o la comunicación de datos con poca o ninguna interacción humana: *“dispositivos que comunican activamente utilizando redes físicas o inalámbricas, que no son computadores en el sentido tradicional, y que utilizan internet de una forma u otra”*.

Internet de las Cosas es la era de los dispositivos inteligentes (sensores, objetos, casas, edificios, ciudades, sensores, ...), tecnologías ponibles (*wearables*, ropa, anillos, relojes...),

¹⁵ Kevin Ashton: *That 'Internet of Things' Thing: in the real world, things matter more than ideas*, *RFID Journal*, 22 July 2009. [en línea] <http://www.rfidjournal.com/articles/view?4986>.

¹⁶ Recomendación serie Y: ITU-T Y.2060 (06/2012)

<https://itunews.itu.int/Es/4503-Internet-de-las-cosas.Maquinas-empresas-personas-todo.note.aspx>. Actualizada en 2016: www.itu.int/recl/T-REC-Y.2060-201206-I/es

¹⁷ EOI. *Internet del futuro. Estado del arte y oportunidades de negocio*. Madrid: EOI, 2015, p. 4.

vehículos autónomos (coches, carros, buses, trenes, aviones, barcos...), altavoces inteligentes...

Internet industrial de las cosas. Con el Internet Industrial de las Cosas, un mayor número de dispositivos (objetos inteligentes) se enriquecerán de la computación y se conectarán a través de estándares tecnológicos. Esto permitirá a los dispositivos de campo comunicarse e interactuar tanto con otros iguales a ellos como con controladores más centralizados, según sea necesario. También descentraliza el análisis y la toma de decisiones, lo que permitirá respuestas en tiempo real.

Tecnologías *wearables* (ponibles). Tecnologías *wearables* (*ponibles*)¹⁸ -o *vestibles*- son tecnologías inteligentes que se llevan puestas en forma de prenda o complemento en la ropa de vestir. Las tecnologías *wearables* llevan dispositivos móviles asociados (*wearables devices*) desde gafas inteligentes hasta relojes inteligentes, chaquetas inteligentes, pasando por pulseras o anillos inteligentes. La expansión de los dispositivos «ponibles» se ha producido por la posibilidad de conversión en dispositivos inteligentes al estilo de los sensores. La Internet de las cosas ha facilitado la citada expansión al convertir a estos dispositivos en objetos inteligentes a los que se les dota de un protocolo de acceso a Internet (ahora protocolo IPv6, aunque también podría ser IPv4) y también conectividad inalámbrica con otros dispositivos inteligentes.

Drones. Un **dron** (término aceptado por la Real Academia Española, RAE) se define como «*Aeronave no tripulada*». El término original en inglés, *drone*, significa “zángano, el macho de la abeja”, pero su significado más usual en la actualidad es para referirse a un vehículo aéreo no tripulado. Los drones tienen su origen en el ámbito militar donde comenzaron a utilizarse en la Segunda Guerra Mundial y luego en posteriores acontecimientos bélicos.

Sin embargo, desde un punto de vista técnico y profesional, los drones tienen diferentes acepciones y significados, tales como: **RPA** (*Remotely Piloted Aircraft*). Aeronave dirigida (pilotada) por control remoto; **RPAS**¹⁹ (*Remotely Piloted Aircraft System*). Aeronave no tripulada y su sistema de control (controladora más estación de control). En el caso de RPA y RPAS se hace referencia al piloto que opera la aeronave en forma remota; es decir, se utiliza un enlace de comunicaciones entre la estación de tierra y la aeronave, y el piloto no se encuentra a bordo de la aeronave.

Otros términos sinónimos de drones son: **UA** (*Unmanned Aircraft*). Aeronave no tripulada; **UAV** (*Unmanned Aerial Vehicles*). Es una aeronave en la que el piloto no está físicamente a bordo. Se refiere a la plataforma de vuelo. Estos aparatos, a diferencia de los RPA, son aparatos no controlados remotamente por una persona; es decir, son capaces de despegar, navegar y aterrizar en forma automática con base en los parámetros que hayan sido establecidos previamente (este tipo de vuelos no son permitidos por la legislación española, ya que se requiere que un piloto se encuentre en todo momento supervisando las operaciones, aunque el vuelo se realice en forma autónoma). **UAS** (*Unmanned Aerial System*). Sistema aéreo no tripulado. Comprende la plataforma –vehículo aéreo–, el enlace de comunicaciones y la estación de tierra.

¹⁸ Al día de la edición de esta obra no existe una traducción al español aceptada por la RAE, una traducción recomendada por FundéuRAE (Fundación Española del Español Urgente, www.fundeu.es) es “ponible”. www.fundeu.es/recomendacion/tecnologia-ponible.mejor-que-wearable-technology

¹⁹ En ocasiones se confunde con el plural en inglés de RPA y no es así, ya que son siglas diferentes.

Las aplicaciones de los drones son hoy muy numerosas. Son herramientas que han adquirido un gran relieve desde la iniciativa Industria 4.0, ya que no solo se utilizan en tareas de entretenimiento y ocio, sector donde han adquirido gran notoriedad, sino y sobre todo en numerosos sectores de beneficio a organizaciones y empresas, así como en distribución logística, salud, transportes, control de seguridad, defensa, etc.

Vehículos autónomos y vehículos conectados. Son términos que se suelen utilizar como sinónimos, pero tienen grandes diferencias entre sí. Un vehículo autónomo (coche o carro, autobús, ...) es un vehículo cuya característica principal es que puede circular sin conductor. Sin embargo, el auto conectado no tiene esta propiedad, pero sí puede tener conectividad a Internet, uso de asistentes virtuales como Siri y Alexa, y comunicarse con los elementos de su entorno y, a través de múltiples avisos y datos proporcionados al conductor, conseguir la llamada *conducción asistida*, el paso previo a la conducción autónoma. Los coches (carros) autónomos son un tipo de automóvil capaz de detectar su entorno y navegar sin intervención humana. Además, de momento es quizás el tipo de robot avanzado por excelencia

Las características fundamentales del vehículo autónomo son: *conectividad* (el despliegue de las redes móviles 5G y de la reciente certificación de las redes Wifi 6 serán determinantes para el éxito del vehículo autónomo) *la baja latencia, densidad de dispositivos controlados, velocidad y la computación en el borde (edge computing)* serán decisivas y prevalecerán sobre otras características en la elección de la tecnología por fabricantes y usuarios.

Fabricación aditiva. «Las compañías acaban de empezar a adoptar la fabricación aditiva, como por ejemplo la impresión 3D, la cual es usada, principalmente, para crear prototipos y producir componentes individuales. Con la Industria 4.0, estos métodos de fabricación aditiva serán ampliamente usados para producir pequeños lotes de productos personalizados que ofrecen ventajas de construcción, como son los diseños ligeros y complejos. Los sistemas de fabricación aditiva descentralizados, de alto rendimiento, reducirán las distancias de transporte y el *stock* de productos».

2.6 Robótica

Robots colaborativos

Muchas industrias han usado robots desde hace mucho tiempo para abordar tareas complejas, pero es ahora cuando los robots industriales están evolucionando para alcanzar una mayor utilidad. Cada vez son más autónomos, flexibles, [colaborativos] y cooperativos hasta tal punto que interactuarán con otros robots y trabajarán lado a lado con humanos de forma segura, aprendiendo de ellos: es un nuevo modelo de robot humanoide, denominado *cobot (robot colaborativo)*. Esta nueva generación de robots está llegando a las fábricas y cadenas de producción, así como muchos otros sectores como el turismo, la medicina, los centros comerciales o los aeropuertos: son los *robots colaborativos* que, a su vez, están originando una nueva tendencia en ingeniería: la *cobótica*. Estos robots costarán menos y tendrán más capacidades que los usados actualmente en la fabricación.

Automatización robótica de procesos (RPA): robots de software

RPA, Automatización Robótica de Procesos (Robotic Process Automation) es una nueva tecnología informática de automatización de procesos de negocio. Es un robot (*bot*) de software virtual diseñado para la ejecución de tareas repetitivas. RPA es un operador virtual o robot que interactúa con los sistemas en la capa de interfaz del usuario de un sistema

informático y es capaz de utilizar tareas mecánicas basadas en reglas de negocio que, normalmente, realizan los profesionales humanos. En la práctica es un programa que interacciona con la interfaz gráfica existente, de modo que el usuario puede configurarlo a medida (y bajo demanda) como un servicio que no requiere implicar a la infraestructura y recursos de TI existentes en la empresa. Deloitte²⁰ define un RPA “como un método de automatización de procesos, principalmente, transaccionales basados en reglas específicas, realizado con software que aprende del usuario de negocios y le asiste en tareas relativamente sencillas”.

Los RPA se combinan con componentes de inteligencia artificial tales como gestión de decisiones basadas en aprendizaje automático, inteligencia conversacional mediante *chatbots* y asistentes virtuales, analítica de datos, en especial de textos y de sentimientos. Las soluciones de RPA se realizan mediante un robot de software que simula las acciones realizadas por los humanos. La inteligencia artificial aplicada a la industria de la robótica ha generado en los últimos años, los **robots virtuales** (*asistentes virtuales*) y *chatbots* cuya presencia en empresas e industrias de todo tipo es cada día mayor, en los diferentes sectores de las organizaciones, así como en aplicaciones de gestión empresarial y para la toma de decisiones.

2.7 Experiencias inmersivas (Realidad extendida): realidad virtual, realidad aumentada y realidad mixta

Las tecnologías inmersivas: realidad virtual, realidad aumentada y realidad mixta están cambiando la forma que tiene el usuario de percibir e interactuar con el mundo digital y están originando las conocidas experiencias inmersivas. El término de **Realidad Extendida**²¹ (*Extended Reality, XR*) es un espectro de tecnologías que enlazan o integran el mundo digital y el mundo real. Engloba las citadas tres tecnologías: **realidad virtual** (*Virtual Reality, VR*), **realidad aumentada** (*Augmented Reality, AR*) y **realidad mixta** (*Mixed Reality*), todas las cuales brindan diferentes grados de inmersión sensorial e interacción entre el mundo real y el contenido digital. La realidad extendida, combina los mundos real y digital para crear nuevos tipos de interactividad y percepción. Todas las tecnologías brindan diferentes grados de inmersión sensorial e interacción entre el mundo real y el contenido digital. Estas tecnologías ofrecen nuevas formas de capacitar, educar, entretener, colaborar y simular, así como potenciar el análisis y exploración de datos. Sin embargo, la realidad extendida también hace que los datos sean más vulnerables y, en consecuencia, aumentan los riesgos y peligros de las experiencias inmersivas sin un uso eficiente y el control de los datos, así como el necesario uso de la ética. La realidad extendida está facilitando el soporte de los universos virtuales “*el metaverso*” donde los usuarios pueden interactuar unos con otros.

El objetivo de la realidad extendida es dar a conocer las distintas realidades, sus diferencias, virtudes y potencialidades. En los últimos años, todas las realidades han crecido y se han desarrollado de forma exponencial: virtual, aumentada y mixta.

Realidad Virtual. Es un sistema informático que genera una representación de la realidad (mundo real) al que reemplaza. Introduce al usuario en un mundo virtual similar al real. Implica la creación de un mundo generado por computadora en el que una persona puede interactuar de tal manera que piensa está en el mundo real en lugar de en un mundo virtual.

²⁰ Deloitte. *Automatización Robótica de Procesos (RPA)*. 2017.

²¹ GAO, U.S. Government Accountability Office. <https://www.gao.gov/assets/gao-22-105541.pdf>

La realidad virtual sustituye a la realidad física pero no sobreimprime datos informáticos como sí hará la realidad aumentada. El usuario está inmerso en un entorno interactivo, generado digitalmente.

En esencia, la realidad virtual reemplaza al mundo real. En ella se sumerge totalmente de modo inmersivo. Las primeras aplicaciones de impacto de la realidad virtual fue el popular juego de simulación *Second Life*²². Los dispositivos de realidad virtual sumergen al usuario en una realidad completamente artificial, pero con audio y vídeo, y en muchos casos se realiza interacción.

Realidad aumentada. Los sistemas basados en realidad aumentada soportan una gran variedad de servicios, como por ejemplo la selección de piezas en un almacén y el envío de instrucciones de reparación a través de dispositivos móviles. Estos sistemas se encuentran aún en sus primeros pasos, pero en el futuro las compañías harán un uso mucho más amplio de la realidad aumentada para proporcionar a los trabajadores información en tiempo real con el objetivo de mejorar la toma de decisiones y los procedimientos de trabajo. Por ejemplo, los trabajadores podrían recibir instrucciones de cómo sustituir una pieza en particular mientras están mirando el propio sistema bajo reparación, a través de unas gafas de realidad aumentada, por ejemplo. Otra aplicación podría ser la formación de trabajadores de forma virtual. Ejemplos de realidad aumentada se ven con mucha frecuencia en acontecimientos deportivos, como los anuncios que se instalan durante la emisión de partidos de fútbol o en las líneas de *off side* (fuera de línea) en los partidos que se retransmiten por televisión. La televisión la usa habitualmente, y esta tecnología está alcanzando gran popularidad, ya que se puede usar en numerosos dispositivos, desde computadoras hasta dispositivos móviles como iPhone, iPad de Apple o dispositivos Android.

Realidad mixta²³, como indica su nombre es un híbrido de ambas realidades; la realidad aumentada superpone el añadido digital al mundo físico que vemos con nuestros ojos y la realidad virtual con la intermediación de unas gafas, unas viseras cerradas o dispositivos análogos, transporta al usuario a otro lugar. Utiliza también óptica avanzada, sensores y una gran potencia de computación, creando un espacio holográfico en tiempo real. En realidad mixta, un auricular dedicado reconoce su entorno y permite la interacción entre el contenido digital y el mundo real en múltiples dimensiones. Por ejemplo, la tecnología de realidad mixta puede ayudar a entrenar a los cirujanos al permitirles interactuar con imágenes 3D de órganos en lugar de imágenes 2D disponibles en pantallas planas.

2.8 Blockchain

La integración de Internet de las Cosas y la ya consolidada tecnología *Blockchain* abre nuevas perspectivas en todo tipo de sectores fundamentales en la industria, infraestructuras, ciudades y edificios, salud, logística, calidad de vida... La revolución de Cloud, Big Data, Internet de las Cosas e Inteligencia Artificial con el soporte de seguridad de *Blockchain* están cambiando el panorama del desarrollo de software. Desde el nacimiento oficial de la

²² Ver página de Second Life: *Acerca del Visor de Second Life*. Para conectarte a Second Life y para explorar y comunicarte en el mundo virtual necesitas nuestro software de navegación en 3D, al que llamamos el Visor de SL. Este programa, que se descarga e instala rápidamente, es gratuito y no contiene programas espía. <https://secondlife.com/support/downloads/?lang=es-ES>

²³ Josep Lluís Micó, ya en un artículo publicado en La Vanguardia, en 2017, anticipaba el futuro de la realidad mixta ¿Tiene más futuro la realidad mixta que la realidad virtual? La Vanguardia. 6 de febrero, 2017 <http://www.lavanguardia.com/tecnologia/20170206/414053748119/internet-en-las-cosas-realidad-mixta-realidad-virtual-comercio-iphone-8.html>

criptomoneda *bitcoin* en 2008, la tecnología *blockchain*, soporte fundamental y nuclear de la mencionada *bitcoin*, se ha ido expandiendo y evolucionando en muchos sectores y casos de estudio, incluyendo internet de las cosas y seguridad, además de áreas industriales, financieras o de salud.

Un *blockchain* es una lista de registros de datos en continuo crecimiento que están enlazados y protegidos mediante técnicas criptográficas. Una cadena de bloques es una estructura de datos básica que propuso Satoshi Nakamoto -creador de *bitcoin* en 2008 para presentar y poner en funcionamiento la criptomoneda *bitcoin* como soporte de las conocidas redes “igual a igual” (*peer-to-peer*, P2P). Una cadena de bloques, en esencia, se compone de un gran número de bloques, los cuales pueden contener cualquier tipo de datos, aunque se usan con mayor frecuencia para mantener un registro de diversas transacciones entre pares (*peers*). Estos bloques se enlazan juntos, hacia atrás, y cada bloque verifica la integridad de su bloque anterior de modo aleatorio con un código *hash*. La manipulación de un bloque anterior invalidará su código *hash*, haciendo fácilmente perceptible la misma. Calcular un nuevo *hash* es una tarea que se realiza con minería, un proceso muy exigente y realizado mediante un minero, de modo que la modificación de un bloque tiene un efecto en cada bloque más joven vinculado a él. Aunque la minería es muy difícil, verificar la validez de un bloque minado es muy fácil para los pares que participan; de esta forma, esta propiedad de *blockchain* disuade a los usuarios de modificar datos del bloque.

2.9 Los gemelos digitales (*digital twins*)

El gemelo digital (*digital twin*) es un concepto conocido desde hace años, pero ha sido a lo largo del bienio 2016-2017 cuando ha adquirido gran resonancia, por el auge del Internet de las Cosas, la fabricación aditiva (impresión 3D) y las técnicas de aprendizaje automático (inteligencia artificial). Los gemelos digitales son la representación virtual de un objeto físico o un sistema a lo largo de su ciclo de vida que utiliza datos en tiempo real para permitir la comprensión, aprendizaje y razonamiento. Los sensores del IoT que reúnen datos e información como el estado en tiempo real, salud y desempeño, posición en vivo, etc. están integrados con un objeto físico. El ecosistema de gemelos digitales comprende las tecnologías *IA*, *Big Data*, *IoT* y *plataformas en la nube* para habilitar este ecosistema gemelo. Las estimaciones del mercado global de gemelos digitales se esperan crezca en 15.600 millones de euros en 2023.

Un gemelo digital, es un concepto de producto software que sirve como plantilla virtual para la producción que crece y se desarrolla en la fase de creación de un producto y permanece vinculado con el producto físico durante su ciclo de vida. Con los gemelos digitales, los fabricantes prueban y validan digitalmente sus productos utilizando software de simulación construido al efecto (copias virtuales de objetos o servicios físicos). Es decir, son capaces de modelar y optimizar los flujos de productos a lo largo de todo el ciclo de vida de las máquinas antes de que se construyan, mediante el intercambio de datos entre ambos “gemelos” y la simulación completa del diseño, producción y uso futuro del producto, incluyendo en su caso a los proveedores. Con el gemelo digital el fabricante podrá, además, prever muchas disfunciones antes de sacar el producto al mercado (Joyanes 2017).

2.10 Ciberseguridad

Muchas compañías todavía dependen de sistemas de gestión y producción desconectados o cerrados, pero con la creciente conectividad y uso de protocolos de comunicación estándar, la necesidad de proteger los sistemas industriales críticos y líneas de

fabricación, las amenazas de ciberseguridad aumentan dramáticamente. Como resultado, son esenciales tanto comunicaciones seguras y fiables, como sofisticados sistemas de gestión de identidad y acceso de máquinas y usuarios. La ciberseguridad es transversal a todas las tecnologías con obligatoriedad de implantación para todo tipo de organizaciones y empresas con datos como soporte.

Recomendación: *Ciberseguridad: La colaboración público-privada (2018)*, Instituto Español de Estudios Estratégicos, Ministerio de Defensa de España, IEEE (www.ieee.es) *Cuadernos de Estrategia*, no. 185²⁴

3 Computación en la nube nativa (*cloud native computing*) y computación sin servidores (*serverless computing*)

La nube moderna, también conocida como la **computación nativa en la nube o nube nativa (*cloud native*)** ha traído nuevos modelos que se integran y adaptan con los modelos tradicionales antes descritos. La Fundación Cloud Native Computing Foundation (CNCF, www.cncf.io) fue creada en 2015 para definir, desarrollar y proporcionar estándares, técnicas y tecnologías de computación nativa en la nube a nivel mundial con innumerables miembros de todo tipo de organizaciones, empresas, y centros académicos y de investigación. CNCF está integrada en la Fundación de Software Libre Linux (Linux Foundation, www.linuxfoundation.org).

3.1 ¿Qué es Computación Nativa en la Nube: Cloud Native Computing Foundation (CNCF)?

Cloud Native Computing Foundation (CNCF) es una base de software de código abierto que promueve la adopción de la computación nativa de la nube. (CNCF)²⁵ en la presentación de la Fundación en la página principal, su Directora General define el objeto principal “***The Foundation of Doers***: *Somos una comunidad de “hacedores” [doers] que permiten proyectos de código abierto, incluidos Kubernetes, Prometheus, Envoy y muchos otros. Kubernetes y otros proyectos de CNCF ganaron adopción rápidamente y aseguraron el apoyo de una comunidad diversa, convirtiéndose en algunos de los proyectos de mayor velocidad en la historia del código abierto*”. En esencia CNCF tiene como objetivo establecer una comunidad de desarrolladores, usuarios finales y proveedores de servicios y tecnología de TI independiente del proveedor para colaborar en proyectos de código abierto. A finales de 2023 tenía proyectos y desarrolladores de 188 países. Miembros destacados de CNCF son todas las empresas de software y computación en la nube más grandes del mundo, así como numerosas nuevas empresas innovadoras, proveedores de nube pública, como *Amazon Web Services*, *Google* y *Microsoft*; empresas de *software* empresarial, como *SAP*, *Oracle* y *Salesforce...* y numerosas empresas de medios sociales, medios de comunicación, fabricantes, etc. así como numerosas empresas innovadoras *startups*.

La CNCF define la **nube nativa** (Cloud Native)²⁶ como el uso de software de código abierto, así como tecnologías tales como contenedores, microservicios, malla de servicios,

²⁴Recomendación: *Ciberseguridad: La colaboración público-privada*. Instituto Español de Estudios Estratégicos, Ministerio de Defensa de España, IEEE (www.ieee.es) *Cuadernos de Estrategia*, no. 185 (2017).

²⁵ **CNCF**. Priyanka Sharma, General Manager, CNCF: <https://www.cncf.io/about/who-we-are/>

²⁶ **CNCF** *Cloud Native Definition v1.0*. <https://www.cncf.io/about/faq/#what-is-cloud-native>. Aprobada por el TOC de CNCF el 11 de junio de 2018.

infraestructuras inmutables y APIs declarativas, para desarrollar e implementar “aplicaciones escalables en entornos dinámicos modernos” como las plataformas de computación en la nube (públicas, privadas e híbridas). El CNCF en su definición oficial considera que estas técnicas permiten sistemas débilmente acoplados que son resilientes, manejables y observables. Combinados con una sólida automatización, permiten a los ingenieros desarrollar sus proyectos y realizar cambios de alto impacto con frecuencia y de manera predecible con un trabajo mínimo, mientras que en las aplicaciones tradicionales sus proyectos exigen gran trabajo y la realización de cambios supone esfuerzos considerables en dedicación profesional y tiempo.

De igual forma también en la definición oficial, la Cloud Native Computing Foundation busca “Impulsar la adopción de este paradigma mediante el fomento y el mantenimiento de un ecosistema de proyectos de fuente abierta (*open source*) e independientes del proveedor, así como la democratización de los patrones de última generación para que estas innovaciones sean accesibles para todos”.

Así se han desarrollado y evolucionado nuevos modelos de la nube como **computación sin servidores** (*serverless computing*) y nuevos componentes de esta nueva arquitectura de computación nativa: **microservicios**, **contenedores** y **orquestradores**.

3.2 Computación sin servidores

La computación sin servidores o informática sin servidores (*serverless computing*) es un modelo de computación en la nube, en la que un proveedor en la nube, en forma dinámica, gestiona el alojamiento de los recursos computacionales. El término *sin servidor* significa que las tareas asociadas con el aprovisionamiento y la administración de la infraestructura son invisibles para el desarrollador. Los desarrolladores de aplicaciones sin servidor ya no necesitan dedicar tiempo y recursos al aprovisionamiento, mantenimiento, actualizaciones, escalado y planificación de la capacidad del servidor. La arquitectura sin servidor no significa que no se necesiten servidores, sino que, en el desarrollo y ejecución de una aplicación lógica, no se necesita servidores para su administración. Contrariamente a lo que su nombre parece indicar, los modelos de computación sin servidor guardan la información en servidores físicos. Su nombre hace referencia a la facilidad de los usuarios de no tener que estar pendientes de la administración de dicho servidor. Los desarrolladores de aplicaciones sin servidor ya no necesitan dedicar tiempo y recursos al aprovisionamiento, mantenimiento, actualizaciones, escalado y planificación de la capacidad del servidor. Los conceptos de arquitectura sin servidor y función como servicio han crecido de la mano con la popularidad de los microservicios, contenedores y las ofertas de nube bajo demanda. Contrariamente a lo que su nombre parece indicar, la computación sin servidores (*Serverless Computing*) si guarda la información en servidores físicos.

En una aplicación tradicional, se administra toda la ejecución lógica de la aplicación en una máquina virtual, un servidor físico o en la nube. La informática sin servidor permite a los desarrolladores crear aplicaciones más rápido, ya que no es necesario que administren la infraestructura. Los servidores siguen existiendo y la única diferencia es que lo hacen debajo de una capa de abstracción (invisible) de modo que la tarea de manejar la infraestructura queda relegada al proveedor del servicio y no al desarrollador y programador. El precio de sus servicios se basa en la cantidad de recursos consumidos por una aplicación.

La computación sin servidor ayuda a los equipos a aumentar su productividad y a comercializar los productos más rápido, además de permitir a las organizaciones optimizar mejor los recursos y seguir centrándose en la innovación. Con este enfoque los desarrolladores se centran más en la lógica del negocio y en aportar más valor al núcleo principal del negocio. Su nombre hace referencia a la facilidad de los usuarios de no tener que estar pendientes de la administración del servidor y se pueden centrar en escribir código y en su desarrollo. Este modelo de la nube permite, por tanto, que al usuario se le facture solo por el tiempo en que se ejecuta el código.

3.3 Microservicios

El término microservicio surgió en una conferencia de arquitectos de software en mayo de 2011 para representar un estilo de arquitectura de sistemas, y no exactamente el tamaño de los servicios que la componen, como el nombre podría dar a entender. El enfoque tradicional para el diseño de aplicaciones de software se centra en la arquitectura monolítica, en la que todos los elementos que pueden implementarse están contenidos en una sola aplicación. Este enfoque tiene sus desventajas: cuanto más grande es la aplicación, más difícil es resolver los problemas que se presentan y agregar funciones nuevas rápidamente. Estos inconvenientes que se presentan en el diseño y desarrollo tradicional de aplicaciones se resuelven, considerablemente, con los microservicios.

¿Qué son los microservicios?

Los microservicios implican un cambio de paradigma respecto a la arquitectura tradicional monolítica, donde todas las funciones del código son compiladas en una sola aplicación. Los microservicios son un enfoque arquitectónico y organizativo para el desarrollo de software donde el software está compuesto por pequeños servicios independientes que se comunican entre sí a través de interfaces de programación de aplicaciones (APIs) bien definidas. Los microservicios pueden ser implementados utilizando cualquier lenguaje de programación y tecnologías de almacenamiento de datos diferentes. Cuando se despliega cada microservicio se hace de forma independiente.

Los microservicios son un estilo de arquitectura, así como un modo de programar software. Con los microservicios, las aplicaciones se dividen en sus elementos más pequeños e independientes entre sí. A diferencia del enfoque tradicional y monolítico de las aplicaciones, en el que todo se compila en una sola pieza, los microservicios son elementos independientes que funcionan en conjunto para llevar a cabo las mismas tareas. Cada uno de esos elementos o procesos es un microservicio. Este enfoque de desarrollo de software valora el nivel de detalle, la sencillez y la capacidad para [compartir un proceso similar en varias aplicaciones](#). Es un elemento fundamental de la [optimización del desarrollo de aplicaciones hacia un modelo nativo de la nube](#).

La arquitectura de microservicios (**MSA**, del inglés *Micro Services Architecture*) es un método de desarrollo de software que permite construir aplicaciones dividiéndolas en un conjunto de servicios (microservicios), simples, pequeños e independientes, interoperables y auto desplegados que tienen interfaces bien definidas. Las arquitecturas de microservicios hacen que las aplicaciones sean más fáciles de escalar y más rápidas de desarrollar. Esta característica permite la innovación y acelera el tiempo de comercialización de los nuevos servicios. Los microservicios intercambian datos y consultas entre sí; son accesibles utilizando interfaces de programación de aplicaciones (APIs) y pueden ser implementados

utilizando cualquier lenguaje de programación o sistemas de almacenamiento de datos. Los microservicios se despliegan en contenedores.

¿Cuáles son las ventajas de utilizar una infraestructura de microservicios?

La gran ventaja es la distribución de sistemas de software de calidad con mayor rapidez, lo cual es posible gracias a los microservicios que son componentes de software que pueden separar y ejecutar las aplicaciones de modo independiente. Sin embargo, dividir las aplicaciones en microservicios no es suficiente; es necesario administrarlos, coordinarlos y gestionar los datos que crean y modifican, para lo que se requieren otros componentes denominados *contenedores* que, a su vez, necesitan un sistema para su gestión y administración denominado *orquestración de contenedores* u *orquestradores*.

3.4 Contenedores

Los contenedores son una tecnología disruptiva y habilitadora de nuevas formas de desarrollo e implementación de sistemas. Los [contenedores de Linux](#) brindan a las aplicaciones basadas en microservicios una unidad ideal de implementación de aplicaciones y un entorno de ejecución autónomo. Esto le permite aprovechar mejor el hardware y coordinar los servicios con facilidad, lo que incluye el almacenamiento, las redes y la seguridad.

Por estas características, la fundación [Cloud Native Computing Foundation](#) afirma que los microservicios y los contenedores conforman, en conjunto, la base para [desarrollar aplicaciones nativas de la nube](#). Este modelo agiliza el desarrollo y facilita la transformación y la optimización de las aplicaciones actuales, y todo comienza con los microservicios en contenedores.

Los microservicios están muy ligados al concepto de *contenedor*, una unidad estándar de software que empaqueta el código junto a todas de sus dependencias para que el servicio o aplicación se ejecute de forma rápida y fiable de un entorno informático a otro.

De todas las soluciones de contenedores que en los últimos años se han presentado, Docker es sin lugar a duda, la solución más utilizada en todo el mundo. *Docker*, es una plataforma de contenedores cuyo fin es ejecutar aplicaciones de forma aislada y eficiente. Uno de los grandes motivos que justifican su popularidad es que trata los contenedores como si fueran máquinas virtuales extremadamente ligeras y modulares lo que ofrece una mejor experiencia de usuario frente al «contenedor clásico» y favorece el trabajo de los desarrolladores, por ejemplo, automatizando la tarea repetitiva de crear y desplegar microservicios dentro de los mismos.

Orquestrador de contenedores: Kubernetes

Kubernetes realiza la gestión de cada uno de los ciclos de vida de los contenedores mediante un sistema de código que permite automatizar la implementación, el escalado y la administración de aplicaciones en contenedores y su misión principal es actuar de orquestrador de contenedores. Kubernetes es un *orquestrador de contenedores de código abierto* para implementar y administrar aplicaciones en contenedores. Su adopción por parte de los equipos de desarrollo ha crecido exponencialmente hasta convertirse en un estándar para el despliegue de contenedores en el *cloud*.

4. *Data lakes* (Lagos de datos): Los nuevos depósitos de almacenamiento de datos

En la actualidad, los repositorios (almacenes de datos) de información de la empresa se dividen en dos grandes grupos:

- Los almacenes de datos (*data warehouses* y *data marts*) soportados por bases de datos relacionales que soportan datos estructurados organizados en filas y columnas (tablas).
- Los sistemas de *Big Data* que soportan grandes volúmenes de datos estructurados, no estructurados y semiestructurados, que se basan esencialmente en marcos de trabajo Hadoop y Spark.

Un almacén de datos (*data warehouse*) es el primer paso natural para almacenar todo tipo de datos de un proyecto, pero su efectividad se reduce cuando el proyecto madura y los datos comienzan a crecer. Todos los datos son estructurados y procesados de una determinada forma, reduciendo su agilidad y coste al gestionar grandes volúmenes de datos.

Con el objeto de resolver los problemas típicos de un almacén de datos, cuando los datos son de gran volumen y no son estructurados, nacieron las infraestructuras de Big Data con el sistema Hadoop y las bases de datos NoSQL y “en memoria” como soluciones eficaces y fiables. Sin embargo, estos sistemas pensados para la recopilación de grandes volúmenes de datos y de diferentes formatos (estructurados, no estructurados y semiestructurados), así como su creación a grandes velocidades, exige procesos de ETL de extracción, transformación y carga de los datos en bruto (*raw data*) para su almacenamiento y posterior análisis con las herramientas adecuadas.

Las empresas en sus estrategias organizativas pueden decidir conservar todos los datos capturados que generan sus diferentes fuentes de información y solo utilizarlos cuando los necesiten. Para tratar de resolver estos problemas, surgieron los *data lakes* (lagos de datos). Un lago de datos resuelve los problemas anteriores, ya que son sistemas de almacenamiento de datos sin procesamiento previo pero que exigen soluciones técnicas muy sofisticadas, y normalmente perfiles profesionales especializados que sean capaces de gestionar esos datos.

4.1 Origen del término *Data Lake*

El origen del término se le atribuye a James Dixon²⁷, CTO de Pentaho (uno de los proveedores más populares de software libre para herramientas de *Big Data*, fundamentalmente Hadoop), que en el blog oficial de la compañía y presentando sus soluciones de *Big Data*, introdujo en un artículo publicado en octubre de 2010, el término *data lake*, para referirse a las mejoras introducidas por Pentaho en los sistemas de *Big Data*. Analizando los factores comunes que había visto en numerosos casos, extrajo la conclusión de que la gran mayoría de las empresas manejaban datos estructurados, no estructurados o semiestructurados provenientes típicamente de una única aplicación o sistema, en un volumen que hacía inviable técnica o económicamente el uso de un sistema de base de datos relacional para su almacenamiento.

Dixon propuso una solución de cómo deberían almacenarse los datos, sin tener que pasar por un preprocesamiento que extrajera los atributos que en principio se pensase que iban a ser utilizados en las consultas, y sin tener que agregarlos necesariamente, para así no perder capacidad de ser usados a posteriori según otras necesidades. De ahí surge la analogía del "lago de datos", donde los datos (el agua del lago, según la analogía) provienen de un origen y "llenen el lago" sin haber sido aún limpiados, procesados o empaquetados a priori. De esa manera, los datos se almacenarán en crudo y en su totalidad, y serán las necesidades de los diferentes grupos de usuarios las que permitirán identificar las versiones filtradas o procesadas para resolver dichas necesidades (bien sea en forma de consultas preprocesadas por ser las más frecuentes, o como entrada para un *data warehouse* con una estructura de datos determinada para agregarlo a otras fuentes).

27 Disponible en línea: <<https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/>>.

4.2 Definición de *Data Lake*

Un *data lake* (**lago de datos**) es un repositorio de almacenamiento que contiene una gran cantidad de datos en bruto en su formato original, incluyendo datos estructurados, semiestructurados y no estructurados que se guardan sin ningún procesamiento (*raw data*). Los datos se guardan en bruto y sin ningún tipo de esquema; su estructura y los requisitos de datos no se definen hasta que estos se necesitan. En resumen, en un *data lake*, se recogen todos los datos (en forma directa o formato nativo), pero no se alteran, limpian o manipulan; su valor se mantiene en bruto (nativo) y no se transforma previamente hasta su análisis y explotación; los datos se vierten por las organizaciones y se recuperan cuando es necesario; y en ese instante; se procede a ordenarlos y darles una estructura que permite análisis posteriores.

Un lago de datos es un depósito de datos masivo y de fácil acceso para almacenar *Big Data*. Hadoop es la tecnología más utilizada para crear lagos de datos. En esencia, es un tipo de almacenamiento en el que la información tiene una estructura variable (diferentes tipos de datos, texto, imágenes, mensajes, audios, ubicaciones físicas), es masiva, de fácil y rápido acceso y resiliencia, sin atender a una lógica de negocio específica. Los lagos de datos almacenan los datos en su formato más básico (en bruto), se actualizan añadiendo más información, pero nunca se modifica la información ya existente. De este modo, los *data lakes* permiten almacenar los datos en bruto y estar disponibles en todo momento y -casi- en tiempo real en su formato original. El uso de fuentes muy variadas permite realizar análisis complejos y modelos predictivos.

En realidad, un *data lake* es un almacén de datos donde se guardan todos los datos que necesita la organización sin ningún tratamiento ni estructura; son almacenes de datos en crudo (archivos, logs, tablas de cualquier sistema ERP, tuits o cookies de los navegadores de usuarios). En esencia y dado que permite el almacenamiento de datos no estructurados, un lago de datos es un repositorio de datos Hadoop. Desde el punto de vista técnico, un lago de datos tiene una arquitectura plana de almacenamiento; a cada elemento dato de un lago se le asigna un identificador único y se le etiqueta con un conjunto de etiquetas de metadatos ampliados. Cuando se realiza una consulta al lago de datos en busca de datos se analiza el conjunto de datos para responder a dicha consulta

En resumen, el *data lake* no sustituye ni elimina los almacenes de datos (*data warehouses*), sino que amplía las capacidades de la organización en lo relativo a la gestión de datos voluminosos y variables. Los *data lake* se pueden desarrollar con múltiples tecnologías tales como: Hadoop, NoSQL, bases de datos relacionales.

4.3 *Data Lakehouse* (Almacenes de lagos de datos)

Un almacén de lago de datos (*data lakehouse*) es un sistema que combina los elementos de un almacén de datos (*data warehouse*) con los de un lago de datos (*data lake*). Los almacenes de lagos de datos implementan las estructuras de datos de los almacenes de datos y las propiedades de administración de datos de los lagos de datos, que, normalmente, son más económicos y rentables para almacenamiento de datos. Los almacenes de lagos de datos permiten almacenar y administra todo tipo de datos (estructurados, no estructurados y semiestructurados) son muy útiles para los científicos de datos ya que permiten realizar tareas de aprendizaje automático y de sistemas de información e inteligencia de negocios. Los almacenes de lagos de datos²⁸ normalmente almacenan todos los tipos de datos, al estilo de los lagos de datos, y, a continuación, se convierten a un formato que permite la administración de todos los tipos de datos, cosa que los lagos de datos solo realizan cuando se extraen datos para su consumo o visibilidad.

²⁸ <https://www.talend.com/resources/data-lake-vs-data-warehouse/>



Figura 4. *Data Warehouse, Data Lake, Data Lakehouse.* **Fuente:** Ben Lorica, Michael Armbrust, Reynold Xin, Matei Zaharia y Ali Ghodsi. *What Is a Lakehouse?* <https://www.databricks.com/blog/2020/01/30/what-is-a-data-lakehouse.html>

5. Inteligencia artificial: conceptos, definiciones y tecnologías

El Parlamento de la Unión Europea, en un artículo publicado²⁹ (en 2020 y actualizado en 2021), en su sitio Web hace una primera definición de inteligencia artificial (IA) “habilidad de una máquina de presentar las mismas capacidades que los seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear”. En el documento, se destaca que *“la IA permite que los sistemas tecnológicos perciban su entorno, se relacionen con él, resuelvan problemas y actúen con un fin específico. La máquina recibe datos (ya preparados o recopilados a través de sus propios sensores, por ejemplo, una cámara), los procesa y responde a ellos. Los sistemas de IA son capaces de adaptar su comportamiento en cierta medida, analizar los efectos de acciones previas y de trabajar de manera autónoma”*.

5.1 Aplicaciones de la IA: una primera revisión

En el mismo artículo la UE describe las aplicaciones de inteligencia artificial más utilizadas en la vida diaria:

- Compras por internet y publicidad
- Búsquedas en la web
- Asistentes personales digitales
- Traducciones automáticas
- Casas, ciudades e infraestructuras inteligentes
- Vehículos inteligentes
- Ciberseguridad
- La inteligencia artificial para luchar contra la Covid-19

También reseña cuales son los sectores o dominios donde mayor impacto está ofreciendo la IA

Salud. Los investigadores estudian cómo usar la IA para analizar grandes cantidades de datos sobre la salud para encontrar patrones que podrían llevar a nuevos descubrimientos en la medicina y a otras formas de mejorar los diagnósticos individuales.

Transporte. La inteligencia artificial podría mejorar la seguridad, velocidad y eficiencia del tráfico en sus diferentes sistemas de transporte.

²⁹ *¿Qué es la inteligencia artificial y cómo se usa?* (UE, Unión Europea). Actualización: 26-03-2021. <https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa>

Manufacturas. La inteligencia artificial puede ayudar a que los productores europeos sean más eficientes y potenciar las fábricas en Europa al usar robots, optimizar los recorridos de ventas o con predicciones puntuales del mantenimiento necesario o de averías en "fábricas inteligentes".

Comida y agricultura. La IA puede usarse para construir un sistema alimentario sostenible: podría garantizar comida más sana al minimizar el uso de fertilizantes, pesticidas y mejorar el riego, la productividad y reducir el impacto medioambiental. Además, los robots podrían quitar las malas hierbas y reducir el uso de herbicidas. En la UE, ya hay muchos granjeros que usan la IA para controlar el movimiento, la temperatura y el consumo de alimentos de sus ganados.

Administración pública y servicios. Al usar enormes cantidades de datos y reconocer patrones, la IA podría prever desastres naturales, permitir una preparación adecuada y reducir sus consecuencias.

5.2. Definición de inteligencia artificial

Aunque han surgido varias definiciones de inteligencia artificial (IA) en las últimas décadas, la referencia fundamental es la de John McCarthy, referencia mundial en IA y uno de los más reconocidos creadores de la disciplina. John McCarthy acuñó el término en 1956 y realizó la siguiente definición en su artículo³⁰ *"What is artificial intelligence?"* publicado en 2004: *"Inteligencia Artificial es la ciencia y la ingeniería de crear máquinas inteligentes, especialmente programas informáticos inteligentes. Está relacionada con la tarea similar de utilizar computadores para comprender la inteligencia humana, pero la IA no se limita a métodos que sean observables biológicamente"*.

Otras definiciones que ayudan a la comprensión de la IA son: IBM (una de las empresas multinacionales que inició sus estudios e investigaciones en sus orígenes y referencia mundial en la I+D+i en Inteligencia Artificial); una nueva definición ya "oficial" de la Comisión Europea y de las prestigiosas consultoras multinacionales en TIC: Gartner (Glossary IT)³¹ y McKinsey³² cuyas reseñas se adjuntan

Definición de IBM

IBM³³ considera que la inteligencia artificial utiliza computadores y máquinas para imitar las capacidades de resolución de problemas y toma de decisiones de la mente humana y, en su forma más simple, define la Inteligencia Artificial como "Una disciplina que combina la informática y sólidos conjuntos de datos para permitir la resolución de problemas. También abarca las subcategorías *machine learning* y *deep learning*, que aparecen mencionadas con frecuencia junto a la inteligencia artificial. Estas disciplinas se componen de algoritmos de IA que buscan crear sistemas expertos que hagan previsiones o clasificaciones en función de los datos de entrada".

Definición de Inteligencia Artificial de la Comisión Europea

³⁰ John McCarthy. *What is artificial intelligence?* <http://www-formal.stanford.edu/jmc/.2004 Nov 24>

³¹ Gartner. <https://www.gartner.com/en/information-technology/glossary/artificial-intelligence>

³² McKinsey. *An executive's guide to AI*.

<https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/An%20executives%20guide%20to%20AI/An-executives-guide-to-AI.ashx>.

³³ IBM Cloud Education. Inteligencia Artificial (IA). <https://www.ibm.com/es-es/cloud/learn/what-is-artificial-intelligence>, 3 de junio, 2020.

“La inteligencia artificial (IA) se refiere a los sistemas que muestran un comportamiento inteligente mediante el análisis de su entorno y la adopción de medidas, con cierto grado de autonomía, para lograr objetivos específicos. Los sistemas basados en IA pueden basarse puramente en software y actuar en el mundo virtual (p. ej., asistentes de voz, software de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento de voz y rostro) o la IA puede integrarse en dispositivos de hardware (p. ej., robots avanzados, automóviles autónomos, drones o aplicaciones de Internet de las Cosas)”³⁴.

5.3 Tipos de inteligencia artificial

Los tipos de Inteligencia Artificial más utilizados y reconocidos se agrupan en siete tipos o categorías, clasificados en dos grandes grupos: *basados en Capacidades* y *basados en Funcionalidades*.

- *Basados en Capacidades*: Artificial Narrow Intelligence (**ANI**); Artificial General Intelligence (**AGI**) y Artificial Super Intelligence (**ASI**). ANI se considera AI “*débil*”, mientras que los otros dos tipos se clasifican como “*fuertes*”.
- *Basados en Funcionalidades*: Máquinas Reactivas, Máquinas de memoria limitada, Teoría de la Mente y Autoconciencia.

Tipos de inteligencia artificial basados en capacidades

Esta clasificación es la más popular y aceptada universalmente en todo tipo de organizaciones y empresas.

Artificial Narrow Intelligence (ANI):

Inteligencia Artificial Estrecha (o Débil) [ANI]. Es básicamente la IA más utilizada en la actualidad. Es inteligencia artificial dedicada a resolver un problema específico o un conjunto de ellos de forma optimizada, pero sin posibilidad de extenderse a problemas generales sin la programación pertinente para ello. Carece de inteligencia de tipo humano.

ANI se define por su capacidad para completar una tarea específica, como ganar un juego de ajedrez o la identificación de individuo específico en una serie de motos. Aplicaciones básicas de ANI son: recomendaciones de productos de Amazon; noticias de Facebook; vehículos autónomos; *aplicaciones de reconocimiento de voz*; asistentes virtuales (**Siri, Alexa, Cortana**, ...); filtros de *spam* en correos electrónicos; sistemas de vehículos autónomos, ...

Artificial General Intelligence (AGI)

Inteligencia Artificial General (o Fuerte) [AGI] es la Inteligencia Artificial capaz de igualar o superar la inteligencia humana en capacidad de razonamiento y deducción, y puede realizar tareas en todos los ámbitos de la manera más hábil y flexible posible. Hoy en día es “un proyecto sólo existente en el sector de investigación y simulación, porque a pesar de que las máquinas ya nos superan en multitud de capacidades (incluso visión y reconocimiento auditivo en algunos ámbitos), no poseen sentimientos reales ni capacidades cognitivas nativas, así como conciencia propia y adaptabilidad a cualquier escenario”. El objetivo es conseguir “máquinas pensantes” con inteligencia comparable a la de la mente humana.

³⁴ Documento oficial de la Comisión Europea publicado el 8 de abril de 2019; “*A definition of AI: Main capabilities and scientific disciplines*”.

https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

En esta década y sobre todo en la segunda mitad se espera que ya existan proyectos en marcha AGI y con resultados prácticos. Empresas como OpenAI, Google o Microsoft ya han lanzado aplicaciones que se pueden considerar de Inteligencia Artificial General y que funcionan en numerosas aplicaciones de reconocimiento de imágenes, generación de textos, etc.

Artificial Super Intelligence, (ASI)

La Inteligencia Artificial SuperInteligente [ASI], es la perfección de la IA, ya que hará realidad la innovación final de la especie humana. Serán capaces de simular la inteligencia humana y serán muy superiores en todos los sentidos. Superaría la inteligencia y la capacidad de un ser humano con capacidad en todos los aspectos de emociones, investigación e incluso de respuestas ante desastres. Sin embargo, esta tecnología tan poderosa podrá implicar graves riesgos en la vida de las personas sino se controla eficientemente y en todo momento en tiempo real. Hoy la Super IA es un concepto hipotético que representa el futuro de la IA, pero en el que se investiga muy profundamente y los informes de grupos de investigación especializados consideran llegará en la década de 2040-2050.

Tipos de inteligencia artificial basados en funcionalidades

El profesor Hintze de la Universidad de Michigan, de Estados Unidos, considera que la IA deberá seguir el mismo proceso de aprendizaje que el de una persona y establece una clasificación de la IA con cuatro tipos de Inteligencia Artificial:

Máquinas Reactivas. Se trata del tipo más básico, ya que *no tienen la capacidad de formar recuerdos*. No existe, por lo tanto, la posibilidad de utilizarlos en la toma de decisiones. Un buen ejemplo de máquina reactiva es Deep Blue de IBM: identifica las piezas del tablero de ajedrez, sabe cómo se mueven y predice los mejores movimientos, pero no tienen en cuenta ningún proceso del pasado.

Memoria Limitada. Un paso por encima de las máquinas reactivas están aquellas que tienen memoria limitada. Un ejemplo de inteligencia artificial de este tipo son los *carros (coches) autónomos*. Identifican y monitorizan objetos en movimiento, como peatones y otros vehículos. Estos pequeños recuerdos son usados en el momento, pero *no son almacenados para ser utilizados en el futuro*, por lo que tampoco aprende de ellos.

Teoría de la mente. En este nivel las máquinas son capaces de *comprender que las personas, criaturas y objetos en el mundo pueden tener pensamientos y emociones* que afectan a su comportamiento. Si las máquinas van a relacionarse con nosotros, tendrán que ajustar su conducta. Deberán saber qué esperamos y cómo queremos que nos traten.

Autoconciencia. El nivel más complejo en inteligencia artificial presenta máquinas capaces de verse a sí mismas con perspectiva en su entorno. Los investigadores tendrán que comprender no solo la conciencia, sino también construir máquinas que la tengan. Es una etapa crucial para entender la inteligencia humana por sí misma.

5.4 Tecnologías de soporte de la IA

La Inteligencia Artificial en la década actual con el crecimiento continuo de los grandes volúmenes de datos (*big data*) y la consolidación de la **Ciencia de Datos** como un campo multidisciplinar ha obligado al conocimiento y uso de las tecnologías de soporte de la IA.

La inteligencia artificial tiene dos componentes principales: Aprendizaje Automático (ML, *Machine Learning*) y Aprendizaje Profundo (DL, *Deep Learning*) como subconjunto del Aprendizaje Automático (Figura 5). Otros componentes importantes son: las Redes Neuronales Artificiales (ANN, *Artificial Neural Networks*) -simulan el cerebro humano a través de un conjunto de algoritmos- (Figura 5); el Lenguaje de Procesamiento Natural (NPL, *Natural Processing Language*) y sus componentes fundamentales, Comprensión del Lenguaje Natural (NLU, *Natural Language Understanding*), Generación del Lenguaje Natural (NLG, *Natural Language Generation*). Sistemas de reconocimiento de texto-voz, TTS y AsR que han dado lugar a la confirmación de otra nueva disciplina la Inteligencia Artificial Conversacional. Los nuevos modelos de lenguaje están aumentando el uso de la Inteligencia Artificial Generativa.



Figura 5. Componentes principales de la Inteligencia Artificial

Aprendizaje automático

Machine Learning (ML) se traduce al español como “*aprendizaje automático*”, aunque también se suele traducir como “*aprendizaje máquina*”. Se suele considerar una rama de la inteligencia artificial que busca construir algoritmos que permiten a los computadores “aprender” a partir de conjuntos de datos y obtener como resultado un modelo que permita realizar predicciones basándose en dichos datos y no en instrucciones estáticas.

El aprendizaje automático es una disciplina que toma experiencias de otras disciplinas como la estadística, la complejidad computacional, ciencias de la computación e ingeniería. La expansión del aprendizaje automático como disciplina complementaria o autónoma de la inteligencia artificial se debe, esencialmente, al diluvio de los datos (*big data*) que se han producido estos últimos años y que hemos tratado ampliamente en capítulos anteriores. Hoy en día el aprendizaje automático está más que nunca al alcance de cualquier programador. Para experimentar con estos servicios plataformas de gran difusión son: IBM Watson Developer Cloud, Amazon Machine Learning, Azure Machine Learning, TensorFlow o BigML.

Aprendizaje profundo

Es una subcategoría del aprendizaje automático. El aprendizaje profundo trata del uso de redes neuronales artificiales para mejorar funcionalidades tales como el reconocimiento de voz, la visión por computador y el procesamiento del lenguaje natural. Rápidamente se está convirtiendo en uno de los campos más solicitados en informática (*ingeniería de sistemas o computación*). En los últimos años, el aprendizaje profundo ha ayudado a lograr avances en áreas tan diversas como la percepción de objetos, la traducción automática y el reconocimiento de voz (todas ellas áreas especialmente complejas para los investigadores en IA).

6 Ciencia de datos

El crecimiento exponencial de los datos en la década pasada y en la actual ha consolidado la disciplina de Ciencia de Datos y ha hecho que *se haya convertido en uno de los campos de más rápido crecimiento en todas las industrias*. El informe anual³⁵ de IDC de 2012 relativo a la presencia de datos en la Tierra, señalaba en aquel entonces los volúmenes de datos previstos para finales del año 2020 a 40 *zettabytes* (ZB) que suponía un crecimiento de más de 50 veces el volumen que estaba disponible a principios de 2010 y 2,8 ZB a finales de 2012 y además anunciaba que cada persona que habitaría en la Tierra en el citado año 2020 sería de aproximadamente 5,247 GB (aprox. 5 terabytes). Las organizaciones dependen cada vez más de los datos y de su información para su interpretación y proporcionar recomendaciones prácticas para mejorar los resultados comerciales.

Ciencia de datos es la extracción de información útil de grandes volúmenes de datos para obtener su valor significativo y poderlos utilizar en la toma de decisiones estratégicas en las organizaciones. La Ciencia de Datos es una especialización creciente que toca muchos de las siguientes áreas: *computación en nube, big data, matemáticas, estadística, métodos de optimización, teoría de negocios, teoría de ciencias de la computación... y las transversales de Internet de las Cosas, Blockchain, Ciberseguridad, ... y la futura integración con la Web 3.0 y el Metaverso*

Definición de Data Science (Ciencia de Datos)

El término Ciencia de Datos ha adquirido gran notoriedad en los últimos años y Ciencia de Datos es una ciencia multidisciplinar, y existe cierta unanimidad en considerar el diagrama de Venn, creado por Drew Conway -importante científico de datos estadounidense- en el año 2010, *The Data Science Venn Diagram* (figura 6), como la definición más empleada de *Data Science* y que, afortunadamente, su autor dejó de libre uso con licencia Creative Commons. Drew Conway hizo una representación gráfica de las disciplinas o áreas que comprenden la Ciencia de Datos, utilizando un diagrama de Venn.

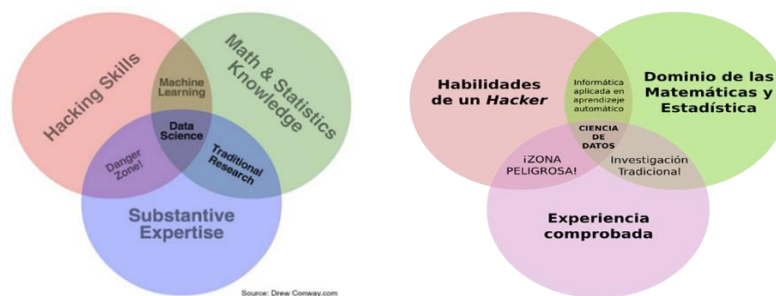


Figura 6. Disciplinas de la Ciencia de Datos (licencia Creative Commons) 2010. **Fuente:** Drew Conway [Traducida y adaptada]. <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

Las tres disciplinas esenciales, según Conway, que conforman la Ciencia de Datos son: **estadística y matemáticas** (*Math & Statistics Knowledge*), **informática y computación** (*Hacking Skills*) y **conocimiento del dominio o experiencia en el entorno** (*Substantive Expertise*), como puede ser el conocimiento que se debe tener del entorno; por ejemplo, sobre

³⁵ <https://www.computerweekly.com/news/2240174381/Data-to-grow-more-quickly-says-IDCs-Digital-Universe-study>

la gerencia, la publicidad o los recursos humanos. Por lo tanto, para que una persona desempeñe el perfil de **científico de datos** (experto en Ciencia de Datos), Conway considera que debe ser capaz de desempeñar estas tres actividades y competencias:

Estadística y Matemáticas (*Math and Statistics Knowledge*). Una vez extraídos los datos, el científico de datos deberá tener los conocimientos matemáticos necesarios para poder interpretarlos y procesarlos mediante las herramientas más adecuadas. La formación matemática es de gran importancia en la Ciencia de Datos.

Habilidades y competencias informáticas (*Hacking Skills*). La mayoría de los datos procederán de fuentes heterogéneas de datos; en consecuencia, deberá tener las habilidades necesarias para poder extraer, ordenar, procesar, analizar y visualizar estos datos. Debe crear los algoritmos necesarios, utilizando distintos lenguajes de programación para cada caso concreto.

Experiencia del entorno (*Substantive Expertise*) o conocimiento del dominio del sector. Para poder diseñar y desarrollar el análisis masivo de datos en diferentes casos de uso y aplicación, es necesario conocer el contexto. El científico de datos debe tener un alto conocimiento del entorno, que lo motive a plantear nuevos escenarios y crear nuevas hipótesis en las que trabajar, siempre cuidando la calidad de los datos, todo ello con la intención de obtener resultados que terminen incrementando el conocimiento del área de trabajo. Es de gran importancia el conocimiento del dominio en la Ciencia de Datos.

Además del conocimiento de las disciplinas anteriores, Conway también plantea la confluencia de otras áreas. En la Ciencia de Datos confluyen también el aprendizaje automático, la investigación tradicional y una tercera materia relacionada con las habilidades y conocimientos de *hacking*, que ya el mismo autor consideraba como una zona importante, pero de peligro o de incertidumbre.

- **Aprendizaje automático**. El conocimiento de algoritmos de aprendizaje automático permitirá obtener resultados adecuados a los objetivos previstos.
- **Investigación tradicional**. La diferencia entre el científico tradicional y el científico de datos radica, en gran parte, en las habilidades informáticas y conocimientos de lenguajes de programación que debe tener el científico de datos, que le permitirán poder manejar mucha más información y procesarla más rápidamente.
- **Zona de incertidumbre o zona de peligro**. Un científico de datos que, por ejemplo, no tenga destreza en los campos de la estadística y de las matemáticas, aunque tenga conocimiento del entorno y las habilidades informáticas así como del aprendizaje automático, es probable que procese los datos incorrectamente o los interprete de forma inadecuada y los resultados de la investigación pueden no tener validez, lo que implicará obtener unas conclusiones erróneas, que incluso podrían perjudicar a futuros proyectos que se pudieran basar en estos resultados incorrectos.

Como se ilustra en el diagrama de Venn, el científico de datos debe ser competente en las tres áreas básicas descritas anteriormente. Si no se tiene habilidad en alguna de estas áreas, entonces no se considera que estemos hablando de Ciencia de Datos. En resumen, la Ciencia de Datos, según el diagrama de Venn, es multidisciplinar (matemáticas y estadística, programación *hacking*, experiencia de dominio) y debe tener presente los principios de investigación tradicional, experiencia comprobada y un área de peligro o incertidumbre. Es un campo multidisciplinar que se encuentra entre ciencias de la computación (informática),

matemáticas y estadística, y comprende el uso de métodos y técnicas científicas para extraer conocimiento y valor de grandes volúmenes de datos estructurados o no estructurados.

Definiciones complementarias de Ciencia de Datos

Una definición más actualizada de Ciencia de Datos (Shah, 2020:5) es: “*Un campo de estudio y prácticas que implica la recolección, almacenamiento y procesamiento de datos, para obtener información importante sobre un problema o fenómeno. Tales datos pueden ser generados por humanos (informes, registros...) o máquinas (datos del tiempo, visión de una carretera...) y pueden estar en diferentes formatos (texto, audio, video, realidad virtual o aumentada, etc.)*”. Shah (2020) también considera que la ciencia de datos es un campo independiente por sí misma en lugar de un subconjunto de otro dominio, tal y como ya se ha señalado, estadística o informática.

Simplilearn³⁶, un prestigioso portal educativo proporciona la siguiente definición: “*La ciencia de datos es el dominio de estudio que se ocupa de grandes volúmenes de datos utilizando herramientas y técnicas modernas para encontrar patrones invisibles, obtener información significativa y tomar decisiones comerciales. La ciencia de datos utiliza complejos algoritmos de aprendizaje automático para crear modelos predictivos. Los datos utilizados para el análisis pueden provenir de muchas fuentes diferentes y presentarse en varios formatos*”.

Ciencia de Datos es el ámbito del conocimiento que engloba las habilidades asociadas a la extracción del conocimiento de datos, esencialmente *Big Data*, que pueda ser comprendido por los expertos del área del dominio o sector. Incorpora diferentes componentes y se basa en métodos, técnicas y herramientas de numerosos campos, que van desde las matemáticas y estadística, informática (ciencias de la computación), almacenamiento y procesamiento de datos, visualización, reconocimiento de patrones, aprendizaje automático y profundo y algoritmos, entre otros sectores. El objetivo central de la Ciencia de Datos es extraer el significado de datos y la creación de productos de datos.

La ciencia de datos combina matemáticas y estadística, programación especializada, análisis avanzado, inteligencia artificial (IA) y aprendizaje automático con experiencia en un tema específico para descubrir información procesable oculta en los datos de una organización. Estos conocimientos se pueden utilizar para impulsar la toma de decisiones y la planificación estratégica.

Definición de *Data Science* de IBM

IBM, una de las empresas a nivel mundial y líder en Big Data y Ciencia de Datos, define *Data Science*³⁷ (**Ciencia de Datos**) como: “*El proceso de utilizar algoritmos, métodos y sistemas para extraer conocimiento e ideas (insights) de datos estructurados y no estructurados. Puede ser utilizada para hacer predicciones y decisiones utilizando analítica y aprendizaje automático*”. IBM ya en sus primeras publicaciones consideraba también

³⁶ Simplilearn. *What is Data Science: Lifecycle, Applications, Prerequisites and Tools*. <https://www.simplilearn.com/tutorials/data-science-tutorial/what-is-data-science>. Última actualización Jul 27, 2022

³⁷ IBM. *What is data science?* <https://developer.ibm.com/technologies/data-science/>; IBM Cloud Education. *Data Science*. 2 Agosto 2022. www.ibm.com/cloud/learn/data-science-introduction.

Ciencia de Datos³⁸ como un área multidisciplinar que estaba cambiando el modo en que las organizaciones resuelven problemas y ganan ventaja competitiva, y que lo concentraba en las tres grandes disciplinas, siguiendo el modelo de Conway: **computer science (informática), matemáticas/estadística y dominio del conocimiento**. Además de estas áreas de conocimiento, se necesitan otras técnicas de computación avanzada e Inteligencia Artificial, como aprendizaje automático (*Machine Learning*), minería de datos, reconocimiento de patrones, almacenamiento de datos, procesamiento avanzado de bases de datos y técnicas y herramientas de visualización de datos.

En esencia, la ciencia de datos se puede considerar como un campo de estudio que implica la recolección, almacenamiento, procesamiento, análisis y la visualización de los datos, con el objeto de deducir información y conocimientos importantes en la resolución de un problema o proyecto específico. Tales datos pueden ser generados por humanos (informes, estudios, logs de la web) o las máquinas (datos del tiempo, visión de una carretera, etc.). Como ya se ha señalado se considerará como un campo o disciplina independientes que está integrado y converge con otros dominios como las matemáticas, la estadística, la programación, las ciencias de la computación, big data, la inteligencia artificial y el aprendizaje automático, o la Internet de las Cosas y la transversalidad de la Ciberseguridad y Blockchain.

En esencia, la Ciencia de Datos es multidisciplinar y contiene otras disciplinas, también independientes que tienen sus propias tecnologías y campos de aplicación, y convergen entre sí en beneficios de todas ellas.

6.1 Dominios y aplicaciones de la ciencia de datos

Shah (2020: 5-12) en su excelente obra, se hace la pregunta “¿Dónde se puede ver la Ciencia de Datos estos días? El gran impacto de la ciencia de datos en la actualidad y en un futuro próximo es que no está limitada a una faceta de la sociedad, un dominio o un departamento de una universidad; está virtualmente en todas partes”. Así Shah enumera un conjunto de aplicaciones y sectores donde se está aplicando con gran éxito la ciencia de datos: Finanzas, Políticas públicas (regulaciones y leyes), Vida diaria, Cuidado de la Salud, Política, Defensa, Planificación urbana, Educación, Bibliotecas, etc. (2020).

¿Por qué razón la industria y la academia incrementan su demanda de ciencia de datos y de científicos de datos (el rol profesional de la ciencia de datos)? se sigue preguntando Shah: ¿Qué ha cambiado en los pasados años y que cambiará en los próximos? La respuesta, señala Shah no es sorprendente: “tenemos muchos datos, continuamos generando una cantidad asombrosa de datos a una velocidad sin precedentes cada vez mayor; el análisis de los datos requiere sabiamente la participación de profesionales competentes y bien capacitados, de modo que el análisis de estos grandes volúmenes de datos proporcionará información procesable y muy eficiente”.

La figura 7 refleja la ciencia de datos como un superconjunto de la inteligencia artificial, aprendizaje automático y aprendizaje profundo (García-Maroto, Santamaría: 2021).

³⁸ M. Tim Jones. IBM. Data science and open source. IBM. Learn about open source tools for converting data into useful information. <https://www.ibm.com/developerworks/opensource/library/os-datascience/>

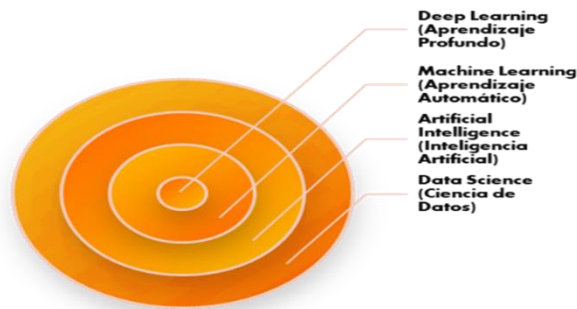


Figura 7. Ecosistema global de Ciencia de Datos

Fuente: María García-Maroto García y Rosalía Santamaría Muñoz, Fundación iS+D. <https://isdfundacion.org/2021/09/01/data-science-ciencia-de-datos-aclaracion-de-conceptos-basicos/>

6.2 El científico de datos

Ciencia de Datos (*Data Science*) tiene sus orígenes al final de la primera década del siglo XXI, y aunque existen muchas teorías sobre su nacimiento, parece que hay cierto acuerdo en que primero se popularizó el término **científico de datos** (*data scientist*) como un rol profesional experto en Ciencia de Datos. El objetivo de la Ciencia de Datos es la extracción de información útil de un conjunto de grandes volúmenes de datos (*big data*). Las compañías han reconocido el valor de los datos como un activo durante mucho tiempo; sin embargo, las enormes cantidades que ahora están disponibles necesitan nuevos medios para darles sentido y gestionarlos eficientemente. Por esta razón, han comenzado a proliferar ingenieros y científicos de datos que están construyendo sistemas para aplicar la Ciencia de Datos a grandes volúmenes de ellos. El advenimiento creciente de datos ha conducido a nuevos perfiles profesionales. Aunque son muchas las nuevas profesiones que han ido emergiendo, sin duda el científico de datos es el profesional más reconocido y demandado por organizaciones y empresas que desean gestionar y explotar los datos existentes en las empresas y fuera de ellas.

Un **científico de datos** (*data scientist*) es un experto de Ciencia de Datos que resuelve problemas complejos de diferentes sectores (negocios, finanzas, marketing, ciencias de la vida, industria, logística, etcétera), haciendo uso de análisis de datos, y extrae conocimiento de valor de las compañías para una toma de decisiones acertada y eficiente. El término fue acuñado por D.J. Patil³⁹, considerado como uno de los grandes científicos de datos actuales, y Jeff Hammerbacher⁴⁰ en 2008, que trabajaban en LinkedIn y Facebook, respectivamente.

Para Microsoft⁴¹ «Un científico de datos dirige proyectos de investigación para extraer información valiosa de macrodatos (*big data*) y tiene aptitudes de tecnología, matemáticas, negocios y comunicaciones. Las organizaciones usan esta información para tomar mejores decisiones, resolver problemas complejos y mejorar sus operaciones».

³⁹ T. D. J. Patil y Hilary Mason, *Data Driven. Creating a Data Culture* en la editorial O'Reilly. <https://oreily.ly/1aKXJw> <http://www.oreilly.com/data/free/files/data-driven.pdf>

⁴⁰ <https://linkd.in/17sqKNZ>

⁴¹ <https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-data-science-scientists/#what-scientist>

Así, un científico de datos es un experto que ha de tener formación multidisciplinar para resolver problemas complejos a partir del análisis de datos, extrayendo conocimiento y conclusiones para la toma de decisiones. Cada día se le requiere también conocimiento de negocios, de la Web y, sobre todo, de sociología e incluso de filosofía, esencialmente de ética empresarial. Pero también una formación avanzada en matemáticas, estadística, programación y sus diferentes lenguajes, Analítica de Datos —básicamente, Analítica de Big Data—, aprendizaje automático y visualización de datos.

Científico de datos: la profesión más sexy del siglo XXI (HBR)

Un artículo muy influyente, que se ha convertido en referencia mundial, se publicó en octubre de 2012 en la prestigiosa revista *Harvard Business Review*⁴²: *Científico de datos. La profesión más sexy del siglo XXI*. Fue escrito por dos expertos mundiales en datos y conocimiento: Tom Davenport (experto mundial en Gestión del Conocimiento y Capital Intelectual) y D. J. Patil (reconocido como uno de los primeros científicos de datos del mundo; de hecho, en 2015, el presidente Obama lo nombró *Chief Data Science* de los Estados Unidos). En julio de 2022 publicaron la continuación del artículo anterior “*Is Data Scientist Still the Sexiest Job of the 21st Century?*” en la misma revista HBR, confirmando y actualizando las actividades del científico de datos en 2022.

Davenport y Patil definen al científico de datos como un profesional que combina conocimientos de matemáticas, estadística y programación de computadoras, que se encarga de analizar los grandes volúmenes de datos (*Big Data*). Señalan que, a diferencia de la estadística tradicional que utiliza muestra de datos, el científico de datos aplica sus conocimientos estadísticos para resolver problemas de negocio, aplicando nuevas tecnologías que permiten realizar cálculos que hasta ahora no se podían realizar y que comprende todos los volúmenes de datos.

Comienzan su primer artículo analizando la situación de ese momento producida por el advenimiento de *Big Data* (el número de HBR estaba dedicado como tema central a *Big Data*) y la figura de Jonathan Goldman, doctor en ciencias físicas que llegó a LinkedIn en junio de 2006 aplicando métodos de científico de datos —de hecho, por esta circunstancia fue contratado— y convirtió a la red social LinkedInⁱ en la referencia mundial en redes sociales y, en particular, de profesionales. A continuación, explican quiénes son los científicos de datos y cómo encontrar al que se necesita en la empresa, en las organizaciones o en las administraciones públicas. Asimismo, describen, a modo de decálogo, cómo encontrar al científico de datos que necesita una compañía y, en consecuencia, sus competencias y características. En este catálogo de recomendaciones figuran desde universidades idóneas para la formación en la disciplina en aquellos momentos —hoy ya no sólo en los Estados Unidos, sino en España y Latinoamérica, proliferan las universidades que imparten cursos y maestrías de *Big Data* y Ciencia de Datos— hasta características profesionales de los candidatos y, sobre todo, empresas que son casos de éxito, donde los científicos de datos las han convertido en empresas líderes y de referencia mundial. También recomiendan las visitas a sitios de referencia en *Big Data* y *Data Science*, como Kaggle o TopCoder, asociaciones

⁴² Thomas Davenport y D. J. Patil, “Data Scientist: The sexiest Job of the 21st Century”, *Harvard Business Review*, October 2012. <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>. HBR (julio 2022). <https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century>

profesionales especializadas en estas disciplinas. Terminan su artículo explicando la razón del título: el nuevo trabajo “caliente” de la década o “la profesión más sexy”.

6.3 Roles profesionales relacionados con datos y certificaciones profesionales

Los términos relacionados con la Administración y Análisis de Datos suelen ser muy variables y, en ocasiones, sinónimos, por lo que es difícil asociar las competencias a los muchos roles profesionales que surgen a menudo en el ámbito empresarial o de investigación. No obstante, es frecuente encontrar los siguientes roles: analista de datos, ingeniero de datos, arquitecto de datos, científico de datos, ingeniero de visualización, además de ingeniero de *big data* y arquitecto de *big data*. Normalmente, el rol profesional va asociado a una, a todas o a algunas de las etapas de la arquitectura de *Big Data* o del proceso de Ciencia de Datos.

Científico de datos. Tiene una visión más horizontal de todo el proceso de Ciencia de Datos. Su tarea principal será la programación de algoritmos para el Análisis de Datos, pero debe conocer bien el negocio de la empresa (su plan de negocio, así como las líneas de negocio fundamentales). En definitiva, un científico de datos debe ser capaz de identificar aquellas variables relevantes para la empresa que ayuden a mejorar resultados, multiplicar el volumen de ventas, fidelizar a los clientes, ahorrar costes, etcétera. Los científicos de datos suelen mezclar, entre otros, conocimientos de matemáticas, estadística e informática, a los que es conveniente unir conocimientos de negocios, administración de empresas, ciencias de la salud y ciencias sociales. El científico de datos se ha consolidado ya como una profesión muy demandada en todo tipo de organizaciones y empresas (grandes y pequeñas). El científico de datos jefe, sobre todo en las grandes multinacionales, se está comenzando a reconvertir en el nuevo rol profesional de director de datos (**CDO**, *Chief Data Officer*), que es el responsable de toda la estrategia y política de datos de las organizaciones y con dependencia directa del presidente o director general.

Así, a los perfiles o roles profesionales ya implantados de Director de Tecnología (**CTO**, *Chief Technology Officer*), con unas competencias muy abiertas y transversales en toda la organización y empresa con independencia del clásico CIO, y también con un alto nivel directivo y estratégico, Director de Informática o de Sistemas de Información (**CIO**, *Chief Information Officer*) y Director de Seguridad de la Información (**CISO**, *Chief Information Security Officer*), hay que sumar dos perfiles profesionales que cada día serán muy demandados en la administración, organizaciones y en empresas de todo tipo:

Chief Data Officer (CDO) o Director de Datos. Es un cargo dependiente del máximo ejecutivo de la empresa, a quien reporta directamente. Une las dos disciplinas hoy imperantes en la transformación digital: *Big Data* y ciberseguridad. Su misión es impulsar el crecimiento de la organización o empresa mediante la transformación digital de la misma. Ha de tener una amplia visión de la gestión de la empresa y del mundo digital, de análisis de datos y de seguridad de la información, con el objeto de diseñar estrategias y políticas de ciberseguridad. Requiere una formación multidisciplinar de ingeniería, estadística, análisis digital, ciberseguridad, además de una amplia visión y conocimiento de la empresa. El director de datos debe velar por el buen Gobierno y la gestión de la Calidad de los datos. El CDO suele tener en su equipo dos roles profesionales muy destacados: el científico de datos (especialista en analítica de datos y negocios) y el ingeniero de datos (desarrollador de plataformas de big data como Hadoop, Spark, NoSQL, etc.).

Data Protection Officer (DPO) Director o Delegado de Protección de Datos. Es un perfil jurídico, pero con formación informática y de seguridad de la información y ciberseguridad,

aunque también podría ser un tecnólogo con formación jurídica en protección de datos y privacidad. Este perfil es exigible en la administración y en determinadas empresas de la Unión Europea desde el 25 de mayo de 2018, que entró en vigor el Reglamento General de Protección de Datos (GDPR).

También comienza a tener bastante fuerza en grandes y medianas empresas el rol del **CTO** (*Chief Transformation Officer*), con las competencias de director de transformación digital de la corporación y con la responsabilidad de liderar dicha transformación digital y la digitalización necesaria.

HABILIDADES DEL CIENTÍFICO DE DATOS: Técnicas (“Duras”), No técnicas (“Blandas”)

El científico de datos⁴³ requiere una formación multidisciplinar profunda y unas habilidades que se suelen agrupar en dos categorías: Técnicas (“Duras”), No técnicas (“Blandas”)

Habilidades Técnicas

1. Matemáticas y Estadística
2. Programación (Python, R, SQL, ...)
3. Bases de datos relacionales, NoSQL, NewSQL, y “en memoria”
4. Aprendizaje Automático y Aprendizaje Profundo
5. Procesamiento y Análisis de datos, Minería de datos
6. Técnicas y herramientas de visualización de datos y narraciones de datos “*Data Storytelling*”
7. Administración y experiencia de proyectos del mundo real.
8. Herramientas y marcos de trabajo de *Big data* (Hadoop, Spark, ...)
9. ...

Habilidades no Técnicas

1. Pensamiento crítico
2. Comunicación efectiva y habilidades de comunicación
3. Solución de problemas proactiva
4. Curiosidad intelectual
5. Visión y perspicacia para los negocios
6. Comprensión y sentido del negocio
7. Ética en el desarrollo de proyectos
8. Trabajo en equipo
9. ...

Algunas empresas refuerzan la necesidad de *habilidades “sociales” (de comunicación)* que le permitan presentar las conclusiones obtenidas de manera exitosa a cualquier “nivel” de la organización y deberá ser por tanto *un buen comunicador capaz de empatizar con la audiencia y habilidades de “negocio” (visión para los negocios)* que le permita ser *un apasionado del negocio y sus datos, conocer bien la empresa, sus objetivos, necesidades, preocupaciones y motivaciones y, por supuesto, conocer la competencia*. Todo ello le proporcionará el contexto adecuado en el que interpretar los datos.

⁴³ KDnuggets. <https://www.kdnuggets.com/2020/10/data-science-minimum-10-essential-skills.html>

NUEVAS CARRERAS UNIVERSITARIAS y CERTIFICACIONES: Ingeniería, Maestrías, Diplomados...

Las universidades y escuelas de negocio están desarrollando nuevas carreras donde se ofrecen formación en inteligencia artificial en convergencia con la ciencia de datos y el conocimiento y dominio de las tecnologías disruptivas junto también con el conocimiento de las tecnologías emergentes de los próximos años. Así nuevas carreras que se pueden ver en centros españoles y latinoamericanos son grados e ingenierías tales como:

- Ingeniería en Inteligencia Artificial
- Ingeniería en Ciencia de Datos
- Ingeniería en *Cloud Computing*
- Ingeniería en Internet de las Cosas
- Ingeniería en Aprendizaje Automático (*Machine Learning*)
- Ingeniería en datos y en Inteligencia Artificial
- Ingeniería en Big Data y Análisis de Datos (Analytics)

En la actualidad, existe un gran número de universidades y escuelas de negocio que imparten cursos de maestría, especializaciones, diplomados en las diferentes materias que componen la Ciencia de Datos. Asimismo, y en beneficio de usuarios personales y empresas, existen muchos cursos gratuitos en plataformas **MOOC** (cursos masivos en línea y abiertos), donde es posible formarse y obtener diplomas y certificaciones que acrediten una formación avanzada. Las plataformas más populares y acreditadas son **Coursera**, **Udemy**, **edX** y **MiriadaX**. Y las certificaciones oficiales son muy numerosas desde las clásicas de seguridad como **CISA**, **CISSP**, **CISM** hasta certificaciones profesionales muy reconocidas impartidas por empresas como SAP, IBM, Microsoft, HP, SAS, EMC y Coursera.

7. Inteligencia artificial conversacional

IBM⁴⁴, define la Inteligencia Artificial Conversacional como “*un conjunto de tecnologías como chatbots o agentes virtuales que permiten a los usuarios conversar entre ellos. Utilizan grandes volúmenes de datos, aprendizaje automático y procesamiento de lenguaje natural para ayudar a imitar interacciones humanas, reconociendo las entradas de voz y de texto, y traduciendo sus mensajes a varios idiomas*”. **Microsoft**⁴⁵ define la IAC como “*aquella que hace referencia a un conjunto de tecnologías que permiten que los dispositivos y aplicaciones, como los bots de chat [chatbots], conversen con personas empleando sus lenguajes naturales*”. **Inteligencia artificial conversacional o IA conversacional** es cualquier máquina con la que una persona puede hablar y, por lo general, está involucrada hoy a través de *chatbots* y asistentes de voz.

Componentes de la IA Conversacional

Los componentes fundamentales de la IA Conversacional son los chatbots, asistentes virtuales y altavoces inteligentes, aunque también se suelen considerar por su importancia en la era actual y, esencialmente, con el despliegue creciente del teletrabajo, a las plataformas virtuales para videoconferencias.

⁴⁴ IBM Cloud Education. *Conversational AI*: [_https://www.ibm.com/cloud/learn/conversational-ai](https://www.ibm.com/cloud/learn/conversational-ai). 31 de agosto 2020.

⁴⁵ Microsoft. *¿Qué es la IA conversacional?* <https://powervirtualagents.microsoft.com/es-es/conversational-ai/>

Bots y chatbots; asistentes virtuales

¿Qué es un *bot*? Un *bot* es un software de inteligencia artificial diseñado para realizar una serie de tareas por su cuenta y sin la ayuda del ser humano, como hacer una reserva en un restaurante, marcar una fecha en el calendario o recoger y mostrar información a los usuarios. El *bot* es capaz de comunicarse con los seres humanos (a través de texto, voz, emociones...) manteniendo una conversación con una persona utilizando el lenguaje natural en dicha conversación. Un *chatbot* (*chatter bot*) o *bot conversacional*, es el modelo de *bot* más popular, capaz de simular una conversación con una persona y se ha integrado en las aplicaciones de mensajería, tipo *chat*. El *chatbot* ofrece un servicio a través de una conversación con el usuario y fundamentalmente por esta razón comienzan a estar muy presentes en aplicaciones de mensajería. Existen numerosas soluciones comerciales de inteligencia conversacional:

Chatbots (presentes al entrar en muchos sitios web de medios de comunicación, de empresa, etc.)... Herramientas de creación de chatbots: Amazon Lex, IBM Watson Assistant, Google “Dialogflow”, ...

Asistentes virtuales. Google Assistant, Alexa de Amazon, Siri de Apple, Cortana de Microsoft.

Altavoces inteligentes. Amazon Echo (Alexa) [Amazon Echo Dot, Amazon Echo Studio]; Apple con Siri; Google Home; Sonos One, ... Apple HomePod 2, presentado en febrero de 2023.

7.1 Arquitectura clásica de un Sistema de Diálogo Hablado (Sistema conversacional)

Los sistemas de diálogo hablado, componente nuclear de la inteligencia conversacional, requieren de diferentes tecnologías para procesar y generar lenguaje natural con una calidad que permita la interacción natural con los humanos, tarea, por otra parte, muy compleja. Las etapas clásicas de un sistema conversacional con una entrada de audio (voz) [también podría ser una entrada de texto] son:

1. *Reconocimiento automático de voz (ASR, Automatic Speech Recognition)* [Asistentes de voz de Googleo Amazon]: salida de texto
2. *Comprensión del lenguaje natural (NLU, Natural Language Understanding)*: salida de semántica
3. *Gestión del diálogo (DM, Dialog Management)*: salida semántica de la respuesta al usuario.
4. *Generación de lenguaje natural (NLG, Natural Language Generation)*: salida de texto
5. *Síntesis de Texto a Voz (TTS, Text-To-Speech synthesis)*. Salida de audio: [Google Translator]

El usuario mediante un dispositivo de IA conversacional, un asistente virtual o un chatbot como Siri o Alexa (teléfono, Web, ...) realiza una entrada de audio (voz) [también podría ser texto como en un *chatbot*] para solicitar una consulta o realizar una pregunta. El *audio* del usuario llega a la etapa 1 en la que el componente de reconocimiento automático de voz (ASR) convierte en salida *texto* que, a su vez, es la entrada a la etapa 2 de comprensión del lenguaje natural (NLU) que lo convierte en semántica del texto, que, a su vez, es la entrada a la etapa 3 de gestión del diálogo (DM). El sistema de gestión de diálogo, con su información disponible puede contestar directamente, pero normalmente necesitará consultar al sistema de información de la organización y empresa que le ayudará en la toma de decisiones para responder a la pregunta o consulta, de forma que producirá una semántica adecuada y fiable que se enviará al componente de generación natural (NLG) que es la etapa

4 y que ya lo convertirá en el texto final correspondiente y este texto final se enviará al componente de síntesis de texto a voz (TTS), etapa 5, que lo convertirá, de nuevo, en mensajes de audio que se enviarán finalmente al usuario que serán reconocidos como respuesta a la pregunta o consulta realizada..

8. Inteligencia artificial generativa

La consultora Gartner fue una de las primeras instituciones que hizo las primeras referencias de Inteligencia Artificial Generativa, allá por los finales de la década pasada, aunque ha sido con ocasión de la publicación de sus dos informes anuales de referencia *Hype Cycle de Tecnologías Emergentes* (ver apartado) publicados en agosto de 2021 y el de *Tendencias Estratégicas para 2022* publicado en octubre de 2021 cuando ya ha anunciado sus predicciones de llegada al mercado de tecnologías y herramientas de IA Generativa. Así Gartner⁴⁶ en dichos informes ya define la Inteligencia Artificial Generativa como “métodos de aprendizaje automático que aprenden sobre contenido u objetos a partir de sus datos y los usan para generar artefactos completamente nuevos, completamente originales y realistas” y anunció que “La inteligencia artificial generativa mejorará la calidad, el rendimiento y la accesibilidad de los productos digitales, al tiempo que reducirá el tiempo de comercialización. Para obtener estas ventajas competitivas, los líderes de ingeniería de software deben introducir ahora la IA generativa mediante la promulgación de cambios en las personas, los procesos y las herramientas. La IA generativa se correlaciona con los programas que permiten que las máquinas usen elementos como archivos de audio, texto e imágenes para producir nuevo contenido plausible”. En enero de 2022⁴⁷ volvió a publicar otro informe sobre IA Generativa en la que destaca que:

“La IA generativa es una tecnología disruptiva que puede generar artefactos que anteriormente dependían de los humanos, lo que ofrece resultados innovadores sin los sesgos de las experiencias humanas y los procesos de pensamiento. Los líderes de TI deben utilizar la gobernanza adecuada para explotar su potencial transformador”.

8.1 Redes generativas adversariales

La tecnología fundamental de soporte de la IA generativa está constituida, esencialmente por las redes GAN (Generative Adversarial Network). Una GAN es un conjunto de dos redes neuronales que compiten en el proceso de información. Una red genera imágenes (la *red generativa*) y la otra (la *red discriminadora*) evalúa las imágenes. De este modo, las redes GAN no requieren intervención humana de ningún tipo. Las máquinas se enseñan así mismas con muestras de imágenes.

Las Redes Generativas Antagónicas son una clase de algoritmos de inteligencia artificial que se utilizan en el aprendizaje no supervisado, implementadas por un sistema de dos redes neuronales que compiten mutuamente en una especie de juego de suma cero.

8.2 Aplicaciones de Inteligencia Artificial Generativa

Las aplicaciones de la IA Generativa son numerosas y de gran aplicación en numerosos campos y sectores de la vida actual, tanto a nivel de usuario como en organizaciones y

⁴⁶ [https:// www.gartner.com/en/documents/4006921](https://www.gartner.com/en/documents/4006921)

⁴⁷ Gartner. 11 January 2022; [https:// www.gartner.com/en/documents/4006921](https://www.gartner.com/en/documents/4006921)
<https://www.gartner.com/en/documents/4010235>

empresas. Una breve descripción de aplicaciones originales y de gran impacto se muestra a continuación.

- **Conversión Imagen a Imagen:** Traduce una imagen a otra. P. e. fotografías en blanco y negro a color, fotos de día a fotos de noche, una foto a una pintura artística o fotos satelitales a las vistas de Google Maps.
- **Traducción de texto a imágenes:** produce fotografías realistas a partir de descripciones textuales de objetos simple como pájaros y flores.
- **Envejecimiento facial:** genera una versión anterior de rostros a partir de una foto joven.
- **Procesamiento de imágenes:** puede confiar en esta tecnología para mejorar las capacidades de procesamiento de imágenes, actualizando imágenes de baja resolución a imágenes de alta resolución.
- **Restauración de películas:** la IA generativa mejora las películas antiguas al ampliarlas a resoluciones de imágenes y vídeo, 4K y más. Elimina el ruido, agrega colores y lo hace nítido.
- **Síntesis de audio:** la IA generativa puede convertir cualquier voz generada por computadora en una que realmente suene como una voz humana.
- **Fotos a emojis: Cambia fotos reales a emojis o caritas de dibujos animados.** (Ejemplo de fotografías de celebridades y emojis generados por GAN. Fuente: machinelearningmastery.com)

La IA Generativa se puede utilizar en una variedad de actividades como la creación de código de software, la facilitación de desarrollo de medicamentos, el marketing dirigido, generación de rostros de personas que no existen con resultados extraordinarios... pero existen grandes peligros también en su utilización, tales como: *Estafas, fraudes, desinformación política, identidades falsas, crean imágenes falsas que se parecen a imágenes reales, etc.*

Por esta razón se requiere un uso ético de la IA Generativa y el cumplimiento de todas las normativas nacionales e internacionales sobre privacidad, protección de datos y normas del uso y aplicaciones de la Inteligencia Artificial

8.3 Modelos de lenguaje en la inteligencia artificial generativa

Un modelo de lenguaje es un sistema de IA que entiende y genera texto. Es una herramienta estadística que analiza el patrón del lenguaje humano para la predicción de palabras. Utiliza diferentes técnicas estadísticas y probabilísticas para determinar la probabilidad de que ocurra una secuencia dada de palabras en una frase. Los modelos de lenguaje analizan los cuerpos de los datos de texto para proporcionar una base para sus predicciones de palabras. Se utilizan normalmente en aplicaciones de NLP, especialmente aquellas que generan texto como salida. Aplicaciones típicas son traducciones y respuestas a preguntas.

Los modelos de lenguaje determinan la probabilidad analizando datos de texto. Interpretan estos datos alimentándolos a través de un algoritmo que establece reglas para el contexto en lenguaje natural. Luego, el modelo aplica estas reglas en tareas de lenguaje para predecir con precisión o producir nuevas oraciones. El modelo esencialmente aprende las funciones y características del lenguaje básico y usa esas funciones para comprender nuevas frases.

Los modelos de lenguaje grande (*LLM, large language model*) tienen, en general, decenas de gigabytes de tamaño y están entrenados con enormes cantidades de datos de texto,

a veces a escala de petabytes. También se encuentran entre los modelos más grandes en términos de recuento de parámetros, donde un "parámetro" se refiere a un valor que el modelo puede cambiar de forma independiente a medida que aprende. Los parámetros son las partes del modelo aprendidas a partir de los datos de entrenamiento históricos y esencialmente definen la habilidad del modelo en un problema, como la generación de texto.

Existen ya numerosos modelos de lenguaje grande, siendo los más populares GPT-3 de OpenAI, PaLM y LaMDA de Google, y OPT-175 de Meta (Facebook). Todos ellos con grandes aplicaciones y abiertos para usuarios. Vamos a detallar GPT-3 por ser uno de los primeros modelos presentados y desarrollados, y ChatGPT un modelo de lenguaje conversacional de gran impacto, ambos creados por la empresa OpenAI.

GPT-3

GPT (*Generative Pre - Trained Transformer*) es un modelo de lenguaje de IA creado por la empresa OpenAI. GPT-3 se publicó en mayo de 2020 y está a disposición de desarrolladores a través de una API, aunque todavía no está "abierto" para su uso general y solo para desarrolladores registrados donde pueden descargar la API correspondiente para su uso en investigación y desarrollo de aplicaciones.

*GPT-3 es un generador de texto ya que permite la creación de texto de forma automática. Es capaz de generar imágenes basadas en un texto. Puede escribir automáticamente correos electrónicos, noticias y otros textos, así como crear un código en diferentes lenguajes de programación*⁴⁸. GPT-3 puede traducir entre idiomas, narrar historias, crear artículos casi indistinguibles de los escritos por humanos. Es capaz de completar un diálogo entre dos personas, continuar una serie de preguntas y respuestas o "terminar un poema o un capítulo de un libro, al estilo Cervantes".

Un modelo de GPT3, muy popular y de gran impacto, es DALL-E2 (creada por la empresa OpenAI), un sistema (modelo) muy avanzado de *creación de imágenes en base a una descripción de texto y/o oral*. Otros modelos de aplicación de creación de imágenes, también muy populares son: MidJourney y Stable Diffusion.

ChatGPT

Una de las mayores innovaciones tecnológicas de los últimos tiempos en inteligencia artificial generativa es el modelo de lenguaje ChatGPT -también creado por la empresa OpenAI y basado en GPT-3; es un chatbot conversacional inteligente entrenado para mantener conversaciones de todo tipo, con petición de preguntas, consultas, informes, traducciones, escritura de código, etc. Proporciona respuestas muy acertadas y completas, se expresa de modo natural y con una información muy completa. Fue lanzado el 30 de noviembre de 2022 y en solo cinco días alcanzó la cifra de un millón de usuarios registrados, lo que demuestra la popularidad alcanzada y su eficacia. Está entrenado con cientos de miles de millones de palabras de entradas de Internet, textos, revistas, etc.

ChatGPT es un modelo de lenguaje grande (LLM) basado en GPT-3.5, la última versión del modelo GPT-3. La definición dada por su empresa creadora y que se puede ver en la página web oficial (www.openai.co/blog/chatgpt) es el siguiente: «Chatbot: Optimización de

⁴⁸ Un artículo muy ilustrativo sobre la tecnología GPT-3 se puede leer en el Blog Oficial del banco BBVA. *¿Qué es GPT-3?: la inteligencia artificial que se encargará de escribir por ti.* <https://www.bbva.com/es/que-es-gpt-3-la-inteligencia-artificial-que-se-encargara-de-escribir-por-ti/>.

modelos de lenguaje para el diálogo: Hemos entrenado un modelo llamado ChatGPT que interactúa de forma conversacional. El formato de diálogo hace posible que ChatGPT responda preguntas de seguimiento, admita sus errores, cuestione preguntas incorrectas y rechace solicitudes inapropiadas. ChatGPT es un modelo hermano de Instructor GPT que está capacitado para seguir una instrucción en un aviso y proporciona una respuesta detallada».

La forma de utilizar ChatGPT es muy fácil: 1. Entrar en la página web oficial (chat.openai.com); 2. crear una cuenta personal y 3. Iniciar la conversación. Otra de sus grandes ventajas es que, por ahora, es totalmente gratis. Es necesario, sin embargo, tener ciertas precauciones, primero como advierte el sitio oficial puede cometer errores, pocos, pero puede cometer, y hay que tener precaución con todo lo que escribe el usuario ya que queda registrado y podría ser revisado por los desarrolladores de OpenAI para seguir entrenando a ChatGPT; por ello (por ahora), no es conveniente incluir información personal del usuario, ni peticiones que sean potencialmente peligrosas o delictivas. La propia OpenAI “exige” que no se pueden generar imágenes de tipo violento o explícito.

Así pues, un resumen de las innumerables actividades que se pueden realizar y solicitar son:

- Realizar consultas, pedir información, diversas.
- Escribir artículos o resúmenes, incluso se le puede pedir un número máximo de caracteres o líneas. Es idóneo para temas culturales, redacciones, realización de trabajos académicos o literarios, etc.
- Escribir textos diversos, como un guion para YouTube o TikTok, o temas de historia.
- Escribir poemas, chistes, letras de canciones...
- Escribir líneas de código de programación para diseño de páginas web
- ...

Es una herramienta que actúa como un buscador o navegador de Internet, aunque todavía no tiene las propiedades que tienen estas aplicaciones, pero si continua el entrenamiento de modo tan eficiente, en algún momento podrá llegar a competir con Google, Bing, ...

Bard de Google y Prometeo de Microsoft

El modelo de lenguaje ChatGPT3 como se ha comentado anteriormente, fue la gran noticia de impacto en modelos de lenguaje del último bimestre de 2023 debido a su gran resonancia mediática. A primeros de febrero de 2023, Google presentó Bard⁴⁹ y Microsoft, Prometeo⁵⁰ (Prometheus Model), un nuevo modelo de lenguaje de última generación desarrollado juntamente con OpenAI que integra su buscador Bing y su navegador Edge con inteligencia artificial y basado en ChatGPT3 dado que Microsoft es socio de la empresa OpenAI creadora del modelo GPT.

Estos modelos unidos a ChatGPT3 y los generadores de video y texto como DALL-E 2, Midjourney, Stable Diffusion y el generador de código Copilot -entre otras innovaciones-, anuncian un año 2023 en el que la inteligencia artificial se desplegará e instalará en un gran

⁴⁹ [acceso 10 febrero, 2023]. <https://blog.google/technology/ai/bard-google-ai-search-updates/>

⁵⁰ [acceso 10 febrero, 2023]. <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>

número de organizaciones y empresas, así como centros de formación universitaria y profesional, y grupos y laboratorios de investigación y facilitaran el acceso a gran número de usuarios, con versiones gratuitas aunque también, lentamente, se irán presentando versiones premium ya de pago.

9. Tendencias tecnológicas de gran impacto en el horizonte 2025-2030

De las tecnologías disruptivas analizadas en el apartado 2 destacamos dos que se están considerando tendencias tecnológicas de gran impacto: experiencias inmersivas y gemelos digitales. A estas tecnologías hay que unir la potencia de los modelos de Inteligencia Artificial Conversacional y Generativa. Sin embargo, otras tecnologías emergentes que cada día se utilizan más y con grandes propiedades mirando al horizonte 2025-2030 son: *NFT (Non-Fungible Token)* integrada en *Blockchain*, el *metaverso*, la *computación cuántica*, la naciente *Web3 (Web 3.0)* y la evolución creciente de la *hiperconectividad inteligente* (futuras Redes 6G y WiFi 7).

9.1 Tecnologías NFT

Las tecnologías **NFT**⁵¹ (*Non Fungible Token, Token no fungibles*) son activos digitales únicos, cifrados y vinculados a la cadena de bloques (*blockchain*). Estos no pueden ser intercambiados por otros, debido a su naturaleza única e irrepetible. Y gracias a sus funciones únicas, los NFT tampoco se pueden copiar ni dividir en unidades fraccionarias. Son simples archivos de computador que contienen una imagen, un vídeo, música, un documento, una pintura, una escultura... cualquier elemento digital puede convertirse en NFT. Otros activos NFT pueden representar: Colecciones de arte, Deportes, Música, Videos, Animación, Documentos Legales, Firmas, Documentación de un coche, Cromos, Simuladores (juegos, videojuegos...), etc. *No Fungible*, define algo único, que no se puede sustituir ni reemplazar por otro. *Activos fungibles* son: bitcoin, copias de archivos .mp3, .jpg, ... , monedas ordinarias...

¿Cómo funcionan los NFT? A los NFT, se les asigna una especie de certificado digital de autenticidad, una serie de metadatos que no se van a poder modificar. En estos metadatos se garantiza su autenticidad, se registra el valor de partida y todas las adquisiciones o transacciones que se hayan hecho, y también a su autor. Es decir, si se compra un contenido digital *tokenizado* con NFT, siempre habrá constancia del primer valor que tenía, y por cuánto se ha comprado. Es como cuando se compra un cuadro y se lleva un registro de por dónde se está moviendo. Los NFT están comenzando a utilizarse en diferentes categorías de simuladores y numerosos y rigurosos estudios reflejan una adopción creciente, tanto por parte de usuarios personales como por organizaciones y empresas.

9.2 El metaverso

El término metaverso (todavía no ha sido reconocido como término en el DRAE) viene de la novela de 1992 *Snow Crash* y es un término que se ha asentado para describir visiones de espacios de trabajo tridimensionales o virtuales. Este metaverso, por lo tanto, significa un mundo virtual al que nos podemos conectar utilizando una serie de dispositivos y con el que podemos interactuar con todos sus elementos, y que ha sido creado para parecerse a una realidad externa. El metaverso se está desarrollando mediante tecnologías disruptivas que

⁵¹ Juan Antonio Pascual Escapé, 13 de marzo de 2022. *Guía y todo lo que debes saber para crear y vender tus propios archivos NFT*. <https://computerhoy.com/reportajes/tecnologia/guia-todo-debes-saber-crear-vender-nft-1002911>

ya están consolidadas, así como la implantación de nuevas tecnologías emergentes. Las tecnologías más empleadas en el diseño y desarrollo del metaverso son:

- Inteligencia Artificial y Aprendizaje Automático.
- Realidad Extendida: Realidad Virtual, Realidad Aumentada y Realidad Mixta (que potenciará la experiencia inmersiva).
- Tecnologías conversacionales con el soporte del Procesamiento de Lenguaje Natural y otras tecnologías de la Inteligencia Artificial Conversacional.
- Diseño en 3D y 4D.
- Sensores que permiten recrear experiencias de olfato y táctiles.

Como grandes novedades de tecnologías disruptivas de uso en el metaverso, son las tecnologías ya descritas anteriormente junto al despliegue creciente y exponencial de la Inteligencia Artificial Conversacional con el soporte de los *Chatbots* y Asistentes Virtuales Inteligentes y comienzan a irrumpir las tecnologías de Inteligencia Generativa

El Metaverso es un mundo virtual, uno al que nos conectaremos utilizando una serie de dispositivos que nos harán pensar que realmente estamos dentro de él, interactuando con todos sus elementos. Será como realmente teletransportarse a un mundo totalmente nuevo a través de gafas de realidad virtual y otros complementos que nos permitirán interactuar con él.

En la realidad digital alternativa que ofrece el metaverso se pueden realizar las mismas actividades que se hacen en el día a día en cualquier entorno y sin moverse de una habitación o despacho y al que se puede acceder con dispositivos especiales como gafas de realidad virtual o aumentada, a través de las cuales será posible interactuar con otros usuarios. Cada uno de estos usuarios tiene un avatar (su personaje en el mundo virtual) y podrá trabajar, tener reuniones sociales e incluso jugar con otros usuarios en mundos inmersivos. Se trata de espacios interactivos autónomos por sí mismos, descentralizados, sin límites y con economías virtuales. El metaverso se puede considerar como una nueva versión de Second Life, o de los Sims o de videojuegos populares como Roblox, Fortnite o similares.

En cuanto al concepto preciso de metaverso por el que apuestan Facebook y otras empresas es la de crear un universo paralelo y completamente virtual, al *que podremos acceder con dispositivos de realidad virtual y realidad aumentada*, de forma que podamos interactuar entre nosotros dentro de él, y desde fuera con el contenido que tenemos dentro. La clave de este metaverso es que pueda ser totalmente inmersivo, o por lo menos mucho más de lo que es la actual realidad virtual. Sí, tendremos unas gafas que posiblemente sean parecidas a las actuales para sumergirnos en él, pero también sensores que registren nuestros movimientos físicos para que nuestro avatar dentro de ese metaverso haga exactamente lo mismo.

Muchas empresas están apostando por el desarrollo de esta tecnología. Por ejemplo, Facebook (con el cambio de nombre de su empresa matriz, Meta), Google y Microsoft están en el proceso. Pero no solo están invirtiendo en el campo empresas tecnológicas, sino también marcas y firmas de lujo. Algunos casos de estudio publicados en el prestigioso portal *Business Insider*⁵² son: “El caso de Nike, la compañía estadounidense con sede en Oregón ha puesto en marcha Nikeland: un metaverso disponible a través de la plataforma de videojuegos Roblox, en ella los usuarios pueden vestir a sus avatares con las prendas más icónicas de la marca. Adidas, por su parte, anunciaba mediante un tuit su alianza con Coinbase, una de las

⁵² <https://www.businessinsider.es/metaverso-ha-convertido-tierra-prometida-moda-977963>

plataformas de intercambio de criptomonedas más conocidas del planeta. Además, el acuerdo traía consigo la construcción de su metaverso propio: *AdiVerse*. Otros metaversos creados en el mundo de la moda son los proyectos de Gucci, Ralph Lauren, Balenciaga e incluso Zara ya han lanzado sus propios proyectos al respecto. Un caso de éxito de estas tecnologías es el caso de la prestigiosa marca de supermercados, Walmart que ha creado un supermercado virtual similar a sus supermercados físicos desplegados a lo largo del mundo y, de modo predominante en Estados Unidos, y con una gran promoción de chatbots y asistentes virtuales. En el citado artículo de *Business Insider*, se publicaba que “Un reciente estudio de la consultora estadounidense Morgan Stanley arrojaba que en 2030 el metaverso acumulará el 10% de las ventas del sector del lujo. Concretamente, hablaba de unos 56.000 millones de dólares”.

9.3 Computación cuántica⁵³

Esta nueva generación de supercomputadores aprovecha el conocimiento de la mecánica cuántica —la parte de la física que estudia las partículas atómicas y subatómicas— para superar las limitaciones de la informática clásica. Aunque la computación cuántica presenta en la práctica problemas evidentes de escalabilidad y de coherencia, permite realizar multitud de operaciones simultáneas y eliminar el efecto túnel que afecta a la programación actual en la escala nanométrica.

La computación cuántica utiliza como unidad básica de información el **qubit** en lugar del bit convencional. La principal característica de este sistema alternativo es que admite la superposición coherente de unos y ceros, los dígitos del sistema binario sobre los que gira toda la computación, a diferencia del bit, que solo puede adoptar un valor al mismo tiempo —uno o cero—. Esta particularidad de la tecnología cuántica hace que un *qubit* pueda ser cero y uno a la vez, y además en distinta proporción. La multiplicidad de estados posibilita que un computador cuántico de apenas 30 *qubits*, por ejemplo, pueda realizar 10 billones de operaciones en coma flotante por segundo, es decir, unos 5,8 billones más que la videoconsola PlayStation más potente del mercado.

9.4 Web3 [Web 3.0]

No existe todavía una definición aceptada universalmente como en su caso la Web 1.0, la Web 2.0 y la Web Semántica de su creador Tim Berners-Lee que también comenzó a llamarse Web 3.0. Sin embargo, el término Web3 se suele considerar un sinónimo de Web 3.0 pero con diferencias sustanciales como ahora veremos, aunque por su gran difusión los dos términos se utilizan indistintamente en la actualidad.

La primera característica diferenciadora e importante de la Web3 es ser “descentralizada” al contrario que las versiones anteriores Web 1.0, Web 2.0 y Web Semántica en las cuales la propiedad es “centralizada” de las corporaciones y, en general, controladas con regulaciones establecidas por los gobiernos. Otras características esenciales de la Web3⁵⁴ son: abierta por lo que se basa en gran medida en software de código abierto, sin confianza y sin permiso.

⁵³ Iberdrola (operadora eléctrica española): <https://www.iberdrola.com/innovacion/que-es-computacion-cuantica>

⁵⁴ <https://www.nytimes.com/es/interactive/2022/03/29/espanol/web3-que-es.html>;
<https://future.com/why-web3-matters/>
<https://www.forbes.com/advisor/investing/cryptocurrency/what-is-web-3-0/>

En esencia es “Una Internet que es propiedad de los desarrolladores y los usuarios, coordinada con tokens”⁵⁵ como la define Packy McCormick. *The New York Times* atribuye a Packy McCormick el haber puesto “de moda” el concepto de Web3.

La Web3 tiene como soportes fundamentales las tecnologías disruptivas emergentes como *blockchain* y NFT y la integración del cifrado y la criptoconurrencia (criptomonedas), todo ello dentro de organizaciones autónomas descentralizadas (DAO) y concurrentes. Sus características funcionales unidas a las tecnologías citadas anteriormente son:

- **Web3 está descentralizada:** en lugar de grandes extensiones de Internet controladas y apropiadas por entidades centralizadas, la propiedad se distribuye entre sus constructores y usuarios, y un gran uso de aplicaciones descentralizadas dApp.
- **Web3 no tiene permisos:** todos tienen el mismo acceso para participar en la Web3, y nadie queda excluido.
- **Web3 no depende de la confianza:** opera utilizando incentivos y mecanismos económicos en lugar de depender de terceros de confianza.
- **Web3 está cifrada e integrada con criptomonedas.** Web3 tiene pagos nativos: utiliza criptomonedas para gastar y enviar dinero en línea en lugar de depender de la infraestructura obsoleta de los bancos y procesadores de pagos.

Web3 Foundation

Con el objetivo de desarrollar, potenciar y definir la Web3 se creó la organización Web3 Foundation (web3.foundation). En su página principal en la opción “About” se destacan los objetivos principales de la Fundación: “Nuestra misión es fomentar aplicaciones de vanguardia para protocolos de software web descentralizados. Nuestra pasión es ofrecer la Web 3.0, una Internet descentralizada y justa donde los usuarios controlan sus propios datos, identidad y destino”.

La importancia de la Fundación se destaca también en el sitio Web oficial del prestigioso World Economic Forum con una página web específica denominada *Web 3.0 Technologies Foundation*⁵⁶ donde se destaca prácticamente la descripción oficial citada anteriormente: “La misión es fomentar aplicaciones de vanguardia para protocolos de software web descentralizados. Su objetivo es ofrecer una Internet descentralizada y justa donde los usuarios controlen sus propios datos, identidad y destino”.

9.5 Hiperconectividad inteligente: redes 6G y Wi-Fi 7

La nueva era de la hiperconectividad viene impulsada por las tendencias tecnológicas de las futuras redes Wi-Fi y móviles que se desplegarán a lo largo de la segunda década y el horizonte 2030: redes 6G y redes Wi-Fi 7.0 que se complementarán con las tecnologías actuales Wi-Fi 6 y redes móviles 5G.

Futuras redes Wi-Fi: Wi-Fi 7

La Wi-Fi Alliance tiene previsto que el futuro estándar Wi-Fi 7 (protocolo 802.11 be) tras su uso experimental como es obligatorio, se apruebe en la primavera de 2024. Wi-Fi 7

⁵⁵ Packy McCormick. <https://www.nytimes.com/es/interactive/2022/03/29/espanol/web3-que-es.html>;
<https://future.com/why-web3-matters/>;
<https://www.forbes.com/advisor/investing/cryptocurrency/what-is-web-3-0/>

⁵⁶ World Economic Forum: <https://www.weforum.org/organizations/web-3-0-technologies-foundation>

podrá alcanzar velocidades de hasta 30 Gbps y podrá utilizar las bandas de frecuencias 2,4 GHz, 5 GHz y como gran novedad, la banda de 6 GHz.

Futuras redes móviles: 6G

El 2019, fue el comienzo del despliegue de redes 5G en Europa y América, numerosos países y empresas tecnológicas están investigando sobre el futuro 6G, que según se conoce podrá llegar a alcanzar velocidad de descarga de hasta 1 TeraByte por segundo (TBps). China, Corea del Sur y los Estados Unidos lideran la investigación en redes 6G, y por lo que se ha publicado en informaciones posteriores, el factor clave del desarrollo de las redes 6G será la inteligencia artificial, lo que facilitará un mundo conectado o hiperconectado en tiempo real. Los investigadores consideran que la inteligencia artificial será el principal impulsor de la tecnología móvil o celular, 6G, y se habilitará una nueva generación de aplicaciones que potenciará el despliegue del ecosistema de internet de las cosas. Las ciudades inteligentes, fábricas inteligentes, vehículos autónomos y transmisión de realidad inmersiva (virtual, aumentada y mixta), entre otras grandes aplicaciones se verán impulsadas por el lanzamiento de las redes 6G.

10. Ética, privacidad y protección de datos (retos, oportunidades y peligros): normativas y regulaciones

La Ética en la Inteligencia Artificial y la Ciencia de Datos, es una obligación y una necesidad de cumplimiento por organizaciones y empresas, así como por usuarios personales y profesionales. Desde la entrada en vigor en la Unión Europea del Reglamento General de Protección de Datos (RGPD, GDPR) (25 de mayo, 2018) se ha dado un gran impulso a su uso permanente afrontando los grandes retos y oportunidades que ofrece así como los peligros que entraña su no aplicación.

10.1 El Reglamento *ePrivacy* de la UE y las cookies | Actualización del Reglamento *ePrivacy* 2021

El Consejo de la Unión Europea (UE) aprobó, el 10 de febrero de 2021, la propuesta de **Reglamento *ePrivacy*** del Parlamento Europeo y del Consejo sobre el respeto de la vida privada y la protección de los datos personales en el sector de las comunicaciones electrónicas y por el que se deroga la Directiva *e-Privacy* de 2002 (también conocida como la ley de *cookies* de la UE). El Reglamento *ePrivacy* sobre la privacidad y las comunicaciones electrónicas de 2021, es un proyecto de ley que aportará importantes actualizaciones al incluir a las nuevas tecnologías en su marco jurídico y regula todas las comunicaciones electrónicas en servicios y redes disponibles públicamente para los individuos dentro de la Unión Europea. La UE pretende reforzar la privacidad de los ciudadanos en Internet y regular la protección de los datos dentro de la propia UE de un modo más estricto. Básicamente se trata de que los ciudadanos recobren la confianza en los canales de comunicación digitales

La propuesta de *Reglamento ePrivacy 2021*, abarca todos los tipos de comunicaciones electrónicas (como textos, emails, mensajes de Facebook, SnapChat, etc.), y protege a los individuos dentro de la UE de la intromisión de terceros en sus comunicaciones privadas, a no ser que den su consentimiento previo. El reglamento *ePrivacy* ofrece plenas garantías de privacidad en la Unión Europea y se basa en cuatro pilares:

- La confidencialidad de las comunicaciones electrónicas.
- La aceptación de *cookies* y otros métodos de seguimiento por parte del usuario.

- Tratamiento de los datos y metadatos de las comunicaciones autorizados por el usuario. El Reglamento extiende la protección a los metadatos que generan los usuarios y sus dispositivos a la hora de navegar o conectarse a Internet.
- Consentimiento expreso para recibir comunicaciones comerciales.

10.2 La ética en la inteligencia artificial: ética digital

Ética y tecnología ¿qué está bien y qué no lo está cuando hablamos de usos de la tecnología? Es un tema de debate a nivel empresarial y que la consultora Gartner, ya, reconoció *la ética digital y privacidad como una de las 10 tendencias tecnológicas estratégicas en 2019* “Las corporaciones han de tomar conciencia de la importancia de establecer una serie de principios éticos a la hora de implantar las estrategias de Transformación Digital”. Carlos Bueso en la prestigiosa revista Forbes⁵⁷ ya aventuraba la necesidad de “Implementar la transformación digital con modelos éticos de uso de información personal y protocolos de seguridad avanzados, asegura grandes beneficios para nuestras empresas”.

Ética y privacidad en la inteligencia artificial y en el gobierno de los algoritmos: *algoritmos* y nuevas fronteras de la ética

La sociedad está “gobernada” por los algoritmos. Se requiere un análisis ético y control de las múltiples formas con las que un algoritmo puede impactar en la sociedad. La inteligencia artificial (IA) y la robótica -especialmente los robots humanoides, robots colaborativos “*cobots*” y los asistentes virtuales “bots” y “*chatbots*”- apoyados en la expansión de *big data*, plantean un desafío ético en el uso de los robots y los límites en el caso de la inteligencia artificial y de los algoritmos. Uno de los aspectos importantes recogidos en el *Reglamento ePrivacy* de la UE es la *privacidad desde el diseño*, lo que obliga a tener presente y utilizar la privacidad en todas las etapas del diseño de proyecto y -en consecuencia, el diseño de los algoritmos- y posteriormente en la implementación, pruebas, despliegue y mantenimiento de este.

Una tendencia imparable es el uso de algoritmos como herramientas de control de todo tipo de máquinas y aplicaciones de software de IA. ¿Qué hacer? ¿Nos fiamos del asesoramiento del algoritmo? Los algoritmos refuerzan las nuevas fronteras de la ética. Muchas de las decisiones cotidianas las toma un modelo matemático basado en un algoritmo (*Machine Learning*) y se requiere una ética especial de los algoritmos, así como para todos los dispositivos en los que los algoritmos son su espina dorsal y el cumplimiento y respuesta a cuestiones como las siguientes, que la UE sigue estudiando y analizando:

- Nuevo modelo en la Nube: *Algorithms as a Service (AaaS)*
- ¿Se requiere una ética especial de los algoritmos?
- ¿Se requiere una ética especial para los robots humanoides, robots colaborativos, robots virtuales? (asistentes virtuales como Siri, Alexa de Amazon...)
- ¿Deben pagar impuestos los robots igual que las personas a las que pueden sustituir? ¿Cómo controlar la toma de decisiones de los algoritmos?

Los dilemas éticos de la Inteligencia Artificial es necesario abordarlos y para ello la mejor solución para organizaciones, empresas y particulares y profesionales es el cumplimiento de los citados reglamentos RGPD y *ePrivacy*.

⁵⁷ Transformación digital, IA y ética corporativa. *Ética y Responsabilidad Social*. 20 de diciembre, 2018. Edición de México: <https://www.forbes.com.mx/transformacion-digital-ia-y-etica-corporativa>

Regulación y uso ético de la inteligencia artificial en la Unión Europea (normalización)

La Unión Europea aprobó en febrero de 2020 el Libro Blanco⁵⁸ sobre la *Inteligencia Artificial: un enfoque europeo orientado a la excelencia y la confianza*. En la misma fecha, la Comisión Europea publicó su *Estrategia Europea de datos (A European strategy for data)* para los próximos años, con el objetivo de delimitar el rumbo del modelo de economía digital de la UE. El 21 de abril de 2021 se publicó en Bruselas el nuevo Reglamento de Inteligencia Artificial de la Unión Europea⁵⁹. “El nuevo Reglamento sobre la IA garantizará que los europeos puedan confiar en lo que la IA puede ofrecer”.

Caso de estudio en España: Estrategia Nacional de Inteligencia Artificial (ENIA)

Presentada el 2 de diciembre, 2020. El presidente del Gobierno de España presentó el 2 de diciembre de 2020, la *Estrategia Nacional de Inteligencia Artificial* de España, con una inversión pública de 600 millones de euros *en el periodo 2021-2023* y se enmarca en el *Plan España Digital 2025*. Para su implantación se crearía la *Oficina del Dato* a nivel nacional y la figura del *Chief Data Officer*; así como la creación de espacios compartidos de datos y repositorios de datos descentralizados, accesibles e integrados en repositorios de datos de la UE.

La estrategia plantea la necesidad de una "digitalización humanista", esto es, de una IA ética que tenga a las personas en el centro, con impacto social y económico positivo (en la productividad de la Administración, en logística, educación...) y que ayude a la recuperación económica. El objetivo de ENIA es, según el ejecutivo, generar un entorno de confianza respecto al desarrollo de una Inteligencia Artificial (IA) “inclusiva, sostenible y que ponga a la ciudadanía en el centro”. ENIA contempla 6 ejes estratégicos y 30 medidas⁶⁰.

Acuerdo mundial de la UNESCO sobre la ética de la inteligencia artificial

El 25 de noviembre de 2021, los 193 Estados miembros de la Conferencia General de la UNESCO adoptaron la *Recomendación sobre la Ética de la Inteligencia Artificial*, el primer instrumento normativo mundial sobre el tema. Los Estados Miembros de la UNESCO, adoptan el primer acuerdo mundial y recomendación sobre la ética de la inteligencia artificial⁶¹, el primer instrumento normativo mundial sobre el tema: “No solo protegerá, sino que también promoverá los derechos humanos y la dignidad humana, y será una brújula guía ética y una base normativa global que permitirá construir un sólido respeto por el estado de derecho en el mundo digital”. Audrey Azoulay, Directora General de la UNESCO⁶², presentó en la citada fecha, la primera norma mundial sobre la ética de la inteligencia artificial. Este texto histórico establece valores y principios comunes que guiarán la construcción de la infraestructura jurídica necesaria para garantizar un desarrollo saludable de la IA.

Marco Ético, Regulación y Normalización de la Inteligencia Artificial en Argentina y LATAM

⁵⁸ https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf.

⁵⁹ https://ec.europa.eu/commission/presscorner/detail/es/ip_21_1682. Manual de 120 páginas.

⁶⁰ ENIA: Ejes estratégicos y medidas. https://www.escudodigital.com/tecnologia/las-30-medidas-fijadas-por-el-gobierno-para-impulsar-la-inteligencia-artificial_34810_102.html.

⁶¹ *Recomendación sobre la Ética de la Inteligencia Artificial*: <https://es.unesco.org/artificial-intelligence/ethics>; <https://es.unesco.org/news/estados-miembros-unesco-adoptan-primer-acuerdo-mundial-etica-inteligencia-artificial>. 25/11/2021

⁶² Audrey Azoulay, Directora General de la UNESCO <https://www.unesco.org/es/articles/los-estados-miembros-de-la-unesco-adoptan-el-primer-acuerdo-mundial-sobre-la-etica-de-la>

En Argentina del 7 al 8 de junio de 2022 se celebró el Seminario internacional sobre inteligencia artificial⁶³ (7-8 de junio 2022) organizado por el Consejo Económico y Social (CES) de Presidencia de la Nación y la Unión Europea en Argentina. Participó la Subsecretaría de Telecomunicaciones y Conectividad, y se debatió “sobre el carácter estratégico y el potencial de la IA para la Argentina”. En este mismo Seminario se celebraron los siguientes talleres: “Pautas éticas para los procesos de IA”, “Gobernanza de los datos” y “Arquitectura institucional del CamIA, Centro Argentino Multidisciplinario de Inteligencia Artificial”. Durante el Seminario se anunció la adhesión de Argentina al Pacto Global de Inteligencia Artificial. De igual forma en Buenos Aires se aprobó el Plan de Inteligencia Artificial de Buenos Aires⁶⁴.

11. CONCLUSIONES

Las tecnologías emergentes siguen creciendo en popularidad -especialmente las recomendadas en los informes de *Hype Cycle* de Gartner- y van madurando y convirtiéndose en tecnologías disruptivas llegando a organizaciones y empresas, y a la sociedad en general. Así en los últimos años se han consolidado tecnologías disruptivas como Inteligencia Artificial, *Big Data*, Computación en la Nube, Internet de las Cosas, *Blockchain* y tecnologías NFT, Experiencias inmersivas (RV, RA, RM), gemelos digitales, etc. De todas ellas sin lugar a duda, es la Inteligencia Artificial, la tecnología disruptiva de mayor impacto y con mayor capacidad de evolución y despliegue. Las dos categorías más reconocidas son: Inteligencia Artificial Conversacional (IAC) y Inteligencia Artificial Generativa (IAG). En particular la IAG está proporcionando gran número de aplicaciones y uso en las más diferentes disciplinas y dominios. Los modelos de lenguaje (LLM) entre los que destacan GPT-3 y una de sus aplicaciones más extendida, como DALL-E 2 de la empresa OpenAI, están dando origen también a potenciar la IA Creativa, facilitando conversión de textos a imágenes o vídeos, generación y redacción de textos, de informes, etc. Otras plataformas de IA creativa son: LaMDA de Google, Stable Diffusion y MidJourney.

Sin género de dudas, la aplicación **ChatGPT** con el modelo GPT-3 está siendo en 2023 uno de los grandes desarrollos; es un chat conversacional inteligente que está entrenado para mantener conversaciones, escribir artículos, responder a preguntas de cualquier tipo, y comienza a convertirse en un buscador que, en su momento, competirá con los grandes buscadores y navegadores actuales. Otros modelos presentados en 2023 y de gran futuro son Bard de Google y Prometeo de Microsoft.

La Inteligencia Artificial y la Ciencia de datos están consolidando nuevas tecnologías emergentes, como la computación cuántica, el metaverso o la Web3, las cuales, a medida vayan llegando a la sociedad conformarán la futura Industria 5.0 y a la Quinta Revolución Industrial.

⁶³ <https://www.argentina.gob.ar/noticias/debaten-sobre-el-caracter-estrategico-y-el-potencial-de-la-ia-para-la-argentina>. Publicado el martes 14 de junio de 2022

⁶⁴ *Plan Estratégico Inteligencia Artificial. SECRETARÍA DE INNOVACIÓN Y TRANSFORMACIÓN DIGITAL. Buenos Aires Ciudad. Documento Oficial, 202 páginas.*
<https://www.buenosaires.gob.ar/jefaturadegabinete/innovacion/plan-de-inteligencia-artificial> (202 páginas).

BIBLIOGRAFÍA

- AEPD. *El uso de las tecnologías en la lucha contra el Covid-19. Un análisis de costes y beneficios*. Unidad de Evaluación y Estudios Tecnológicos. Informe de la Agencia Española de Protección de datos, mayo 2020. Disponible en: <https://www.aepd.es/sites/default/files/2020-05/analisis-tecnologias-COVID19.pdf>.
- Barrio, Moisés, *Internet de las Cosas*, Madrid, Editorial Reus, 2018.
- Caballero, María Ángeles y Cilleros, Diego, *Ciberseguridad y Transformación Digital. Cloud, Identidad Digital, Blockchain, Agile, Inteligencia Artificial*, Madrid, Anaya, 2019.
- Joyanes, Luis, *Inteligencia de Negocios y Analítica de Datos. Una visión global de Business Intelligence & Analytics*. CDMX, México DF, Alfaomega, 2019.
- Joyanes, Luis, *Big Data. El análisis de los grandes volúmenes de datos*, Barcelona: Marcombo; México DF: Alfaomega, 2014.
- Joyanes, Luis, *Sistemas de Información. Un enfoque dirigido a la empresa*, Barcelona, Marcombo; México DF: Alfaomega, 2016.
- Joyanes, Luis, *INDUSTRIA 4.0: La Cuarta Revolución Industrial*. Barcelona: México DF, Alfaomega. 2018.
- Joyanes, Luis (Coordinador), *Cibersociedad. La colaboración público-privada*. Madrid: Instituto Español de Estudios Estratégicos. Cuaderno de Estrategia no. 185. Departamento de Seguridad Nacional. Ministerio de Defensa. 2018. .
- Joyanes, Luis. *Computación en la nube. Estrategias de Cloud Computing en las empresas*. 2ª ed. Barcelona, Marcombo; México DF, Alfaomega, 2022.
- Joyanes, Luis, *Ciencia de Datos. Un enfoque práctico de tecnologías, herramientas y aplicaciones*. Barcelona, Marcombo; México DF, Alfaomega, 2023.
- Laudon, Kennet C. y Laudon, Jane P., *Management Information Systems. Managing the Digital Firm*, Seventeenth edition, Pearson Education, 2022.
- Marr, Bernard, *Big Data en la práctica*. Editorial TEELL, 2017.
- Marr, Bernard, *Data Strategy. Como beneficiarse de un mundo de big data, Analytics e internet de las cosas*. Editorial TEELL, 2017
- Preukschat, Alex y LAGE, Oscar, *Blockchain. La revolución industrial de Internet*. Barcelona: Ediciones Deusto. 2017.
- Quesada, J. F., Calleja, Z. y Griol, D., “Informe sobre sistemas conversacionales multimodales multilingües. Plan TL. Secretaría de Estado para el Avance Digital”, Ministerio de Economía y Empresa de España, Noviembre, 2019.
- Rodriguez, Pablo, *Inteligencia Artificial. Cómo cambiará el mundo (y tu vida)*, Ediciones Deusto. 2018
- Rouhiainen, Lasse. *Inteligencia Artificial. 101 cosas que debes saber hoy sobre nuestro futuro*. Barcelona: Alienta Editorial-Editorial Planeta. 2018.
- Rusell, Stuart y Norvig, Peter. *Artificial Intelligence: A Modern Approach, Cuarta Edición*, Pearson. US. (Referencia en la Universidad de Berkeley, <http://aima.cs.berkeley.edu/>), 2021.
- Shah, Chirag, *A Hands-On Introduction to Data Science*, Cambridge University Press, 2020.
- Sharda, R., Delen, D. y Turban, E., *Analytics, Data Science, & Artificial Intelligence. Systems for Decision Support, 11th Edition*. Pearson. 2020.

RECURSOS WEB

- Encyclopedia Britannica. Descripción de *Artificial Intelligence*. *The Editors of Encyclopedia Britannica*. <https://www.britannica.com/technology/artificial-intelligence>. (20 pp.)
- Gartner: *What Is Artificial Intelligence?* www.gartner.com/en/topics/artificial-intelligence.

- IBM, IBM Cloud Education, *Artificial Intelligence (AI)*, 3 de junio de 2020. <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence>. IBM Cloud Education. *Machine Learning*, 15 July 2020. <https://www.ibm.com/cloud/learn/machine-learning>.

PRONÓSTICO DE SERIES TEMPORALES APLICADA A LA PREDICCIÓN DE EVENTOS CLIMÁTICOS SEVEROS PARA LA REGIÓN DEL VALLE DE CATAMARCA

Pucheta, Julián^{(1)*}, Salas, Carlos⁽²⁾, Herrera, Martín⁽²⁾, Patiño, H. Daniel⁽³⁾ y Rodríguez Rivero, Cristian⁽⁴⁾

⁽¹⁾ *Departamento de Electrónica, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba, Av. Vélez Sarsfield 1611, Córdoba, Argentina.*

⁽²⁾ *Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca, Catamarca, Argentina.*

⁽³⁾ *Instituto de Automática, Facultad de Ingeniería, Universidad Nacional de San Juan, Av. Libertador General San Martín (Oeste) N° 1109. San Juan, Argentina.*

⁽⁴⁾ *Cardiff School of Technologies, Cardiff Metropolitan University, Western Avenue, Cardiff, CF5 2YB, Wales.*

**Correo Electrónico: jpucheta@unc.edu.ar.*

RESUMEN

En la Región de la ciudad de San Fernando del Valle de Catamarca hay una serie de antecedentes de eventos climáticos severos que han quedado registradas por sus impactos negativos en la sociedad. Desde la disciplina de control automático se trata de proponer una solución al hecho de estar alerta y prever acciones que mitiguen los daños mediante un sistema de modelado de procesos dinámicos. Se han generado diversas soluciones al modelado de procesos climáticos partiendo de series temporales de valores de lluvia, y se proponen algoritmos de ajuste, validación y prueba para el caso particular del Valle de Catamarca. Se emplea la serie de Mackey Glass para realizar el diseño preliminar, y luego se emplean las series de lluvia obtenidas desde una central meteorológica del Ministerio de Ciencia e Innovación Tecnológica del Gobierno de Catamarca. El desafío en el proceso de cálculo surge a partir de que las series de datos tienen pocos años de historia y los eventos severos aparecieron menos de diez veces, sabiendo que las variables climáticas tienen estacionalidad marcada. Los algoritmos de predicción propuestos tienen métricas de evidencia de modelo definidas mediante la autocorrelación de cada serie temporal, y se busca que tengan una firma similar. Se diseñaron algoritmos basados en modelos dinámicos lineales, ya que evidencian un buen desempeño para realizar extrapolaciones con simulación de Monte Carlo que emplea ruidos de distribución gaussiana, gamma y Poisson. Se muestran resultados de predicción para las series de Mackey Glass 17 y 30 con horizonte de predicción de 24 pasos adelante, y para predicción de series de lluvia con horizonte de predicción de cuatro días. Se detallan las simulaciones Monte Carlo en cada caso.

Palabras Clave — pronóstico de series temporales, modelos dinámicos, predicción de series temporales, Simulación Monte Carlo.

TIME SERIES FORECAST FOR APPLICATION ON THE PREDICTION OF SEVERE WEATHER EVENTS IN CATAMARCA VALLEY REGION

ABSTRACT

In the Region of San Fernando del Valle de Catamarca there is various antecedents of severe climatic events that have been recorded due to their negative impacts on society. An approach from the automatic control discipline to being aware and foreseeing actions that mitigate the damage through a dynamic process modeling system is proposed. Several solutions have been generated for modeling climatic processes based on rainfall values time series, tuning, validating, and testing algorithms are proposed for the case of the Catamarca Valley. The Mackey Glass series is used to carry out the preliminary design, and then the rainfall time series obtained from a meteorological center of the Ministry of Science and Technological Innovation of the Government of Catamarca are used. The challenge in the calculation process arises from the fact that the data series have a few years of history and the severe events appeared less than ten times, knowing that the climatic variables have marked seasonality. The proposed prediction algorithms have model evidence metrics defined by autocorrelation of each time series and are tuned to have a similar signature. Algorithms based on linear dynamic models were designed, since they show a satisfactory performance to perform extrapolations with Monte Carlo simulation by using Gaussian, Gamma, and Poisson distribution noises. Prediction results are shown for the Mackey Glass time series 17 and 30 with a forecast horizon of 24 steps ahead, and for rainfall time series forecast a prediction horizon of four days was used. The Monte Carlo simulations are detailed in each case.

Keywords — time series forecasting, dynamic models, prediction, Monte Carlo Simulation.

INTRODUCCIÓN: PRESENTACIÓN DEL PROBLEMA

En la Región de la ciudad de San Fernando del Valle de Catamarca hay una serie de antecedentes de inclemencias de eventos climáticos que han quedado registradas por sus impactos en la sociedad, por ejemplo en el año 2020 [1] [2] y en el año 2021 [3] [4] en la propia Ciudad. Estos eventos generan una demanda de un sistema de alerta que tenga capacidad de predecir el evento con una antelación útil para la toma de decisiones.

Para ello, se requiere de realizar sistemas de predicción y pronóstico enfocados en el evento de interés, con la historia de valores climáticos que dan información para proponer un sistema con exactitud adecuada a la cantidad de tiempo previo que se requiera en los cuales ya se tienen diversos resultados que sirven como referencia [5] [6] [7] [8]. En la Fig. 1 se muestran los datos obtenidos desde la central de adquisición que tiene el Gobierno de Catamarca [9]. El objetivo se puede sintetizar en generar un valor probable de precipitación para los dos días siguientes al actual. En este trabajo se genera un pronóstico para los cuatro días siguientes, con fines de facilitar el análisis. Para el caso de las MG el horizonte es de 24 días adelante.

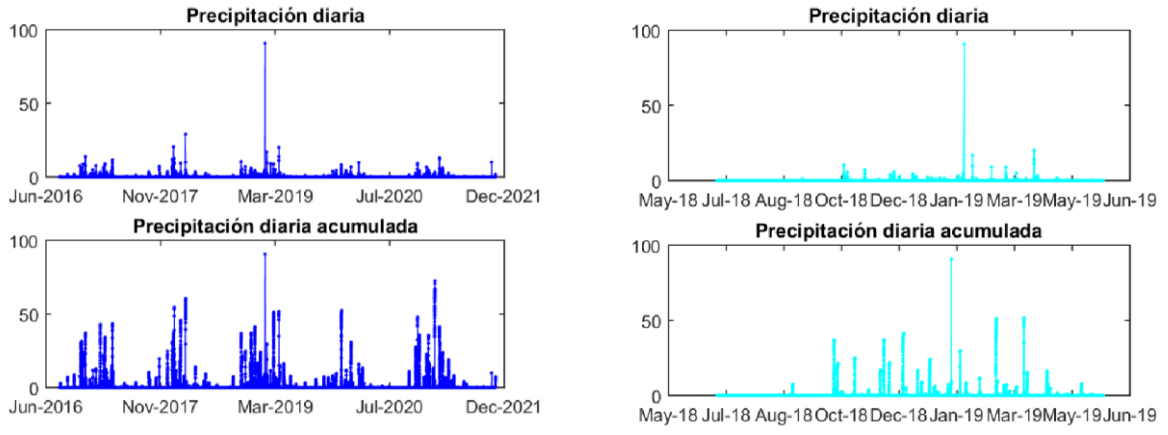


Fig. 1. Visualización de las series temporales de lluvia obtenidas desde [9]. A la izquierda se detallan los valores de precipitación disponibles, y a la derecha los valores de precipitación de un año hidrológico, que serán usados en este trabajo como prueba.

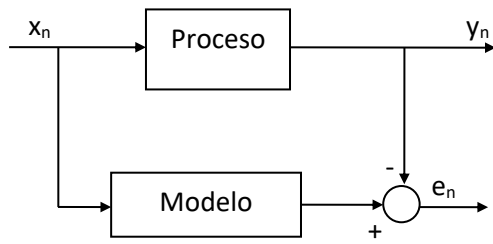


Fig. 2. Representación de la señal de error entre un proceso real y su modelo. El modelo tiene parámetros que son ajustados mediante la secuencia de error.

MATERIALES Y MÉTODOS

Se propone realizar un modelado que toma las entradas y salidas para el cálculo como se muestra en la Fig. 2. El modelo contiene parámetros que deben sintonizarse. Para ello se propone el uso del ajuste por el gradiente descendente, donde el error de ajuste está dado por

$$e_n = y_n - f(\theta_n, x_n, y_{n-1}) = y_n - \hat{y}_n \quad (1)$$

donde y_n es el valor actual de la salida del proceso, $f(\cdot)$ indica la salida del modelo que se está ajustando con parámetros θ_n a definir en cantidad y en valores, y x_n es el valor de la entrada actual. Se asume que se dispone de un total de datos muestreados con intervalos unitarios desde 0 hasta n .

Aquí $f(\cdot)$ se implementa como un sistema dinámico lineal, definiendo al vector de los parámetros θ como

$$\theta^T = [a_2 \quad a_3 \quad \dots \quad a_{o_y} \quad b_1 \quad b_2 \quad \dots \quad b_{o_x}] \quad (2)$$

se define que la función $f(\cdot)$ es

$$\hat{y}_n(\theta, x_n, y_{n-1}) = b_1 x_n + b_2 x_{n-1} + b_{o_x} x_{n-o_x+1} - a_2 y_{n-1} - \dots - a_{o_y} y_{n-o_y+1} \quad (3)$$

donde los parámetros serán ajustados en el sentido de minimizar el error en donde interviene la medición actual y_n , que no aparece. Para el ajuste de dichos parámetros, se define el funcional a minimizar respecto de los parámetros como

$$J_n = \frac{1}{2} \sum_{n=n_i}^{n_f} e_n^2 = \frac{1}{2} \sum_{n=n_i}^{n_f} \{y_n - f(\theta, x_n, y_{n-1})\}^2 \quad (4)$$

donde están definidos dos tiempos específicos de ajuste que son n_i y n_f como tiempo inicial y tiempo final para el ajuste del modelo. Este ajuste se realiza una vez por cada iteración, durante la que se hará variar cada componente de θ . Los parámetros serán ajustados durante una etapa de tiempo, pero luego quedarán fijos para generar la predicción.

Para ajustar los parámetros del modelo, que son las componentes a_i y b_i del vector θ , se procede empleando el método del gradiente descendente. Para ello, se empieza minimizando respecto de dichas componentes hallando sus derivadas parciales, por lo tanto

$$\frac{\partial J_n}{\partial a_i} = -e_n y_{n-i+1}, i = 2, 3 \dots o_y. \quad (5)$$

$$\frac{\partial J_n}{\partial b_i} = e_n x_{n-i+1}, i = 1, 2 \dots o_x. \quad (6)$$

Con éstos incrementos se definen las cantidades del algoritmo de gradiente descendente para el instante n , como

$$b_i := b_i + \gamma e_n x_{n-i+1}, i = 1, 2 \dots o_x. \quad (7)$$

$$a_i := a_i - \gamma e_n y_{n-i+1}, i = 2, 3 \dots o_y. \quad (8)$$

con el valor de γ como ganancia de ajuste o paso de ajuste. Con las variables $\{x, y\}$ correctamente condicionadas, éste valor puede estar en el orden de 10^{-3} a 10^{-6} para los casos aquí estudiados. Por lo tanto, dados un conjunto de datos de mediciones hasta un instante n , $\{x_n, y_n\}$ se requiere elegir la relación de correspondencia dinámica entre la entrada x y la salida y de tal manera que el modelo muestre máxima evidencia. El algoritmo expresado depende de las condiciones iniciales de θ , ya que es incremental, por lo que se proponen valores de orden reducido en dimensión y luego se procede a aumentar los valores de o_x, o_y si el modelo no muestra la suficiente evidencia.

La predicción de los valores siguientes al instante n estarán dados por

$$\hat{y}_{t+D} = [-\hat{y}_{t+D-1} \quad \dots \quad -\hat{y}_{t+D-n} \quad u_{t+D-1} \quad \dots \quad u_{t+D-1-m}] \cdot \theta = \phi(t+D)^T \cdot \theta. \quad (9)$$

En la predicción sí aparece la señal de entrada u que tendrá una distribución Gaussiana, Gamma o de Poisson y se genera una simulación Monte Carlo.

El criterio para medir la evidencia del modelo con el vector θ actual consiste en estudiar el comportamiento la autocorrelación. Así, el modelo obtenido que presente la función de autocorrelación más parecida a la que presentan los datos originales, será el modelo elegido.

Se supone disponible una determinada cantidad de datos observados, por lo que existirán errores debido a que la cantidad es finita. La función de autocorrelación para sistemas ergódicos se define como

$$\phi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)x(t+\tau) dt \quad (10)$$

y para intervalos discretos unitarios se tiene

$$\phi(j) = \frac{1}{N-j} \sum_{i=0}^{N-j} x(i)x(i+j) \quad (11)$$

con $j=0, 1, \dots, M$, y además $i=0, 1, \dots, N$. Donde M es el número de puntos de la función autocorrelación, N es el número de puntos medidos de la señal. Cabe aclarar que normalmente M es al menos la mitad de N , ya que la función intercorrelación está definida para N mucho mayor que M para asegurar una buena representación. Esto se ve en la definición, donde la integral va desde $-\infty$ a ∞ donde deberían estar definidas las señales y en la práctica se trabajan con valores discretos limitados.

RESULTADOS

Se implementa el algoritmo detallado para probarlo en series de dato Mackey Glass 17 y 30 siguiendo el método de modelado de series de lluvia de [8] y luego en las series de lluvia de la Región del Valle de Catamarca [9]. Se aplicó el método descrito y se obtuvieron los resultados de predicción que se muestran en las Fig. 3 para la serie MG17, Fig. 4 para las MG30, y en las Fig. 5 y Fig. 6 para el caso de las series de lluvia. Para éstas últimas se ha agregado una figura extra debido a la necesidad de ver los detalles. En todos los casos, se muestran las firmas de autocorrelación a la derecha, con las funciones superpuestas.

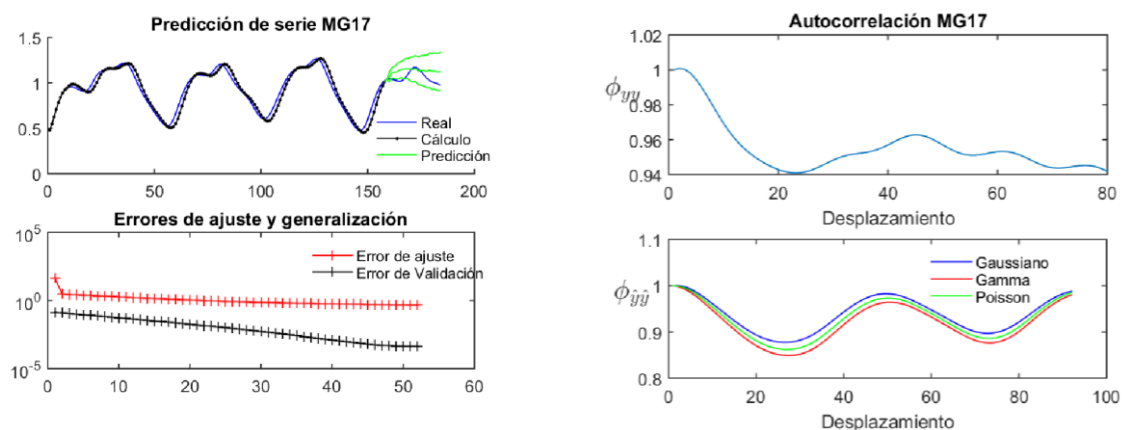


Fig. 3. Ajuste del algoritmo para correr el Monte Carlo con ruido Gaussiano, Gamma y Poisson. Para el primero fue de media nula con desviación de 0,31, para el Gamma con un factor de forma 0,03 y para Poisson con un λ de 0,08.

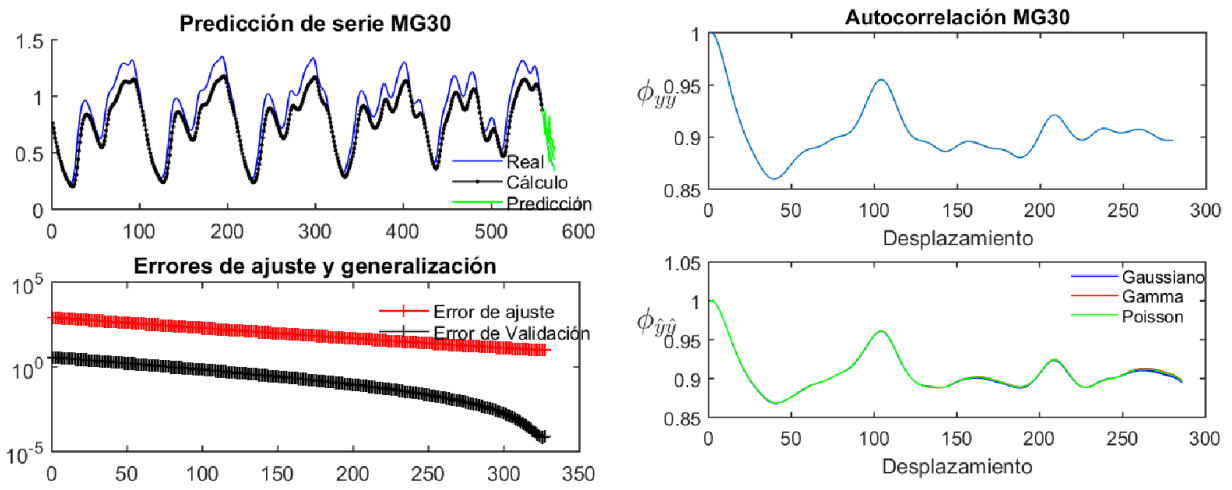


Fig. 4. Ajuste del algoritmo para correr el Monte Carlo con ruido Gaussiano, Gamma y Poisson. Para el primero fue de media nula con desviación de 0,44, para el Gamma con un factor de forma 0,03 y para Poisson con un λ de 0,08.

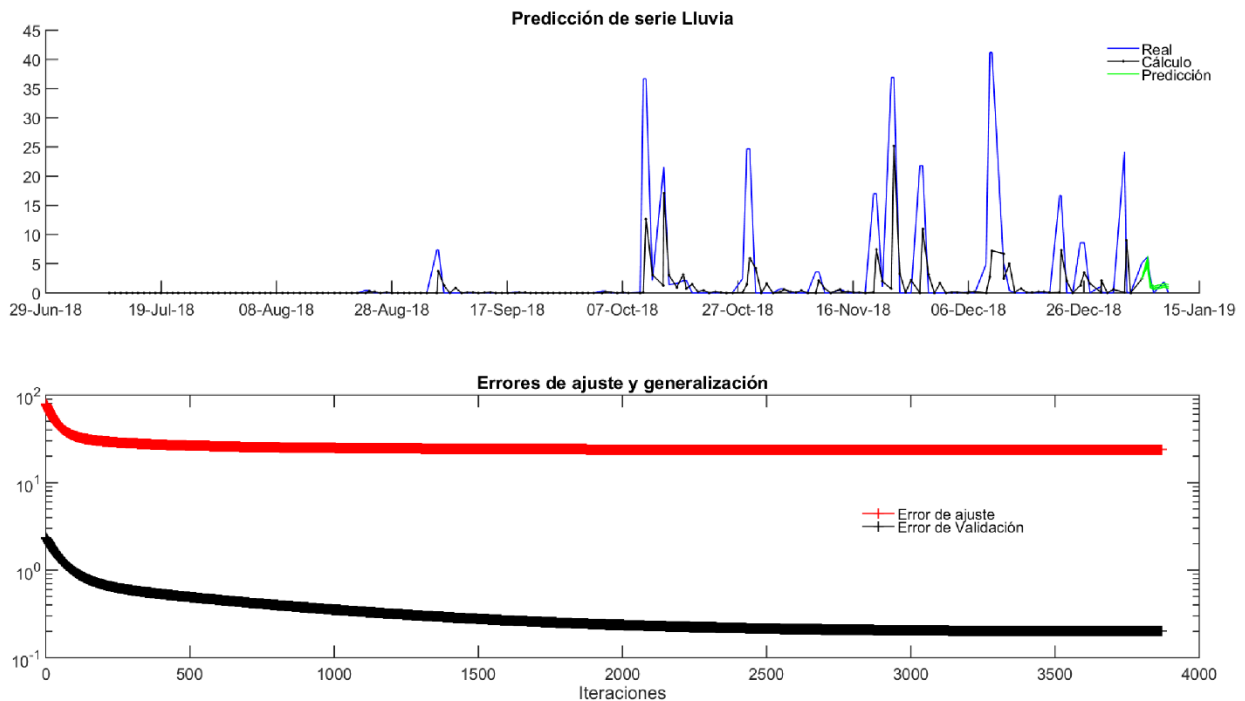


Fig. 5. Visualización de la predicción las series temporales de lluvia obtenidas desde [9].

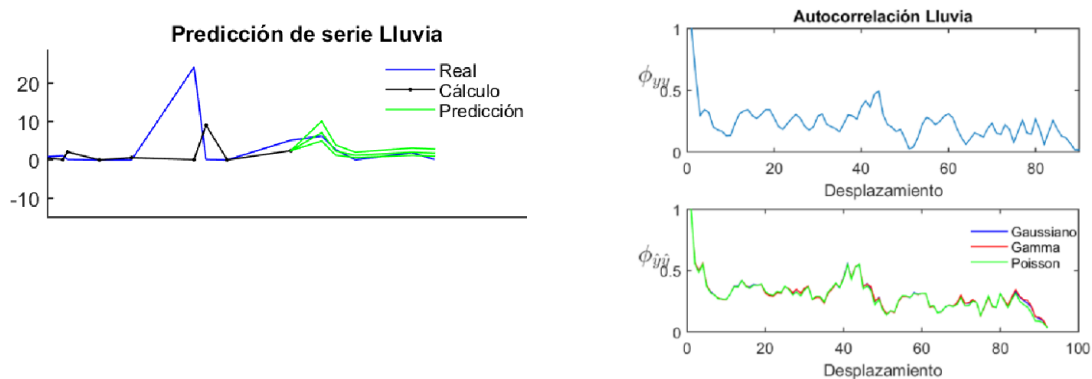


Fig. 6. Visualización de las series temporales de lluvia obtenidas desde [9].



Fig. 7. Sistema de visualización de variables ambientales disponible en la web [9].

CONCLUSIONES

Se ha mostrado un método para pronóstico de series temporales que permite generar la predicción de eventos climáticos severos con un tiempo de antelación de cuatro días. Así planteado, se tiene que realizar el ajuste del algoritmo hasta hallar la máxima evidencia del modelo mediante las gráficas de autocorrelación. Este procedimiento es consumidor de tiempo de cálculo, pero se espera lograr minimizar este costo mediante técnicas de aprendizaje automático, como en [5]. Este método asume a los datos con errores acotados y que no hay faltantes. El método de evidencia de modelo empleado aquí es más simple que el usado previamente en casos similares [5], que requieren de grandes volúmenes de cómputo.

Agradecimientos

Los autores agradecen a la Universidad Nacional de Catamarca, al Ministerio de Ciencia e Innovación Tecnológica del Gobierno de Catamarca, a la Universidad Nacional de Córdoba, a la Universidad Nacional de San Juan y a la Cardiff Metropolitan University por su apoyo en el desarrollo de proyectos involucrados en el Control y modelado de procesos de dinámica lenta, media y rápida.

REFERENCIAS

- [1] Crecida Enero 2020. <https://www.catamarcaactual.com.ar/informacion-general/2020/1/15/impresionante-crecida-del-rio-el-tala-dejo-gente-aislada-193936.html>
- [2] Crecida Enero 2020. <https://www.elancasti.com.ar/info-gral/2020/1/4/crecio-un-rio-algunos-autos-quedaron-varados-422962.html>
- [3] Crecida Marzo 2021. <https://www.lanacion.com.ar/sociedad/catamarca-un-vehiculo-fue-arrastrado-por-un-rio-y-su-conductor-desaparecio-nid03032021/>
- [4] Crecida Marzo 2021. <https://www.pagina12.com.ar/326964-la-capital-catamarquena-quedo-anegada-por-una-tormenta>
- [5] Julián Pucheta, Martín Herrera, Carlos Salas, Víctor Sauchelli, C. Rodríguez Rivero, H. Daniel Patiño, Capítulo titulado "Non-parametric methods for forecasting time series from cumulative monthly rainfall" del libro "Rainfall: Behavior, Forecasting and Distribution" Pp 204, Editors: Olga E. Martín and Tricia M. Roberts, Nova Science Publishers, Inc. ISBN: 978-1-62081-551-9. Pp. 69-89 (2012). Ed Nova Publishers. Nova Science Publishers, Inc. 415 Oser Avenue, Suite N Hauppauge, NY, 11788 USA. https://www.novapublishers.com/catalog/product_info.php?products_id=30548.
- [6] Cristian Rodríguez Rivero; Julian Pucheta; Alvaro Orjuela Canon; Leonardo Franco; Yvan Tupac Valdivia; Paula Otano, Victor Sauchelli., "Noisy Chaotic time series forecast approximated by combining Reny's entropy with Energy associated to series method: application to rainfall series," in IEEE Latin America Transactions, vol. 15, no. 7, pp. 1318-1325, 2017, doi: 10.1109/TLA.2017.7959353. Electronic ISSN: 1548-0992.
- [7] Cristian Rodríguez Rivero, Julián Pucheta, Sergio Laboret and Víctor Sauchelli. "Energy associated tuning method for short-term series forecasting by complete and incomplete datasets". Journal of Artificial Intelligence and Soft Computing Research. Volume 7, Issue 1, Pages 5–16, ISSN (Online) 2083-2567, DOI: <https://doi.org/10.1515/jaiscr-2017-0001>, 2017.
- [8] Cristian Rodriguez Rivero, Julián Pucheta, Victor Sauchelli, H. Daniel Patiño, "Short time series prediction: Bayesian Enhanced modified Approach with application to cumulative rainfall series", International Journal of Innovative Computing and Applications, Inderscience Publishers Ltd, 2016 Vol.7, No.3, pp.153 - 162. ISSN online 1751-6498, ISSN print: 1751-648X DOI: <http://dx.doi.org/10.1504/IJICA.2016.078730>. 2016.
- [9] Ministerio de Ciencia e Innovación Tecnológica, Gobierno de Catamarca, Argentina. Sitio: <http://meteo.sucytca.gob.ar/>
- [10] Oppenheim, Alan V., Ronald W. Schafer, and John R. Buck. Discrete-Time Signal Processing. Upper Saddle River, NJ: Prentice-Hall, 1999.
- [11] Cox, D. R. and V. Isham. "Point Processes". Chapman & Hall, London 1980.

MIGRACIÓN INTELIGENTE DE PROCESOS PARA BALANCEAR LA CARGA DE TRABAJO EN SISTEMAS DISTRIBUIDOS

DAVID LUIS LA RED MARTINEZ¹
Federico Agostini²
Terecio Diosnel Marecos Brizuela³

Resumen

En los sistemas de procesamiento distribuido existen múltiples grupos de procesos que comparten recursos y compiten entre sí para acceder a ellos. Además de que los procesos pueden o no requerir sincronización, se debe llegar a un acuerdo para lograr el acceso a los recursos de tal manera que se establezca un orden riguroso para asegurar la exclusión mutua en el acceso a recursos compartidos. Se propone un modelo de migración inteligente de procesos para la gestión de recursos y procesos en sistemas distribuidos considerando aspectos de balanceo de carga, que permitirá resolver los inconvenientes relacionados a la sobrecarga que pueden presentarse en los distintos nodos de un sistema distribuido.

Abstract

In distributed processing systems, there are multiple groups of processes that share resources and compete to access them. In addition to whether processes require synchronization or not, an agreement must be reached to achieve access to resources in such a way that a strict order is established to ensure mutual exclusion in accessing shared resources. A model of intelligent process migration is proposed for resource and process management in distributed systems, considering aspects of load balancing. This model will help address the issues related to overload that may arise in different nodes of a distributed system.

Palabras Clave

Sistemas distribuidos, migración de procesos, balanceo de carga.

Introducción

En los sistemas de procesamiento distribuido, es común la necesidad de coordinar la asignación de recursos compartidos para los procesos en un entorno de exclusión mutua. Esto implica la tarea de determinar el orden en el cual los recursos compartidos deben asignarse a los procesos que los requieren (La Red Martínez, 2017, 2022).

Cada nodo en este sistema tiene una capacidad específica de recursos computacionales, de datos y de comunicación. Los procesos solicitantes de recursos pueden clasificarse en diferentes categorías, como cálculo intensivo, datos intensivos, y comunicación intensiva, según sus necesidades de recursos.

En muchos casos, los nodos pueden experimentar sobrecarga al tratar de manejar múltiples asignaciones de recursos a los procesos, lo que resulta en una respuesta lenta del

¹ Doctor, Facultad de Ciencias Exactas y Naturales y Agrimensura – Universidad Nacional del Nordeste, lrmdavid@exa.unne.edu.ar.

² Magíster, Facultad de Ciencias Exactas y Naturales y Agrimensura – Universidad Nacional del Nordeste, fagostini@exa.unne.edu.ar.

³ Magíster, Facultad de Ciencias Aplicadas – Universidad Nacional de Pilar, ecosmar097@hotmail.com.

nodo a las solicitudes de los usuarios. Por otro lado, algunos nodos pueden permanecer ociosos durante gran parte del tiempo. Por todas estas razones, se busca lograr un equilibrio ⁴

adecuado en la carga de trabajo realizando migraciones inteligentes de procesos para mejorar el rendimiento.

El logro de un equilibrio de carga efectivo implica redistribuir la carga de trabajo entre los nodos del sistema distribuido con el fin de mejorar la utilización de los recursos y los tiempos de respuesta. Un algoritmo de equilibrio de carga debe contar con un conocimiento previo sobre el comportamiento de los trabajos y el estado global del sistema. Las decisiones de equilibrio de carga se basan en información global sobre el estado actual del sistema, y su desarrollo eficaz abarca aspectos importantes como la estimación de la carga, la comparación de niveles de carga y rendimiento, la estabilidad del sistema, el intercambio de información entre nodos, la estimación de los requisitos de recursos de los trabajos, la selección de trabajos para su transferencia, la elección de nodos remotos, entre otros.

En base a lo expuesto, el objetivo de este trabajo es proponer un modelo de decisión para un sistema distribuido que integre la migración inteligente de procesos a un operador de agregación. Esto permitirá la redistribución de la carga de trabajo entre los nodos del sistema, mejorando la utilización de los recursos, los tiempos de respuesta de los procesos y aumentando el rendimiento.

Trabajos previos

Entre los métodos considerados tradicionales para la asignación de recursos en sistemas distribuidos respetando la exclusión mutua en el acceso a los recursos compartidos, se destacan los mencionados en (Ricart & Agrawala, 1981), (Cao & Singhal, 2001) y (Lodha & Kshemkalyani, 2000).

Métodos innovadores con respecto a la asignación de recursos a procesos se mencionan en (La Red Martínez, 2017) y (La Red Martínez et al., 2018), y una alternativa con respecto a la asignación de recursos, considerando niveles estrictos de consenso y ordenamiento por grupos de procesos se puede ver en (Agostini & La Red Martínez, 2019), (La Red Martínez et al., 2017), (Agostini et al., 2019) y (Agostini, 2019).

Existen diferentes estudios sobre los métodos de migración de procesos, balanceo de carga y consumo de energía. En (Beiruti & Ganjali, 2020), desarrollan un nuevo protocolo que se utiliza para una amplia gama de aplicaciones de red, como el equilibrio de carga, el ahorro de energía y la optimización de recursos.

En (Upadhyay & Lakkadwala, 2015) desarrollan un algoritmo de migración para gestionar procesos en máquinas sobrecargadas derivándolos a otra en el mismo entorno de la nube teniendo en cuenta la utilización de los recursos.

En (Deshmukh & Deshmukh, 2015) se adopta un enfoque modificado para crear un nuevo equilibrador de carga dinámico distribuido para el sistema de archivos distribuido, el algoritmo tiene en cuenta restricciones como la carga de la red, la carga de IO (entrada/salida) del disco, la capacidad del disco y la carga que no se consideraba en otros enfoques.

En (Junaidi et al., 2020) se propone un algoritmo de aprendizaje automático que utiliza los recursos del servidor, CPU y la memoria, para predecir el futuro de las cargas del servidor.

En (Liu et al., 2013) se investigan metodologías de diseño para predecir cuantitativamente el rendimiento de la migración y el consumo de energía.

En (Asadi et al., 2020) se proponen modelos analíticos, basados en las redes de recompensa estocástica (SRN), para analizar el impacto de los algoritmos de asignación de recursos y los métodos de migración de procesos en el consumo de energía y el rendimiento de los sistemas virtualizados.

Algunos modelos de decisión utilizando operadores de agregación específicos se muestran en (Chiclana et al., 2001) y (Dong et al., 2016).

En (Sohrabi & Mousavi Khaneghah, 2020) se examinan los desafíos de la ocurrencia de eventos dinámicos e interactivos en los migradores de procesos basados en máquinas virtuales, mediante el análisis de la función de rendimiento de los migradores.

En (Rathore & Chana, 2014) se presenta un amplio estudio de las técnicas existentes de equilibrio de carga y migración de trabajos, donde se incluye una clasificación detallada basada en diferentes parámetros que dependen del análisis de las técnicas existentes, y se propone una nueva técnica de equilibrio de carga.

En (Chang et al., 2020) se propone un esquema de migración de procesos de borde consciente de los recursos (REM) que es capaz de optimizar la decisión de migración de procesos.

En (Marecos et al., 2022) se propone la aplicación de un modelo basado en competencias y en el aprendizaje centrado en el alumno para el estudio de la migración controlada de procesos, y donde el estudiante debe analizar el comportamiento de los recursos y procesos bajo diferentes cargas de trabajo, teniendo en cuenta el balanceo de carga como objetivo principal.

En (Fornerón Martínez et al., 2023) se propone un operador de agregación para gestión de recursos compartidos en sistemas distribuidos utilizando 2-tuplas asociadas a etiquetas lingüísticas.

Premisas

Se trata de grupos de procesos distribuidos en nodos de procesamiento que acceden a recursos críticos compartidos en la modalidad de exclusión mutua distribuida, debiendo decidirse, ante la demanda de recursos por parte de los procesos, cuáles serán las prioridades para asignar los recursos a determinados procesos o grupos de procesos que los requieren (sólo intervendrán como alternativas de asignación a los procesos aquellos recursos disponibles, es decir, no asignados aún a determinados procesos):

- El permiso de acceso a los recursos compartidos propios de un nodo no dependerá sólo de si los nodos los están utilizando o no, sino del valor de agregación de las preferencias (prioridades) de los distintos nodos respecto de otorgar el acceso a los recursos compartidos (alternativas).

- Las opiniones (prioridades) de los distintos nodos respecto de otorgar el acceso a los recursos compartidos (alternativas) dependerá de la consideración del valor de variables que representen el estado de cada uno de los distintos nodos. Cada nodo deberá expresar sus prioridades para la asignación de los distintos recursos compartidos respecto de los requerimientos de recursos de cada proceso.

En un entorno de sistemas distribuidos, representado en la Figura 1, es común encontrarse con múltiples nodos, algunos de ellos con alta carga de trabajo, con carga media y otros con carga baja. Cada nodo se encarga de ejecutar procesos que requieren acceso a ciertos recursos, ya sea de manera local o remota.

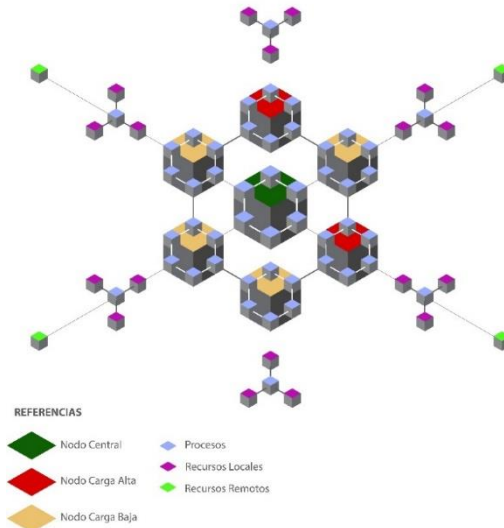


Figura 1. Representación de un sistema distribuido con cargas nodales.
Fuente: Elaboración propia.

El escenario planteado consiste en que el Runtime (software tiempo de ejecución) del Nodo Central, reciba la información proveniente de los distintos nodos y a través de un operador de agregación determine el orden de asignación de recursos a procesos y de procesos a nodos, según la siguiente premisa:

- Que los procesos accedan a recursos compartidos en la modalidad de exclusión mutua sin constituir grupos de procesos, los procesos no requieren sincronización (estar activos en sus respectivos procesadores en un mismo lapso de tiempo) y sin exigencias estrictas de consenso para lograr el acceso (no se requiere un consenso para asignar de manera consecutiva los recursos solicitados por un proceso o grupo de procesos, es decir, que iniciada la secuencia de asignación de recursos a un proceso, la misma puede ser interrumpida para asignar recursos a otro proceso).

- Bajo las circunstancias mencionadas anteriormente se tendrá en cuenta la aplicación de la migración de procesos para lograr un mejor balanceo de carga de trabajo suponiendo homogeneidad (respecto de compatibilidad de código, SO (Sistema Operativo), arquitectura y set de instrucciones).

El algoritmo de balanceo de carga corresponde a una serie de cálculos que se realizan para lograr la migración de procesos de nodos cargados a nodos más aliviados, partiendo de la estructura de datos utilizados en (La Red Martínez, 2017) y (La Red Martínez et al., 2018).

Propuesta de solución

Considerando lo desarrollado en (La Red Martínez, 2017) y (La Red Martínez et al., 2018), se presentará una variante de un enfoque innovador para la gestión de recursos compartidos en sistemas distribuidos, donde se desarrolla un operador de agregación para la asignación de recursos en un sistema distribuido.

Las premisas, estructuras de datos y operadores mencionados en la publicación anterior se utilizan como punto de partida para este trabajo. Para afrontar la problemática planteada, se define un Runtime (software en tiempo de ejecución) central que gestiona los procesos y recursos compartidos e interactúa con sus similares de los nodos distribuidos para intercambiar información. En el nodo central se recopila la información de todos los nodos, se aplica el proceso de agregación y se obtiene la lista de asignaciones de recursos a procesos.

Durante el funcionamiento del sistema distribuido, cada cierto tiempo se reúne y evalúa información de control referida a las cargas nodales para obtener una imagen global del sistema (macro imagen), que refleja el estado de los nodos, recursos y procesos en un determinado momento, lo que permite evaluar posibles migraciones para balancear la carga de trabajo de los distintos nodos del sistema. A continuación, se mencionan las estructuras de datos utilizadas:

Nodos que alojan procesos: 1, ..., n . El conjunto de nodos se representa de la siguiente

manera:

$$nodos = \{n_1, \dots, n_n\}$$

Procesos alojados en cada uno de los n nodos: $1, \dots, p$. El conjunto de procesos se representa de la siguiente manera:

$procesos = \{p_{ij}\}$ con $i = 1, \dots, n$ (n° de nodos en el sistema distribuido) y $j = 1, \dots, p$ (n° máximo de procesos en cada nodo), lo que se puede expresar mediante la Tabla 1.

Tabla 1: Procesos en cada nodo

Nodos	Procesos			
1	p_{11}	p_{12}	p_{1p}
....
i	p_{i1}	p_{i2}	p_{ip}
....
n	p_{n1}	p_{n2}	p_{np}

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Grupos de procesos distribuidos: $1, \dots, g$. El conjunto de grupos de procesos distribuidos se representa de la siguiente manera:

$$grupos = \{p_{ij}\}$$
 con i indicando el nodo y j el proceso en dicho nodo.

Tamaño de cada uno de los g grupos de procesos. El n° de procesos en cada grupo indica la cardinalidad del grupo y se representa de la siguiente manera:

$$card = \{card(g_i)\}$$
 con $i = 1, \dots, g$ indicando el grupo.

Prioridad grupal de cada uno de los g grupos de procesos. Estas prioridades se pueden fijar según distintos criterios; en esta propuesta se considerará que es función de la cardinalidad de cada grupo y se representa de la siguiente manera:

$$prg = \{prg_i = card(g_i)\}$$
 con $i = 1, \dots, g$ indicando el grupo.

Recursos compartidos en la modalidad de exclusión mutua distribuida disponibles en los n nodos: $1, \dots, r$. El conjunto de recursos se representa de la siguiente manera:

$$recursos = \{r_{ij}\}$$
 con $i = 1, \dots, n$ (n° de nodos en el sistema distribuido) y $j = 1, \dots, r$ (n° máximo de recursos en cada nodo), lo que se puede expresar mediante la Tabla 2.

Tabla 2. Recursos compartidos disponibles en cada nodo

Nodos	Recursos			
1	r_{11}	r_{12}	r_{1r}
....
i	r_{i1}	r_{i2}	r_{ir}
....
n	r_{n1}	r_{n2}	r_{nr}

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Estos recursos compartidos disponibles alojados en distintos nodos del sistema distribuido podrán ser requeridos por los procesos (agrupados o independientes) en ejecución en los nodos, estas solicitudes de recursos por parte de los procesos se muestran en la Tabla 3.

Estados posibles de cada uno de los p procesos:

- Proceso independiente.
- Proceso perteneciente a un grupo de procesos.

Tabla 3. Recursos solicitados por los procesos

Recursos		Procesos			
r_{11}	p_{11}	p_{kl}	p_{np}
....

r_{ij}	p_{11}	p_{kl}	p_{np}
....
r_{nr}	p_{11}	p_{kl}	p_{np}

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Estado posible de cada uno de los n nodos:

- Número de procesos.
- Prioridades de los procesos.
- Uso de CPU.
- Uso de memoria principal.
- Uso de memoria virtual.
- Memoria adicional necesaria para cada recurso solicitado por cada proceso (si se dispone del dato).
 - Carga adicional de procesador estimada necesaria para cada recurso solicitado por cada proceso (si se dispone del dato).
 - Carga adicional de E/S (entrada/salida) estimada necesaria para cada recurso solicitado por cada proceso (si se dispone del dato).
 - Estado de cada uno de los r recursos compartidos en la modalidad de exclusión mutua distribuida existentes en el nodo:
 - Asignado a un proceso local o remoto.
 - Disponible.
 - Predisposición (prioridad nodal) para otorgar el acceso a cada uno de los r recursos compartidos en la modalidad de exclusión mutua distribuida (resultará de la consideración de las variables representativas del estado del nodo, de la prioridad de los procesos y de la carga computacional adicional que significaría asignar el recurso al proceso solicitante).
 - Carga actual del nodo, que se podrá calcular como el promedio de los porcentajes de uso de CPU, memoria y entrada/salida en un momento dado (estos indicadores de carga podrán variar según los casos, pudiendo agregarse otros o cambiarse algunos de los puestos como ejemplo); también habrán de definirse las categorías de carga actual, por ejemplo, Alta, Media y Baja, señalándose los rangos de valores para cada categoría.

Una vez verificado e identificado la existencia de nodos accesibles y no accesibles, se atienden las solicitudes de los procesos y recursos correspondientes a los nodos accesibles. Aquellas asignaciones que no han podido ser resueltas, se agrupan con las nuevas solicitudes y se vuelve a tener una macro imagen (entendiéndose por macro imagen a la información referida al estado de los nodos respecto de la carga computacional actual y la demanda de asignación de recursos a procesos) para verificar si el nodo que estaba inaccesible ahora ya no lo está.

En base a la información recogida del sistema se puede detectar la presencia de nodos muy cargados con respecto a nodos que están inactivos o con baja carga de trabajo, una inadecuada distribución de la carga de trabajo, provocando un aumento en el tiempo de ejecución de los procesos que, a su vez, podría generar un incremento en el consumo de energía y perjudicar al rendimiento general del sistema.

En el siguiente apartado, se considerarán y evaluarán los estados referentes a los nodos, procesos y recursos y las opciones disponibles. Se establecerán las configuraciones iniciales del algoritmo, que podrán ser distintas para cada caso en particular y que permitirán adecuar una solución en base a la carga de trabajo de los nodos distribuidos y la locación de los procesos y recursos.

Para cubrir estas necesidades se plantea un conjunto de propuestas que consideran la migración de procesos para lograr un mejor balanceo de carga de trabajo suponiendo homogeneidad de los nodos (respecto de compatibilidad de código, SO, arquitectura, set de instrucciones, etc.), incluyendo migración de procesos que significarán el acceso a datos distribuidos con alto impacto en E/S y tráfico de red (consumo de ancho de banda), y una posible migración de proceso consecuencia de cambios en el estado global del sistema y de los requerimientos variables de los procesos a lo largo del tiempo, esta verificación se hará cada

cierto lapso Δt .

El modelo se desarrolla en cuatro etapas bien definidas como se muestra en la Figura 2.



Figura 2. Etapas del cálculo de migración de procesos

Las etapas corresponden a la evaluación de los nodos, clasificándolos según su nivel de carga (alta, media, baja), seguidamente se evalúan los procesos de los nodos con carga alta y los recursos vinculados a ellos. Se priorizan los nodos remotos que contengan mayor peso de recursos para cada proceso. Se obtiene un listado de los procesos candidatos a ser migrados que serán ordenados por su impacto en el ancho de banda de red, priorizándose los procesos con menor peso de transferencia.

Descripción del operador de agregación

En cada nodo se define una interfaz entre las aplicaciones y el sistema operativo, que a través de un *Runtime* (software en tiempo de ejecución complementario del sistema operativo) incluido en esa interfaz, gestiona los procesos y recursos compartidos y define el escenario correspondiente. Además, los *Runtime* interactúan entre sí para intercambiar información y en uno de los nodos hay un coordinador global de *Runtime* que evalúa y ejecuta el modelo de decisión y el operador de agregación correspondiente.

El operador propuesto consta de las siguientes etapas:

A. Aplicación del operador de agregación desarrollado en (La Red Martínez, 2017), obteniéndose la carga computacional actual de los nodos, las categorías de carga computacional y de los vectores de pesos asociados a las mismas, las prioridades o preferencias de los procesos teniendo en cuenta el estado del nodo y las prioridades o preferencias de los procesos para acceder a los recursos compartidos disponibles.

Estos cálculos, considerados preliminares, son el punto de partida para la etapa B., en la que se determinarán los procesos que serán migrados.

En la Figura 3 se muestra un diagrama que indica los pasos referidos a la etapa A., y cómo ello interactúa con los pasos de la etapa B. La secuencia de iteración de los distintos módulos es la siguiente: módulo de nodos, módulo de procesos y módulo de nodos destino, que se describen posteriormente.

En el punto (a) se realiza la configuración de valores iniciales, en donde se establecen los parámetros para indicar los porcentajes de carga y prioridades que se tendrán en cuenta para la migración. Para los nodos con carga alta se calcula el promedio de carga del porcentaje de uso de CPU, memoria y entrada/salida. Se consideran cargados los nodos cuyo promedio sea mayor a 70, carga media entre 40 y 70, y como carga baja a los nodos que como valor promedio tenga menor que 40.

(b) El primer paso consiste en analizar el escenario a nivel de nodos del sistema, considerando que cada uno puede tener diferentes procesos y recursos. Estos procesos pueden competir no solo con el acceso exclusivo a recursos locales (*RL*) sino también por aquellos que se encuentran en otros nodos, es decir, que se encuentran de manera remota. Para cada situación se considerarán diferentes alternativas. Esta evaluación se repetirá siempre y cuando existan nodos por evaluar, una vez terminada la evaluación de todos los nodos, se obtendrá un listado final de procesos y los nodos destinos.

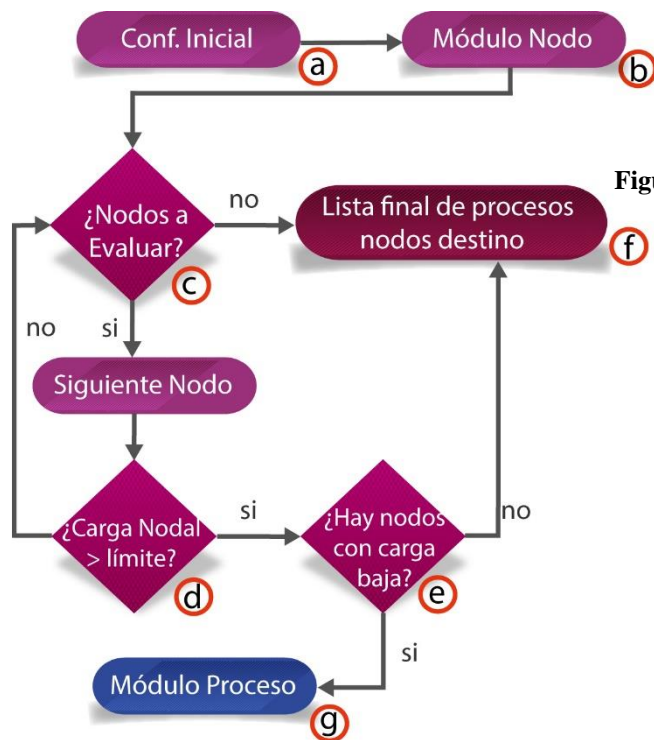


Figura 3. Módulo de nodo

(c) Situación: se analiza si existen nodos a evaluar dentro del sistema distribuido.

- *Primera alternativa* (sí): dado que existen nodos a evaluar se pasa a evaluar la carga nodal en el punto (d).

- *Segunda alternativa* (no): al no existir nodos a evaluar se pasa a evaluar el listado final de procesos y nodos destinos en el punto (f).

(d) Situación: se verifica si la carga del nodo supera un cierto límite, éste se refiere al valor mínimo de carga que debe superar el nodo para considerarse como cargado (a modo de ejemplo el valor mínimo de carga podría ser un 70%).

- *Primera alternativa* (sí): dado que la carga nodal supera el límite establecido se pasa al punto (e) para verificar si dentro del SD existen nodos con carga baja.

- *Segunda alternativa* (no): dado que la carga nodal no supera el límite se vuelve al punto (c) para evaluar otro nodo.

(e) Situación: se verifica si existen nodos con carga baja a donde puedan migrarse procesos.

- *Primera alternativa* (sí): si existen nodos con carga baja se pasa al módulo de procesos en (g) para analizar cada proceso con que cuenta el nodo. Esto corresponde a la etapa B., que se describe posteriormente.

- *Segunda alternativa* (no): si no quedan nodos con carga baja se vuelve al punto (f). Esto corresponde a la etapa B., que se describe posteriormente.

(f) Situación: se guarda la tabla de los procesos candidatos a ser migrados con sus posibles nodos destino e impacto correspondiente.

B. Al considerar aspectos relacionados con el balanceo de carga, se debe planificar una política efectiva de migración, que considere información sobre la prioridad de los procesos, limitaciones en la migración (para determinar el número de procesos que pueden pasar de un nodo a otro), el peso de los recursos cuyo estado puede ser local, pero podría cambiar a remoto al realizar la migración de procesos que los requieren.

Para un buen diseño en el balanceo de carga se consideran algunas políticas muy utilizadas para esta tarea, entre las cuales se pueden mencionar, políticas de:

- Locación, para determinar el destino del proceso seleccionado.
- Intercambio de información, que permita intercambiar información entre los nodos.
- Asignación, que determine la prioridad de ejecución de los procesos locales y remotos en un determinado nodo.
- Migración, que limite la cantidad de procesos migrables por nodo, y la cantidad de procesos que puede recibir un nodo destino.

Los procesos pueden ser locales o remotos, los segundos se diferencian en que se procesarán en un nodo distinto al original.

Los recursos serán considerados locales si están en el mismo nodo donde se aloja el proceso que los utilizará, y remotos si están en nodos distintos.

Para obtener información de la carga de nodos y estado global del sistema distribuido

se deben realizar diferentes observaciones, de las que surgen varios escenarios que se representan mediante un diagrama de flujo que muestra una serie de situaciones en las que se deben tomar decisiones que derivan en dos opciones posibles: "Sí" y "No". Cada opción se define como "Primera alternativa" y "Segunda alternativa" respectivamente.

Dependiendo del resultado de la evaluación, el flujo del diagrama se dirigirá hacia otra situación, donde se tomará otra decisión. Este proceso se repetirá hasta que se alcance una solución final, cuando ya no existan evaluaciones de recursos, procesos y nodos pendientes.

Se considerará la aplicación de las siguientes etapas adicionales para lograr balanceo de carga aplicando migración de procesos:

1. Determinación de procesos migrables mediante el cálculo de prioridad promedio de los procesos.
2. Selección de los posibles nodos destino mediante la evaluación de su carga.
3. Evaluación del peso de los recursos y su posible impacto ante la migración de procesos.
4. Determinación del orden de migración de procesos y de los nodos destino correspondientes.
5. Evaluación del estado de las cargas computacionales una vez migrados los procesos.

Determinación de procesos migrables mediante el cálculo de prioridad promedio de los procesos

La información resultante de la evaluación del estado de los nodos ayuda a identificar aquellos con carga alta, sus procesos y los recursos tanto locales como remotos vinculados con dichos procesos. En el punto (g) de la Figura 4, se da inicio al módulo de procesos, y es en donde se analiza el escenario a nivel de procesos, dentro de cada nodo.

(h) Situación: se verifica dentro del nodo si existen procesos que aún no han sido evaluados.

- *Primera alternativa* (sí): si existen procesos por evaluar se pasa a (i) para verificar la prioridad del proceso.
- *Segunda alternativa* (no): si ya no quedan procesos por evaluar se vuelve al módulo nodos en el punto (a).

(i) Situación: para el cálculo del promedio de prioridades de cada proceso se tendrán en cuenta el criterio de prioridad de procesos para todas las asignaciones de recursos (La Red Martínez, 2017).

Por cada solicitud de recursos de cada proceso, se establece un criterio de prioridad, que es asignado por el Runtime del nodo al momento de hacer la solicitud. La sumatoria de este valor de prioridad, para cada recurso solicitado, dividido el total de solicitudes de cada proceso, indicará la prioridad promedio del proceso.

Se consideran candidatos a ser migrados los procesos cuya prioridad supere un límite fijado arbitrariamente; se consideran solo aquellos procesos cuyo promedio de prioridad supere 0.6.

$$\text{promedio de prioridad de proceso } (p_{ij}) = (\Sigma cp_{ijk} / o) \quad (1)$$

con p_{ij} indicando los procesos, donde $i = 1, \dots, n$ (nº de nodos en el sistema distribuido) y $j = 1, \dots, m$ (nº de procesos de cada nodo),

con cp_{ijk} indicando el criterio prioridad de proceso relacionado a cada recurso, donde $i = 1, \dots, n$ (nº de nodos en el sistema distribuido), $j = 1, \dots, m$ (nº de procesos de cada nodo), y $k = 1, \dots, o$ (nº de recursos para cada proceso),

con $o =$ cantidad de asignaciones asociadas a cada proceso, lo que se puede expresar mediante la Tabla 4.

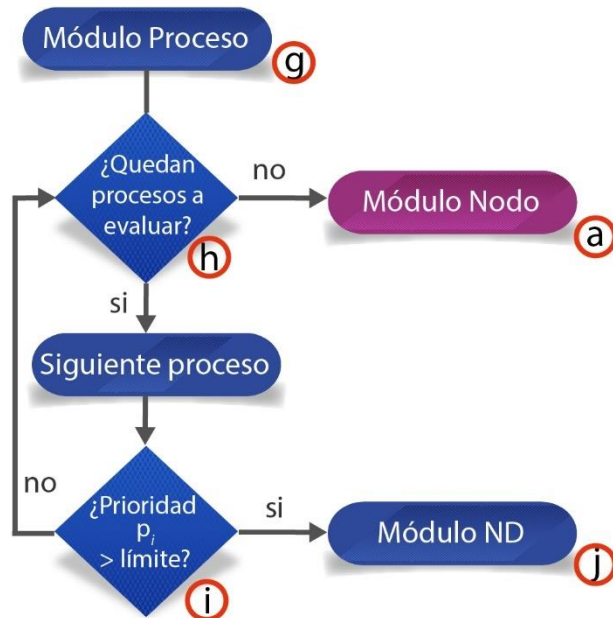


Figura 4. Módulo procesos

Tabla 4. Promedio de prioridad de procesos

Procesos	r_1	...	r_k	...	r_o	Promedio
p_{11}	cp_{111}	...	cp_{11k}	...	cp_{11o}	ppp_{11}
...
p_{ij}	cp_{ij1}	...	cp_{ijk}	...	cp_{ijo}	ppp_{ij}
...
p_{nm}	cp_{nm1}	...	cp_{nmk}	...	cp_{nmo}	ppp_{nm}

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Con r_k indicando los recursos, donde $k = 1, \dots, o$ (nº de recursos para cada proceso), con ppp_{ij} indicando los promedios de prioridad de los procesos, donde $i = 1, \dots, n$ (nº de nodos en el sistema distribuido) y $j = 1, \dots, m$ (nº de procesos de cada nodo).

Para todos los procesos de los nodos con carga alta se utilizará el mismo método para seleccionar los procesos candidatos a ser migrados.

- *Primera alternativa* (sí): si supera el límite establecido se pasa a evaluar a los posibles nodos destino en el punto (j).
- *Segunda alternativa* (no): si el promedio de prioridad de proceso no supera el límite establecido se vuelve a evaluar otro proceso en el punto (h).

Selección de los posibles nodos destino mediante la evaluación de su carga

En esta etapa se evalúa si existen nodos con carga baja que puedan alojar procesos de otros nodos. Los procesos que serán evaluados son aquellos que clasificaron en la etapa anterior. Se denotan los nodos que se encuentran dentro de la categoría de carga baja:

nodos destino con carga baja = {NDB} = $\{n_1, \dots, n_n\}$ con n indicando el nº total de posibles nodos destino según se muestra en la Tabla 5.

Para cada proceso puede haber varios nodos con carga baja disponibles. Teniendo en cuenta el listado de la Tabla 5 se verifican los recursos solicitados por cada proceso p_{ij} para determinar el nodo candidato.

Tabla 5: Nodos con carga baja

Nodos con carga baja
n_1
....
n_i
....
n_n

Evaluación del peso de los recursos y su posible impacto ante la migración de procesos

Una vez identificado los procesos del nodo con carga alta y los posibles nodos destinos, se realiza la comparación de los recursos locales en relación con los remotos, con el fin elaborar la lista preliminar de los procesos migrables y sus posibles nodos destinos.

En la Figura 5 se puede apreciar la distribución de los recursos de un proceso en un determinado momento. La opción a) representa al nodo origen con sus procesos y recursos, la opción b) representa al nodo remoto, que no tiene recursos asignados, la opción c) también representa al nodo remoto, pero al tener recursos asignados en ese nodo se lo denomina nodo destino.

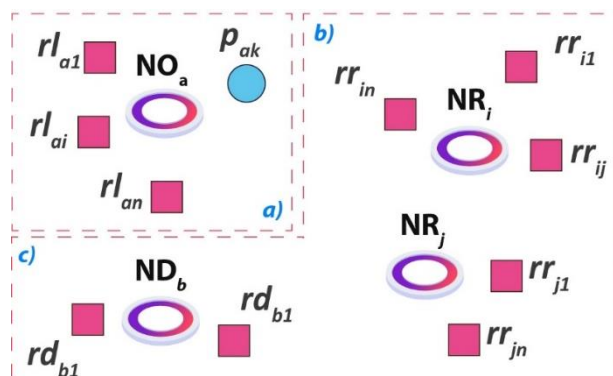


Figura 5. Recursos locales vs. Recursos remotos.
Fuente: Elaboración propia

NO_a , NO representa el nodo origen, a el número de nodo origen.

p_{ak} , p representa el proceso, a el nodo origen y k el número de proceso.

rl_{ai} , donde rl representa al recurso local, con a indicando el nodo origen, con i de 1 a n indicando el número de recursos.

NR_i , NR representa los nodos remotos con carga baja, con i el número de nodo remoto.

rr_{ij} , donde rr representa el recurso remoto que solicita el proceso, con i indicando el nodo remoto, con j de 1 a n indicando el número de recursos.

ND_b , ND representa el posible nodo destino que cumple con los requisitos de carga baja, compatibilidad y peso mínimo de transferencia de recursos (entendiéndose dicho peso como el impacto en el consumo de ancho de banda); b es el número de nodo destino.

rd_{bg} , donde rd representa el recurso remoto solicitado por un proceso, alojado en un posible nodo destino, con b indicando el nodo destino, con g de 1, ..., n , indicando el número de recurso.

En el punto (j) de la Figura 6, se activa el módulo de nodos destino, donde se buscan nodos remotos que puedan albergar recursos solicitados por cada proceso evaluado. Se busca generar el menor peso de transferencia al pasar los recursos locales (que pueda tener en un momento dado un proceso), a remotos (al migrar el proceso). Cualquier nodo remoto compatible con el nodo origen será evaluado, y podrá ser candidato a convertirse en un nodo destino.

(k) Situación: se verifica dentro del sistema distribuido si existen posibles nodos destino (ND) que aún no han sido evaluados.

- *Primera alternativa* (sí): si existen nodos destino que aún no han sido evaluados se selecciona el siguiente nodo destino y se pasa al punto (d).

- *Segunda alternativa* (no): si no existen nodos destino para evaluar entonces se vuelve al módulo de procesos en el punto (g).

(j) Situación: se evalúa si para el proceso en cuestión existen nodos destino con carga baja (**NDB**) considerando compatibilidad de códigos, arquitectura, set de instrucciones, etc.

- *Primera alternativa* (sí): si para el proceso existen nodos con carga baja se pasa al punto (m).

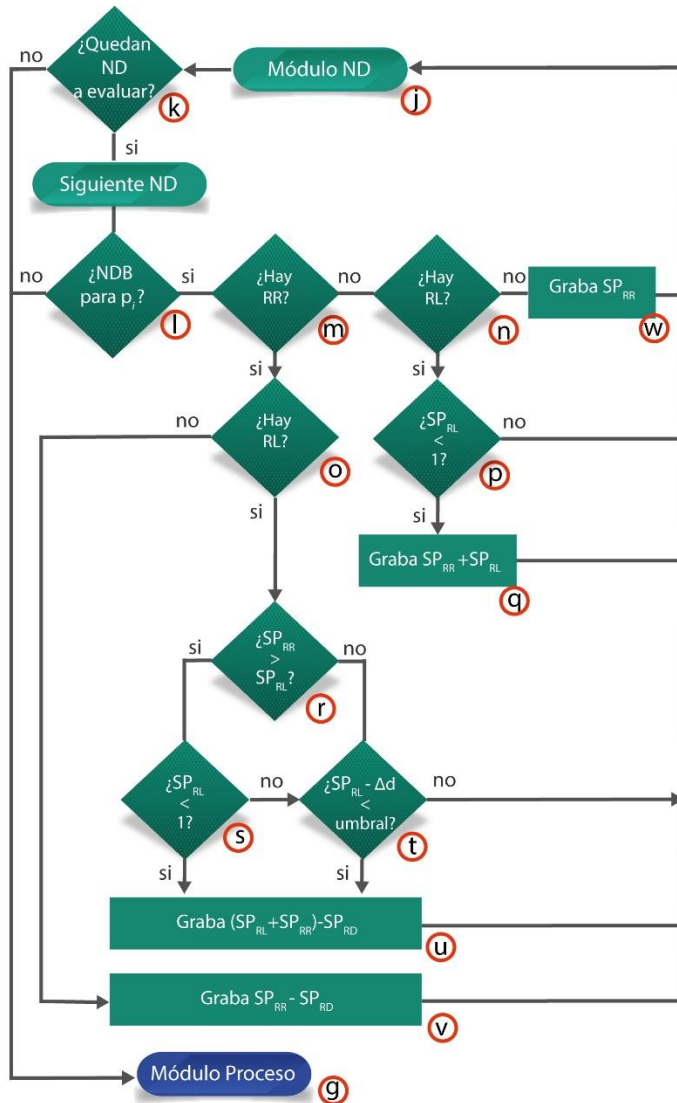


Figura 6. Módulo nodo destino

Fuente: Elaboración propia

- *Segunda alternativa* (no): si no existen nodos con carga baja entonces se vuelve al punto (g).

(m) Situación: se verifica si el proceso tiene asociados recursos remotos (**RR**)

- *Primera alternativa* (sí): si el proceso tiene recursos remotos se realiza (o).

- *Segunda alternativa* (no): si el proceso no tiene recursos remotos entonces se evalúa (n).

(n) Situación: se verifica si el proceso en cuestión tiene recursos locales.

- *Primera alternativa* (sí): si el proceso tiene recursos de forma local se evalúa (p).

- *Segunda alternativa* (no): si el proceso no tiene recursos de forma local entonces se vuelve a (j).

(o) Situación: se verifica si el proceso en cuestión tiene recursos locales.

- *Primera alternativa* (sí): si el proceso tiene recursos de forma local se realiza (r).

- *Segunda alternativa* (no): si el proceso no tiene recursos de forma local entonces se ejecuta el punto (v).

(p) Situación: se verifica si la sumatoria de pesos de los recursos locales (**SP_{RL}**) es menor a un límite; sin pérdidas de generalidad, en este caso se considera que el valor límite es 1 (uno).

- *Primera alternativa* (sí): si la sumatoria de pesos de los recursos locales es menor a 1 (uno) entonces se ejecuta el punto (q).

- *Segunda alternativa* (no): si la sumatoria de pesos de los recursos locales es no menor a 1 (uno) entonces se vuelve al punto (j).

(q) Situación: en el vector de migración de procesos (Figura 7) se almacena el valor resultante de la suma de pesos de recursos locales y los pesos de recursos remotos, peso de transferencia que generaría la migración del proceso al nodo destino.

(r) Situación: Se verifica si la suma de los pesos de los recursos remotos (**SP_{RR}**) es mayor a la suma de los pesos de recursos locales.

- *Primera alternativa* (sí): Si la sumatoria de los pesos de recursos remotos es mayor al peso de los recursos locales se realiza (*s*).
- *Segunda alternativa* (no): Si la sumatoria de los pesos de recursos remotos es menor al peso de los recursos locales se realiza (*t*).
- (*s*) Situación: Se verifica si la suma de los pesos del recurso local es menor a un límite (para este caso límite es igual a uno en base a (La Red Martínez, 2017))
 - *Primera alternativa* (sí): si la suma de los pesos del recurso local es menor al límite entonces se realiza (*u*).
 - *Segunda alternativa* (no): si la suma de los pesos del recurso local es mayor al límite se evalúa (*t*).
- (*t*) Situación: se verifica si la suma de los pesos de recursos locales menos Δd (Δd es el peso de los recursos destino más un 20% del total de pesos de recursos del proceso) es menor a un umbral (sin pérdidas de generalidad, en este caso se considera que el umbral es 1 (uno)).
 - *Primera alternativa* (sí): si la suma de los pesos de recursos locales menos Δd es menor a un umbral, entonces se realiza el punto (*u*).
 - *Segunda alternativa* (no): si la suma de los pesos de recursos locales menos Δd es mayor a un umbral, se vuelve a (*j*).
- (*u*) Situación: En el vector de migración de procesos (Figura 7) se almacena el valor resultante de la suma de pesos de recursos locales más la suma de los pesos de los recursos remotos menos la suma de pesos de recursos destino (SP_{RD}), peso de transferencia que generaría la migración del proceso al nodo destino.
- (*v*) Situación: En el vector de migración de procesos (Figura 7) se almacena el valor resultante de la suma de pesos de recursos locales menos la suma de pesos de recursos destino, peso de transferencia que generaría la migración del proceso al nodo destino.
- (*w*) Situación: En el vector de migración de procesos (Figura 7) se almacena el valor resultante de la suma de pesos de recursos remotos que tiene asociado el proceso, peso de transferencia que generaría la migración del proceso al nodo destino.

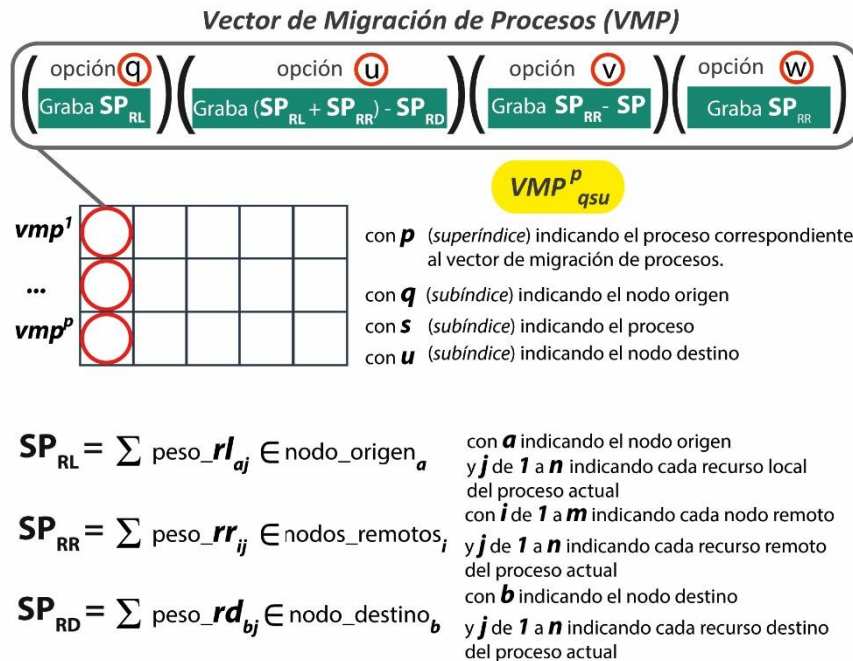


Figura 7. Vector de migración de proceso.

Fuente: Elaboración propia

Determinación del orden de migración de procesos y de los nodos destino correspondientes

En la Tabla 6 se observa la matriz general de procesos (MGP), resultado de la evaluación de las diferentes situaciones que pueden presentarse dentro de un sistema

distribuido para la migración inteligente de procesos. Se verifica qué procesos migrar y adónde migrar, considerando el peso de transferencia que se generaría en la migración. Para cada proceso se guarda el conjunto de nodos con carga baja y para cada nodo evaluado el valor del peso de transferencia del vector vmp obtenido en la Figura 7.

Tabla 6: Matriz general de procesos

MGP	n_1	...	n_k	...	n_m
mgp_1	$vmp^{l_{111}}$...	$vmp^{l_{11u}}$...	$vmp^{l_{11m}}$
...
mgp_i	$vmp^{p_{qs1}}$...	$vmp^{p_{qsu}}$...	$vmp^{p_{qsm}}$
...
mgp_o	$vmp^{n_{rt1}}$...	$vmp^{n_{rtu}}$...	$vmp^{n_{rtm}}$

Con i en mgp_i variando de 1, ..., o , indicando el proceso correspondiente en la matriz general de procesos.

Con p en vmp (*superíndice*) variando de 1, ..., n , indicando el proceso correspondiente al vector de migración de procesos.

Con q, s, u en vmp (*subíndices*) variando q de 1, ..., r , indicando el nodo origen, s de 1, ..., t , indicando el número de proceso y u de 1, ..., m , indicando el nodo destino, del vector de migración de procesos.

Con k en n variando de 1, ..., m , indicando el nodo con carga baja para cada proceso.

El primer paso que habrá que realizar a partir de los datos de la matriz general es dividirla en dos matrices, una primaria (primera ronda de asignación) y otra secundaria (segunda ronda de asignación). La diferencia radica en que primero se intentará favorecer a los procesos que tienen recursos en los nodos destinos, es decir, cuya migración implicaría una reducción en el impacto de entrada salida (transferencia) con respecto a aquellos que no tengan recursos en nodos destino; para ello se utiliza la matriz primaria. La segunda matriz se analizará luego de haber evaluado la matriz primaria.

Matriz Primaria

En la Tabla 7 se observa la matriz primaria (MP) que se obtiene de la matriz general de procesos. Esta matriz contiene datos solamente de aquellos procesos que tienen asignados recursos en los nodos destino.

Tabla 7. Matriz primaria

MP	n_1	...	n_k	...	n_m
mp_1	$vmp^{l_{111}}$...	$vmp^{l_{11u}}$...	$vmp^{l_{11m}}$
...
mp_i	$vmp^{p_{qs1}}$...	$vmp^{p_{qsu}}$...	$vmp^{p_{qsm}}$
...
mp_o	$vmp^{n_{rt1}}$...	$vmp^{n_{rtu}}$...	$vmp^{n_{rtm}}$

Con i en mp_i variando de 1, ..., o , indicando el proceso correspondiente en la matriz general de procesos.

Con p en vmp (*superíndice*) variando de 1, ..., n , indicando el proceso correspondiente al vector de migración de procesos.

Con q, s, u en vmp (*subíndices*) variando q de 1, ..., r , indicando el nodo origen, s de 1, ..., t , indicando el número de proceso y u de 1, ..., m , indicando el nodo destino, del vector de migración de procesos.

Con k en n variando de 1, ..., m , indicando el nodo con carga baja para cada proceso.

Se considerará para cada proceso el menor peso de transferencia en la matriz primaria, es decir, se verifica todos los pesos de la matriz primaria y se ordena de menor a mayor según el peso de transferencia, el proceso que esté asociado al menor peso tendrá prioridad en la selección del nodo destino. La evaluación continúa con los procesos pendientes. Por cada ronda de evaluación se verifica que el porcentaje de migración del nodo origen sea

menor al límite establecido, ya que el nodo puede migrar procesos mientras no supere ese límite. Este cálculo se realiza sobre la cantidad total de procesos por nodo.

Se inicia la elección del nodo destino para el primer proceso con menor peso de transferencia. Al terminar cada ronda, se eliminan de la matriz primaria el proceso y el nodo asignado recientemente. Los procesos cuyos recursos destino estaban únicamente en el nodo seleccionado pasan a formar parte de la matriz secundaria (*MS*) (Tabla 9).

Los procesos que se encuentran en nodos origen que hayan alcanzado el límite de migración establecido, son descartados y pasan a formar parte de una matriz de procesos no migrables (*MPNM*) (Tabla 11). Se continúa la evaluación mientras existan procesos a evaluar. El orden de asignación de los procesos a nodos destino de la matriz primaria se indica mediante la Función de Asignación para la Migración en Sistemas Distribuidos (*FAMSD*) primaria.

$$FAMSD^p \text{ primaria} = vmp^p_{qsu} \text{ del mínimo}(MP^p) \quad (2)$$

Con p indicando el proceso correspondiente a cada proceso de la matriz primaria.

Calculando la *FAMSD* primaria para todos los procesos de la matriz primaria se obtendrá un vector que indica el orden de migración de procesos a nodos destino. Esto se indica en la Tabla 8.

Matriz Secundaria

En la Tabla 9 se observa la matriz secundaria que resulta de la matriz general de procesos. Esta tabla contiene datos de procesos que no tienen asociados recursos (requerimientos de accesos) en nodos destino.

Tabla 8. Función de asignación para migración en sistemas distribuidos (*FAMSD*) primaria

Peso de transferencia del proceso	Proceso seleccionado
1º: vmp^p_{qsu} del Mínimo(MP^p)	p_{ij} del MP seleccionado
2º: vmp^p_{qsu} del Mínimo(MP^p) para los p_{ij} no asignados	p_{ij} del MP seleccionado
...	...
último: vmp^p_{qsu} del Mínimo(MP^p) del p_{ij} no asignado	p_{ij} del MP seleccionado

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Tabla 9. Matriz secundaria

<i>MS</i>	n_1	...	n_k	...	n_m
ms_1	$vmp^{l_{111}}$...	$vmp^{l_{11u}}$...	$vmp^{l_{11m}}$
...
ms_i	vmp^p_{qs1}	...	vmp^p_{qsu}	...	vmp^p_{qsm}
...
ms_o	vmp^n_{rt1}	...	vmp^n_{rtu}	...	vmp^n_{rtm}

Con i en ms_i variando de 1, ..., o , indicando el proceso correspondiente en la matriz general de procesos.

Con p en vmp (*superíndice*) variando de 1, ..., n , indicando el proceso correspondiente al vector de migración de procesos.

Con q, s, u en vmp (*subíndices*) variando q de 1, ..., r , indicando el nodo origen, s de 1, ..., t , indicando el número de proceso y u de 1, ..., m , indicando el nodo destino, del vector de migración de procesos.

Con k en n variando de 1, ..., m , indicando el nodo con carga baja para cada proceso.

La matriz secundaria se analiza una vez que la primera ronda de evaluación correspondiente a la matriz primaria haya terminado. Se aplica el mismo criterio de

evaluación que en la matriz primaria, pero para procesos que no tienen requerimientos de acceso a recursos de otros nodos. El orden de asignación de los procesos a nodos destino de la matriz secundaria se indica a continuación mediante la Función de Asignación para la Migración en Sistemas Distribuidos (*FAMSD*) secundaria.

$$FAMSD^p \text{ secundaria} = vmp^{p_{qsu}} \text{ del mínimo } (MS^p) \quad (3)$$

Calculando la *FAMSD* para todos los procesos de la matriz secundaria se obtendrá un vector que indica el orden de asignación de procesos a nodos destino. Esto se indica en la Tabla 10.

Tabla 10. Función de asignación para migración en sistemas distribuidos (*FAMSD*) secundaria

Peso de transferencia del proceso	Proceso seleccionado
1°: $vmp^{p_{qsu}}$ del Mínimo (MS^p)	p_{ij} del MS seleccionado
2°: $vmp^{p_{qsu}}$ del Mínimo (MS^p) para los p_{ij} no asignados	p_{ij} del MS seleccionado
....
último: $vmp^{p_{qsu}}$ del Mínimo(MS^p) del p_{ij} no asignado	p_{ij} del MS seleccionado

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Matriz de procesos no migrables

En la Tabla 11 se observa la matriz de procesos no migrables, que resulta de los procesos que se encuentran en nodos origen, que hayan alcanzado el límite de migración establecido y de procesos cuyos nodos destino fueron seleccionados por otros procesos.

Con i en $mpnm_i$ variando de 1, ..., o , indicando el proceso correspondiente en la matriz general de procesos.

Con p en $mpnm$ (*superíndice*) variando de 1, ..., n , indicando el proceso correspondiente al vector de migración de procesos.

Con q,s,u en $mpnm$ (*subíndices*) variando q de 1, ..., r , indicando el nodo origen, s de 1, ..., t , indicando el número de proceso y u de 1, ..., m , indicando el nodo destino, del vector de migración de procesos.

Con k en n variando de 1, ..., m , indicando el nodo con carga baja para cada proceso.

Tabla 11. Matriz de procesos no migrables

$MPNM$	n_1	...	n_k	...	n_m
$mpnm_1$	$mpnm^1_{111}$...	$mpnm^1_{11u}$...	$mpnm^1_{11m}$
...
$mpnm_i$	$mpnm^p_{qs1}$...	$mpnm^p_{qsu}$...	$mpnm^p_{qsm}$
...
$mpnm_o$	$mpnm^n_{rt1}$...	$mpnm^n_{rtu}$...	$mpnm^n_{rtm}$

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

De la evaluación de las matrices primaria y secundaria se obtiene la Función de Asignación para Migración en Sistemas Distribuidos (*FAMSD*) primaria y la Función de Asignación para Migración en Sistemas Distribuidos (*FAMSD*) secundaria. La unión de estas dos asignaciones conforma lo que se denomina Función de Asignación para la Migración en Sistemas Distribuidos Concatenada (*FAMSDC*), como se muestra en la Figura 8 y en la Tabla 12, que contiene la lista final de las asignaciones de cada ronda, proceso, nodo origen, nodo destino y peso de transferencia.

Evaluación del estado de las cargas computacionales una vez migrado los procesos

Para obtener un indicador de la carga computacional actual de cada nodo se pueden adoptar distintos criterios; en esta propuesta los criterios serán el % de uso de la CPU, el % de uso de la memoria y el % de uso de operaciones de entrada / salida.

Se aplica el operador de agregación desarrollado en (La Red Martínez, 2017) para obtener carga computacional actual de los nodos, las categorías de carga computacional y de los vectores de pesos asociados a las mismas, las prioridades o preferencias de los procesos teniendo en cuenta el estado del nodo y las prioridades o preferencias de los procesos para acceder a los recursos compartidos disponibles.



Figura 8. Función de Asignación para la Migración en Sistemas Distribuidos Concatenada

Fuente: Elaboración propia.

Tabla 12. Concatenación de las tablas de asignación (*FAMS D* primaria y *FAMS D* secundaria)

<i>FAMS DC</i>	Iteraciones
<i>FAMS D primaria</i>	Filas de 1 a n n = número de filas de la <i>FAMS D primaria</i> en la primera iteración
<i>FAMS D secundaria</i>	Filas de $n+1$ a m m = número de filas de la <i>FAMS D secundaria</i> en la segunda iteración

Para evaluar y comparar los nuevos valores de cargas computacionales se realiza en dos etapas, la primera es aplicar la sobrecarga sin realizar la migración de procesos y la segunda es aplicar la sobrecarga una vez migrado los procesos, se utilizarán los siguientes criterios asociado a cada proceso migrado: sobrecarga de memoria (memoria adicional que tendrá el nodo solicitado), sobrecarga de procesador (uso adicional de procesador que tendrá el nodo solicitado) y sobrecarga de entrada / salida (entrada / salida adicional que tendrá el nodo solicitado).

La carga computacional de cada nodo se calculará de dos maneras.

a. Cálculo de sobrecarga sin migración

- Establecimiento del n° de criterios para determinar la carga de los nodos: $card(\{criterios\}) = c$
- Establecimiento de los criterios que se aplicarán (podrán diferir de un nodo a otro):

$criterios = \{c_{ij}\}$ con $i = 1, \dots, n$ (n° de nodos en el sistema distribuido) y $j = 1, \dots, c$ (n° máximo de criterios para cada nodo), lo que se puede expresar mediante la Tabla 13.

Al promedio de carga computacional que se obtiene mediante el operador de agregación desarrollado en (La Red Martínez, 2017) se le agrega un criterio adicional que es la sobrecarga, que se obtiene con la suma de los promedios de sobrecargas de cada proceso del nodo.

Cálculo de la carga computacional actual de cada nodo:

$carga_i = (\text{valor_criterio}(c_{i1}) + \dots + \text{valor_criterio}(c_{ic})) / m$ con $i = 1, \dots, n$ (n° de nodos en el sistema distribuido) y $j = 1, \dots, c$ (n° de criterios para cada nodo), con m cantidad de criterios del nodo. (4)

Cálculo de la sobrecarga estimada que generaría la nueva asignación de recursos a procesos:

$$sobrecarga_i = \sum(\text{prom_sobrec}(p_{ij})) \quad (5)$$

con $i = 1, \dots, n$ (n° de nodos) y $j = 1, \dots, k$ (n° de procesos en el nodo).

Cálculo de la carga computacional de cada nodo contemplando la sobrecarga de las nuevas asignaciones:

$$carga_final_sin_migración_i = carga_i + sobrecarga_i \quad (6)$$

Tabla 13. Criterios para medir la carga computacional en cada nodo

Nodos		Criterios			
1	c_{11}	c_{12}	...	c_{1c}	
...	
i	c_{i1}	c_{i2}	...	c_{ic}	
...	
n	c_{n1}	c_{n2}	...	c_{nc}	

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

b. Cálculo de sobrecarga con migración

Se realiza el mismo procedimiento que para el cálculo de sobrecarga sin migración, pero se tiene en cuenta la sobrecarga que generan los procesos que no se migran del nodo actual y la sobrecarga que generan los procesos migrados que éste recibe.

En relación con la migración de procesos, estos pueden tener diferentes estados. Los procesos migrados entrantes son aquellos procesos que fueron migrados desde un nodo origen hacia el nodo actual. Los procesos migrados salientes son los procesos que fueron migrados del nodo actual hacia un nodo destino. Los procesos no migrados son aquellos procesos que permanecen en el nodo actual.

Cálculo de la carga computacional actual de cada nodo:

$carga_i = (\text{valor_criterio}(c_{i1}) + \dots + \text{valor_criterio}(c_{ic})) / m$ con $i = 1, \dots, n$ (n° de nodos en el sistema distribuido) y $j = 1, \dots, c$ (n° de criterios para cada nodo), con m cantidad de criterios del nodo. (7)

Cálculo de la sobrecarga estimada que generaría la nueva asignación de recursos a los procesos que no se migran del nodo actual:

$$sobrecarga_no_migrados_i = \sum(\text{prom_sobrec}(p_{ij} \text{ no migrados})) = \sum(\text{prom_sobrec}(p_{ij})) - \sum(\text{prom_sobrec}(p_{ij} \text{ salientes})) \quad (8)$$

Cálculo de la sobrecarga estimada que generaría la nueva asignación de recursos a los procesos que son migrados hacia el nodo actual:

$$sobrecarga_inmigrados_i = \sum(\text{prom_sobrec}(p_{ij} \text{ inmigrados})) \quad (9)$$

Cálculo de la carga computacional de cada nodo contemplando la sobrecarga de las nuevas asignaciones:

$$carga_final_con_migración = carga_i + sobrecarga_no_migrados_i + sobrecarga_inmigrados_i \quad (10)$$

Métrica propuesta para evaluar el balanceo de carga en el sistema distribuido

La aplicación de una métrica es esencial para la evaluación de la investigación. Al tener una medida objetiva de los resultados, se pueden comparar los resultados de una investigación con otros de investigaciones similares, lo que permite evaluar su relevancia.

Para evaluar la eficacia del modelo propuesto, es necesario medir los resultados para determinar si se comporta según lo deseado.

Para evaluar el impacto de la migración de procesos y cómo esta afecta el estado global del sistema, se utilizará la desviación estándar. Esta medida considerará la carga final de los nodos del sistema distribuido para evaluar la variabilidad o dispersión de los datos.

La desviación estándar se aplicará a los promedios finales obtenidos de la aplicación de sobrecarga sin migración y de la aplicación de sobrecarga con migración. Si la desviación estándar de los valores de los promedios finales de sobrecarga sin migración es mayor que la desviación estándar de los mismos con migración, esto indica que hay una mayor variabilidad en los resultados. Por lo tanto, los valores de sobrecarga después de la migración estarán más balanceados porque estarán más cerca de la media.

Para el cálculo de la desviación estándar sin considerar migración se procede de la siguiente manera:

desvest_sin_migración (carga_final_sin_migración) = s

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}} \quad (11)$$

Donde $\{x_1, x_2, \dots, x_n\}$ son los valores de carga final de cada uno de los nodos (sin contemplar migración), \bar{x} es el valor medio de todos estos elementos y N es el número total de nodos en el sistema distribuido.

Para el cálculo de la desviación estándar considerando la migración se procede de manera similar considerando los valores de sobrecarga luego de la migración de procesos, aplicando la fórmula (11).

Ejemplo

En esta sección se explicará en detalle un ejemplo de aplicación del operador de agregación propuesto. El sistema de procesamiento distribuido, las estructuras de datos, los recursos y los procesos que se ejecutan en los diferentes nodos, grupos, cardinalidades, criterios y categorías para evaluar las diferentes cargas y cálculos necesarios, son los mencionados en (La Red Martínez, 2017), (Agostini et al., 2018) y Agostini et al. (2019a, 2019b, 2019c).

Para generar la entrada de datos para el ejemplo desarrollado se ha utilizado un simulador (La Red Martínez, 2020), haciendo posible la representación de los procesos, recursos y nodos en un modelo dinámico, simulándose diferentes cargas de trabajo para el sistema distribuido.

El sistema de procesamiento distribuido tiene diez nodos:

$$nodos = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Los procesos que se ejecutan en los nodos son los siguientes: cinco procesos en el nodo 1, cinco procesos en el nodo 2 y siete procesos en el nodo 3. Los nodos del 4 al 10 son nodos que se agrega al modelo original mencionados en (La Red Martínez, 2017), (Agostini et al., 2018) y Agostini et al. (2019a, 2019b, 2019c).

procesos = {p_{ij}} con i indicando el nodo y j indicando el proceso, lo que se puede expresar mediante la Tabla 14.

Tabla 14. Procesos en cada nodo

Nodos	Procesos						
1	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{17}
2	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}		
3	p_{31}	p_{32}	p_{33}	p_{34}	p_{35}	p_{36}	p_{37}

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Los recursos compartidos disponibles en los nodos son los siguientes: cuatro recursos en el nodo 1, cuatro en el nodo 2, tres en el nodo 3, tres en el nodo 4, tres en el nodo 5, dos en el nodo 6, cuatro en el nodo 7 y tres en el nodo 8.

$recursos = \{p_{ij}\}$ con i indicando el nodo y j indicando el recurso, lo que se puede expresar mediante la Tabla 15.

Las solicitudes de recursos por parte de los procesos se muestran en la Tabla 16. A continuación, se describirá cada una de las etapas de cálculo.

Tabla 15. Recursos compartidos disponibles en cada nodo

Nodos	Recursos			
1	r_{11}	r_{12}	r_{13}	r_{14}
2	r_{21}	r_{22}	r_{23}	r_{24}
3	r_{31}	r_{32}	r_{33}	
4	r_{41}	r_{42}	r_{43}	
5	r_{51}	r_{52}	r_{53}	
6	r_{61}	r_{62}		
7	r_{71}	r_{72}	r_{73}	r_{74}
8	r_{81}	r_{82}	r_{83}	

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Tabla 16. Recursos solicitados por los procesos

Recursos	Procesos
r_{11}	$p_{11}, p_{12}, p_{13}, p_{14}, p_{16}, p_{17}, p_{24}, p_{32}, p_{33}, p_{36}, p_{37}$
r_{12}	$p_{11}, p_{12}, p_{13}, p_{14}, p_{16}, p_{17}, p_{21}, p_{23}, p_{24}, p_{32}, p_{33}, p_{34}, p_{35}, p_{36}, p_{37}$
r_{13}	$p_{13}, p_{14}, p_{16}, p_{17}, p_{21}, p_{31}, p_{32}, p_{33}, p_{34}, p_{35}, p_{36}$
r_{14}	p_{16}
r_{21}	$p_{11}, p_{12}, p_{13}, p_{14}, p_{15}, p_{16}, p_{22}, p_{25}, p_{33}, p_{36}, p_{37}$
r_{22}	$p_{11}, p_{12}, p_{13}, p_{14}, p_{16}, p_{21}, p_{22}, p_{33}, p_{34}, p_{35}, p_{36}$
r_{23}	$p_{11}, p_{21}, p_{24}, p_{32}, p_{33}, p_{34}$
r_{24}	$p_{11}, p_{23}, p_{24}, p_{34}, p_{35}, p_{36}$
r_{31}	$p_{12}, p_{13}, p_{16}, p_{21}, p_{22}, p_{23}, p_{31}, p_{34}, p_{35}, p_{36}$
r_{32}	$p_{13}, p_{15}, p_{23}, p_{33}, p_{34}, p_{35}, p_{36}, p_{37}$
r_{33}	$p_{12}, p_{13}, p_{15}, p_{16}, p_{21}, p_{22}, p_{23}, p_{31}, p_{33}, p_{34}, p_{35}, p_{36}, p_{37}$
r_{41}	p_{14}
r_{42}	p_{14}
r_{43}	p_{14}
r_{51}	p_{15}
r_{52}	p_{15}
r_{53}	p_{15}
r_{61}	p_{15}
r_{62}	p_{15}
r_{71}	p_{31}
r_{72}	p_{31}
r_{73}	p_{35}
r_{74}	p_{35}
r_{81}	p_{36}
r_{82}	p_{36}
r_{83}	p_{36}

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Cálculo de la carga computacional actual de los nodos

Para obtener un indicador de la carga computacional actual de cada nodo se adoptarán los mismos tres criterios en los tres nodos:

$$\text{card}(\{\text{criterios}\}) = 3$$

$\text{criterios} = \{\% \text{ de uso de la CPU, el \% de uso de la memoria, \% de uso de operaciones de entrada / salida}\}$.

Los valores que se asumirán para los indicadores de carga computacional de los nodos y el cálculo de carga promedio para cada nodo se muestran en la Tabla 17.

Establecimiento de las categorías de carga computacional y de los vectores de pesos asociados a las mismas

En esta propuesta las categorías serán las mismas para todos los nodos: Alta (si la carga es mayor al 70%), Media (si la carga está entre el 40% y el 70% inclusive) y Baja (si la carga es menor al 40%).

$$\text{card}(\{\text{categorías}\}) = 3$$

$\text{categorías} = \{\text{Alta, Media, Baja}\}$.

Los valores que se obtienen para las categorías de carga en base a los promedios mostrados en la Tabla 17, se indican en la Tabla 18.

Tabla 17. Valores de los criterios de carga computacional

Nodos	Valores de los Criterios			Promedio
1	80	90	75	81.67
2	30	30	50	36.67
3	77	49	85	70.33
4	6	7	9	7.33
5	5	5	6	5.33
6	7	9	7	7.67
7	7	10	15	10.67
8	9	6	7	7.33
9	10	9	15	11.33
10	8	7	8	7.67

Nota: Valores de los criterios para medir la carga computacional en cada nodo.

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Para establecer los vectores de pesos asociados a las categorías de carga computacional actual de cada nodo se utilizarán, para todos los nodos y para todas las categorías de carga, los siguientes criterios: N° de procesos en el nodo, % de uso de CPU, % de uso de memoria, % de uso de memoria virtual, prioridad del proceso (prioridad del proceso en el nodo donde se ejecuta), peso del recurso (peso de transferencia que genera sobre el ancho de banda), sobrecarga de memoria (memoria adicional que requerirá disponer el recurso solicitado, si el dato está disponible), sobrecarga de procesador (uso adicional de procesador que requerirá disponer el recurso solicitado, si el dato está disponible), sobrecarga de entrada / salida (entrada / salida adicional que requerirá disponer el recurso solicitado, si el dato está disponible).

Tabla 18. Valores de las categorías para medir la carga computacional

Nodos	Valores de las Categorías
1	Alta
2	Baja
3	Alta
4	Baja
5	Baja
6	Baja
7	Baja
8	Baja
9	Baja
10	Baja

Nota: Valores de las categorías para medir la carga computacional en cada nodo.

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

$$card(\{critpref\}) = 9$$

criterios para preferencias = {N° de procesos en el nodo, % de uso de CPU, % de uso de memoria, % de uso de memoria virtual, prioridad del proceso, peso del recurso, impacto transferencia sobre ancho de banda, sobrecarga de memoria, sobrecarga de procesador, sobrecarga de entrada / salida}

Seguidamente se deben establecer los valores correspondientes a los criterios constituyendo así los vectores de pesos para las distintas categorías de carga, que serán iguales para todos los nodos, lo cual se indica en la Tabla 19 que se continúa en la Tabla 20.

La sumatoria de los pesos asignados a los distintos criterios es 1 para cada una de las categorías, o lo que es lo mismo, que la suma de elementos del vector de pesos de cada categoría es 1.

Tabla 19. Pesos asignados a los criterios para calcular la prioridad (N° Procesos, %CPU, %Memoria, %MV, Prioridad Proceso)

Categorías	Pesos				Prioridad Proceso
	N° Procesos	% CPU	% Memoria	% MV	
Alta	0.05	0.1	0.1	0.3	0.1
Media	0.04	0.08	0.4	0.1	0.03
Baja	0.1	0.2	0.1	0.2	0.1

Tabla 20. Pesos asignados a los criterios para calcular la prioridad (Peso del Recurso, Impacto Transferencia, Sobrecarga Memoria, Sobrecarga Procesador, Sobrecarga E/S)

Categorías	Pesos				
	Peso Recurso	Impacto Transfer.	Sobrecarga Memoria	Sobrecarga Procesador	Sobrecarga E/S
Alta	0.1	0.05	0.1	0.05	0.05
Media	0.2	0.05	0.05	0.025	0.025
Baja	0.2	0.01	0.025	0.025	0.04

Nota: Pesos asignados a los criterios para calcular la prioridad o preferencia que cada nodo otorgará a cada requerimiento de cada proceso según la carga del nodo.

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Cálculo de las prioridades o preferencias de los procesos teniendo en cuenta el estado del nodo

Los vectores de valoraciones se aplican para cada requerimiento de un recurso hecho por un proceso, según los criterios establecidos para la determinación de la prioridad que en cada caso y momento fija el nodo en el cual se produce el requerimiento; cada vector de valoraciones de cada requerimiento se multiplica escalarmente por el vector de pesos correspondiente a la categoría de carga actual del nodo para obtener la prioridad según cada criterio y la prioridad nodal otorgada a cada requerimiento; esto se muestra en la Tabla 21 que se continúa en la Tabla 22.

Tabla 21. Valoraciones asignadas a los criterios para calcular la prioridad (N° proceso, %CPU, %Memoria, %MV, Prioridad Proceso)

Rec.	N° Proceso	%CPU	%Memoria	%MV	Prioridad Proceso
$p_{11}r_{11}$	0.7	0.5	0.7	0.9	0.8
$p_{11}r_{12}$	0.8	0.7	0.4	0.5	0.3
$p_{11}r_{21}$	0.3	0.4	0.5	0.2	0.9
$p_{11}r_{22}$	0.5	0.5	0.7	0.4	0.8
$p_{11}r_{23}$	0.5	0.6	0.8	0.8	0.9
$p_{11}r_{24}$	0.3	0.5	0.9	0.2	0.6
$p_{12}r_{11}$	0.4	0.7	0.5	0.9	1.0
$p_{12}r_{12}$	0.2	0.7	0.3	0.7	0.8
$p_{12}r_{21}$	0.7	0.4	0.3	0.7	0.8
$p_{12}r_{22}$	0.9	0.6	0.7	0.7	0.8
$p_{12}r_{31}$	0.2	0.5	0.7	0.7	0.3
$p_{12}r_{33}$	0.4	0.5	0.7	0.9	0.3
$p_{13}r_{11}$	0.5	0.7	0.7	0.8	0.6
$p_{13}r_{12}$	0.7	0.8	0.7	0.4	0.9
$p_{13}r_{13}$	0.7	0.6	0.7	0.8	0.9
$p_{13}r_{21}$	0.7	0.4	0.9	0.3	0.5
$p_{13}r_{22}$	0.5	0.9	0.8	0.3	0.5
$p_{13}r_{31}$	0.5	0.7	0.3	0.6	0.8
$p_{13}r_{32}$	0.6	0.9	0.3	0.6	0.4
$p_{13}r_{33}$	0.6	0.2	0.4	0.6	0.9
$p_{14}r_{11}$	0.5	0.7	0.6	0.7	0.7
$p_{14}r_{12}$	0.6	0.7	0.6	0.0	0.8
$p_{14}r_{13}$	0.6	0.5	0.6	3.0	0.8
$p_{14}r_{21}$	0.6	0.3	0.8	0.3	0.4
$p_{14}r_{22}$	0.4	0.8	0.7	0.3	0.4
$p_{14}r_{41}$	0.4	0.6	0.4	0.5	0.7
$p_{14}r_{42}$	0.6	0.8	0.2	0.5	0.8
$p_{14}r_{43}$	0.5	0.3	0.3	0.7	0.7
$p_{15}r_{21}$	0.6	0.8	0.8	0.9	0.7
$p_{15}r_{32}$	0.8	0.9	0.8	0.5	0.8
$p_{15}r_{33}$	0.7	0.3	0.5	0.7	0.6
$p_{15}r_{51}$	0.8	0.4	0.8	0.4	0.6
$p_{15}r_{52}$	0.6	0.8	0.9	0.4	0.6
$p_{15}r_{53}$	0.8	0.7	0.8	0.9	0.8
$p_{15}r_{61}$	0.6	0.8	0.4	0.7	0.8
$p_{15}r_{62}$	0.7	0.8	0.4	0.7	0.5

$p_{16}r_{11}$	0.4	0.7	0.5	0.9	1.0
$p_{16}r_{12}$	0.2	0.7	0.3	0.7	0.8
$p_{16}r_{13}$	0.3	0.6	0.3	0.6	0.7
$p_{16}r_{14}$	0.4	0.7	0.2	0.5	0.6
$p_{16}r_{21}$	0.7	0.4	0.3	0.7	0.8
$p_{16}r_{22}$	0.9	0.6	0.7	0.7	0.8
$p_{16}r_{31}$	0.2	0.5	0.7	0.7	0.3
$p_{16}r_{33}$	0.4	0.5	0.7	0.9	0.3
$p_{17}r_{11}$	0.3	0.6	0.5	0.8	0.6
$p_{17}r_{12}$	0.4	0.6	0.4	0.6	0.8
$p_{17}r_{13}$	0.5	0.7	0.6	0.9	0.7
$p_{21}r_{12}$	0.2	0.1	0.3	0.8	0.7
$p_{21}r_{13}$	0.2	0.4	0.8	0.9	0.7
$p_{21}r_{22}$	0.3	0.5	0.8	0.9	0.5
$p_{21}r_{23}$	0.7	0.5	0.6	0.8	0.5
$p_{21}r_{31}$	0.7	0.5	0.8	0.6	0.9
$p_{21}r_{33}$	0.7	0.4	0.9	0.8	0.4
$p_{22}r_{21}$	0.5	0.4	0.7	0.8	0.6
$p_{22}r_{22}$	0.5	0.9	0.6	0.8	0.9
$p_{22}r_{31}$	0.5	0.4	0.5	0.2	0.4
$p_{22}r_{33}$	0.8	0.4	0.3	0.2	0.8
$p_{23}r_{12}$	0.5	0.6	0.8	0.3	0.5
$p_{23}r_{24}$	0.6	0.2	0.1	0.7	0.3
$p_{23}r_{31}$	0.3	0.1	0.4	0.8	0.7
$p_{23}r_{32}$	0.4	0.4	0.8	0.2	0.4
$p_{23}r_{33}$	0.3	0.6	0.8	0.2	0.9
$p_{24}r_{11}$	0.4	0.6	0.7	0.9	0.5
$p_{24}r_{12}$	0.3	0.6	0.7	0.8	0.9
$p_{24}r_{23}$	0.4	0.9	0.4	0.8	0.8
$p_{24}r_{24}$	0.5	0.8	0.7	0.9	0.3
$p_{25}r_{21}$	0.2	0.8	0.8	0.9	0.4
$p_{31}r_{13}$	0.6	0.9	0.6	0.9	0.7
$p_{31}r_{71}$	0.8	0.3	0.9	0.5	0.7
$p_{31}r_{72}$	0.4	0.7	0.7	0.5	0.9
$p_{32}r_{11}$	0.6	0.9	0.8	0.5	0.9
$p_{32}r_{12}$	0.7	0.4	0.6	0.9	0.8
$p_{32}r_{13}$	0.8	0.9	1.0	0.7	0.9
$p_{32}r_{21}$	0.8	0.8	1.0	0.9	0.6
$p_{33}r_{11}$	0.2	0.7	0.9	0.8	0.6
$p_{33}r_{12}$	0.9	1.0	0.9	0.7	0.3
$p_{33}r_{13}$	0.9	0.5	0.7	0.9	0.3
$p_{33}r_{21}$	0.4	0.6	0.7	0.8	0.8
$p_{33}r_{22}$	0.9	0.4	0.7	0.8	0.7
$p_{33}r_{23}$	0.4	0.6	0.9	1.0	0.6
$p_{33}r_{32}$	0.6	0.6	0.7	0.8	0.4
$p_{33}r_{33}$	0.6	1.0	0.7	0.4	0.6
$p_{34}r_{12}$	0.8	1.0	0.3	0.5	0.7
$p_{34}r_{13}$	0.2	1.0	0.8	0.5	0.8
$p_{34}r_{22}$	0.9	0.8	0.6	0.8	0.9

p_{34I23}	0.4	0.6	0.7	0.8	0.9
p_{34I24}	0.9	0.4	0.7	0.4	0.7
p_{34I31}	0.5	0.6	0.7	0.9	0.7
p_{34I32}	0.8	0.6	0.9	0.5	0.6
p_{34I33}	0.4	0.6	0.8	0.5	0.9
p_{35I31}	0.9	0.8	0.7	0.4	0.8
p_{35I32}	0.9	0.7	0.8	0.6	0.4
p_{35I33}	0.5	0.7	0.6	0.9	0.8
p_{35I61}	0.2	0.4	0.7	0.4	0.9
p_{35I62}	0.8	0.9	0.7	0.3	0.9
p_{35I72}	0.3	0.8	0.9	0.4	0.8
p_{35I74}	0.5	0.8	0.9	0.4	0.6
p_{36I11}	0.8	0.9	0.6	0.5	0.8
p_{36I12}	0.7	0.9	0.4	0.9	0.8
p_{36I13}	0.9	0.9	0.8	0.7	0.8
p_{36I21}	0.8	0.9	0.5	0.3	0.7
p_{36I22}	0.8	0.2	0.1	0.3	0.8
p_{36I24}	0.4	0.7	0.9	0.3	0.8
p_{36I81}	0.5	0.9	0.9	0.7	0.8
p_{36I82}	0.9	0.8	0.6	0.7	0.8
p_{36I83}	0.2	0.7	0.4	0.8	0.8
p_{37I11}	0.9	0.6	0.8	0.6	0.9
p_{37I12}	0.7	0.9	0.8	0.7	0.5
p_{37I31}	0.9	0.7	0.8	0.6	0.6
p_{37I42}	0.8	0.9	0.5	0.3	0.8
p_{37I43}	0.8	0.4	0.6	0.8	0.8

Tabla 22. Valoraciones asignadas a los criterios para calcular la prioridad (Peso del Recurso, Impacto Transferencia, Sobrecarga Memoria, Sobrecarga Procesador, Sobrecarga E/S)

Rec.	Peso del Recurso	Impacto Transf.	Sobrecarga Memoria	Sobrecarga Procesador	Sobrecarga E/S
p_{11I11}	0.3	0.5	0.2	0.3	0.4
p_{11I12}	0.3	0.5	0.7	0.2	0.4
p_{11I21}	0.8	0.6	0.2	0.5	0.7
p_{11I22}	0.4	0.3	0.3	0.5	0.6
p_{11I23}	0.7	0.5	0.9	0.7	0.6
p_{11I24}	0.2	0.2	0.6	0.7	0.4
p_{12I11}	0.7	0.6	0.9	0.8	0.8
p_{12I12}	0.3	0.2	0.3	0.8	0.9
p_{12I21}	0.3	0.3	0.9	0.5	0.2
p_{12I22}	0.9	0.7	0.2	0.5	0.4
p_{12I31}	0.8	0.3	0.2	0.7	0.8
p_{12I33}	0.5	0.9	0.4	0.5	0.8
p_{13I11}	0.5	0.3	0.5	0.7	0.8
p_{13I12}	0.8	0.9	0.2	0.9	0.7
p_{13I13}	0.8	0.9	0.4	0.8	0.7

$p_{13}r_{21}$	0.4	0.9	0.7	0.2	0.3
$p_{13}r_{22}$	0.6	0.4	0.4	0.9	0.3
$p_{13}r_{31}$	0.6	0.4	0.9	0.9	0.5
$p_{13}r_{32}$	0.6	0.6	0.8	0.7	0.8
$p_{13}r_{33}$	0.8	0.7	0.8	0.5	0.8
$p_{14}r_{11}$	0.6	0.2	0.4	0.6	0.7
$p_{14}r_{12}$	0.7	0.8	0.2	0.8	0.6
$p_{14}r_{13}$	0.9	0.8	0.3	0.9	0.6
$p_{14}r_{21}$	0.5	0.8	0.8	0.3	0.2
$p_{14}r_{22}$	0.7	0.5	0.5	0.8	0.2
$p_{14}r_{41}$	0.5	0.3	0.8	0.8	0.6
$p_{14}r_{42}$	0.7	0.6	0.9	0.6	0.7
$p_{14}r_{43}$	0.7	0.5	0.9	0.6	0.7
$p_{15}r_{21}$	0.6	0.4	0.6	0.8	0.7
$p_{15}r_{32}$	0.9	0.8	0.3	0.8	0.8
$p_{15}r_{33}$	0.9	0.8	0.9	0.6	0.9
$p_{15}r_{53}$	0.9	0.8	0.5	0.9	0.8
$p_{15}r_{51}$	0.5	0.8	0.8	0.3	0.4
$p_{15}r_{52}$	0.7	0.5	0.5	0.8	0.4
$p_{15}r_{61}$	0.7	0.5	0.8	0.8	0.6
$p_{15}r_{62}$	0.7	0.7	0.9	0.8	0.9
$p_{16}r_{11}$	1.0	0.6	0.8	0.7	0.7
$p_{16}r_{12}$	1.0	0.2	0.4	0.7	0.8
$p_{16}r_{13}$	0.7	0.3	0.3	0.5	0.4
$p_{16}r_{14}$	0.8	0.4	0.5	0.6	0.7
$p_{16}r_{21}$	0.3	0.3	0.9	0.5	0.2
$p_{16}r_{22}$	0.9	0.7	0.2	0.5	0.4
$p_{16}r_{31}$	0.8	0.3	0.2	0.7	0.8
$p_{16}r_{33}$	0.5	0.9	0.4	0.5	0.8
$p_{17}r_{11}$	0.3	0.5	0.8	0.9	0.9
$p_{17}r_{12}$	0.4	0.3	0.4	0.7	0.8
$p_{17}r_{13}$	0.2	0.3	0.3	0.8	0.6
$p_{21}r_{12}$	0.8	0.7	0.7	0.5	0.8
$p_{21}r_{13}$	0.4	0.3	0.4	0.5	0.2
$p_{21}r_{22}$	0.6	0.2	0.6	0.2	0.4
$p_{21}r_{23}$	0.4	0.8	0.2	0.9	0.4
$p_{21}r_{31}$	0.2	0.8	0.4	0.9	0.9
$p_{21}r_{33}$	0.7	0.7	0.8	0.6	0.6
$p_{22}r_{21}$	0.8	0.8	0.4	0.6	0.9
$p_{22}r_{22}$	0.3	0.6	0.4	0.2	0.1
$p_{22}r_{31}$	0.5	0.7	0.8	0.2	0.9
$p_{22}r_{33}$	0.2	0.7	0.9	0.5	0.8
$p_{23}r_{12}$	0.2	0.7	0.7	0.5	0.5
$p_{23}r_{24}$	0.8	0.7	0.8	0.9	0.4
$p_{23}r_{31}$	0.5	0.3	0.9	0.4	0.6
$p_{23}r_{32}$	0.5	0.6	0.6	0.6	0.6
$p_{23}r_{33}$	0.6	0.8	0.6	0.4	0.2
$p_{24}r_{11}$	0.7	0.9	0.7	0.9	0.5
$p_{24}r_{12}$	0.4	0.5	0.7	0.6	0.9
$p_{24}r_{23}$	0.4	0.8	0.7	0.6	0.3
$p_{24}r_{24}$	0.6	0.4	0.4	0.8	0.7

$p_{25r_{21}}$	0.9	0.7	0.5	0.6	0.8
$p_{31r_{13}}$	0.7	0.8	0.4	0.9	0.8
$p_{31r_{71}}$	0.6	0.9	0.3	0.9	0.7
$p_{31r_{72}}$	0.9	0.5	0.9	0.7	0.7
$p_{32r_{11}}$	0.2	0.6	0.7	0.3	0.7
$p_{32r_{12}}$	0.4	0.3	1.0	0.9	0.7
$p_{32r_{13}}$	0.3	0.5	0.3	0.5	0.9
$p_{32r_{21}}$	0.6	0.6	0.8	0.4	0.9
$p_{33r_{11}}$	0.4	0.2	0.9	1.0	0.3
$p_{33r_{12}}$	0.8	0.5	0.5	0.7	0.3
$p_{33r_{13}}$	0.7	0.5	0.4	0.5	0.8
$p_{33r_{21}}$	0.4	0.9	0.4	0.8	0.8
$p_{33r_{22}}$	0.9	0.5	0.4	0.9	0.7
$p_{33r_{23}}$	0.3	0.8	0.4	0.7	0.8
$p_{33r_{32}}$	0.3	0.2	0.5	0.8	0.8
$p_{33r_{33}}$	0.2	0.2	0.8	0.8	0.9
$p_{34r_{12}}$	0.5	0.6	0.3	0.8	0.7
$p_{34r_{13}}$	0.7	0.7	0.6	0.9	0.7
$p_{34r_{22}}$	0.5	0.6	0.8	0.5	0.6
$p_{34r_{23}}$	0.3	0.3	0.5	0.9	0.6
$p_{34r_{24}}$	0.4	0.6	0.5	0.9	0.7
$p_{34r_{31}}$	0.4	0.5	0.9	0.6	0.7
$p_{34r_{32}}$	0.4	0.2	0.9	0.3	0.9
$p_{34r_{33}}$	0.7	0.8	0.9	0.4	0.3
$p_{35r_{31}}$	0.2	0.5	0.3	0.4	0.8
$p_{35r_{32}}$	0.7	0.7	0.9	0.4	0.7
$p_{35r_{33}}$	0.6	0.7	0.5	0.9	0.7
$p_{35r_{61}}$	0.3	0.7	0.5	0.7	0.9
$p_{35r_{62}}$	0.9	0.9	0.8	0.7	0.4
$p_{35r_{72}}$	0.9	0.4	0.6	0.3	0.6
$p_{35r_{74}}$	0.9	0.5	0.9	0.4	0.7
$p_{36r_{11}}$	0.7	0.6	0.9	0.9	0.6
$p_{36r_{12}}$	0.8	0.9	0.9	0.4	0.7
$p_{36r_{13}}$	0.3	0.5	0.7	0.9	0.7
$p_{36r_{21}}$	0.5	0.4	0.7	0.9	0.9
$p_{36r_{22}}$	0.5	0.7	0.7	0.6	0.9
$p_{36r_{24}}$	0.6	0.2	0.5	0.8	0.9
$p_{36r_{81}}$	0.7	0.7	0.8	0.8	0.7
$p_{36r_{82}}$	0.3	0.6	0.5	0.6	0.7
$p_{36r_{83}}$	0.2	0.4	0.9	0.4	0.5
$p_{37r_{11}}$	0.4	0.9	0.7	0.9	0.8
$p_{37r_{12}}$	0.7	0.5	0.8	0.8	0.6
$p_{37r_{31}}$	0.8	0.5	0.9	0.5	0.4
$p_{37r_{42}}$	0.5	0.9	0.7	0.4	0.6
$p_{37r_{43}}$	0.8	0.3	0.6	0.9	0.3

Determinación de procesos migrables mediante el cálculo de prioridad promedio de los procesos

Por cada solicitud de recursos de cada proceso, se establece un criterio de prioridad, que es asignado por el Runtime al momento de hacer la solicitud. La sumatoria de este valor de prioridad, para cada recurso solicitado, dividido el total de solicitudes de cada proceso,

indicará la prioridad promedio del proceso. Los nodos con carga alta son el 1 y 3, y sus procesos se evalúan según la siguiente fórmula:

$$\text{promedio de prioridad de proceso } (p_{ij}) = (\sum cp_{ijk} / o) \quad (12)$$

Nodo 1

promedio de prioridad de proceso $(p_{11}) = (0.8+0.3+0.9+0.8+0.95+ 0.6) / 6 = \mathbf{0.725}$
 promedio de prioridad de proceso $(p_{12}) = (1+0.8+0.8+0.8+0.3+0.3) / 6 = \mathbf{0.667}$
 promedio de prioridad de proceso $(p_{13}) = (0.6+0.9+0.9+0.5+0.5+ 0.8+0.4+0.9) / 8 = \mathbf{0.687}$
 promedio de prioridad de proceso $(p_{14}) = (0.7+0.8+0.8+0.4+0.4+ 0.7+0.8+0.7) / 8 = \mathbf{0.662}$
 promedio de prioridad de proceso $(p_{15}) = (0.7+0.8+0.6+0.8+0.6+ 0.6+0.8+0.5) / 8 = \mathbf{0.675}$

Nodo 3

promedio de prioridad de proceso $(p_{31}) = (0.7+0.7+0.9) / 3 = \mathbf{0.767}$
 promedio de prioridad de proceso $(p_{32}) = (0.9+0.8+0.9+0.6) / 4 = \mathbf{0.800}$
promedio de prioridad de proceso $(p_{33}) = (0.6+0.3+0.3+0.8+0.7+ 0.6+0.4+0.6) / 5 = \mathbf{0.538}$
 promedio de prioridad de proceso $(p_{34}) = (0.7+0.8+0.9+0.9+0.7+ 0.7+0.6+0.9) / 8 = \mathbf{0.775}$
 promedio de prioridad de proceso $(p_{35}) = (0.9+0.9+0.8+0.6+0.8+ 0.4+0.8) / 7 = \mathbf{0.743}$
 promedio de prioridad de proceso $(p_{36}) = (0.8+0.8+0.8+0.7+0.8+ 0.8+0.8+0.8+0.8) / 9 = \mathbf{0.789}$
 promedio de prioridad de proceso $(p_{37}) = (0.9+0.5+0.6+0.8+0.8) / 5 = \mathbf{0.720}$

El proceso p_{33} no se evalúa en este ciclo porque para el modelo propuesto sólo se consideran aquellos procesos cuyo promedio de prioridad de proceso supere el límite establecido, para este caso 0.6. Los procesos que clasifican para evaluación para posible migración a otro nodo se listan en la Tabla 23.

Tabla 23. Lista de procesos con prioridad de procesos superior al límite

Nodos	Procesos	Promedio de prioridad de proceso
1	p_{11}	0.725
1	p_{12}	0.667
1	p_{13}	0.687
1	p_{14}	0.662
1	p_{15}	0.675
3	p_{31}	0.767
3	p_{32}	0.800
3	p_{34}	0.775
3	p_{35}	0.743
3	p_{36}	0.789
3	p_{37}	0.720

Fuente: Elaboración propia. Basado en La Red Martínez (2017).

Selección de los posibles nodos destino mediante la evaluación de su carga

En esta etapa se evalúa si existen nodos con carga baja que puedan alojar procesos de otros nodos. De la Tabla 18 que contiene los valores de la carga computacional, se filtran solamente los nodos que se encuentran dentro de la categoría de carga baja.

$$\text{nodo destino con carga baja} = \{2, 4, 5, 6, 7, 8, 9, 10\}$$

Para cada proceso puede haber varios nodos con carga baja disponible. Teniendo en cuenta el conjunto de nodos con carga baja se verifican los recursos solicitados por los procesos p_{ij} para determinar el nodo candidato para una posible migración.

Evaluación del peso de los recursos y su posible impacto ante la migración de procesos

Una vez identificados los nodos con carga alta, sus procesos y los posibles nodos destinos, se realiza la comparación de los recursos locales en relación con los remotos, con el fin de elaborar la lista preliminar de los procesos migrables y sus posibles nodos destinos.

En la Figura 7 se muestra el vector de migración de procesos donde se almacenan valores del peso de migración que generaría el proceso si se migrara a un determinado nodo destino, habrá un vector de migración de procesos para cada proceso en relación con el nodo destino evaluado.

Para cada nodo con carga alta se realiza la evaluación de sus procesos para las diferentes situaciones que pueden presentarse según la Figura 6 y la Tabla 22.

En la Tabla 24 se muestra el vector de migración de proceso para el proceso p_{11} que tiene los siguientes pesos de recursos asociados: el peso de recursos locales es de 0.6 y como peso de recursos remotos en el nodo 2 es de 2.1. Analizando la Figura 6 en relación con el nodo 2 su valor se graba dada la situación (u), y en relación con los nodos 4 al 10 su valor se graba dada la situación (q).

Tabla 24. Vector de migración del proceso p_{11}

vmp ¹			Opciones			
q	s	u	(q)	(u)	(v)	(w)
1	1	2	-	0.6	-	-
1	1	4	2.7	-	-	-
1	1	5	2.7	-	-	-
1	1	6	2.7	-	-	-
1	1	7	2.7	-	-	-
1	1	8	2.7	-	-	-
1	1	9	2.7	-	-	-
1	1	10	2.7	-	-	-

En la Tabla 25 se muestra el vector de migración de proceso para el proceso p_{12} que tiene los siguientes pesos de recursos asociados: el peso de recursos locales es de 1.0, el peso de recursos remotos está distribuido en dos nodos, en el nodo 2 es de 1.2 y en el nodo 3 es de 1.3. Analizando la Figura 6 en relación con el nodo 2 su valor se graba dada la situación (u). El proceso p_{12} podría migrar únicamente al nodo 2 ya que hacerlo a otro nodo generaría un alto impacto en el ancho de banda.

En la Tabla 26 se muestra el vector de migración de proceso para el proceso p_{13} que tiene los siguientes pesos de recursos asociados: el peso de recursos locales es de 2.1, el peso de recursos remotos está distribuido en dos nodos, en el nodo 2 es de 1.0 y en el nodo 3 es de 2.0. Analizando la Figura 6 en relación con el nodo 2 su valor se graba dada la situación (u). El proceso p_{13} podría migrar únicamente al nodo 2 ya que hacerlo a otro nodo generaría un alto impacto en el ancho de banda.

En la Tabla 27 se muestra el vector de migración de proceso para el proceso p_{14} que tiene los siguientes pesos de recursos asociados: el peso de recursos locales es de 2.2, el peso de recursos remotos está distribuido en dos nodos, en el nodo 2 es de 1.2 y en el nodo 4 es de 1.9. Analizando la Figura 6 en relación con los nodos 2 y 4 su valor se graba dada la situación (u). El proceso p_{14} podría migrar únicamente al nodo 2 o nodo 4 ya que hacerlo a otro nodo generaría un alto impacto en el ancho de banda.

Tabla 25. Vector de migración del proceso p_{12}

vmp ²			Opciones			
q	s	u	(q)	(u)	(v)	(w)
1	2	2	-	2.3	-	-
1	2	4	-	-	-	-
1	2	5	-	-	-	-

1	2	6	-	-	-	-
1	2	7	-	-	-	-
1	2	8	-	-	-	-
1	2	9	-	-	-	-
1	2	10	-	-	-	-

Tabla 26. Vector de migración del proceso p_{13}

vmp ³			Opciones			
q	s	u	(q)	(u)	(v)	(w)
1	3	2	-	4.1	-	-
1	3	4	-	-	-	-
1	3	5	-	-	-	-
1	3	6	-	-	-	-
1	3	7	-	-	-	-
1	3	8	-	-	-	-
1	3	9	-	-	-	-
1	3	10	-	-	-	-

Tabla 27. Vector de migración del proceso p_{14}

vmp ⁴			Opciones			
q	s	u	(q)	(u)	(v)	(w)
1	4	2	-	4.1	-	-
1	4	4	-	3.4	-	-
1	4	5	-	-	-	-
1	4	6	-	-	-	-
1	4	7	-	-	-	-
1	4	8	-	-	-	-
1	4	9	-	-	-	-
1	4	10	-	-	-	-

En la Tabla 28 se muestra el vector de migración de proceso para el proceso p_{15} que tiene los siguientes pesos de recursos asociados: el peso de recursos remotos está distribuido en cuatro nodos, en el nodo 2 es de 0.6, en el nodo 3 es de 1.8, en el nodo 5 es de 2.1 y en el nodo 6 es de 1.4. Analizando la Figura 6 se observan dos grupos de nodos posibles, el primer grupo integrados por los nodos 2, 5, 7, 8, 9 y 10, su valor se graba dada la situación (v), y el segundo grupo integrados por los nodos 4 y 6, su valor se graba dada la situación (q).

El proceso p_{16} se descarta de la posibilidad de migración porque no cumple las condiciones para migrarse a otros nodos ya que hacerlo generaría un alto impacto en el ancho de banda.

En la Tabla 29 se muestra el vector de migración de proceso para el proceso p_{17} , el peso de recursos locales de 0.9, no posee recursos remotos y dado que su peso no supera el límite establecido en la situación (p) de la Figura 6, cualquiera de los nodos con carga baja del sistema distribuido es candidato para una posible migración y su valor se graba dada la situación (q).

En la Tabla 30 se muestra el vector de migración de proceso correspondiente al proceso p_{31} del nodo 3 que es el siguiente con carga alta.

Tabla 28. Vector de migración del proceso p_{15}

vmp ⁵			Opciones			
q	s	u	(q)	(u)	(v)	(w)
1	5	2	-	-	5.3	-
1	5	4	5.9	-	-	-
1	5	5	-	-	3.8	-

1	5	6	4.5	-	-	-
1	5	7	-	-	5.9	-
1	5	8	-	-	5.9	-
1	5	9	-	-	5.9	-
1	5	10	-	-	5.9	-

Tabla 29. Vector de migración del proceso p_{17}

vmp ⁷			Opciones			
q	s	u	(q)	(u)	(v)	(w)
1	7	2	0.9	-	-	-
1	7	4	0.9	-	-	-
1	7	5	0.9	-	-	-
1	7	6	0.9	-	-	-
1	7	7	0.9	-	-	-
1	7	8	0.9	-	-	-
1	7	9	0.9	-	-	-
1	7	10	0.9	-	-	-

Respecto del proceso p_{31} , su peso de recursos remotos está distribuido en dos nodos, en el nodo 1 es de 0.7 y en el nodo 7 es de 1.5. Analizando la Figura 6 y teniendo en cuenta que el proceso p_{31} no posee recursos locales, en relación con los nodos 2, 4, 5, 6, 8, 9, 10, su valor se graba dada la situación (w) y en relación con el nodo 7 su valor se graba dada la situación (v).

Tabla 30. Vector de migración del proceso p_{31}

vmp ¹			Opciones			
q	s	u	(q)	(u)	(v)	(w)
3	1	2	-	-	-	2.2
3	1	4	-	-	-	2.2
3	1	5	-	-	-	2.2
3	1	6	-	-	-	2.2
3	1	7	-	-	0.7	-
3	1	8	-	-	-	2.2
3	1	9	-	-	-	2.2
3	1	10	-	-	-	2.2

En la Tabla 31 se muestra el vector de migración de proceso para el proceso p_{32} que tiene los siguientes pesos de recursos asociados: el peso de recursos remotos está distribuido en dos nodos, en el nodo 1 es de 0.9 y en el nodo 2 es de 0.6. Analizando la Figura 6 y teniendo en cuenta que el proceso p_{32} no posee recursos locales, en relación con los nodos 4, 5, 6, 7, 8, 9, 10 su valor se graba dada la situación (w) y en relación con el nodo 2 el valor se graba dada la situación (v).

En la Tabla 32 se muestra el vector de migración de proceso para el proceso p_{34} que tiene los siguientes pesos de recursos asociados, el peso de recursos locales es de 1.5, el peso de recursos remotos está distribuido en dos nodos, en el nodo 1 es de 1.2 y en el nodo 2 es de 1.2. Analizando la Figura 6 en relación con el nodo 2 su valor se graba dada la situación (u).

El proceso p_{34} podría migrar únicamente al nodo 2 ya que hacerlo a otro nodo generaría un alto impacto en el ancho de banda. En la Tabla 33 se muestra el vector de migración de proceso para el proceso p_{35} que tiene los siguientes pesos de recursos asociados: el peso de los recursos locales es de 1.5, el peso de recursos remotos está distribuido en dos nodos, en el nodo 6 es de 1.2 y en el nodo 7 es de 1.8. Analizando la Figura 6 en relación con el nodo 6 y nodo 7, su valor se graba dada la situación (u). El proceso p_{35} podría migrar únicamente al nodo 6 o nodo 7 ya que hacerlo a otro nodo generaría un alto impacto en el ancho de banda.

Tabla 31. Vector de migración del proceso p_{32}

vmp ²				Opciones			
q	s	u	(q)	(u)	(v)	(w)	
3	2	2	-	-	0.9	-	
3	2	4	-	-	-	1.5	
3	2	5	-	-	-	1.5	
3	2	6	-	-	-	1.5	
3	2	7	-	-	-	1.5	
3	2	8	-	-	-	1.5	
3	2	9	-	-	-	1.5	
3	2	10	-	-	-	1.5	

Tabla 32. Vector de migración del proceso p_{34}

vmp ⁴				Opciones			
q	s	u	(q)	(u)	(v)	(w)	
3	4	2	-	2.7	-	-	
3	4	4	-	-	-	-	
3	4	5	-	-	-	-	
3	4	6	-	-	-	-	
3	4	7	-	-	-	-	
3	4	8	-	-	-	-	
3	4	9	-	-	-	-	
3	4	10	-	-	-	-	

Tabla 33. Vector de migración del proceso p_{35}

vmp ⁵				Opciones			
q	s	u	(q)	(u)	(v)	(w)	
3	5	2	-	-	-	-	
3	5	4	-	-	-	-	
3	5	5	-	-	-	-	
3	5	6	-	3.3	-	-	
3	5	7	-	2.7	-	-	
3	5	8	-	-	-	-	
3	5	9	-	-	-	-	
3	5	10	-	-	-	-	

En la Tabla 34 se muestra el vector de migración de proceso para proceso p_{36} que tiene los siguientes pesos de recursos asociados: el peso de recursos remotos está distribuido en tres nodos, en el nodo 1 es de 1.8, en el nodo 2 es de 1.6 y en el nodo 8 es de 1.8. Analizando la Figura 6 se observan dos grupos de nodos posibles, el primer grupo integrado por los nodos 2 y 8, su valor se graba mediante la situación (v), y el segundo grupo integrados por los nodos 4, 5, 6, 7, 9 y 10, su valor se graba dada la situación (w).

Tabla 34. Vector de migración del proceso p_{36}

vmp ⁶			Opciones			
q	s	u	(q)	(u)	(v)	(w)
3	6	2	-	-	3	-
3	6	4	-	-	-	4.6
3	6	5	-	-	-	4.6
3	6	6	-	-	-	4.6
3	6	7	-	-	-	4.6
3	6	8	-	-	3.4	-
3	6	9	-	-	-	4.6
3	6	10	-	-	-	4.6

En la Tabla 35 se muestra el vector de migración de proceso para el proceso p_{37} que tiene los siguientes pesos de recursos asociados, el peso de recursos locales es de 0.8, el peso de recursos remotos está distribuido en dos nodos, en el nodo 1 es de 1.1 y en el nodo 4 es de 1.3. Analizando la Figura 6 en relación con los nodos 2, 5, 6, 7, 8, 9 y 10, su valor se graba dada la situación (q), y en relación con el nodo 4, su valor se graba dada la situación (u).

Una vez evaluadas todas las situaciones de cada proceso en relación con cada nodo destino posible, se genera la matriz general de procesos (MGP) que se observa en la Tabla 36.

Tabla 35. Vector de migración del proceso p_{37}

vmp ⁷			Opciones			
q	s	u	(q)	(u)	(v)	(w)
3	7	2	3.2	-	-	-
3	7	4	-	1.9	-	-
3	7	5	3.2	-	-	-
3	7	6	3.2	-	-	-
3	7	7	3.2	-	-	-
3	7	8	3.2	-	-	-
3	7	9	3.2	-	-	-
3	7	10	3.2	-	-	-

Tabla 36. Matriz general de procesos (MGP)

Procesos	Nodos							
	2	4	5	6	7	8	9	10
p_{11}	0.6	2.7	2.7	2.7	2.7	2.7	2.7	2.7
p_{12}	2.3	-	-	-	-	-	-	-
p_{13}	4.1	-	-	-	-	-	-	-
p_{14}	4.1	3.4	-	-	-	-	-	-
p_{15}	5.3	5.9	3.8	4.5	5.9	5.9	5.9	5.9
p_{17}	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
p_{31}	2.2	2.2	2.2	2.2	0.7	2.2	2.2	2.2
p_{32}	0.9	1.5	1.5	1.5	1.5	1.5	1.5	1.5
p_{34}	2.7	-	-	-	-	-	-	-
p_{35}	-	-	-	3.3	2.7	-	-	-
p_{36}	3	4.6	4.6	4.6	4.6	3.4	4.6	4.6
p_{37}	3.2	1.9	3.2	3.2	3.2	3.2	3.2	3.2

Determinación del orden de migración de procesos y de los nodos destino correspondientes

En la matriz general de procesos se almacenan datos de procesos y los pesos de transferencia (migración) que se generarían al ser migrado cada proceso. Las celdas verdes indican que el proceso tiene asignado algún recurso en ese nodo, las celdas en rosa que no lo tiene.

El primer paso que habrá que realizar a partir de los datos de la matriz general de procesos es dividirla en dos matrices, una primaria (ver Tabla 37) y otra secundaria (ver Tabla 38). La diferencia radica en que primero se intenta favorecer a los procesos que tienen recursos en los posibles nodos destino, es decir, cuya migración implicaría una reducción en el impacto de transferencia (consumo de ancho de banda) con respecto a aquellos que no tengan recursos en nodos destinos. La segunda matriz se analizará luego de concluir las evaluaciones de solicitudes de la primera matriz.

Para el modelo propuesto, cada nodo con una carga alta podrá migrar hasta el 40% de sus procesos, con el fin de evitar migrar todos sus procesos y sobrecargar otros nodos.

Tabla 37. Matriz primaria (MP). Primera ronda

Procesos	Nodos							
	2	4	5	6	7	8	9	10
p_{11}	0.6	2.7	2.7	2.7	2.7	2.7	2.7	2.7
p_{12}	2.3	-	-	-	-	-	-	-
p_{13}	4.1	-	-	-	-	-	-	-
p_{14}	4.1	3.4	-	-	-	-	-	-
p_{15}	5.3	5.9	3.8	4.5	5.9	5.9	5.9	5.9
p_{31}	2.2	2.2	2.2	2.2	0.7	2.2	2.2	2.2
p_{32}	0.9	1.5	1.5	1.5	1.5	1.5	1.5	1.5
p_{34}	2.7	-	-	-	-	-	-	-
p_{35}	-	-	-	3.3	2.7	-	-	-
p_{36}	3	4.6	4.6	4.6	4.6	3.4	4.6	4.6
p_{37}	3.2	1.9	3.2	3.2	3.2	3.2	3.2	3.2

Tabla 38. Matriz secundaria (MS). Primera ronda

Procesos	Nodos							
	2	4	5	6	7	8	9	10
p_{17}	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9

El orden de asignación de los procesos a nodos destino de la matriz primaria se indica mediante la Función de Asignación para la Migración en Sistemas Distribuidos (FAMSD) primaria.

El primer proceso asignado es aquel que tiene menor peso de transferencia. En este caso el proceso p_{11} es el que implica menor peso de transferencia, que corresponde al nodo 2.

Lo anterior define la primera ronda de asignación (Ver Tabla 39); se continua la evaluación mientras existan procesos a evaluar.

Tabla 39. Función de Asignación para Migración en Sistemas Distribuidos (FAMSD) primaria. Primera ronda

Peso de transferencia del proceso	Proceso seleccionado	Nodo destino
0.6	p_{11}	2

El siguiente paso es eliminar el proceso p_{11} y el nodo 2 de la matriz primaria porque fueron asignados en la primera ronda. Aquellos procesos que tenían recursos destino en el nodo 2 (ya eliminado) pero tienen como opciones otros nodos destinos (que no poseen recursos destino de ese proceso) pasan a formar parte de la matriz secundaria (Ver Tabla 40 y Tabla

41). Los demás procesos cuya única opción es el nodo 2 se descartan y pasan a formar parte de la matriz de procesos no migrables (Ver Tabla 42).

Tabla 40. Matriz primaria (MP). Segunda ronda

Procesos	Nodos						
	4	5	6	7	8	9	10
p_{14}	3.4	-	-	-	-	-	-
p_{15}	5.9	3.8	4.5	5.9	5.9	5.9	5.9
p_{31}	2.2	2.2	2.2	0.7	2.2	2.2	2.2
p_{35}	-	-	3.3	2.7	-	-	-
p_{36}	4.6	4.6	4.6	4.6	3.4	4.6	4.6
p_{37}	1.9	3.2	3.2	3.2	3.2	3.2	3.2

Tabla 41. Matriz secundaria (MS). Segunda ronda

Procesos	Nodos						
	4	5	6	7	8	9	10
p_{17}	0.9	0.9	0.9	0.9	0.9	0.9	0.9
p_{32}	1.5	1.5	1.5	1.5	1.5	1.5	1.5

Tabla 42. Matriz de procesos no migrables. Segunda ronda

Procesos	Nodos							
	2	4	5	6	7	8	9	10
p_{12}	2.3	-	-	-	-	-	-	-
p_{13}	4.1	-	-	-	-	-	-	-
p_{34}	2.7	-	-	-	-	-	-	-

Se continúa con la matriz primaria hasta agotar opciones de procesos mientras los nodos no alcancen el porcentaje de migración fijado.

El segundo proceso de la matriz primaria de la segunda ronda con menor peso de transferencia es el proceso p_{31} , que corresponde al nodo 7.

Lo anterior define la segunda ronda de asignación (Ver Tabla 43), se continua la evaluación mientras existan procesos a evaluar.

Tabla 43. Función de Asignación para Migración en Sistemas Distribuidos (FAMSD) primaria. Segunda ronda

Peso de transferencia del proceso	Proceso seleccionado	Nodo destino
0.6	p_{11}	2
0.7	p_{31}	7

El siguiente paso es eliminar el proceso p_{31} y el nodo 7 de la matriz primaria porque fueron asignados en la segunda ronda. Aquellos procesos que tenían recursos destino en el nodo 7 (ya eliminado) pero tienen como opciones otros nodos destinos (que no poseen recursos destino de ese proceso) pasan a formar parte de la matriz secundaria (Ver Tabla 44 y Tabla 45).

Los demás procesos cuya única opción es el nodo 7 se descartan y pasan a formar parte de la matriz de procesos no migrables (Ver Tabla 46).

Tabla 44. Matriz primaria (MP). Tercera ronda

Procesos	Nodos
----------	-------

	4	5	6	8	9	10
p_{14}	3.4	-	-	-	-	-
p_{15}	5.9	3.8	4.5	5.9	5.9	5.9
p_{35}	-	-	3.3	-	-	-
p_{36}	4.6	4.6	4.6	3.4	4.6	4.6
p_{37}	1.9	3.2	3.2	3.2	3.2	3.2

Tabla 45. Matriz secundaria (MS). Tercera ronda

Procesos	Nodos					
	4	5	6	8	9	10
p_{17}	0.9	0.9	0.9	0.9	0.9	0.9
p_{32}	1.5	1.5	1.5	1.5	1.5	1.5

Tabla 46. Matriz de procesos no migrables. Tercera ronda

Procesos	Nodos							
	2	4	5	6	7	8	9	10
p_{12}	2.3	-	-	-	-	-	-	-
p_{13}	4.1	-	-	-	-	-	-	-
p_{34}	2.7	-	-	-	-	-	-	-

Se continúa con la matriz primaria hasta agotar opciones de procesos mientras que los nodos no alcancen el porcentaje de migración establecido.

El tercer proceso de la matriz primaria de la tercera ronda con menor peso de transferencia es el proceso p_{37} , que corresponde al nodo 4.

Lo anterior define la tercera ronda de asignación (Ver Tabla 47), se continúa la evaluación mientras existan procesos a evaluar.

Tabla 47. Función de Asignación para Migración en Sistemas Distribuidos (FAMSD) primaria. Tercera ronda

Peso de transferencia del proceso	Proceso seleccionado	Nodo destino
0.6	p_{11}	2
0.7	p_{31}	7
1.9	p_{37}	4

El siguiente paso es eliminar el proceso p_{37} y el nodo 4 de la matriz primaria porque fueron asignados en la tercera ronda. Aquellos procesos que tenían recursos destino en el nodo 4 (ya eliminado) pero tienen como opciones otros nodos destinos (que no poseen recursos destino de ese proceso) pasan a formar parte de la matriz secundaria (Ver Tabla 48 y Tabla 49). Los demás procesos cuya única opción es el nodo 4 se descartan y pasan a formar parte de la matriz de procesos no migrables (Ver Tabla 50).

Tabla 48. Matriz primaria (MP). Cuarta ronda

Procesos	Nodos				
	5	6	8	9	10
p_{14}	-	-	-	-	-
p_{15}	3.8	4.5	5.9	5.9	5.9
p_{35}	-	3.3	-	-	-
p_{36}	4.6	4.6	3.4	4.6	4.6

Tabla 49. Matriz secundaria (MS). Cuarta ronda

Procesos	Nodos				
	5	6	8	9	10
p_{17}	0.9	0.9	0.9	0.9	0.9

p_{32}	1.5	1.5	1.5	1.5	1.5
----------	-----	-----	-----	-----	-----

Tabla 50. Matriz de procesos no migrables. Cuarta ronda

Procesos	Nodos							
	2	4	5	6	7	8	9	10
p_{12}	2.3	-	-	-	-	-	-	-
p_{13}	4.1	-	-	-	-	-	-	-
p_{34}	2.7	-	-	-	-	-	-	-
p_{14}	4.1	3.4	-	-	-	-	-	-

Se continúa con la matriz primaria hasta agotar opciones de procesos mientras que los nodos no alcancen el porcentaje de migración establecido.

El cuarto proceso de la matriz primaria de la cuarta ronda con menor peso de transferencia es el proceso p_{35} , que corresponde al nodo 6.

Lo anterior define la tercera ronda de asignación (Ver Tabla 51); se continua la evaluación mientras existan procesos a evaluar.

Tabla 51. Función de Asignación para Migración en Sistemas Distribuidos (FAMSD) primaria. Cuarta ronda

Peso de transferencia del proceso	Proceso seleccionado	Nodo destino
0.6	p_{11}	2
0.7	p_{31}	7
1.9	p_{37}	4
3.3	p_{35}	6

El siguiente paso es eliminar el proceso p_{35} y el nodo 6 de la matriz primaria porque fueron asignados en la cuarta ronda. Aquellos procesos que tenían recursos destino en el nodo 6 (ya eliminado) pero tienen como opciones otros nodos destinos (que no poseen recursos destino de ese proceso) pasan a formar parte de la matriz secundaria (Ver Tabla 52 y Tabla 53). Los demás procesos cuya única opción es el nodo 6 se descartan y pasan a formar parte de la matriz de procesos no migrados (Ver Tabla 54).

Teniendo en cuenta que el nodo 3 ya alcanzó el porcentaje máximo de migración de procesos, sólo se evaluarán los procesos del nodo 1. Los procesos del nodo 3 que no fueron asignados en la matriz primaria y secundaria, se almacenan en una matriz de procesos no migrables (Ver Tabla 54).

Tabla 52. Matriz primaria (MP). Quinta ronda

Procesos	Nodos			
	5	8	9	10
p_{15}	3.8	5.9	5.9	5.9

Tabla 53. Matriz secundaria (MS). Quinta ronda

Procesos	Nodos			
	5	8	9	10
p_{17}	0.9	0.9	0.9	0.9

Tabla 54. Matriz de procesos no migrables. Quinta ronda

Procesos	Nodos							
	2	4	5	6	7	8	9	10
p_{12}	2.3	-	-	-	-	-	-	-
p_{13}	4.1	-	-	-	-	-	-	-
p_{34}	2.7	-	-	-	-	-	-	-
p_{14}	4.1	3.4	-	-	-	-	-	-

p_{32}	0.9	1.5	1.5	1.5	1.5	1.5	1.5	1.5
p_{36}	3	4.6	4.6	4.6	4.6	3.4	4.6	4.6

Se continúa con la matriz primaria hasta agotar opciones de procesos mientras que los nodos no alcancen el porcentaje de migración establecido.

El quinto proceso de la matriz primaria de la quinta ronda con menor peso de transferencia es el proceso p_{15} , que corresponde al nodo 5.

Lo anterior define la tercera ronda de asignación (Ver Tabla 55); se continua la evaluación mientras existan procesos a evaluar.

Tabla 55. Función de Asignación para Migración en Sistemas Distribuidos (*FAMSD*) primaria. Quinta ronda

Peso de transferencia del proceso	Proceso seleccionado	Nodo destino
0.6	p_{11}	2
0.7	p_{31}	7
1.9	p_{37}	4
3.3	p_{35}	6
3.8	p_{15}	5

El siguiente paso es eliminar el proceso p_{15} y el nodo 5 de la matriz primaria porque fueron asignados en la quinta ronda. Aquellos procesos que tenían recursos destino en el nodo 5 (ya eliminado) pero tienen como opciones otros nodos destinos (que no poseen recursos destino de ese proceso) pasan a formar parte de la matriz secundaria (Ver Tabla 56). Los demás procesos cuya única opción es el nodo 5 se descartan y pasan a formar parte de la matriz de procesos no migrables (Ver Tabla 57).

Tabla 56. Matriz secundaria (*MS*). Sexta ronda

Procesos	Nodo 8	Nodo 9	Nodo 10
p_{17}	0.9	0.9	0.9

Tabla 57. Matriz de procesos no migrables. Sexta ronda

Procesos	Nodo 2	Nodo 4	Nodo 5	Nodo 6	Nodo 7	Nodo 8	Nodo 9	Nodo 10
p_{12}	2.3	-	-	-	-	-	-	-
p_{13}	4.1	-	-	-	-	-	-	-
p_{34}	2.7	-	-	-	-	-	-	-
p_{14}	4.1	3.4	-	-	-	-	-	-
p_{32}	0.9	1.5	1.5	1.5	1.5	1.5	1.5	1.5
p_{36}	3	4.6	4.6	4.6	4.6	3.4	4.6	4.6

Teniendo en cuenta que en la matriz primaria se agotaron las opciones de migración de procesos, sólo se evaluarán las opciones de la matriz secundaria.

El proceso de la matriz secundaria de la sexta ronda con menor peso de transferencia es el proceso p_{17} , que corresponde al nodo 8.

Lo anterior define la sexta ronda de asignación (Ver Tabla 58); se continua la evaluación mientras existan procesos a evaluar.

Tabla 58. Función de Asignación para Migración en Sistemas Distribuidos (*FAMSD*) secundaria. Sexta ronda

Peso de transferencia del proceso	Proceso seleccionado	Nodo destino
0.9	p_{17}	8

Al finalizar la evaluación de los nodos 3 y 1 y habiéndose alcanzado el porcentaje máximo de migración de procesos establecido, se da por finalizado el presente macrociclo de evaluación.

Seguidamente se procede a la concatenación de la Función de Asignación para Migración en Sistemas Distribuidos (*FAMSD*) primaria y la Función de Asignación para Migración en Sistemas Distribuidos (*FAMSD*) secundaria. Se deben ordenar los procesos conforme a sus pesos de transferencia, de menor a mayor, para favorecer en la migración a los procesos que generan menor impacto. Esto da lugar a la Función de Asignación para Migración en Sistemas Distribuidos Concatenada (*FAMSDC*) que se muestra en la Tabla 59.

Tabla 59. Función de Asignación para Migración en Sistemas Distribuidos concatenada (*FAMSDC*)

Peso de transferencia del proceso	Proceso seleccionado	Nodo destino
0.6	p_{11}	2
0.7	p_{31}	7
0.9	p_{17}	8
1.9	p_{37}	4
3.3	p_{35}	6
3.8	p_{15}	5

Evaluación del estado de las cargas computacionales una vez migrados los procesos

Al promedio de carga computacional que se obtiene mediante el operador de agregación desarrollado en (La Red Martínez, 2017) se le agrega un criterio adicional que es la sobrecarga, se obtiene de la suma de los promedios de sobrecargas de cada proceso del nodo.

En la Tabla 60 se muestra la evaluación de los promedios de sobrecarga del nodo 1; el promedio A hace referencia al promedio de sobrecarga de los procesos que no se migran, el promedio B hace referencia al promedio de sobrecarga de los procesos migrables a otro nodo, este último promedio se evalúa por cada nodo destino.

Tabla 60. Evaluación de promedios de sobrecarga del nodo 1

Procesos	Nodos destinos	Promedios de sobrecarga
Promedio A		17.60
Promedio B p_{11}	2	2.97
Promedio B p_{15}	5	5.53
Promedio B p_{17}	8	2.07
Promedio A+		28.17
Promedio B		

En la Tabla 61 se muestra la evaluación de los promedios de sobrecarga del nodo 2, el promedio A hace referencia al promedio de sobrecarga de los procesos locales, el promedio C hace referencia al promedio de sobrecarga de los procesos inmigrantes. El proceso p_{11} pasa a ser proceso p_{26} en su nodo receptor.

Tabla 61. Evaluación de promedios de sobrecarga del nodo 2

Procesos	Nodos origen	Promedios de sobrecarga
Promedio A		11.70
Promedio C $p_{11} \rightarrow p_{26}$	1	2.97
Promedio A+		14.67
Promedio C		

En la Tabla 62 se muestra la evaluación de los promedios de sobrecarga del nodo 3, el promedio A hace referencia al promedio de sobrecarga de los procesos que no se migran, el promedio B hace referencia al promedio de sobrecarga de los procesos migrables a otro nodo, este último promedio se evalúa por cada nodo destino.

Tabla 62. Evaluación de promedios de sobrecarga del nodo 3

	Procesos	Nodos destinos	Promedios de sobrecarga
Promedio A			14.50
Promedio B	p_{31}	7	2.10
Promedio B	p_{35}	6	3.73
Promedio B	p_{37}	4	3.30
Promedio A+			23.63
Promedio B			

En la Tabla 63 se muestra la evaluación de los promedios de sobrecarga del nodo 4, el promedio C hace referencia al promedio de sobrecarga de los procesos inmigrantes. El proceso p_{37} pasa a ser proceso p_{41} en su nodo receptor.

Tabla 63. Evaluación de promedios de sobrecarga del nodo 4

	Procesos	Nodos origen	Promedios de sobrecarga
Promedio C	$p_{37} \rightarrow p_{41}$	3	3.30

En la Tabla 64 se muestra la evaluación de los promedios de sobrecarga del nodo 5, el promedio C hace referencia al promedio de sobrecarga de los procesos inmigrantes. El proceso p_{15} pasa a ser proceso p_{51} en su nodo receptor.

Tabla 64. Evaluación de promedios de sobrecarga del nodo 5

	Procesos	Nodos origen	Promedios de sobrecarga
Promedio C	$p_{15} \rightarrow p_{51}$	1	5.53

En la Tabla 65 se muestra la evaluación de los promedios de sobrecarga del nodo 6, el promedio C hace referencia al promedio de sobrecarga de los procesos inmigrantes. El proceso p_{35} pasa a ser proceso p_{61} en su nodo receptor.

Tabla 65. Evaluación de promedios de sobrecarga del nodo 6

	Procesos	Nodos origen	Promedios de sobrecarga
Promedio C	$p_{35} \rightarrow p_{61}$	3	3.73

En la Tabla 66 se muestra la evaluación de los promedios de sobrecarga del nodo 7, el promedio C hace referencia al promedio de sobrecarga de los procesos inmigrantes. El proceso p_{31} pasa a ser proceso p_{71} en su nodo receptor.

Tabla 66. Evaluación de promedios de sobrecarga del nodo 7

	Procesos	Nodos origen	Promedios de sobrecarga
Promedio C	$p_{31} \rightarrow p_{71}$	3	2.10

En la Tabla 67 se muestra la evaluación de los promedios de sobrecarga del nodo 8, el promedio C hace referencia al promedio de sobrecarga de los procesos inmigrantes. El proceso p_{17} pasa a ser proceso p_{81} en su nodo receptor.

Tabla 67. Evaluación de promedios de sobrecarga del nodo 8

	Procesos	Nodos origen	Promedios de sobrecarga
Promedio C	$p_{17} \rightarrow p_{81}$	1	2.07

Los nodos 9 y 10 no se evalúan porque no emigraron ni inmigraron procesos, se quedan con sus promedios de carga iniciales, calculados en la Tabla 17.

Para realizar la comparación de los nuevos valores de cargas computacionales se consideran dos etapas, la primera es aplicar la sobrecarga sin realizar la migración de procesos (ver Tabla 68) y la segunda es aplicar la sobrecarga una vez migrados los procesos (ver Tabla 69).

Tabla 68. Carga computacional sin migración

Nodo	%CPU	%Mem.	%E/S	Prom.	Prom. de Sobrecar.	Prom. Final
1	80	90	75	81.67	28.17	109.83
2	30	30	50	36.67	11.70	48.37
3	77	49	85	70.33	23.63	93.97
4	6	7	9	7.33	0	7.33
5	5	5	6	5.33	0	5.33
6	7	9	7	7.67	0	7.67
7	7	10	15	10.67	0	10.67
8	9	6	7	7.33	0	7.33
9	10	9	15	11.33	0	11.33
10	8	7	8	7.67	0	7.67

Nota: Los nodos del 4 a 10 no tienen procesos activos, por lo tanto, tienen valor de sobrecarga 0.

Tabla 69. Carga computacional con migración

Nodo	%CPU	%Mem.	%E/S	Prom.	Prom. de Sobrecar.	Prom. Final
1	80	90	75	81.67	17.60	99.27
2	30	30	50	36.67	14.67	51.33
3	77	49	85	70.33	14.50	84.83
4	6	7	9	7.33	3.30	10.63
5	5	5	6	5.33	5.53	10.87
6	7	9	7	7.67	3.73	11.40
7	7	10	15	10.67	2.10	12.77
8	9	6	7	7.33	2.07	9.40
9	10	9	15	11.33	0	11.33
10	8	7	8	7.67	0	7.67

Nota: Los nodos 9 y 10 no tienen procesos activos, por lo tanto, su valor de sobrecarga es 0.

Métrica propuesta para evaluar el balanceo de carga en el sistema distribuido

Para evaluar el impacto de la migración de procesos y cómo ésta afecta el estado global del sistema, es necesario aplicar algún mecanismo de medición de los estados de los nodos

antes y después de la migración. Para ello, se utiliza la desviación estándar, considerando la carga final de los nodos del sistema distribuido.

Se utilizarán los siguientes criterios asociados a cada proceso migrado: sobrecarga de procesador, sobrecarga de memoria y sobrecarga de entrada / salida, que se mostraron en la Tabla 22.

La desviación estándar del nivel de carga de los nodos sin considerar migración aplicando la fórmula (11) es la siguiente:

Se tiene en cuenta la Tabla 68: 109.83; 48.37; 93.97; 7.33; 5.33; 7.67; 10.67; 7.33; 11.33 y 7.67, cuyo promedio es de 30.95.

$$s = \sqrt{\frac{((109.83-30.95)^2 + (48.37-30.95)^2 + (93.97-30.95)^2 + (7.33-30.95)^2 + (5.33-30.95)^2 + (7.67-30.95)^2 + (10.67-30.95)^2 + (7.33-30.95)^2 + (11.33-30.95)^2 + (7.67-30.95)^2)}{(10-1)}}$$

s = 39.65

La desviación estándar del nivel de carga de los nodos con migración aplicando la fórmula (11) es la siguiente.

Se tiene en cuenta la Tabla 69: 99.27; 51.33; 84.83; 10.63; 10.87; 11.40; 12.77; 9.40; 11.33 y 7.67, cuyo promedio es de 30.95.

$$s = \sqrt{\frac{((99.27-30.95)^2 + (51.33-30.95)^2 + (84.83-30.95)^2 + (10.63-30.95)^2 + (10.87-30.95)^2 + (11.40-30.95)^2 + (12.77-30.95)^2 + (9.40-30.95)^2 + (11.33-30.95)^2 + (7.67-30.95)^2)}{(10-1)}}$$

s = 34.81

A partir de los resultados obtenidos se observa que la desviación estándar del conjunto de promedios finales sin migración es de 39.65 y la desviación estándar del conjunto de promedios finales con migración es de 34.81. Esto indica que se ha logrado balancear la carga porque los valores del conjunto de promedios finales con migración están más concentrados alrededor de la media, mientras que los valores de promedios finales sin migración están más dispersos.

Comentarios finales

El artículo propone un modelo de migración inteligente de procesos para la gestión de recursos y procesos en sistemas distribuidos, con el objetivo de resolver los inconvenientes relacionados con la sobrecarga que pueden presentarse en los distintos nodos de un sistema distribuido. Se presenta una variante de un enfoque innovador para la gestión de recursos compartidos en sistemas distribuidos que considera el balanceo de carga de trabajo entre los distintos nodos.

La propuesta se basa en la evaluación del estado de las cargas computacionales, lo que permite detectar nodos muy cargados con respecto a nodos inactivos o con baja carga de trabajo y corregir la distribución de carga de trabajo en el sistema.

La solución propuesta se basa en un *Runtime* centralizado que gestiona los procesos y recursos compartidos, y que interactúa con los *Runtimes* de los nodos distribuidos para intercambiar información (todos ellos interactúan con sus respectivos sistemas operativos). Además, se establecen configuraciones iniciales del algoritmo para cada situación en particular, lo que permite adecuar la solución planteada a distintos entornos.

Se recopila información de todos los nodos en un nodo central, se aplica el proceso de agregación y se obtiene la lista de asignaciones de recursos a procesos, y la carga adicional de trabajo que ello significaría sobre los distintos nodos distribuidos.

La evaluación del estado de las cargas computacionales se realiza de manera periódica, lo que permite considerar posibles migraciones para balancear la carga de trabajo de los distintos nodos del sistema.

Como métrica para evaluar la efectividad de la solución propuesta se utiliza la desviación estándar, que es una medida de dispersión muy útil y ampliamente utilizada en

el análisis de datos debido a su fácil interpretación, sensibilidad a todos los valores de los datos y propiedades matemáticas útiles.

Conclusiones y líneas futuras

La propuesta presentada es una solución innovadora y efectiva para la gestión de recursos compartidos en sistemas distribuidos, que permite resolver problemas relacionados con la sobrecarga y el balanceo de carga de trabajo. La evaluación del estado de las cargas computacionales y la comparación final utilizando la desviación estándar son herramientas útiles para detectar y corregir desequilibrios en el sistema. Además, la posibilidad de establecer configuraciones iniciales del algoritmo para cada situación en particular permite adecuar la solución a las necesidades específicas de cada sistema.

Si bien nuestro estudio ha proporcionado información valiosa sobre la migración controlada de procesos en un sistema distribuido de entorno homogéneo, aún existen algunas brechas en el conocimiento que deben abordarse en el futuro. Una de las áreas que requiere más investigación es la migración inteligente de procesos en un sistema distribuido heterogéneo.

Se planea trabajar en varias líneas de investigación y agregar nuevas funcionalidades al simulador disponible (La Red Martínez et al., 2020), con el objetivo de incorporar al mismo los nuevos modelos teóricos desarrollados.

Además, se considerarán los siguientes aspectos: definición e incorporación de escenarios complejos, adaptándolos a las investigaciones más recientes, propias y de otros grupos de investigación; planeamiento de capacidad para que el simulador indique las características ideales de los diferentes nodos del sistema distribuido en función de determinadas cargas de trabajo; incorporación de aspectos vinculados a la seguridad en la ejecución de los procesos, en el acceso a los recursos y en la comunicación entre los nodos; e incorporación del soporte para sistemas de tiempo real con requerimientos perentorios de finalización de tareas.

Se espera que la incorporación de estas nuevas funcionalidades, sumadas a las ya existentes, permita disponer de un simulador complejo, flexible y fácil de utilizar para la investigación y la docencia.

Agradecimientos

Este trabajo ha sido apoyado por el Proyecto: “Modelos de decisión para la gestión de recursos y procesos en sistemas distribuidos considerando la migración de procesos, la imputación de datos y la lógica difusa en nuevos operadores de agregación.”, código 20F005 de la Universidad Nacional del Noreste (Argentina), y el Proyecto: “Desarrollo de un simulador para la evaluación de algoritmos clásicos y nuevos para la gestión de recursos compartidos en sistemas distribuidos contemplando exclusión mutua.”, código PI 126/20 de la Universidad Nacional del Chaco Austral (Argentina).

Un sincero agradecimiento a la Academia Nacional de Ciencias de Buenos Aires y al Centro de Estudios en Tecnologías Inteligentes, por la posibilidad de publicar este trabajo.

Bibliografía

- Agostini, Federico, “*Nueva Propuesta para la Administración de Recursos y Procesos en Sistemas Distribuidos*”, Universidad Nacional del Nordeste, 2019.
- Agostini, Federico & La Red Martínez, David Luis, “Allocation of Shared Resources”, *14th Iberian Conference on Information Systems and Technologies - CISTI 2019* 1–6, 2019.
- Agostini, Federico, La Red Martínez, David Luis & Acosta, Julio César, “Assignment of Resources in Distributed Systems with Strict Consensus Requirements”, *IMCIC 2019 - 10th International Multi-Conference on Complexity, Informatics and Cybernetics, Proceedings*, 1, 2019.

- Agostini, Federico, La Red Martínez, David Luis & Acosta, Julio, “Nueva Propuesta Para La Administración de Recursos y Procesos En Sistemas Distribuidos.” *Workshop de Investigadores en Ciencias de la Computación*, 2018.
- Asadi, Ali Naghash, Azgomi, Mohammad & Entezari-Maleki, Reza, “Analytical Evaluation of Resource Allocation Algorithms and Process Migration Methods in Virtualized Systems”, *Sustainable Computing: Informatics and Systems*, 25, 2020, p 100370.
- Beirut, Mohanmmad Amin. & Ganjali, Yashar, “Load Migration in Distributed SDN Controllers”, *NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium*, 2020, 1-9.
- Cao, Guohong. & Singhal, Mukesh, “A Delay-Optimal Quorum-Based Mutual Exclusion Algorithm for Distributed Systems”, *IEEE Transactions on Parallel and Distributed Systems*, 12, 2001, pp 1256 - 1268.
- Chang, Chii, Hadachi, Amir & Srirama, Satish Narayama., “Adaptive Edge Process Migration for IoT in Heterogeneous Fog and Edge Computing Environments”, *International Journal of Mobile Computing and Multimedia Communications*, 11, 3, 2020, pp 21.
- Chiclana, Francisco, Herrera, Francisco & Herrera-Viedma, Enrique, “Integrating Multiplicative Preference Relations in a Multipurpose Decision-Making Model Based on Fuzzy Preference Relations”, *Fuzzy Sets and Systems* 122, 2, 2001, pp 277-291.
- Deshmukh, Shyam & Deshmukh, Sudarshan, “Improved Load Balancing for Distributed File System Using Self Acting and Adaptive Loading Data Migration Process”, *4th International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions*, 2015.
- Dong, Yucheng, Zhang, Hengjie. & Herrera-Viedma, Enrique, “Consensus Reaching Model in the Complex and Dynamic MAGDM Problem.” *Knowledge-Based Systems*, 106, 2016, pp. 206-219.
- Fornerón Martínez, Jorge Tomás, Agostini, Federico, La Red Martínez, David Luis, “Resource and Process Management With a Decision Model Based on Fuzzy Logic”. *International Journal of Interactive Multimedia and Artificial Intelligence - IJIMAI*, (2023), In Press.
- Junaidi, Junaidi, Wibowo, Prasetyo, Yuniasri, Dini, Damayanti, Putri, Shiddiqi, Ary Mazharuddin & Pratomo, Baskoro Adi, “Applied Machine Learning in Load Balancing” *JUTI: Jurnal Ilmiah Teknologi Informasi*, 18, 2, 2020, pp. 76
- La Red Martínez, David Luis. “Aggregation Operator for Assignment of Resources in Distributed Systems”, *International Journal of Advanced Computer Science and Applications* 8, 10, 2017.
- La Red Martínez, David Luis. “Nuevos modelos de decisión y operadores de agregación - gestión de recursos y procesos”; capítulo de libro; en *Selección de escritos sobre inteligencia artificial - Inteligencia Artificial: Algunos Aspectos de su Impacto*; Compilador Juan Carlos Ferreri - Centro de Estudios en Tecnologías Inteligentes (CETI), Academia Nacional de Ciencias de Buenos Aires (ANCBA), Argentina, ISBN 978-987-537-171-2, 2022, pp. 156-200.
- La Red Martínez, David Luis, Acosta, Julio César. & Agostini, Federico, “Assignment of Resources in Distributed Systems”, *IMCIC 2018 - 9th International Multi-Conference on Complexity, Informatics and Cybernetics, Proceedings, International Institute of Informatics and Systemics*, IIIS, 2, 2018, pp. 122-127.
- La Red Martínez, David Luis, Agostini, Federico & Primorac, Carlos Roberto, “Modelo de Asignación de Recursos Para La Enseñanza de Los Procesos Distribuidos”, *Primer Congreso Latinoamericano de Ingeniería - CLADI*, 2017.
- La Red Martínez, David Luis, Agostini, Federico, Acosta, Julio César, Gerzel, Stella & Latyn, Leandro, “Simulador para la evaluación de algoritmos para la gestión de recursos compartidos en sistemas distribuidos”, *Revista de Investigación en Tecnologías de la Información: RITI*, 10, 20, 2022, pp. 62-79.
- Liu, Haikum, Xu, Cheng-Zhong, Jin, Hai, Jiayu Gong & Liao, Xiaofei, “Performance and Energy Modeling for Live Migration of Virtual Machines”, *Cluster Computing* 16, 2, 2013, pp. 249-64.

- Lodha, Sandeep. & Kshemkalyani, Ajay, “A Fair Distributed Mutual Exclusion Algorithm”, *IEEE Transactions on Parallel and Distributed Systems*, 11, 6, 2000, pp. 537-549.
- Marecos, Terecio Diosnel, Agostini, Federico, La Red Martínez, David, “Migración controlada de procesos en sistemas distribuidos”, *Memorias del Encuentro Argentino de Ingeniería, 6º Congreso Argentino de Ingeniería y 12º Congreso Argentino de Enseñanza de Ingeniería*, 2022, pp. 173-180.
- Rathore, Neeraj & Chana, Inderveer, “Load Balancing and Job Migration Techniques in Grid: A Survey of Recent Trends.” *Wireless Personal Communications*, 79, 3, 2014, pp. 2089-2125.
- Ricart, Glenn & Agrawala, Ashok K., “An Optimal Algorithm for Mutual Exclusion in Computer Networks”, *Communications of the ACM* 24, 1, 1981, pp. 9-17.
- Sohrabi, Zeinab & Mousavi Khaneghah, Ehan, 2020. “Challenges of Using Live Process Migration in Distributed Exascale Systems.” *Azerbaijan Journal of High Performance Computing*, 3, 2, 2020, pp. 151–63.
- Upadhyay, Ankit & Lakkadwala, Prashant, “Migration of over Loaded Process and Schedule for Resource Utilization in Cloud Computing”, *4th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2015, pp. 1-4.

1. Introducción

La ética es la parte de la filosofía que estudia el bien y el mal y sus relaciones con la moral y el comportamiento humano. La ética trata de los principios generales que rigen tales comportamientos mientras que la moral se ocupa de los sistemas reales que aplican los humanos para regular sus relaciones.

La inteligencia artificial puede definirse como aquella parte de la ciencia de la computación que crea algoritmos planteados con el propósito de crear máquinas que presenten capacidades parecidas al ser humano. Objetivo obviamente inalcanzable pero que en el camino muestra que en algunas cosas la IA puede ser más precisa y contener más datos de los que un humano puede elaborar.

El encuentro de ambas disciplinas apareció ineludible desde muy temprano, pues desde entonces se vio que era una tarea tanto fructífera cuanto peligrosa.

Siempre que aparece el peligro nos acordamos de la ética. La inteligencia artificial se ha introducido en toda nuestra vida, desde el uso de celulares hasta las armas autónomas, pasando por los vehículos autogobernados, la producción realizada por robots, el flujo incesante de datos como nunca en la historia de la humanidad.²

Como siempre cuando alguna creación humana es muy original y creativa, se torna peligrosa: en el siglo pasado la energía atómica e Internet, en este la I.A. Recién cuando se vuelve peligrosa se piensa en la ética que debe acompañar toda acción humana.

Voy a obviar los riesgos que algunos autores atribuyen a la IA, como ser que sus algoritmos rápidamente sobrepasaran (si no lo han hecho ya) a la inteligencia humana porque, en general son debidos a miedos.³ El robot provoca miedo y evoca miedos ancestrales. Desde el Golem de la tradición hebrea, que aparece citado una sola vez en el Antiguo Testamento, esto es, en salmo 139, verso 16, allí se dice “inconcluso o esbozado me vieron tus ojos. Tus ojos vieron mi embrión, y en tu libro se escribieron todos los días que me fueron dados, cuando no existía ni uno solo de ellos”. Ese embrión, según la interpretación, es el Golem, pero para otros Adán. En hebreo, idioma original del Antiguo Testamento, se coloca el origen de la figura del Golem, de donde la Qabbaláh (tradición) busca sus significados.

¹ Doctor en Derecho y Ciencias sociales por la Universidad de Buenos Aires, Profesor emérito de las universidades de Pisa, Italia y Salvador, Argentina, miembro de la Academia Nacional de Derecho y Ciencias Sociales de Córdoba.

² Es tan grande el uso de chips para todas las actividades que actualmente hay escases de los mismos y algunas empresas deben limitar o paralizar su producción.

³ Los seres humanos tenemos miedos ancestrales. Hay 6 que son universales: a la muerte, a la falta de autonomía, al hambre, a la soledad, a las pandemias, a la mutilación física o psíquica. También a la muerte.

Vale la pena ocuparse de los peligros serios que han aparecido ya y que están siendo tratados; podemos llamarlos riesgos y son debidos al hecho que se trata de una metodología nueva a la cual la sociedad no ha tenido tiempo de habituarse. Los sistemas inteligentes no van a volver atrás, por lo tanto, la solución no es demonizarlos como si no existieran o fuéramos a desterrarlos sino ver cuáles son los riesgos que crean, diferenciarlos pues algunos son menores y otros mayores, colocarlo en los justos contextos, estudiar las medidas para contrarrestarlos y tener un ojo vigilante a una sociedad siempre más compleja, siempre más cruzada por todas las líneas de las nuevas tecnologías.

Todo lo que es nuevo supone riesgo y en los últimos 18 años se han producido más novedades tecnológicas que en cualquier otro periodo de la historia.

Hay muchas teorías éticas, pero vamos a tratar solo dos contrapuestas: la teoría kantiana del imperativo categórico. Dice Emanuel Kant en *Fundamentación de la metafísica de las costumbres* “Obra sólo según aquella máxima por la cual puedas querer que al mismo tiempo se convierta en ley universal. Obra como si la máxima de tu acción pudiera convertirse por tu voluntad en una ley universal de la naturaleza” (AA IV: 421) y la de un autor inglés Charles Ross, quien en su libro *Fundamentos de ética*, dice “que tenemos un deber prima facie de ayudar a los demás, otro de mantener nuestras promesas, otro de devolver los actos de amabilidad anteriores y otro de no defraudar a las personas que confían en nosotros”. Y habla de deberes prima facie porque nuestros deberes cambian con el cambio de informaciones que tenemos⁴.

Esta, me parece, es la gran controversia ética que aun hoy perdura: la rigidez del imperativo categórico contra la laxitud de los deberes prima facie de Ross.

2 . Casos

2.1 El domingo 18 de marzo de 2018, a las 21.59 hs en Tempe, Arizona, Elaine Herzberg fue embestida y muerta por un auto autónomo de prueba, de la firma Uber, conducido por los sistemas desarrollados para ese fin, con una conductora de seguridad llamada Rafaela Vásquez.

El automóvil disponía al momento del accidente, de cámaras con visión de toda la periferia, múltiples sensores, GPS, ubicación en mapas electrónicos, radar para la determinación de la velocidad relativa de cada objeto en movimiento, del LIDAR (Light Detection and Ranging) que permite una visión completa 300m de distancia y a 360°, con imágenes tridimensionales con muy alta definición debido a la utilización de láser. Todo esto controlado y comandado por un sistema inteligente de conducción.

La tecnología recibía los distintos datos y los procesaba con sus algoritmos, calculando continuamente y elaborando predicciones de las distintas alternativas de ocurrencia en los siguientes pocos segundos, en base a los objetos, personas, señalización y reglas de tránsito vigentes. El auto se desplazaba a 69 Km/hora, el sistema de detección de imágenes percibe 6 segundos antes de la colisión un objeto que no identifica en el medio de la calzada. Elaine Herzberg vestía un abrigo negro y llevaba a pie su bicicleta cargada.

El sistema detecta, un objeto no reconocido, luego interpreta un auto y finalmente una bicicleta, todo esto en concordancia con nuevas y sucesivas imágenes. Los sistemas de

⁴ Ross, David, *Fundamentos de ética*, Eudeba, Buenos Aires, 2003. El original inglés es de 1930

identificación necesitan varios miles de muestras de imágenes de un determinado objeto para aprender a identificarlo.

El algoritmo encuentra un objeto que desconocía, porque no estaba comprendido dentro los que habían aprendido, Estamos entonces en presencia de lo que en inteligencia artificial se denomina un “falso positivo”, es decir no identifica (falso) un objeto persona (verdadero). Los sistemas de reconocimiento de imagen expresan el resultado como una probabilidad que sea un objeto determinado, por ejemplo 95% casi certeza o un dudoso 15%. Cuatro segundos después de la detección de un objeto no identificado, el sistema decide frenar para evitar el daño posible en el caso que la consistencia del objeto sea contundente, esto ocurre a 1,3 segundos de la colisión, la velocidad al momento del impacto fue de 63 km/hora.

Esta larga descripción relata la concientización de la aparición de los autos autónomos: tiene lugar cuando hay un trágico accidente. Es una constante en la vida humana: percibimos las cosas cuando nos procuran una gran alegría o un enorme miedo. Pero hay también una segunda consecuencia tal vez más importante de la humana condición: en esos momentos de conmoción y riesgo comenzamos a preguntarnos si es justo que algo así suceda, si está bien o está mal y aquí aparece la dimensión ética.

A partir de ese hecho comenzaron las preguntas que tal vez habría que haberse hecho antes del experimento sobre la eticidad de comportamiento y volvió al ruedo un viejo dilema que Filippa Foot había enunciado en los años sesenta del siglo pasado: el dilema del tranvía.

2.2 Reconocimiento facial

En los años 60 del siglo pasado Woodrow Bledsoe (1921-1995) y el dispositivo, la tableta RAND. Woodrow Wilson *Woody* Bledsoe, pionero de esta disciplina, fue además uno de los fundadores de la inteligencia artificial. Y contribuyó de forma importante al campo de reconocimiento de patrones, un aspecto determinante para la técnica que desarrolló en la década de los 60 y que entonces se llamó "reconocimiento facial de máquina".

El procedimiento consistía en clasificar una serie de retratos fotográficos de diferentes personas que previamente debía digitalizar. Aún sorprende la herramienta de la que se valió para realizarlo: la tableta RAND, un dispositivo electrónico bastante sofisticado para la época, abuelo de las actuales tabletas gráficas, que disponía de un lápiz que se comunicaba con una cuadrícula de 10 x10 pulgadas a través de impulsos electromagnéticos.

El programa de reconocimiento de escritura a mano que usaba la tableta se denominaba GRAIL (Graphical Input Language) y permitía dibujar formas y texto que el software suavizaba y renderizaba en un monitor en tiempo real. Lo más innovador era que incluso reconocía gestos y era posible eliminar, mover o cambiar el tamaño de los elementos. Bledsoe aprovechó estas funciones para marcar las características faciales mediante coordenadas, como la ubicación de la línea del cabello, los ojos o la nariz. Una vez digitalizados, estos registros numéricos se asociaban al nombre del individuo y se guardaban en una base de datos. El proceso inverso consistía en buscar en el fichero aquella cara que coincidiera en rasgos faciales, a partir de una imagen desconocida,



Tableta Rand en una demostración RAND.ORG

El sistema adolecía de las limitaciones técnicas de la época. El propio Bledsoe reconocía la incapacidad de las máquinas para distinguir incluso al mismo individuo en dos fotografías diferentes o desde distintos ángulos o iluminaciones. No obstante, supuso la primera piedra de un método que fue desarrollándose en los siguientes años gracias al progreso de los algoritmos y la capacidad de las computadoras.

El método tiene sus raíces en el siglo XIX: el oficial de policía francés Alphonse Bertillon creó un método para identificar criminales en función de sus características físicas. Había clasificado 11 tipologías y a partir de retratos fotográficos estandarizados, tomaba notas de sus rasgos por escrito.

Un siglo después, en 1994, el “padre fundador” de la identidad digital, Joseph Stick, planteó que las computadoras podían procesar información biométrica. No estaba equivocado. La forma que tiene el software de identificar una cara en concreto es a través de un conjunto de referencias o puntos concretos (suelen ser aproximadamente 68) cuya distancia entre ellos y configuración son distintos para cada persona, confiriendo *patrones únicos como una huella digital*.

También existen otros métodos más modernos de identificación como el análisis de textura superficial que mapea y cataloga la textura de la piel, como si cartografiase cada poro y cada arruga de la cara.

La historia de esta tecnología está atada a China, uno de los primeros países en utilizar el reconocimiento facial para patrullar las calles. Y cuando decimos que es sorprendente nos referimos a este tipo de datos: más de 300 millones de cámaras de seguridad detectan, en China, caras individuales entre multitudes.

El *reconocimiento facial* se ha convertido en los últimos años en un área de investigación activa que abarca diversas disciplinas, como procesado de imágenes, reconocimiento de patrones, visión por ordenador y redes neuronales. Involucra a investigadores de áreas de informática como a neurocientíficos y psicólogos. Se podría considerar también dentro del campo de reconocimiento de objetos, donde la cara es un objeto tridimensional sujeto a variaciones de iluminación, pose, etc., y ha de ser identificada basada en su proyección 2D (excepto cuando se utilizan técnicas 3D).

El objetivo de un sistema de reconocimiento facial es el siguiente: dada una imagen de una cara "desconocida", o imagen de test, encontrar una imagen de la misma cara en un conjunto de imágenes "conocidas", o imágenes de entrenamiento. La gran dificultad añadida es la de conseguir que este proceso se pueda realizar en tiempo real. El sistema identificará las caras presentes en imágenes o videos automáticamente. Puede operar en dos modos:

1. *Verificación o autenticación de caras:* compara una imagen de la cara con otra imagen con la cara de la que queremos saber la identidad. El sistema confirmará o rechazará la identidad de la cara. 2. *Identificación o reconocimiento de caras:* compara la imagen de una cara desconocida con todas las imágenes de caras conocidas que se encuentran en la base de datos para determinar su identidad.

Se utiliza principalmente en sistemas de seguridad para el reconocimiento de usuarios. En estos sistemas se utiliza un lector que define las características del rostro, y cuando este solicita el acceso, se verifica comparando los datos obtenidos con la base de datos. Sin embargo, estos sistemas no son útiles a largo plazo ya que, a medida que pasan los años, los rasgos faciales varían y al solicitar el acceso ya no coinciden con la imagen en la base de datos. Para solucionar este problema se puede utilizar un algoritmo que interprete el paso de los años, aunque igualmente sigue sin ser del todo fiable), o bien, renovar frecuentemente la base de datos.

También se utiliza en aplicaciones de interacción persona-ordenador, en gestión multimedia, y en software como Google's Picasa, Apple iPhoto, Sony's Picture Motion Browser (PMB), Facebook y Asus Smart Logon.

Una aplicación de reconocimiento facial futura se basa en establecer esta técnica a nivel de usuario. Por ejemplo, en un supermercado o en un establecimiento pequeño se podría llevar a cabo un control sobre quien abre la caja registradora mediante un reconocimiento facial previo, de esta manera también se pueden evitar intentos de robo ya que, al no reconocer el rostro, la caja permanecería cerrada. Un caso más extremo sería en los cajeros automáticos donde, para poder operar, fuese necesario un reconocimiento facial en vez del actual PIN.

Las aplicaciones que ya existen de reconocimiento facial se aplican en China para tener en todo tiempo un control absoluto sobre los ciudadanos, esto ¿está bien? ¿O es un gran reto para nuestras libertades?

Que haya una permanente visión de nuestros actos ¿es un atentado contra la privacidad a la cual, hemos aprendido, tenemos derecho?

Fuera del ámbito de la seguridad, ya vimos cómo la inteligencia artificial aplicada a este ámbito asomaba en nuestras redes sociales como algo inocente y original allá por el 2010, cuando Facebook lo incorporó para reconocer las caras de nuestros amigos en las fotografías que subíamos y las etiquetas que le proporcionábamos. Su uso se extendió rápidamente y hoy se encuentran en muchos de los smartphones y aplicaciones que utilizamos a diario.

Las polémicas sobre el uso de la biometría se han alzado principalmente al descubrirse que algunas empresas y organismos utilizaban la información recopilada para otro propósito diferente del autorizado. Este asunto ha estado en el centro del debate sobre ética y privacidad desde el 2001, propiciado a su vez por un vacío legal en la aplicación de las nuevas tecnologías de la información.

El Sistema de Reconocimiento Facial de la INTERPOL (IFRS) almacena imágenes faciales enviadas por más de 160 países, lo que la convierte en una base de datos única en el ámbito policial. Este sistema, puesto en marcha a finales de 2016 ha logrado identificar a más de 650 delincuentes, prófugos y desaparecidos. Pero su web ya advierte que la calidad de las imágenes es un aspecto esencial y que aquellas que solo posean una resolución media o baja pueden no conseguir o influir negativamente en la precisión de la búsqueda.

En los últimos meses, las controversias sobre su uso han aumentado, impulsadas en parte por el movimiento "Black Lives Matter"⁵, por lo que diferentes organismos han empezado a retroceder acusando excesivo control y posible promoción del racismo y la injusticia social. Amazon o IBM dieron un paso atrás. San Francisco ya se había convertido el pasado año en la primera gran ciudad de EE UU en prohibir a todas las agencias locales,

⁵Movimiento surgido en EEUU frente a abusos contra la comunicad afroamericana y que significa literalmente "la vida negra importa".

entre ellas la policía, el uso de técnicas de reconocimiento facial, por miedo a que se las utilice para discriminar.

En nuestro país hubo muchos casos de detenciones erróneas, como uno en Bahía Blanca en 2019. No es sólo en Argentina que hay organizaciones que se oponen a este sistema. A mediados del año pasado, por ejemplo, hubo un caso en Estados Unidos que despertó fuertes polémicas. Un afroamericano había sido detenido por un error en un programa de reconocimiento facial, y esto reavivó la preocupación sobre los riesgos de sesgo en la inteligencia artificial, en medio de una ola de protestas contra el racismo y la violencia policial.

Según la Unión Americana de Libertades Civiles (ACLU), que presentó una denuncia en su nombre el 24 de junio de 2020, "aunque es el primer caso conocido, probablemente no sea la primera persona que ha sido detenida e interrogada erróneamente sobre la base de un problema de reconocimiento facial".

Clara Joy Buolamwini, fundadora del grupo activista Algorithmic Justice League, el caso está revelando "cómo el racismo sistémico puede codificarse y reflejarse en las inteligencias artificiales (IA)".

En ciudades como San Francisco se prohibieron estas técnicas de reconocimiento facial para identificar a criminales. Alemania lo vetó para su policía y en Francia los colegios no pueden usar este sistema. En las antípodas de estas medidas están países como China e Inglaterra. En la Policía Metropolitana londinense implementó desde el año pasado la tecnología de reconocimiento facial en las calles. El debate está abierto porque los derechos civiles pueden verse fuertemente vulnerados.

2.3. Armas autónomas

La tarde del 2 de enero de 2020 dos misiles alcanzaron en los exteriores del aeropuerto de Bagdad a un vehículo en el que viajaban el comandante de una milicia iraquí Abu Mehdi al Muhandis y Quasem Soleimani, líder de la Fuerza Quds, unidad élite de la Guardia Revolucionaria Islámica. Fuentes oficiales de Washington confirmaron horas más tarde la autoría del ataque (Guimón, 2020, 3 de enero).

Este golpe militar fue significativo no solamente porque se produjo en un momento de máxima tensión política entre Irán y Estados Unidos, sino también por su modus operandi. Primero, el ataque fue ejecutado con un dron MQ-9 Reaper, un vehículo aéreo pilotado a control remoto desde bases aéreas que se encuentran normalmente lejos de la zona de operación. Segundo, la zona de la operación, el aeropuerto de Bagdad se encuentra en un país, con el que Estados Unidos oficialmente no está en guerra. Tercero, las víctimas del ataque fueron líderes militares a los que Estados Unidos consideran enemigos que, según las mismas fuentes, estaban preparando ataques contra el personal diplomático y militar estadounidense (IBID.). Se trataba, por tanto, de un ataque selectivo.

No es este el origen de las armas autónomas, sino que es uno más de una larga serie que Estados Unidos está ejecutando desde la primera década del nuevo milenio en el exterior, principalmente en países de Oriente Medio y forma parte de la doctrina de guerra preventiva, cuya artífice es la anterior Secretaria de Estados de Estados Unidos, Condoleezza Rice.

Desde la publicación del informe del Christof Heyns, Relator Especial sobre ejecuciones extrajudiciales, sumarias o arbitrarias explica en su informe, del 9 de abril de 2013, se han multiplicado las voces que advierten de los peligros que podrían manifestarse en el futuro, cuando los drones adquieren tanta autonomía que podrían elegir los objetivos y ejecutar los ataques sin la intervención del hombre. El Secretario General del Consejo de

Seguridad Naciones Unidas, advierte en su informe Protección de civiles en conflictos armados de la necesidad de que se regule sistemas de armas autónomas letales (SAALs en adelante) en cumplimiento con el Derecho Internacional Humanitario.

En la actualidad, ya son nueve países que están desarrollando armas autónomas letales. Siete de ellos son: Estados Unidos, China, Rusia, Reino Unido, Francia, Israel y Corea del Sur⁶A estos hay que añadir Irán, cuyos drones se han encontrado en los ataques a las instalaciones de Saudi Aramco, y Turquía que está usando sus drones en Libia. Este dato revela la importancia del área y no ha pasado de forma desapercibida en el mundo académico y la opinión pública.

Las preguntas que se plantean son ¿quién es el responsable en el uso de armas autónomas letales y de las bajas por un ataque? qué ocurre cuando un arma actúa de forma tan autónoma que para un comandante ya no es posible frenar el ataque? ¿Cómo afectarían armas autónomas en los ámbitos del Derecho Internacional Humanitario y el Derecho de Conflictos Armados?

Todos los sistemas manifiestan cierta probabilidad de fallar, y si no existe ningún control - sea otro sistema o el hombre – las consecuencias pueden ser dramáticas. Uno de los ejemplos más notorios es el del Teniente Coronel Stanislav Petrov quien evitó una guerra nuclear el 26 de septiembre de 1983, cuando confió más en su intuición en lugar del sistema alerta temprana del que era responsable. El nuevo satélite Oko había interpretado los reflejos de la luz solar en las nubes como el lanzamiento de misiles intercontinentales. Petrov decidió contrastar la información del sistema con la vigilancia aérea y sus superiores en lugar de iniciar el contraataque que exigía el protocolo de acción.

¿Qué hubiera ocurrido si un sistema automatizado hubiera actuado sin la intervención humana? No es probable que en un escenario de estas envergaduras el hombre esté fuera del bucle, pero ¿quién puede garantizar que esto no cambie cuando un conflicto se escale? ¿Quién puede garantizar que un país no despliegue estas armas en el modo de autonomía supervisada para responder con más rapidez a un ataque masivo de enemigo? Críticos de las armas autónomas alegan también una posible mayor predisposición para la entrada en conflicto, debido al menor riesgo para las personas.

La razón que alegan muchos críticos es la ingente cantidad de ejecuciones extrajudiciales que han estado ejecutando las fuerzas de seguridad estadounidenses durante los últimos años en Oriente Medio. El informe de Heyns, antes citado, estaba motivado precisamente por este tipo de ataques selectivos.

Estados Unidos llevó a cabo en diez años más de 300 ataques aéreos contra combatientes talibanes y cuadros de Al Qaeda en zonas fronterizas Pero estos ataques no fueron operaciones militares stricto sensu. Al contrario, se trataba de una operación encubierta de la Agencia Central de Inteligencia. Según un recuento de la New America Foundation, entre 2004 y marzo de 2013, los ataques con drones acabaron con la vida de entre 1535 y 2660 militantes de talibanes⁷. La misma fuente señala que solamente el dos por ciento de las víctimas fueron cuadros de Al Qaeda o de otros grupos terroristas, mientras más de 80 por ciento eran militares rasos o colaboradores que apoyaban la yihad de forma no combativa (IBID.).

En febrero de 2020, después de 15 meses de consultas con expertos de la industria, la academia, el gobierno y el público americano, el Departamento de Defensa adoptó cinco

⁶ *Report: Kill the idea of 'killer robots' before they kill us*, Center for a New American Security, ápod Aznar, 2020, p. 26.

⁷ New America Foundation, ápod Montoya, 2014

principios éticos para la inteligencia artificial (DoD, 2020). Estos principios forman parte de la Iniciativa Americana de Inteligencia Artificial que trata de promover el uso innovador de la IA, mientras protege las libertades civiles, la privacidad y los valores.

El análisis de la estrategia nacional de defensa se basa en el resumen ejecutivo, porque el documento entero no está disponible en la página del departamento de defensa americanos.

⁸

Los principales marcos regulatorios bilaterales y multilaterales para la regulación de armas se crearon cuando las armas aún no disponían de la autonomía actual. Su objeto de regulación eran armas químicas (Convención sobre Armas Químicas, de 1993), armas biológicas (Convención sobre las Armas Biológicas de 1975), armas nucleares (Tratado de No Proliferación Nuclear de 1968 y el Tratado sobre Fuerzas Nucleares de Rango Intermedio de 1987) y armas en el espacio (Tratado sobre el Espacio Ultraterrestre de 1967). De fechas más recientes son la Convención sobre la prohibición de minas antipersonales, de 1997 y el Tratado de comercio de armas de 2014. Ninguno de estos tratados regula sistemas de armas autónomas. Una excepción sería el Régimen de Control de Tecnología de Misiles, de 1987, que regula al menos la exportación de vehículos aéreos no tripulados.

Un sistema de armas, por muy sofisticado y autónomo que sea, no puede ser un agente moral porque no tiene ningún vínculo emocional con los hombres. Es por este motivo por el que la mayoría de los expertos rechazan la idea que un sistema deba tener la opción de decidir sobre vida o muerte. Las implicaciones jurídicas se refieren al respeto de algunos principios claves en el Derecho Internacional Humanitario, en concreto los principios de distinción, proporcionalidad y precaución existen varios niveles donde se toman decisiones. Especialmente en la planificación de un ataque existe la posibilidad de respetar estos principios. Un sistema no funcionará nunca con plena seguridad. Esta es la principal razón por la que el hombre siempre debería mantener el control sobre el sistema, reivindicación que repiten numerosos críticos de los sistemas de armas autónomas. Los dos incidentes descritos en este capítulo muestran el peligro cuando un sistema malinterpreta una información e induce a activar un protocolo de contraataque.

El art. 36 del Protocolo adicional a los acuerdos de Ginebra obligan a los estados a revisar si el uso de armas que producen o adquieren respetan las reglas del derecho internacional. Se empieza a considerar que las decisiones de herir, matar o traumatizar seres humanos no debe ser aprendido por sistemas autónomos bélicos. Se va afianzando la idea que en la ciberguerra también el control humano significativo sea la piedra angular del todo el desarrollo. Esta es una tarea necesaria e inminente.

⁸ Según este documento, la inteligencia artificial debe cumplir los siguientes Principios: 1. Responsable: El personal del Departamento de Defensa será responsable del desarrollo, el empleo y el uso de las capacidades de la IA, manteniendo niveles apropiados de juicio y precaución. 2. Equitativo: El departamento procurará minimizar sesgos no intencionados en las capacidades de IA. 3. Trazable: El personal involucrado deberá disponer el entendimiento apropiado de la tecnología y de los métodos operacionales que se pueden aplicar en las capacidades de la IA. Estas incluyen metodologías transparentes y auditables, las fuentes de datos, el diseño de los procesos y de la documentación. 4. Fiable: Las capacidades de la IA tendrán usos bien definidos y explícitos, y la seguridad y eficiencia de estas capacidades. Estas serán objeto de pruebas de seguridad durante todo su ciclo de vida. 5. Gobernable: Con este principio, el departamento se reserva el derecho de desconectar o desactivar sistemas desplegados que muestran un comportamiento no intencionado.

Los métodos de reconocimiento facial dedicados a vigilancia masiva tuvieron que esperar hasta fines del siglo XX. La primera vez que se utilizó en un acontecimiento de grandes proporciones fue en la Super Bowl de 2002, que resultó, en general, un fracaso. Numerosos falsos positivos mostraron que la tecnología no estaba aún preparada para las grandes multitudes. A partir del 2015, las fuerzas policiales del Reino Unido empezaron a probarla también. El Sistema de Reconocimiento Facial de la Interpol (IFRS) almacena imágenes faciales enviadas por más de 160 países, lo que la convierte en una base de datos única en el ámbito policial. Este sistema, puesto en marcha a finales de 2016 ha logrado identificar a más de 650 delincuentes, prófugos y desaparecidos.

Pero también es utilizada en países autocráticos o con tendencias autoritarias a vigilar a la población y saltar sus derechos de intimidad.

El cuestionamiento del *ethos* comunicacional e informativo gira actualmente en torno a temas tan variados como la privacidad, la propiedad intelectual, el acceso libre al conocimiento, el derecho a la expresión en las redes digitales, la censura, las nuevas definiciones de género, la identidad digital, las comunidades digitales, el plagio digital, la sobrecarga informacional, la brecha digital y el control social digital. El filósofo italiano Luciano Floridi distingue entre una “ética de la comunicación global” y una “ética de la información global”.⁹

2.4 Noticias falsas

El periódico neoyorquino *The Sun* en 1835 e informaba de seres que habitaban la Luna. La noticia causó un enorme impacto en EEUU gracias a tres factores: la aparición de las prensas de alta capacidad, la caída del precio de los periódicos (la penny press), y la llegada de los nuevos medios de transporte que superaban la velocidad de los caballos por primera vez en la historia: los trenes y los barcos de vapor. Esos factores ayudaron a difundir a gran escala una información falsa y sensacionalista bajo el disfraz de una noticia verdadera. Es lo que hoy denominamos fake news.

La propaganda busca convencer, ser eficaz, y para eso puede recurrir a todo tipo de instrumentos, desde el arte hasta el cine, los pasquines o las redes sociales. Las noticias falsas, una de las ramas de la propaganda, son diferentes: buscan engañar, crear otra realidad. La preocupación por la forma en la que estos engaños cuajan y por los mecanismos a través de los que se crean y se multiplican no es nueva.

Las noticias falsas envenenan el mundo de las noticias y provocan desequilibrios morales.

El cuestionamiento del *ethos* comunicacional e informativo —o sea, de las normas, principios y valores que fundamentan las formas de comunicación y la información en una determinada sociedad— gira actualmente en torno a temas tan variados como la privacidad, la propiedad intelectual, el acceso libre al conocimiento, el derecho a la expresión en las redes digitales, la censura, las nuevas definiciones de género, la identidad digital, las comunidades digitales, el plagio digital, la sobrecarga informacional, la brecha digital y el control social digital

El filósofo italiano Luciano Floridi distingue entre una «ética de la comunicación global» (*ethics of global communication*) y una «ética de la información global» (global

⁹ Floridi, Luciano, *The Blackwell Guide to the Philosophy of Computing and Information* (editor), Blackwell, O

information ethics)¹⁰ la primera se refiere al diálogo y las prácticas consensuales en la interacción entre diversas culturas y generaciones. Pero tratándose de una postura meramente pragmática deja de lado, según Floridi, preguntas más básicas como, por ejemplo: ¿cuáles son en este caso los principios éticos que se toman como base de dicho diálogo y en qué se fundamenta dicha elección? o ¿existe una «macroética» en el sentido de un cierto tipo de consecuencialismo o deontologismo o contractualismo? ¹¹

2.5 Las maquinas reemplazan trabajo humano

Microsoft ha decidido sustituir a la mayoría de los periodistas que tenía contratados para producir noticias para el sitio web MSN.com y en su lugar emplear robots. La medida supondrá el despido de unos 50 redactores, aunque la compañía ha señalado que también se quedarán otros muchos profesionales en plantilla.

Entre las funciones que desempeñaban los periodistas que serán sustituidos estaba la de organizar las noticias en la portada de MSN, elegir los titulares y las imágenes que acompañan el contenido, y muchas otras. Sin embargo, a partir de finales de junio estas funciones de producción de noticias las realizarán sistemas automatizados utilizando la Inteligencia artificial.

La compañía no ha precisado si hará más cambios en otros departamentos que también impliquen sustituir el trabajo desempeñado por seres humanos por robots. Microsoft ha asegurado que adopta esta decisión tras evaluar la evolución de su negocio y la necesidad de reducir las inversiones para abaratar costes. De todas maneras, la firma aclara que los cambios no se vinculan con la pandemia de COVID-19.

La decisión ha provocado las críticas de algunos periodistas afectados, que consideran «desmoralizante» que una máquina pueda reemplazarlos y realizar su trabajo. Además, hay que tener en cuenta que la Inteligencia Artificial debe pasar por un proceso de formación y aprendizaje para poder sustituirles y esto podría llevar a la generación de contenido inapropiado en el portal de noticias.

Las estimaciones con respecto a la cantidad de **trabajos que se destruirán con la llegada de las máquinas** (robots, inteligencia artificial o cómo le quieran llamar) son variadas, pero la gran mayoría apuntan a una pérdida importante que afectará principalmente a aquellos puestos que requieren trabajo físico.

El Foro Económico Mundial (WEF) acaba de publicar su informe *['Future of Jobs 2018'](#)*¹² donde se lanzan con cifras que podrían sonar igual de alarmantes, pero que en esta ocasión nos plantean un escenario un poco más alentador. Y es que sí, el estudio pronostica que se perderán empleos, pero afirman que **se crearán nuevas funciones que requerían un mayor grado de especialización.**

Un ejemplo concreto de conflicto entre las visiones económicas y éticas: la muy debatida cuestión de la pérdida de puestos de trabajo inducida por la automatización, incluida la IA. Un aspecto crucial, especialmente en el futuro, que puede conducir a un aumento

¹⁰Luciano Floridi, *Sextus Empiricus, The Recovery and Transmission of Pyrrhonism*. Oxford: Oxford University Press, 2002.

¹¹Luciano Floridi, *The Blackwell Guide to the Philosophy of Computing and Information*. (a cura di) Oxford: Blackwell, 2003.

¹² http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

sustancial de las desigualdades, a menos que se adopten medidas, principalmente políticas, pero no sólo, para evitarlo.

"¿Es correcto introducir nuevas tecnologías que conduzcan a la pérdida de puestos de trabajo?", pregunta el representante de la Oficina Nacional de Investigación Económica de Estados Unidos. Y trata de responder: "los economistas podrían estar tentados a responder inmediatamente 'sí' a mi pregunta. Pueden observar que en un mercado que funciona bien, los salarios reflejan perfectamente el valor social del trabajo; si a un nivel determinado de salarios, una empresa encuentra conveniente innovar de manera que ahorre costes de mano de obra, libera mano de obra para emplearla en otras actividades más útiles para la sociedad.

En la práctica, si un determinado trabajo ya no es necesario, se vuelve obsoleto, pierde su utilidad y, por lo tanto, su valor, por lo que es mejor no hacerlo más y cambiarlo. Después de todo, siempre ha funcionado así. "Pero también hay que tener en cuenta que la pérdida de puestos de trabajo es socialmente más cara de lo que sugiere una visión eficiente del mercado".

No es sólo el mercado el que puede decidir

2.6 Pandemia y permisos verdes

El diciembre del 2019 aparecieron los primeros casos de Covid 19 la epidemia abrazó 2020, 2021 y parte del 2022 y no parece tener un final cercano. Esto ha modificado nuestro modo de vivir, aislados, distanciados, con barbijos, vacunados y ahora con obligación de presentar un certificado que se llama *Green pass*. Para viajar los países europeos han creado el *European Union digital Passenger Locator Form (EUdPLF)*¹³ pero se lo exige también en las escuelas, en los medios de transportes, en los lugares de comidas y en el trabajo. Esto está dando lugar a aguerridas disputas morales con intervenciones de jefes políticos (el Premier italiano Draghi dijo que no vacunarse es suicida y promover no vacunarse a los demás una instigación al suicidio) y hasta del Papa quien proclamó que vacunarse es un acto de amor.

La discusión seria es sobre el pedido de certificación para realizar actividades como trabajar: no solo los docentes han protestado sino también las centrales obreras sobre todo porque después que lo solicitaran las empresas públicas ahora también los privados piden certificados para entrar en las fábricas o lugares de trabajo.

En Francia se discute porque es obligatorio para los gendarmes, pero no para la policía y en general si este requisito no vulnera la libertad de los ciudadanos para disponer de las terapias sobre su propio cuerpo.

2.7 Sesgos cognitivos.

Trabajando con máquinas hemos descubierto que transmitimos nuestros sesgos a los programas que creamos, y por ende son prejuiciosos. Pero no somos los únicos que nos ocupamos del problema. En psicología se ocupan de sesgos cognitivos y los efectos psicológicos que dificultan procesar la información y llegan a tergiversar interpretaciones y a juicios inexactos, es la irracionalidad que distancia al *Homo sapiens* del *homo economicus*. Se han hecho estudios y se han podido verificar alrededor de 200 errores que cometemos.

Pero ¿el *Homo sapiens* no es la especie cuya capacidad intelectual lo lleva a la cima de la naturaleza? Parece que muchos de los errores de sesgos cognitivos son conductas

¹³ Permiso con formas estrictas que se obtienen en la plataforma de la UE.

adaptativas que surgieron cientos de miles de años atrás y que fueron útiles hasta no hace mucho. Cada cambio importante en la cultura llevó a cambios en nuestros sesgos cognitivos. Y si esto fuese así todos tendríamos más o menos los mismos sesgos. Pero no es el caso: el antropólogo Joseph Henrich publicó un libro¹⁴ donde los más extraños del mundo son los WEIRD (western educando, industrializado, rico y democrático) ¿Te identificas por tu profesión o logros, más que por tu red familiar? ¿Cultivas tus atributos y metas únicas? Si es así, tal vez usted es RARO: criado en una sociedad que es occidental, educada, industrializada, rica y democrática. A diferencia de la mayoría de los que han vivido alguna vez, las personas WEIRD son altamente individualistas, inconformistas, analíticas y orientadas al control. ¿Cómo las poblaciones WEIRD llegaron a ser tan psicológicamente peculiares? ¿Qué papel jugaron estas diferencias en nuestra historia y qué significan para nuestra identidad colectiva?

Los sesgos son diferentes según la cultura y si los WEIRD están en un extremo hay otros que están el opuesto y obviamente intermedios. Con un ejemplo: Ud. Acompaña una amiga en auto, doblan por una calle casi desierta a mayor velocidad de la permitida, aparece un transeúnte y lo atropellan, la pregunta es si quien conduce tiene derecho a pedir al amigo que mienta con respecto a la velocidad del auto. Los estadounidenses, canadienses, australianos y escandinavos en un 90 % dicen “no”. Venezuela, Corea del Sur o Rusia, el 50 % dice “sí”. No hay un dato seguro pero en Argentina estaríamos en un 65 % de “no”. El WEIRD tiene mejor economía, educación y gobierno, pero son prejuiciosos con respecto a los que no son como ellos. Los opuestos tienen mala economía, gobiernos despóticos, menor educación y menor tasa de ahorro, pero son menos prejuiciosos para con los demás.

Y todo esto tiene consecuencias morales: los que son más estrictos moralmente tienen más prejuicios y esto se transmite también a los programas que crean. En el reconocimiento facial hay problemas cuando la persona no es de tez blanca.

Un ejemplo concreto de conflicto entre las visiones económicas y éticas: la muy debatida cuestión de la pérdida de puestos de trabajo inducida por la automatización, incluida la IA

Este año el ganador del premio más interesante es en la categoría “innovación tecnológica al servicio del periodismo”. Ganó el proyecto realizado por un periódico canadiense, The Globe and Mail, cuyo sitio web ha sido dirigido por un sistema inteligente para elegir los artículos durante algún tiempo y los lectores nunca se han quejado de ello; juran, de hecho, que ni siquiera lo notarían.

3 . Documentos poniendo límites

3.1 La Unión Europea

Fue la primera en plantear el problema de la inteligencia artificial no sólo desde el punto de vista ético (punto de vista que ya ha seguido el Reino Unido en 2016), sino también desde el punto de vista reglamentario, especialmente desde los mecanismos de imputación de la responsabilidad civil.

En su resolución de 16 de febrero de 2017, que contiene recomendaciones a la Comisión Europea sobre las normas de derecho civil relativas a la robótica, en las que también se tienen en cuenta los sistemas de inteligencia artificial, el Parlamento Europeo ya había indicado como temas de atención los relativos a las repercusiones éticas y sociales,

¹⁴ Joseph Henrich *The WEIRDest People in the World*, Penguin Libros Ltd, 2021

subrayando que el desarrollo de la robótica y la inteligencia artificial debe tener por objeto integrar las capacidades humanas y no sustituirlas.

Esta resolución, si bien tiene por objeto proporcionar directrices en el ámbito civil, ofrece muchas ideas para comprender el alcance de los problemas jurídicos y sociales que la inteligencia artificial trae consigo.

Desde el punto de vista de la responsabilidad de los agentes inteligentes, el desafío surge cuando consideramos que el criterio tradicional de imputabilidad está vinculado a una conducta del agente sujeto ¿pone una inteligencia artificial su conducta autónoma en el mundo físico? Sin embargo, los sistemas de inteligencia artificial parecen recordar los principios de la responsabilidad del productor, que deben adaptarse al hecho de que esos sistemas asumen una capacidad de decisión autónoma, y podrían también implicar a sujetos distintos del productor únicamente, como los programadores o los que elaboran los algoritmos de decisión.

Otro criterio de imputación de responsabilidad podría encontrarse también en la culpa en la vigilancia de la persona que utiliza el sistema, cuando puede tener conocimiento de la adopción de decisiones erróneas por parte del sistema.

El documento del Parlamento Europeo enumera una serie de puntos de atención y también sugerencias para una correcta solución normativa de los mismos.

Desde un punto de vista jurídico, están apareciendo trabajos sobre algunos casos concretos: La IA y la competencia, la IA y los derechos de autor directos, la IA y las armas letales, la IA y los coches que se conducen solos, la IA y el mercado financiero, la IA y el reconocimiento facial, la IA y el mercado laboral, la IA y la asistencia sanitaria.

3.2 El Vaticano

El pasado 28 de febrero 2021 se firmó en el Vaticano un documento que enfatiza los aspectos éticos de la inteligencia artificial y que coloca otra pieza en la base de la regulación. Firmaron representantes de las grandes empresas de distribución de datos Amazon, Google, Facebook, Microsoft, Apple, representantes vaticanos, del gobierno italiano y de la Unión Europea. Concretamente firmaron Paola Pisano, Ministra de Innovación Tecnológica de Italia, Monseñor Vincenzo Paglia, Presidente de la Academia Pontificia para la Vida, el Director General de Fao Dongyu Qu y los presidentes de Microsoft e IBM. David Sassoli, Presidente del Parlamento Europeo, también estuvo presente en el evento¹⁵

3.3. El Parlamento Europeo (PE),

Tras una propuesta de borrador (junio 2016) y un informe razonado (febrero 2017), el PE ha aprobado un informe sobre Robótica en el que se establece un Código Ético de Conducta¹⁶.

La propuesta de resolución del PE establece que es necesario establecer “un marco de guía ético para el diseño, producción y uso de los robots” que sirva de complemento a las diferentes recomendaciones puramente legales que se realizan. Es decir, profundizar en una nueva disciplina que aborde los problemas éticos de la Inteligencia Artificial, la “roboética”.

¹⁵https://www.wired.it/attualita/tech/2020/02/28/intelligenza-artificiale-etica-vaticano-microsoft-ibm/?refresh_ce=

¹⁶ P8_TA (2017)0051 Normas de Derecho civil sobre robótica 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))

La idea de fondo es que los estándares éticos deberían ir dirigidos a la humanidad – esto es, los diseñadores, productores y usuarios de los robots– y no tanto a los robots en sí mismos. Como indica la profesora Nathalie Nevejans, responsable del informe encargado por el propio PE, no hay que confundir la ética en la robótica con la ética en las máquinas, es decir, una ética que obligue a los propios robots a adherirse a reglas éticas. Existen varios principios fundamentales que han sido recogidos por la resolución que incluyen la protección de la dignidad humana, la privacidad, la libertad, la igualdad de acceso o los efectos sociales, entre otros.

Proteger la privacidad y el uso de datos: especialmente cuando avancen los coches autónomos, los drones, los asistentes personales o los robots de seguridad.

Protección de la humanidad ante el riesgo de manipulación por parte de los robots: Especialmente en ciertos colectivos –ancianos, niños, dependientes– que puedan generar una empatía artificial.

Evitar la disolución de los lazos sociales haciendo que los robots monopolicen, en un cierto sentido, las relaciones de determinados grupos. Igualdad de acceso al progreso en robótica: Al igual que la brecha digital, la brecha robótica puede ser esencial. Restricción del acceso a tecnologías de mejora regulando la idea del transhumanismo y la búsqueda de mejoras físicas y/o mentales.

La difusión de los sistemas inteligentes ha sido percibida y ya países, asociaciones, entidades universales y regionales han comenzado a redactar y difundir documentos de alarma y orientación sobre la necesidad de límites éticos.

3.4 Los 36 países miembros de la OCDE, junto con Argentina, Brasil, Colombia, Costa Rica, Perú y Rumanía han suscrito hoy en París los Principios de la OCDE sobre la Inteligencia Artificial en el marco de la Reunión del Consejo de Ministros de la Organización, con el lema “La transición digital al servicio del desarrollo sostenible”.

3.5. El documento más impresionante es el de Unesco.

Desde noviembre de 2021, adoptada en por los 193 Estados Miembros de la Organización la [Recomendación de la UNESCO sobre la ética de la inteligencia artificial \(IA\)](#) proporciona un marco normativo común para todos los países del mundo. Cerca de 30 países ya han comenzado a utilizar esta Recomendación para establecer una legislación nacional que garantice que la inteligencia artificial respete las libertades fundamentales y los derechos humanos, y beneficie a toda la humanidad.¹⁷

Es el primer código que trata el tema y crea una ética universal para resolverlo. Eso y la manera de gestionarlo en tiempo record son los puntos salientes de este acuerdo

Este año La UNESCO y la Comisión Europea acaban de firmar un acuerdo para acelerar la aplicación mundial de la Recomendación de la UNESCO sobre la ética de la inteligencia artificial, se dedicará un presupuesto de cuatro millones de euros a apoyar a los países menos desarrollados en el establecimiento de su legislación nacional.

¹⁷ https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Conclusiones

Sinceramente creo que los humanos tenemos una noción bastante clara de lo que está bien y lo que está mal de la vida en sociedad y además de su propia vida. Dicho de otro modo, que la moral social y la ética individual han acompañado al hombre desde su aparición en la Tierra. Es cierto que cambiando el mundo hay nuevos desafíos éticos.

¿Qué es el mal? ¿Y su antítesis, el bien —o será al revés—? ¿De dónde procede el mal? ¿Cuándo se originó? ¿Qué sentido tiene? ¿Por qué tanto dolor y sufrimiento aparentemente inútiles? ¿Es que carece de límites espaciales y temporales el mal? ¿Por qué parece, en ocasiones, cebarse con los más justos y los inocentes —véase si no el caso paradigmático de Job—? ¿Qué implicaciones o consecuencias se derivan del mal? ¿Cómo erradicarlo de una vez por todas de la vida humana? ¿Es, acaso, esto último posible? Estas y otras muchas cuestiones de gran calado contenidas en el problema han agitado por centurias las mentes y vidas de insignes filósofos —y, por qué no decirlo, de todo ‘homo viator’ que ha transitado alguna vez por la faz de la Tierra—, que han intentado encontrar respuestas —en ocasiones de lo más variopintas, todo sea dicho— a las causas y consecuencias posibles planteadas por la diatriba de la presencia del mal en la naturaleza y en el núcleo de la vida misma.

Estas preguntas atraviesan toda la historia de la humanidad y se la puede rastrear en cualquier cultura, incluyendo la religión. En el nuevo testamento el mal tiene un nombre: el diablo y cuentan las escrituras que lo tentó a Jesús en el Desierto. Sin embargo, esa identificación con el Maligno, no es universalmente aceptada. Por lo poco que sé no aparece en el budismo y dentro de la misma iglesia católica es ejemplar la posición de algunos pensadores, por ejemplo, Leibniz: tanto en la Teodicea y en páginas posteriores como La presencia de Dios en el mundo. Leibniz no creía en el Diablo, por lo tanto, define la presencia de Dios en el mundo como bondad. Y ¿la maldad? Ah, eso es la ausencia de Dios en el mundo. Y de un plumazo elimino al Diablo, no solo agrega que esos temidos pueden ser representados numéricamente con 1, bondad y 0 ausencia de Dios en el mundo.

Pero va más lejos, dice que con esos dos términos puede escribir cualquier cosa y para demostrarlo comienza a escribir con 0 y 1 los ocho primeros números de la sucesión fundamental algo así como 000011010101111100, 10000110101011001, 0110000010100, etc. Uds. dirán pero eso es el sistema binario, sí, lo es. Y Leibniz se da cuenta de la fuerza de expresión de esos términos tanto que le escribe a un Jesuita que está en China

El jesuita queda maravillado pues el sistema de Leibniz puede ser comparado con el IChin compuesto por 64 hexagramas con rayas completas y rayas quebradas. Los dos sistemas son binarios y de algún modo reflejan la manera de proceder de nuestra mente tanto en occidente como en oriente.

Ambos le escribieron al Papa para pedirle que autorice la enseñanza de la religión católica en China a través del I Chin. Probablemente Clemente XI quien sentenció que la religión católica debía enunciarse en latín.

Siendo tanto la moral (mores) como la ética (ethos) dos palabras que significan costumbre necesariamente tienen que adaptarse a los tiempos. Pero no es que haya una gran diferencia en el transcurso de la historia. Ciertamente es que la moral de un pitecantropus erectus es distinta de la del cosmonauta Yuri Gagarin, porque el mundo a su alrededor es distinto, pero si se va a excavar que es lo que está bien y que lo que está mal se pueden enumerar las acciones que luego las religiones habrían de establecer como virtuosas o pecaminosas.

La ética tiene que ver con valores. Nuestra vida puede ser concebida como una sucesión de elecciones (conscientes o inconscientes) que la van conformando. Una cualidad del valor es que no es un sustantivo sino un adjetivo no es independiente si se habla de la belleza de una mujer o de un cuadro, otra cualidad del valor es su polaridad: siempre que se habla de un valor se puede predicar un disvalor correspondiente: bello y feo, bueno, malo, etc. Una tercera característica es la jerarquía, pero no del valor sino como los valores se van colocando en una organización jerárquica de valores donde a veces priman algunos sobre otros, pero esto puede cambiar con el tiempo, por ejemplo, en caso de Pandemia.

“Podría decirse que una doctrina ética es un conjunto sistematizado de proposiciones relativas a la vida moral, que recomiendan o aprueba una serie de actividades humanas consideradas valiosas”

La necesidad de controlar las decisiones producidas por la inteligencia artificial en un momento en que las soluciones digitales a los problemas sociales se invocan en toda sociedad tecnológicamente avanzada. Y de hecho si vemos la primera regulación jurídica de la Unión Europea trata de regular el uso de esta conforme a riesgos que produce: nacen así las prácticas imposibles, por lo extremadamente peligroso de hacerlo, las riesgosas pero aceptables, las regularmente riesgosas y las que crean un riesgo mínimo.

¿Cuál es el marco para una IA confiable?: 1 Acción y supervisión humana. 2. Solidez técnica y seguridad. 3. Gestión de la privacidad de los datos. 4. Transparencia 5. No discriminación y equidad, 6. Bienestar social y ambiental. 7. Rendición de cuentas y responsabilidades. La Unesco pone los cimientos de una declaración universal sobre ética en la I.A.

Nace una ética global, aparecen morales ontocéntricas, biocéntricas y antropocéntricas y es necesario informarse, reflexionar, tomar partido. Se tiene clara conciencia que nuestras decisiones no solo afectan nuestra vida y nuestro futuro inmediato sino también el largo plazo y las próximas generaciones y es por eso que, a las tres reglas clásicas de Asimov, se agrega una cuarta en el cual el robot debe cuidar de no afectar la perdurabilidad de la especie humana, como el mal tratamiento del ambiente o la afectación de los recursos naturales.

La economía tiene que rendirle cuentas a la moral, no es posible que solo las leyes del mercado decidan la falta de trabajo de millones de personas, la hambruna y la angustia de muchos. Con cada situación, con cada mundo que se va creando con los sistemas inteligentes es necesario plantearse hasta donde las acciones que se generan y las consecuencias que derivan pueden traer felicidad o dolor a los seres humanos que serán implicados por las mismas porque, en definitiva, como decía Kant: la moral no es realmente la doctrina de cómo hacernos felices, sino de cómo debemos hacernos dignos de la felicidad.

Para terminar nada mejor que la paradoja de Fermi: él se preguntaba cómo era posible que cuando se buscaban vidas inteligentes estos siempre aparecían en la Tierra cuando era estadísticamente improbable que no hubiese vida inteligente en el vasto universo. Su hipótesis era que cuando la tecnología va más rápido de la razón, la vida inteligente se destruye a sí misma. Que nos sirva de advertencia.

SEMBLANZAS DE LOS AUTORES

Juan Carlos Ferreri se graduó como Ingeniero Aeronáutico en la Universidad de La Plata en 1967 y ha dedicado su carrera profesional al campo particular de la Mecánica de los Fluidos computacional y la transferencia de calor y materia y es considerado un pionero en dicha actividad en la Argentina. En los últimos treinta y cinco años (hasta su retiro) ha dedicado su trabajo a la Seguridad e Ingeniería Nuclear en la Autoridad Regulatoria Nuclear de Argentina. También ha contribuido investigaciones en la arqueometría de hogares de combustión prehistóricos desde hace más de tres décadas.

Ha sido investigador del CONICET, desde 2009 es miembro de la Academia Nacional de Ciencias de Buenos Aires y llegó a presidirla desde 2017 hasta 2021. Recibió el premio 2004, categoría Senior, a la Trayectoria Profesional, de Investigación y Docencia en Argentina de la Asociación de Mecánica Computacional de la Argentina (AMCA). Ha sido miembro y presidente del Comité Argentino de Transferencia de Calor y Materia (CONICET). Ha sido también profesor de posgrado en diferentes universidades por períodos cortos y miembro de comités asesores en universidades, en el CONICET y en otras instituciones. Es revisor de trabajos para algunas revistas internacionales. Ha sido miembro de numerosos comités organizadores y científicos en reuniones nacionales e internacionales. Ha dictado cursos de posgrado en Argentina, EEUU, Italia, Perú, China y Francia y dictado decenas de seminarios y conferencias invitadas en Argentina y el extranjero. Ha publicado más de 110 trabajos en su área de conocimiento.



Daniela López De Luise es Licenciada en Análisis de Sistemas por la Facultad de Ingeniería de la Universidad de Buenos Aires (1989), luego continuó su formación en el sector de sistemas inteligentes obteniendo el título de Especialista en Ingeniería de Sistemas Expertos por el Instituto Tecnológico de Buenos Aires (2000), homologando su título al de Ingeniera en Informática (Ced Esp 2004/H04041). En 2008 recibe el título de Doctora en Ciencias informáticas por la Universidad Nacional de La Plata. Finalmente adquiere el título de Comunicadora Pública de la Ciencia y la Tecnología otorgado en conjunto por la Facultad de Ciencias Exactas, de Sociología y de Filosofía y letras de la Universidad de Buenos Aires (2022).

Su trayectoria profesional comienza en sistemas inteligentes, y se expande a otras temáticas relacionadas habiendo colaborado para grandes firmas nacionales e internacionales. Gradualmente incorpora las actividades científicas y académicas llegando a liderar el Computational Intelligence and Information Systems Labs (desde 2013) y el IDTI Lab de la Facultad de Ciencia y tecnología en la Universidad Autónoma de Entre Ríos. Asimismo, participa como docente e investigadora en varias entidades educativas. Actualmente es la coordinadora académica del Centro de Estudios en



<p>Tecnologías Inteligentes de la Academia Nacional de Ciencias de Buenos Aires. Ha sido fundadora y steering committee de varias redes y laboratorios. Fundó, fue primer presidente y actualmente continúa perteneciendo al staff del IEEE Computational Intelligence Society de Argentina (primer capítulo mundial del IEEE en estas temáticas), fue presidente de IEEE Argentina, fundadora y socio gerente de Enterpriseware S.R.L., fundadora y líder de IEEE Women in Computational Intelligence, y pertenece activamente a la cátedra latinoamericana Matilda para las Mujeres en Ciencia. Tiene en su haber varios reconocimientos nacionales e internacionales, entre ellos el premio SADOSDY y el de Ingeniera Eminente Latinoamericana del IEEE R9. Actualmente coordina el IEEE Games Technical Committee en Argentina, y la Red de Desarrollo y Producción de Videojuegos (consorcio internacional latinoamericano). También preside algunas diplomaturas y coordina proyectos relacionados con el razonamiento lingüístico, chatterbots, y Wavelets Morfosintácticas. Realiza tareas de consultoría para varias empresas nacionales e internacionales. Su área de investigación base es la generación automática de lenguaje y su relación matemática con el habla natural.</p>	
<p>Antonio Anselmo Martino Nacido el 29 de agosto de 1937. Se gradúa de abogado en la Facultad de Derecho de la Universidad de Buenos Aires, 1962, de Doctor en Derecho y Ciencias Sociales en la misma universidad, 1974. En 1971 gana el concurso de Profesor Adjunto ordinario en la misma facultad de Introducción al Derecho. En 1969 es nombrado profesor titular de filosofía del derecho en la Universidad de Belgrano. Desde 1975 miembro de la Sociedad Científica Argentina. En 1976, gana la cátedra de Ciencia política en la Facultad de Ciencias políticas de la Universidad de Pisa, Italia, en 1978 es nombrado <i>Maitre de Conférences</i> en filosofía del derecho en la Universidad de Rouen, Francia. En 1982 es nombrado Director del Instituto para la Documentación Jurídica del Consejo Nacional de Investigaciones de Italia, con sede en Florencia. 1982, Profesor del Curso de Estado Mayor de la Escuela Superior de Guerra Naval de Livorno. 1985, Profesor de Ciencias Políticas del Curso de Estado Mayor de las tres fuerzas armadas italianas 1986 es nombrado representante italiano en la Sede de Ediforum de la Comunidad Económica Europea, con sede en Bruselas, Bélgica, 1986, Miembro de la Comisión de la Informática del Consejo Nacional de las Investigaciones en Italia, 1987, Presidente del FIRILITE (Federation of International Research Institutes on Law and Information Technology in Europe), 1991 es nombrado Académico correspondiente de Academia Nacional de Derecho y Ciencias Sociales de Córdoba. 1992, .Miembro ordinario del ISLE (Institute for Legislative Studies and Documentation) Miembro del AIIA (Associazione Italiana di Intelligenza Artificiale), 1994, .Miembro asociado del Center for Artificial Intelligence and Cognate learning of the University of Greenwich, 1994, .Asesor Científico de la Obra de la Convención Nacional Constituyente 1994, Recopilación de la documentación de la última reforma constitucional, La Ley, Buenos Aires I, II, III, IV, V, VI, VII y VIII tomo, 1996-1998.1995, Miembro Honorario del Circolo di Giuristi telematici, Italia. 1995 Director del Master en “Ciencia de la Legislación” promovido por las Universidades de Pisa y del Salvador, en la facultad de Ciencias Jurídicas de la Universidad del Salvador en Buenos Aires, 1996 Presidente de la Comisión italiana EBES</p>	



<p>(European Board for EDI standards), 1997, Director académico de Escuela de altos estudios para el Mercosur. Montevideo, 2004. Miembro del Instituto de Derecho Constitucional de la Academia de Derecho y Ciencias Sociales de Buenos Aires, 2005, Miembro del Instituto de Filosofía política y moral de la Academia Academia Nacional de Ciencias Morales y Políticas, 2006. Miembro Vocal del GESI, Grupo de Estudios de Sistemas Integrados 2007. Miembro del Capítulo Argentino de ISOC, Sección Argentina de Internet Society, 2008, Profesor emérito de la Universidad de Pisa, Italia. 2009, 55. Evaluador de la Agencia española Agencia Nacional de Evaluación y Prospectiva (ANEP). 2010. Profesor emérito de la Universidad del Salvador, Argentina, 2010, presidente del Comité Académico de la VIII Conferencia Internacional de Derecho e Informática de La Habana, 2011. Secretario para Europa de la Red Académica de Protección de Datos Personales, Monterrey, México. 2012, Declarado por Ley 4286 Personalidad Destacada de la Ciudad Autónoma de Buenos Aires en el ámbito de las Ciencias Jurídicas, 2014, Miembro de la International Association Artificial Intelligence and Law. 2015, Miembro del Consejo Asesor del progetto Re.Cri.Re de la Unión Europea. 2017. Profesor del Master in istituzioni parlamentari “Mario Galizia” per consulenti d’assemblea, Università di Roma. Miembro del Consejo Científico de las Revistas: -Del derecho industrial, Buenos Aires; - Computer/Law Series, Amsterdam; - Artificial Intelligence and Law, Boston; - Law and Information Technology, Oxford; - Derecho de la alta tecnología, Buenos Aires; - Forum per la Tecnologia dell’Informazione, Roma. - Legisprudence. International Journal of the study of legislation, London. Derecho del Ciberespacio, Madrid, Revista de Derecho de las Telecomunicaciones, Internet y Medios Audiovisuales, Ciencia de la Legislación (Buenos Aires) RLPDP – Revista Latinoamericana de Protección de Datos Personales. México. - Editor de la colección: Logica, Informatica, Diritto del CEDAM, Milán. Autor de 53 libros, y 762 artículos, ensayos y obras menores</p>	
<p>Julián Antonio Pucheta recibió el grado de Ingeniería Eléctrica de la Universidad Tecnológica Nacional - Facultad Regional de Córdoba, Argentina, el MS y Ph.D. títulos de la Universidad Nacional de San Juan, Argentina, en 1999, 2002 y 2006, respectivamente. Actualmente es investigador docente en el Laboratorio de investigación en matemáticas aplicadas para control, Universidad Nacional de Córdoba, Argentina. Sus intereses de actuación incluyen el modelado de procesos reales mediante métodos estocásticos y el control óptimo para procesos dinámicos, el pronóstico de series temporales y el aprendizaje automático. Entre 2009 y 2016 ha sido consultor de la OEA en el Proyecto SAOCOM de la CONAE. Ha dirigido trabajos de final de carrera de Grado en Ingeniería Electrónica, a nueve maestrandos en Ciencia de Datos y uno Telecomunicaciones, y han terminado su formación doctoral tres investigadores bajo su Dirección. Actualmente está dirigiendo a tres maestrandos en la disciplina del Control automático y dos doctorandos en Ingeniería.</p>	

David L. la Red Martínez se graduó de Experto en Estadística y Computación en la Universidad Nacional del Nordeste (UNNE), Argentina. Posteriormente completó estudios de postgrado de Especialista en Docencia Universitaria en la misma universidad y de Magister en Informática y Computación en dicha universidad en convenio con la Universidad de Cantabria, España. Luego obtuvo el título de Doctor en Ingeniería de Sistemas y Computación por la Universidad de Málaga, España. Es Profesor Titular de la UNNE y de la Universidad Tecnológica Nacional (UTN), Argentina, desempeñándose como docente e investigador. Es Director de la Maestría en Informática y Computación, de la Universidad Nacional de Pilar (UNP), Paraguay y de la Universidad Nacional del Este (UNE), Paraguay. Es Director del Doctorado en Informática de la UNNE en convenio con la UTN y la Universidad Nacional de Misiones (UNaM), Argentina. Está Categorizado I en el Programa de Incentivos a la Investigación de la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación, Argentina. Está Categorizado A en la Carrera de Docente Investigador de la UTN. Es Evaluador de la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) para carreras de grado y de postgrado de Sistemas. Es Revisor para numerosos Congresos y Revistas nacionales e internacionales. Es Miembro del Comité Académico de la Maestría en Sistemas y Redes de Telecomunicaciones de la UNNE, de la Maestría en Tecnologías de la Información de la UNNE en convenio con la UNaM y del Doctorado en Informática de la UNNE en convenio con la UNaM y la UTN. Es Miembro Titular de la Comisión de Postgrado de la Facultad de Ciencias Exactas y Naturales y Agrimensura (FaCENA) de la UNNE. Es autor de numerosas publicaciones en revistas y congresos nacionales e internacionales (especialmente Europa y Estados Unidos de América) y nacionales con referato e indexación en temas de sistemas operativos, gestión de recursos y procesos en sistemas distribuidos, minería de datos educacional y aplicada a la salud, imputación de datos, etc. Ha dictado numerosas conferencias y cursos de postgrado en el país y en el exterior. Ha dirigido y dirige proyectos de investigación y tesis de postgrado. Se ha desempeñado como Coordinador de Área, Director de Departamento, miembro del Consejo Directivo de la FaCENA y ha integrado numerosas comisiones evaluadoras de docentes y de proyectos de investigación en varias universidades. Durante más de veinte años ha dirigido el Centro de Cómputos del Gobierno de la Provincia de Corrientes.



Mgter. Federico Agostini se graduó de Licenciado en Sistemas de Información en la Universidad Nacional del Nordeste (UNNE), Argentina. Posteriormente completó estudios de postgrado como Magister en Sistemas y Redes de Telecomunicaciones en la misma universidad. Se encuentra en la etapa de desarrollo de tesis del Doctorado en Informática (UNNE). Actualmente se desempeña como CPA Bioinformático en el Instituto de Botánica del Nordeste - CONICET. Fue Jefe de Trabajos Prácticos dedicación Simple de la Asignatura Comunicaciones de Datos y en contra cuatrimestre de Sistemas Operativos, de la carrera Licenciatura en Sistemas de Información de la Universidad Nacional del Nordeste. Ha participado y participa en proyectos de investigación relacionados con la Informática en FaCENA-UNNE y en Bioinformática en IBONE-CONICET. Actualmente se encuentra dirigiendo una tesis correspondiente a la Maestría en



<p>Informática y Computación, de la Universidad Nacional de Pilar (UNP), Paraguay. Está Categorizado IV en el Programa de Incentivos a la Investigación de la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación, Argentina.</p> <p>Ha finalizado una beca en la que desarrolló una estancia de investigación en bioinformática en la Universidad de Málaga, Málaga, España. Ha colaborado en el proyecto de investigación P12-AGR-0482 (titulado «Nuevas herramientas para la mejora de la adaptación del trigo en Andalucía») financiado por la Consejería de Economía, Innovación y Ciencia de la Junta de Andalucía (España) y el Fondo Europeo de Desarrollo Regional (Unión Europea), esta investigación se realiza en colaboración con miembros del grupo de investigación Biotecnologías Agroalimentarias (PAIDI AGR248). En 2019 ya había realizado una estancia de investigación en Bioinformática en el departamento de Lenguajes y Ciencias de la computación de la Universidad de Málaga, España, en el seno del grupo de investigación “Sistemas de Información colaborativos (PAI TIC-160)”, bajo la dirección del Dr. Sergio Gálvez Rojas. También ha participado en diferentes publicaciones en congresos internacionales, y revistas de alto impacto.</p>	
<p>Mgter. Terecio Marecos Brizuela se graduó en la carrera de Licenciatura en Análisis de Sistemas en la Facultad de Ciencias Aplicadas – Universidad Nacional de Pilar (UNP), Paraguay. Posteriormente completó estudios de postgrado de Especialista en Docencia Universitaria en la misma universidad y de Magister en Ciencias de la Educación con Énfasis en Investigación Científica en la Facultad de Humanidades y Ciencias de Educación de Universidad Nacional de Pilar (UNP), Paraguay. Actualmente es Tesista en la Maestría en Informática y Computación de la Facultad de Ciencias Aplicadas de la Universidad Nacional de Pilar (UNP). Es profesor Titula en la cátedra de Programación de Aplicaciones en Redes y Tutor/Capacitador Tecnológico en el área de Educación a Distancia (plataforma Moodle). Se desempeña como Auxiliar Técnico en la Dirección de Tecnología e Informática en la Facultad de Derecho, Ciencias Políticas y Sociales de la Universidad Nacional de Pilar (UNP).</p>	
<p>Luis Joyanes Aguilar es Presidente de la Fundación I+D del Software Libre (Fidesol, Centro de Apoyo a la Innovación Tecnológica, CaIT), Granada (España). Dr. Ingeniero en Informática por la Universidad de Oviedo y Dr. en Sociología por la Universidad Pontificia de Salamanca. Catedrático de Lenguajes y Sistemas Informáticos de la Universidad Pontificia de Salamanca (UPSA). Dr. Honoris Causa por la Universidad Privada Antenor Orrego de Trujillo, UPAO, (Perú), por la Universidad San Martín de Porres, Lima (Perú) y por la Universidad Inca Garcilaso de la Vega, Lima (Perú). Líder Académico del TEC de Monterrey, México, campus Querétaro. Teniente Coronel de Artillería del Ejército Español (en la Reserva). Licenciado en Ciencias Físicas con Grado por la Universidad Complutense de Madrid y Licenciado de Enseñanza Superior Militar de la Academia General Militar de Zaragoza y por la Academia de Artillería de Segovia (en la actualidad Teniente Coronel de Artillería en la Reserva). Director de la</p>	

Cátedra Open Ingenium de Transformación Digital de Fidesol. En abril de 2018 recibió la Mención Honorífica del Doctorado en Ingeniería de la Universidad Distrital Francisco José de Caldas, de Bogotá (Colombia). El 30 de abril de 2022 fue nombrado Miembro de Honor de la Asociación de Directores de Servicio AFSM de España. Director del Departamento de Lenguajes y Sistemas Informáticos (1.996-2008), Vicedecano de la Facultad de Informática (1996-2002) y Decano de la Facultad de Informática y Director de los Programas de Doctorado en Ingeniería Informática de la Universidad Pontificia de Salamanca campus Madrid, años 2002 a 2008 y de los Másteres en Ingeniería de Software y de Sistemas de Información Geográfica (GIS) durante el periodo 2002 a 2010. Conferenciante habitual en congresos, simposios, cursos, talleres, jornadas a nivel nacional e internacional, con énfasis especial en Latinoamérica y El Caribe, donde ha sido Profesor Visitante, Asociado e Investigador en numerosas universidades españolas, europeas y -esencialmente- latinoamericanas; en alguna de ellas todavía continua de Profesor Asociado. Durante los años 2021 a 2023 ha impartido numerosas videoconferencias, seminarios, talleres y miembro de Tribunales de tesis doctorales en modalidad on-line, en numerosas universidades de Argentina, Ecuador, Colombia, México, Perú, República Dominicana, entre otros países. Ha dirigido más de 50 tesis doctorales de estudiantes españoles, portugueses y latinoamericanos; así como numerosas tesis de Máster. Ha escrito más de 40 libros de TIC y más de 150 artículos científicos y profesionales. Sus últimos libros publicados en la editorial Alfaomega (México, Argentina) y Marcombo (España) son: en 2023, Ciencia de Datos: Un enfoque práctico de tecnologías, herramientas y aplicaciones; en 2022, segunda edición; Computación en la Nube. Estrategia de Cloud Computing en la empresa; en 2021: Internet de las Cosas: Un futuro hiperconectado; 5G, Inteligencia Artificial, Big Data, Cloud, Blockchain, Ciberseguridad. Libros anteriores publicados en las citadas editoriales: “Inteligencia de Negocios y Analítica de Datos. Una visión global de Business Intelligence & Analytics” (2019); “Industria 4.0. La Cuarta Revolución Industrial” (2018); Sistemas de Información en la Empresa (2015) y Big Data. Análisis de grandes volúmenes de datos en organizaciones (2014). En la editorial McGraw-Hill (México, España) sus últimos libros publicados son: En 2022. Fondamenti di programmazione in C++. Algoritmi, strutture dati e oggetti (Scienze). 2ª Edición (Edizione revisionata) Secondi edizione. Edición italiana de Aldo Franco Dragoni, Università Politecnica delle Marche. En 2021 “Fundamentos de Programación. Algoritmos, Estructuras de Datos y Objetos”, quinta edición. Así mismo ha sido Coordinador y Coautor, de los libros: Ciberseguridad: La colaboración público-privada (2018) y Ciberseguridad. Restos y amenazas a la Seguridad Nacional en el Ciberespacio (2012), ambos publicados por el Instituto Español de Estudios Estratégicos del Ministerio de Defensa de España. Ha sido Investigador del Grupo de Investigación de “Ética en la Nube” de la Facultad de Filosofía de la Universidad Complutense de Madrid y Miembro del Instituto Universitario “Agustín Millares” de Biblioteconomía y Documentación de la Universidad Carlos III de Madrid.

Líneas de investigación actuales: Inteligencia Artificial; Computación en la nube (Cloud Computing y Edge Computing), Big Data, Ciencia de Datos; Internet de las Cosas, Ciberseguridad, Inteligencia de Negocios, Ingeniería de Software y Programación de Computadoras, Sistemas de Información y Transformación Digital.



ANCBA 2023