

AGENTES DE INTELIGENCIA ARTIFICIAL PARA LOS NEGOCIOS Y LA CIBERSEGURIDAD

**Gustavo Eduardo Fernández Villacrés
Freddy Patricio Baño Naranjo
Wilfrido Amilcar Trujillo Quinto
José Luis Vásquez Salazar**

2026



ISBN 978-9907-0-1798-4

DOI 10.64584/aian1456

AGENTES DE INTELIGENCIA ARTIFICIAL PARA LOS NEGOCIOS Y LA CIBERSEGURIDAD

AUTORES

PhD. Gustavo Eduardo Fernández Villacrés
Instituto Superior Tecnológico España, Ambato, Ecuador
gustavo.fernandez@iste.edu.ec



<https://orcid.org/0000-0003-1028-1224>

Mg. Freddy Patricio Baño Naranjo
Instituto Superior Tecnológico Mayor Pedro Traversari, Quito, Ecuador
freddy.bano@istpet.edu.ec



<https://orcid.org/0000-0001-7289-7505>

Mg. Wilfrido Amilcar Trujillo Quinto
Instituto Superior Tecnológico Mayor Pedro Traversari, Quito, Ecuador
wilfrido.trujillo@istpet.edu.ec



<https://orcid.org/0000-0001-8580-1322>

Mg. Jose Luis Vásconez Salazar,
Universidad Estatal de Bolívar.
jvasconez@ueb.edu.ec
Facultad de las Ciencias de la Educación, Guaranda-Ecuador



<https://orcid.org/0000-0002-3588-1448>

AGENTES DE INTELIGENCIA ARTIFICIAL PARA LOS NEGOCIOS Y LA CIBERSEGURIDAD

Primera edición, agosto 31 del 2026

ISBN: 978-9907-0-1798-4

(e-book)

DOI: 10.64584/aian1456



Editado por:

Ing. Marco Salazar C.

Telf. 0960677758

Ambato-Ecuador



Este libro ha sido sometido a un proceso de evaluación por pares externos con base a la normativa editorial. También dispone de revisión antiplagio.

Ver anexos

Prohibida su reproducción total o parcial..

Diseño y diagramación.

Diseño, montaje y producción editorial. **MS Editorial**



ISBN: 978-9907-0-1798-4



Hecho en Ambato, Ecuador

Made in Ambato, Ecuador.

PRÓLOGO

La humanidad atraviesa una de las transformaciones tecnológicas más significativas de su historia. La convergencia entre la Inteligencia Artificial (IA), la automatización inteligente, el análisis masivo de datos y la ciberseguridad está redefiniendo la forma en que las organizaciones operan, compiten y generan valor en un entorno digital cada vez más complejo e interconectado. En este contexto, los agentes de inteligencia artificial emergen como una de las innovaciones más disruptivas del siglo XXI, al permitir que sistemas computacionales perciban información, razonen, aprendan y ejecuten acciones de manera autónoma para alcanzar objetivos específicos.

Los agentes inteligentes representan una evolución natural de la Inteligencia Artificial tradicional. Ya no se trata únicamente de sistemas capaces de procesar información o realizar predicciones, sino de entidades digitales que pueden interactuar con personas, aplicaciones, bases de datos y entornos organizacionales para resolver problemas, optimizar procesos y apoyar la toma de decisiones en tiempo real. Su aplicación se extiende desde la automatización de tareas empresariales y la atención al cliente hasta la detección de amenazas cibernéticas, la respuesta a incidentes de seguridad y la protección de infraestructuras críticas.

En el ámbito de los negocios, los agentes de inteligencia artificial están impulsando una nueva generación de organizaciones inteligentes. Empresas de todos los tamaños utilizan estas tecnologías para mejorar la experiencia del cliente, optimizar recursos, automatizar procesos operativos, fortalecer

la gestión del conocimiento y desarrollar ventajas competitivas sostenibles. La capacidad de analizar grandes volúmenes de información y generar respuestas oportunas convierte a los agentes inteligentes en aliados estratégicos para la innovación y el crecimiento empresarial.

Paralelamente, la ciberseguridad enfrenta desafíos sin precedentes. El incremento de los ataques informáticos, el crecimiento exponencial de los datos digitales y la sofisticación de las amenazas requieren mecanismos de defensa más rápidos, precisos y adaptativos. En este escenario, los agentes de inteligencia artificial desempeñan un papel fundamental al fortalecer la detección temprana de riesgos, automatizar la respuesta ante incidentes y contribuir a la protección de activos digitales esenciales para las organizaciones modernas.

El presente libro, **“Agentes de Inteligencia Artificial para los Negocios y la Ciberseguridad”**, ha sido concebido con una visión integral que combina fundamentos teóricos, aplicaciones prácticas y perspectivas estratégicas. Su propósito es proporcionar a estudiantes, docentes, investigadores, profesionales y directivos una comprensión sólida sobre los principios que sustentan la Inteligencia Artificial y los agentes inteligentes, así como sobre su impacto en los procesos empresariales y en la seguridad de la información.

La obra se estructura en cuatro capítulos fundamentales. El **Capítulo I**, *“Fundamentos de la Inteligencia Artificial”*, introduce al lector en la evolución histórica de la IA, sus conceptos esenciales, tecnologías habilitadoras y tendencias actuales. El **Capítulo II**, *“Fundamentos de los Agentes Inteligentes”*, profundiza en la naturaleza, arquitectura,

funcionamiento y clasificación de los agentes de inteligencia artificial. El **Capítulo III**, “*Agentes de Inteligencia Artificial para los Negocios*”, aborda las aplicaciones empresariales de estas tecnologías como motores de transformación digital, innovación y competitividad. Finalmente, el **Capítulo IV**, “*Agentes de Inteligencia Artificial en la Ciberseguridad*”, analiza su contribución a la prevención, detección y mitigación de amenazas en entornos digitales cada vez más complejos.

Más que una simple recopilación de conocimientos, este libro constituye una invitación a reflexionar sobre el papel que desempeñará la Inteligencia Artificial en el futuro de las organizaciones y de la sociedad. La adopción responsable, ética y estratégica de estas tecnologías será determinante para construir ecosistemas digitales seguros, sostenibles e innovadores que contribuyan al desarrollo económico y social.

Esperamos que esta obra se convierta en una herramienta de consulta, aprendizaje e inspiración para quienes buscan comprender y aprovechar el potencial transformador de los agentes de inteligencia artificial en los negocios y la ciberseguridad, contribuyendo así a la formación de profesionales capaces de liderar los desafíos de la nueva economía digital.

Los autores

INDICE GENERAL

CAPITULO I	1
FUNDAMENTOS DE INTELIGENCIA ARTIFICIAL.....	1
1.1 Introducción y Marco General	1
1.1.1 Importancia estratégica de la Inteligencia Artificial en la sociedad actual.....	1
1.1.2 Impacto de la IA en la transformación digital.....	2
1.1.3 Relación tripartita: Inteligencia Artificial, negocios y Ciberseguridad	3
2.1 Evolución de la Inteligencia Artificial.....	4
2.1.1 Origen y fundamentos teóricos de la IA	4
2.1.2 Etapas de desarrollo y ciclos de invierno.....	6
2.1.3 La era de la Inteligencia Artificial generativa	7
3.1 Conceptos fundamentales y clasificación de la IA.....	8
3.1.1 Definición formal de la Inteligencia Artificial	8
3.1.2 Delimitación conceptual	9
3.1.3 Capacidades cognitivas principales de los sistemas inteligentes	11
3.1.4 Taxonomía y niveles de desarrollo de la IA	12
4.1 Fundamentos del aprendizaje automático (Machine Learning)	13
4.1.1 Concepto y filosofía del aprendizaje automático.....	13
4.1.2 Ciclo de vida del aprendizaje basado en datos	14
4.1.3 Aprendizaje supervisado	15
4.1.4 Aprendizaje no supervisado	17
4.1.5 Aprendizaje por refuerzo	18
4.1.6 Retos y limitaciones del Machine Learning	19
5.1 Deep Learning y redes neuronales artificiales.....	20
5.1.1 Introducción al aprendizaje profundo	20
5.1.2 Inspiración biológica de las redes neuronales	21
5.1.3 Componentes matemáticos de una red neuronal	22
5.1.4 Arquitecturas modernas de Deep Learning.....	23

6.1	Procesamiento del lenguaje natural (NLP) y modelos generativos	25
6.1.1	Definición y alcance del procesamiento del lenguaje natural	25
6.1.2	Modelos de lenguaje a gran escala (LLM).....	26
6.1.3	Modalidades de la IA Generativa.....	27
6.1.4	Aplicaciones de NLP en negocios y Ciberseguridad.....	28
7.1	Ecosistema tecnológico de la inteligencia artificial	29
7.1.1	Los Datos como activo estratégico y combustible de la IA	29
7.1.2	Infraestructura Hardware y computacional	30
7.1.3	Principales actores y plataformas de IA.....	31
8.1	Beneficios, riesgos, ética y regulación de la IA	33
8.1.1	Beneficios organizacionales directos.....	33
8.1.2	Riesgos tecnológicos, desafíos e impacto laboral	33
8.1.3	Marcos regulatorios internacionales.....	35
CAPÍTULO II		36
FUNDAMENTOS DE LOS AGENTES INTELIGENTES.....		36
2.1	¿Qué son los agentes de inteligencia artificial?.....	36
2.1.1	Concepto y definición de los Agentes de Inteligencia Artificial	37
2.1.2	Características principales.....	38
2.1.3	Componentes de un agente inteligente.....	40
2.1.4	Ciclo de percepción, razonamiento y acción	42
2.2.	Tipología de agentes Inteligentes	44
2.2.1	Agentes reactivos	44
2.2.2	Agentes basados en objetivos.....	46
2.2.3	Agentes basados en utilidad	47
2.2.4	Agentes autónomos	49
2.2.5	Agentes colaborativos o multiagente	50
2.3.	Arquitectura y funcionamiento de los Agentes IA	52
2.3.1	Sensores y actuadores	52

2.3.2	Memoria a corto y largo plazo	53
2.3.3	Herramientas y conectores	55
2.3.4	Modelos de lenguaje (LLM).....	56
2.4.	Plataformas y frameworks para la el desarrollo de agentes inteligentes.....	57
2.5.	Aspectos éticos y regulatorios	62
2.5.1	Transparencia algorítmica	63
2.5.2	Privacidad y protección de datos	63
2.5.3	Sesgos en IA	64
2.5.4	Gobernanza de IA	65
CAPÍTULO III.....		66
AGENTES DE INTELIGENCIA ARTIFICIAL PARA LOS NEGOCIOS		66
3.1	Transformación digital e industria impulsada por agentes de IA	66
3.1.1	Contextualización de la transformación digital	66
3.1.2	El concepto de “Negocios Inteligentes”	67
3.1.3	La evolución industrial: De la industria 4.0 a la industria 5.0 ..	68
3.1.4	Automatización empresarial de siguiente generación.....	69
3.2	Agentes de IA en marketing, ventas y comercialización	71
3.2.1	Segmentación inteligente de audiencias	71
3.2.2	Hyper-Personalización de campañas de marketing	72
3.2.3	Asistentes y Chatbots comerciales avanzados	73
3.2.4	Motores y sistemas de recomendación inteligentes	74
3.3	Agentes de IA para la gestión operativa corporativa.....	75
3.3.1	Gestión documental e inteligencia de datos no estructurados	75
3.3.2	Agentes de IA en la gestión del talento humano	77
3.3.3	Automatización de la gestión financiera y contable	78
3.3.4	Inteligencia de Negocios (Business Intelligence) asistida por agentes	80
3.4	Agentes de IA en el ecosistema de microempresas y PYMES ..	81

3.4.1	Democratización de la tecnología para pequeños negocios ...	81
3.4.2	Automatización integral de procesos de bajo costo	83
3.4.3	Comercio electrónico inteligente para PYMES	84
3.5	Agentes de IA en la toma de decisiones estratégicas.....	85
3.5.1	Análítica predictiva y prescriptiva	85
3.5.2	Dashboards e interfaces inteligentes.....	87
3.5.3	Sistemas expertos modernos y soporte decisiones	89
3.6	Modelos de Negocio Emergentes Basados en Agentes de IA ...	90
3.6.1	AI as a service (AIaaS)	90
3.6.2	Agentes como asistentes y empleados corporativos virtuales	92
3.6.3	Ecosistemas inteligentes y redes de agentes interconectados.	93
3.7	Indicadores de impacto empresarial y retorno de inversión	95
3.7.1	Medición de impacto en la productividad operativa	95
3.7.2	Incremento de la rentabilidad financiera.....	96
3.7.3	Metodologías para el cálculo del retorno de inversión (ROI) en proyectos de IA.....	98
CAPÍTULO IV.....		99
AGENTES DE INTELIGENCIA ARTIFICIAL EN LA CIBERSEGURIDAD.....		99
4.1.	Introducción a la Ciberseguridad.....	99
4.1.1	Conceptos fundamentales de Ciberseguridad.....	100
4.1.2	Amenazas digitales actuales.....	101
4.2.	Principios de seguridad de la información	105
4.2.1	Confidencialidad	105
4.2.2	Integridad	106
4.2.3	Disponibilidad	107
4.3	Aplicación de IA en Ciberseguridad.....	108
4.3.1	Detección de anomalías	109
4.3.2	Respuesta automatizada	109
4.3.3	Inteligencia de amenazas.....	111

4.4. Agentes IA para la protección de infraestructuras	112
4.4.1 Redes empresariales	112
4.4.2 Sistemas en la nube	113
4.4.3 Dispositivos IoT	114
4.5 Agentes para detección y respuesta a incidentes	116
4.5.1 SIEM inteligente	117
4.5.2 SOC automatizado	118
4.5.3 Threat Hunting	120
4.6. IA ofensiva y riesgos emergentes	121
4.6.1 Deepfakes	122
4.6.2 Phishing inteligente	123
4.6.3 Ingeniería social automatizada	124
4.7.1 ISO 27001	126
4.7.2 NIST Cybersecurity Framework	127
4.7.3 Protección de datos personales	129
BIBLIOGRAFIA	131

CAPITULO I

FUNDAMENTOS DE INTELIGENCIA ARTIFICIAL

1.1 Introducción y Marco General

1.1.1 Importancia estratégica de la Inteligencia Artificial en la sociedad actual

Es evidente que en la sociedad actual la inteligencia artificial ha dejado de ser un asunto exclusivo de los filmes de ciencia ficción o un campo limitado a los laboratorios académicos, y se ha convertido en una de las fuerzas tecnológicas más poderosas e influyentes en la historia del hombre. Actualmente, la IA funciona como una tecnología de propósito general; esto es, su influencia no se restringe a un solo sector económico, sino que se expande y modifica horizontalmente a la educación, la medicina, el transporte, el comercio internacional y la industria manufacturera (Cockburn, Henderson, & Stern, 2018). Por lo tanto, entender cómo funcionan estos sistemas ya no es una tarea exclusiva de los ingenieros en sistemas o programadores, sino que se ha vuelto un requisito analítico esencial para cualquier profesional que quiera dirigir organizaciones en este siglo.

Al analizar el escenario mundial, se entiende que lo que hace a la inteligencia artificial importante desde una perspectiva estratégica es su capacidad sin igual para procesar grandes cantidades de datos a velocidades sobrehumanas y obtener de ellos información útil. En un mundo en el que los sensores, las transacciones digitales y los teléfonos

móviles producen billones de datos por segundo, las habilidades humanas convencionales son completamente inadecuadas para examinar esta inmensa cantidad de datos. En esta línea, la inteligencia artificial disminuye de manera significativa el coste económico de realizar pronósticos acertados, lo cual posibilita que las organizaciones tomen decisiones estratégicas fundamentadas en evidencia matemática y no solo en simples intuiciones, como apuntan Agrawal et al. (2018). Por lo tanto, se infiere que el elemento crucial que define la competitividad, la innovación y la soberanía tecnológica de los países y corporaciones en tiempos contemporáneos es la implementación de estas tecnologías.

1.1.2 Impacto de la IA en la transformación digital

La inteligencia artificial se ha convertido en el verdadero impulsor y acelerador de la transformación digital de las empresas a medida que las organizaciones globales avanzan en sus procesos de modernización. Hace no tanto, digitalizar una empresa significaba, casi siempre, digitalizar los documentos en papel que tenía a archivos PDF, o bien crear una simple página web donde exponer sus productos. Sin embargo, el entorno empresarial actual demanda un cambio de paradigma mucho más radical, donde las herramientas digitales no se limitan a almacenar datos pasivamente, sino que también interactúan, aprenden y toman decisiones autónomas para optimizar los procesos de trabajo (Brynjolfsson & McAfee, 2014). La inteligencia artificial es exactamente la capa cognitiva que convierte los sistemas de información tradicionales en plataformas inteligentes y flexibles.

Así pues, el impacto real de la inteligencia artificial en la transformación digital se plasma en la completa reorganización de los modelos operativos de las empresas. Las organizaciones se piensan hoy en día a partir de algoritmos y de flujos de datos continuos que hacen posible responder de forma inmediata a las necesidades del mercado, en vez de apoyarse en estructuras burocráticas lentas y rígidas. Según Davenport y Westerman (2018), las empresas que logran combinar la transformación digital con capacidades avanzadas de IA no solo automatizan las tareas rutinarias, sino que además reinventan la experiencia del cliente y crean nuevas fuentes de ingresos. Así pues, la IA deja de ser un complemento tecnológico de segunda categoría para convertirse en el factor clave que asegura la agilidad y la supervivencia comercial en la economía digital.

1.1.3 Relación tripartita: Inteligencia Artificial, negocios y Ciberseguridad

Para comprender de forma completa cómo funcionan los mercados digitales actuales, es absolutamente esencial examinar el vínculo tripartito que hay entre la ciberseguridad, la inteligencia artificial y la administración de negocios, que es estrecho e inseparable. En el contexto de las empresas actuales, la inteligencia artificial posibilita que los ejecutivos optimicen la cadena de valor, pronostiquen tendencias en el consumo y personalicen los servicios con métodos que antes parecían inalcanzables. No obstante, esta enorme dependencia de los sistemas informáticos pone a las compañías ante riesgos digitales sin precedentes. Conforme las operaciones de negocio se vuelven más automatizadas e interconectadas, el área de ataque para los cibercriminales se incrementa

de manera exponencial, lo que hace que la protección de datos sea una prioridad estratégica para la alta dirección (Taddeo & Floridi, 2018).

En este complejo entorno, la inteligencia artificial desempeña un doble rol que es atractivo y arriesgado al mismo tiempo: actúa como el escudo defensivo más fuerte de las compañías, pero también como el arma más avanzada de los atacantes informáticos. Por un lado, las herramientas de ciberseguridad impulsadas por IA tienen la capacidad de examinar miles de millones de eventos por segundo, identificar irregularidades en la red y neutralizar ataques cibernéticos antes de que puedan perjudicar al patrimonio de la empresa. Por otro lado, según lo afirman Buczak y Guven (2015), los atacantes emplean modelos inteligentes con el fin de automatizar la elaboración de código dañino, sortear los antivirus y crear engaños de ingeniería social altamente persuasivos. En consecuencia, se puede sostener que las empresas actuales solo pueden ser sostenibles si equilibran la innovación comercial impulsada por inteligencia artificial con una postura proactiva y sólida en términos de ciberseguridad.

2.1 Evolución de la Inteligencia Artificial

2.1.1 Origen y fundamentos teóricos de la IA

Para poder comprender bien la naturaleza de la IA actual es completamente indispensable revisar sus orígenes teóricos, los cimientos filosóficos y matemáticos que permitieron que naciera. Ya mucho antes de que existieran las computadoras electrónicas modernas, pensadores y matemáticos de la antigüedad se preguntaban si se podía construir máquinas capaces de razonar como los seres humanos. Las bases de la

lógica formal establecidas por Aristóteles, más los desarrollos algebraicos de Gottfried Leibniz y George Boole, sentaron los fundamentos teóricos para demostrar que el pensamiento humano podía ser traducido a símbolos matemáticos y manipulado por reglas mecánicas (Russell & Norvig, 2020). Este hallazgo fue crucial, pues mostró que el razonamiento no era una cualidad mística, sino un cálculo que podía reproducirse artificialmente.

Fue en la década de 1930 cuando se dio el verdadero salto cualitativo a la era informática gracias al genial trabajo del matemático británico Alan Turing. Con la creación de su modelo teórico de la “Máquina Universal de Turing” se probó que un dispositivo mecánico era capaz de realizar cualquier cálculo matemático computable siempre que se le suministraran las instrucciones lógicas adecuadas (Turing, On Computable Numbers, with an Application to the Entscheidungsproblem, 1936). Años más tarde, en su famoso artículo de 1950, “Computing Machinery and Intelligence”, Turing planteó la histórica pregunta de si las máquinas eran capaces de pensar, diseñando el famoso “Juego de la Imitación” (conocido hoy como la Prueba de Turing) para evaluar si un ser humano era capaz de distinguir entre las respuestas de una computadora y las de una persona real (Turing, 1950). Estos aportes, que convirtieron a Turing en el verdadero padre teórico de la disciplina, establecieron la hoja de ruta que guiaría a la ciencia computacional durante los años posteriores.

Figura N° 1 Línea de tiempo del desarrollo de la IA



Nota. Fuente: Imagen creada con IA

2.1.2 Etapas de desarrollo y ciclos de invierno

En 1956 se produjo el nacimiento formal de la Inteligencia Artificial como disciplina científica independiente, durante la famosa Conferencia de Dartmouth organizada por jóvenes visionarios como John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon. En este hito de la historia no sólo se acuñó formalmente el término “Inteligencia Artificial”, sino que se planteó la hipótesis de que todo aspecto del aprendizaje o de la inteligencia podría describirse con tanta precisión que una máquina pudiera programarse para simularlo (McCarthy & al., 2006). A partir de este encuentro, se abrió una época de oro llena de un enorme entusiasmo en la que surgieron los primeros sistemas inteligentes capaces de resolver teoremas matemáticos o jugar al ajedrez, lo que hizo que los científicos prometieran que en menos de dos décadas las máquinas serían capaces de superar al ser humano en todas sus capacidades.

Pero las expectativas desmedidas chocaron pronto con las duras realidades de las limitaciones tecnológicas de la época, dando origen a los oscuros periodos conocidos como los "Inviernos de la IA". En las décadas de 1970 y 1980, la falta de poder de procesamiento en las computadoras y la escasez de datos impidieron que los proyectos cumplieran con las ambiciosas promesas hechas a los gobiernos e inversionistas, lo que resultó en recortes masivos de financiamiento y en un descrédito generalizado de la disciplina (Crevier, 1993). La recuperación definitiva del campo no se produjo hasta finales del siglo XX y principios del XXI, gracias a la aparición del Machine Learning, al incremento exponencial de la capacidad de cómputo y al nacimiento de internet, factores que permitieron pasar de los rígidos sistemas basados en reglas a modelos estadísticos capaces de aprender automáticamente a partir de datos reales.

2.1.3 La era de la Inteligencia Artificial generativa

Hoy estamos viviendo el episodio más disruptivo y veloz en la historia de la computación: la cristalización de la Era de la Inteligencia Artificial Generativa. Por décadas, los sistemas de IA se enfocaron principalmente en analizar, clasificar o predecir patrones de datos existentes, como determinar si un correo electrónico es spam o reconocer una cara en una foto. Sin embargo, la llegada de los modelos fundacionales y de las arquitecturas de aprendizaje profundo, cambiaron por completo el panorama, permitiendo que las computadoras no sólo comprendan el mundo, sino que sean capaces de crear contenido totalmente original en

forma de texto, imágenes, código de programación, audio y vídeo con una calidad indistinguible de la producción humana (Vaswani & al., 2017).

Esta nueva era ha sido impulsada principalmente por la invención de la arquitectura Transformer en 2017 y el desarrollo de Modelos de Lenguaje de Gran Escala (LLM), que han democratizado el acceso a la tecnología a un ritmo jamás visto en la historia de la humanidad. Actualmente, millones de personas y empresas interactúan a diario con la inteligencia artificial mediante el uso del lenguaje natural, saltándose las barreras técnicas de la programación tradicional. Como afirman Bommasani et al. (2021), la IA generativa no es una mera mejora incremental del software, sino un nuevo paradigma de infraestructura cognitiva sobre el cual se están reescribiendo los flujos de trabajo de la industria, la educación y la ciencia a nivel mundial.

3.1 Conceptos fundamentales y clasificación de la IA

3.1.1 Definición formal de la Inteligencia Artificial

Cuando se intenta establecer un marco conceptual riguroso en las ciencias de la computación, definir la inteligencia artificial puede ser, sorprendentemente, una labor muy complicada. Esto es así porque el concepto de "inteligencia" en los seres humanos sigue siendo objeto de acalorados debates a nivel filosófico y biológico. La inteligencia artificial, en términos estrictamente académicos e industriales, se define como la disciplina de la informática que tiene la responsabilidad de crear, elaborar y establecer sistemas informáticos (software y hardware) que

puedan llevar a cabo tareas que requerirían procesos cognitivos sofisticados si las realizara una persona (Russell & Norvig, 2020). Estas habilidades incluyen la solución de problemas abstractos, el entendimiento del lenguaje natural, la identificación de patrones visuales complejos, el aprendizaje a partir de experiencias pasadas y la capacidad autónoma para tomar decisiones en situaciones caracterizadas por la incertidumbre.

Desde el punto de vista práctico de la ingeniería y los negocios, la definición formal de IA ha cambiado desde el viejo concepto de "imitar el pensamiento humano" hasta una perspectiva más pragmática que se refiere a "actuar racionalmente". Esto quiere decir que un sistema inteligente contemporáneo no busca replicar los sesgos emocionales y las imperfecciones de la biología del cerebro humano, sino más bien procesar la información existente para determinar qué acción es la más adecuada para maximizar las posibilidades de alcanzar una meta específica. Según Poole y Mackworth (2017), la inteligencia artificial contemporánea debe ser interpretada como la ciencia que se encarga de diseñar agentes inteligentes autónomos, con la capacidad de captar su entorno por medio de sensores digitales y realizar acciones en el mismo a través de actuadores lógicos, para solucionar dificultades técnicas y comerciales muy complejas.

3.1.2 Delimitación conceptual

En el lenguaje cotidiano y el marketing corporativo contemporáneo, es frecuente observar una utilización indiscriminada e incorrecta de una variedad de términos tecnológicos, lo que lleva a confundir la automatización tradicional con la inteligencia artificial adaptativa y la IA generativa con la discriminativa. La automatización clásica, que se basa en reglas deterministas escritas de forma manual por un programador, es un sistema inflexible del tipo "si ocurre X, ejecuta Y" sin capacidad de aprendizaje. Si un sistema automatizado recibe información con un formato inesperado, se interrumpirá inmediatamente o generará un error. En cambio, la inteligencia artificial adaptativa utiliza modelos probabilísticos para manejar datos no estructurados y modifica de manera autónoma sus parámetros a medida que recibe nueva información, sin necesitar que el ser humano intervenga en la alteración del código fuente (Jordan & Mitchell, 2015).

Por otra parte, es fundamental distinguir entre la IA generativa y la IA discriminativa (o predictiva) dentro del propio ecosistema de la inteligencia artificial. La inteligencia artificial discriminativa se ha creado sobre todo para analizar, clasificar o etiquetar un conjunto de datos ya existente; por ejemplo, para determinar si una operación financiera es fraudulenta o para calcular la posibilidad de que un cliente cancele un servicio. En cambio, la inteligencia artificial generativa no se restringe a clasificar datos ya existentes, sino que identifica la distribución probabilística inherente a ellos con el fin de generar información completamente nueva, manteniendo las propiedades estadísticas del conjunto original (Goodfellow, Bengio, & Courville, 2016). Al comprender esta barrera conceptual, las organizaciones tienen

la capacidad de seleccionar la herramienta apropiada para cada necesidad operacional y así evitar que surjan expectativas tecnológicas falsas.

3.1.3 Capacidades cognitivas principales de los sistemas inteligentes

Es necesario que un sistema informático exhiba una serie de habilidades cognitivas interconectadas que emulen los procesos elevados de la mente para que pueda ser considerado de manera legítima parte del modelo de inteligencia artificial. La capacidad inicial es el aprendizaje continuo, que se refiere a la posibilidad de que el software ajuste sus parámetros internos al ser expuesto a datos nuevos sin necesidad de reprogramarlo. Después, se encuentra la habilidad de razonamiento lógico, que permite al sistema llegar a conclusiones válidas, deducir relaciones de causa y efecto y solucionar problemas usando bases de conocimiento organizadas o normas de inferencia estadística (Nilsson, 2014).

Asimismo, la percepción multimodal y la toma de decisiones contribuyen a definir el perfil cognitivo de la inteligencia artificial actual. La percepción posibilita que las computadoras manejen e interpreten señales provenientes del mundo real, como por ejemplo imágenes, audio, texto escrito y datos de sensores de la industria. Esta transformación convierte las señales analógicas en representaciones matemáticas abstractas. Finalmente, la toma de decisiones proporciona al sistema la autonomía requerida para seleccionar el camino más apropiado en un espacio de estados complicado. La IA es tan revolucionaria en la economía digital gracias a esta combinación de habilidades, sumada a la reciente facultad

de crear contenido en muchos formatos por su propia cuenta, lo que hace posible el desarrollo de agentes que pueden operar en entornos laborales complicados con poca supervisión humana (LeCun, Bengio, & Hinton, 2015).

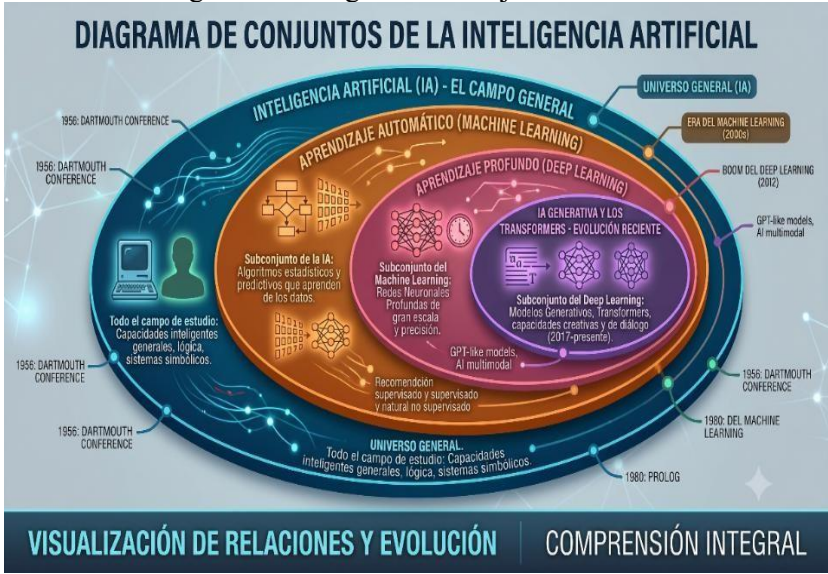
3.1.4 Taxonomía y niveles de desarrollo de la IA

Según la ciencia de la computación, el progreso de la inteligencia artificial se divide en tres niveles jerárquicos o fases evolutivas claramente definidas, y la primera es la Inteligencia Artificial Estrecha o Débil (ANI - Artificial Narrow Intelligence). La ANI se refiere a todos los sistemas inteligentes que existen y funcionan hoy en día; estos son modelos creados, entrenados y perfeccionados para llevar a cabo una sola tarea concreta o un grupo muy limitado de funciones, como diagnosticar imágenes médicas, traducir escritos o jugar al ajedrez. Aunque un sistema de ANI tiene la capacidad de superar ampliamente el rendimiento humano en su campo específico, no puede trasladar su conocimiento a otra tarea; por ejemplo, un modelo de diagnóstico médico no está capacitado para responder preguntas relacionadas con contabilidad o tener una conversación lógica sobre deportes (Goertzel, 2014).

La Inteligencia Artificial General o Fuerte (AGI) es el segundo nivel en esta clasificación teórica. Esta describe una condición futura del progreso tecnológico en la que una máquina tendrá habilidades cognitivas que se asemejan a las de un ser humano en todos los campos del saber. Una AGI podría llevar a cabo razonamientos abstractos, aprender cualquier disciplina sin requerir entrenamiento previo, ajustarse a circunstancias inesperadas y tener conciencia de sí misma. Finalmente, la idea de la

Superinteligencia artificial (ASI - por sus siglas en inglés) hace referencia a un periodo que sigue a la AGI, cuando el intelecto de las máquinas excederá en varios órdenes de magnitud la inteligencia acumulada de todos los seres humanos en campos como resolver problemas científicos, tener sabiduría general y ser creativos. La comunidad científica global tiene que investigar desde ahora los mecanismos de alineamiento y seguridad de estos sistemas, ya que, como advierte Bostrom (2014), el paso de la AGI a la ASI podría suceder mediante un proceso acelerado de automejora recursiva.

Figura N°2 Diagrama de conjuntos de la IA



Nota. Fuente: Imagen creada con IA

4.1 Fundamentos del aprendizaje automático (Machine Learning)

4.1.1 Concepto y filosofía del aprendizaje automático

El Machine Learning (ML), o aprendizaje automático, es la columna vertebral sobre la que descansa toda la IA moderna, y es una forma completamente nueva de escribir software. En la programación convencional, los humanos ingresaban datos de entrada y tenían que escribir las reglas lógicas para que la computadora pudiera calcular una respuesta o salida. La filosofía del Machine Learning da la vuelta a la ecuación: los ingenieros introducen en la máquina los datos de entrada y las respuestas deseadas, y ésta descubre por sí misma mediante algoritmos matemáticos las reglas lógicas que relacionan ambas magnitudes (Mitchell, 1997).

Esta nueva filosofía permite resolver problemas de gran complejidad donde la lógica es demasiado compleja para ser codificada a mano por un ser humano. No es necesario que el programador tenga que basarse en su intuición, sino que el aprendizaje automático se encarga de encontrar patrones ocultos en grandes cantidades de datos mediante la estadística computacional. Tal y como bien argumenta Murphy (2012), el Machine Learning debería ser conceptualizado como un conjunto de métodos para encontrar patrones en los datos de forma automática, y luego utilizar esos patrones no descubiertos para predecir datos futuros o para tomar decisiones bajo condiciones de incertidumbre, siendo la herramienta analítica más potente del entorno corporativo contemporáneo.

4.1.2 Ciclo de vida del aprendizaje basado en datos

Una organización no logra una correcta implementación de un modelo de aprendizaje automático simplemente ejecutando un algoritmo

matemático, sino que debe seguir un ciclo de vida estructurado y disciplinado a través de diversas etapas críticas. La primera fase del ciclo es la definición del problema de negocio, donde se deja claro qué métrica comercial se quiere optimizar. A esta etapa le sigue de forma obligatoria la recolección y preparación de datos, un proceso que suele consumir hasta el 80% del tiempo total del proyecto e incluye la limpieza de datos ruidosos, el manejo de valores faltantes y la eliminación de registros duplicados que podrían distorsionar el entrenamiento (Géron, 2019).

Cuando la información ya está limpia, se pasa a la fase de ingeniería de características (feature engineering), donde los especialistas convierten los datos raw en representaciones numéricas para que el modelo pueda extraer patrones con facilidad. Posteriormente, se divide los datos en conjuntos independientes de entrenamiento, validación y prueba, evitando que el sistema memorice los datos en vez de aprender patrones generales (fenómeno conocido como sobreajuste o overfitting). El ciclo termina con la fase de entrenamiento, evaluación de métricas (precisión, recall, F1-score) y despliegue en producción, proceso que en las empresas modernas se realiza de manera continua a través de las metodologías de MLOps para monitorear que el modelo no pierda rendimiento con el tiempo (Kreuzberger & al., 223).

4.1.3 Aprendizaje supervisado

En el aprendizaje supervisado, el algoritmo se entrena con un conjunto de datos que incluye tanto las entradas como las salidas deseadas. Con

este ejemplo, el modelo aprende a asociar las entradas con las salidas correctas, lo que permite que, una vez entrenado, pueda predecir la salida para nuevas entradas que nunca antes ha visto. El aprendizaje supervisado es la modalidad más usada y, con gran diferencia, la más madura comercialmente dentro del ecosistema del Machine Learning. En este entrenamiento, el algoritmo es alimentado con un conjunto de datos que ha sido previamente “etiquetado” por expertos humanos, esto es, cada registro de entrada viene acompañado de la respuesta correcta o variable a explicar. El trabajo del algoritmo será encontrar la función matemática que mejor aproxime la relación entre los datos de entrada y la etiqueta de salida, de manera que cuando el modelo reciba datos completamente nuevos en el futuro sea capaz de predecir la etiqueta correcta con un alto porcentaje de precisión (Hastie & al., 2009).

En el aprendizaje supervisado hay dos grandes familias de problemas técnicos: los problemas de clasificación y los problemas de regresión. Los algoritmos de clasificación se emplean cuando la variable objetivo es discreta o categórica, como determinar si un correo es “spam” o “no spam”, o identificar si una imagen pertenece a una categoría específica. La regresión, por su parte, se aplica cuando se trata de predecir un valor numérico continuo, como el precio de venta de una propiedad inmobiliaria o la demanda futura de energía eléctrica en una ciudad. James et al. (2021) señalaron que los algoritmos clásicos como la Regresión Logística, los Árboles de Decisión, las Máquinas de Vectores de Soporte (SVM) y los Bosques Aleatorios (Random Forests) son la columna vertebral de la toma de decisiones analíticas en el sector bancario, industrial y de seguros a nivel mundial.

4.1.4 Aprendizaje no supervisado

A diferencia del modelo supervisado que acabamos de ver, el aprendizaje no supervisado tiene que resolver el problema de tratar con conjuntos de datos que no disponen de etiquetas, ni de respuestas anteriores de ningún tipo. La idea detrás de este enfoque es permitir que el algoritmo busque totalmente por su cuenta la estructura subyacente de la información, buscando patrones, agrupaciones naturales, anomalías o relaciones ocultas que ni siquiera conocían los analistas humanos antes (Bishop, 2006). Es una herramienta analítica extraordinariamente potente para tratar el enorme volumen de datos no estructurados que generan las corporaciones modernas en Internet.

La técnica más importante dentro de esta categoría es el clustering o agrupamiento, que consiste en dividir un conjunto de datos heterogéneo en grupos o clusters más pequeños de forma que los elementos que conforman un grupo sean muy similares entre sí y muy diferentes con los elementos de otros grupos. En el ámbito empresarial, el clustering se aplica ampliamente para la segmentación inteligente de clientes, lo que permite a las compañías detectar distintos patrones de consumo y elaborar estrategias de marketing totalmente personalizadas. Una aplicación vital adicional es la reducción de dimensionalidad (mediante algoritmos como PCA), que permite simplificar conjuntos de datos gigantescos con cientos de variables, sin perder información relevante, facilitando la visualización de datos y optimizando la velocidad de cálculo de otros sistemas (Aggarwal, 2015).

Figura N°3 Flujo de MLOPS y entrenamiento de modelos



Nota. Fuente: Imagen creada con IA

4.1.5 Aprendizaje por refuerzo

El aprendizaje por refuerzo (Reinforcement Learning - RL) constituye un paradigma completamente diferente al supervisado y al no supervisado, tomando como referencia directa la psicología del conductismo y la manera en que los seres vivos aprenden interactuando con su entorno mediante prueba y error. En este esquema no existen datos estáticos ni etiquetas, sino un agente inteligente que toma acciones dentro de un entorno dinámico con el fin de alcanzar una meta específica. Cada vez que el agente ejecuta una acción, el entorno cambia de estado y emite una señal de retorno en forma de recompensa o castigo dependiendo de si la acción fue positiva o negativa respectivamente, enseñándole al sistema qué decisiones debe tomar a largo plazo (Sutton & Barto, 2018).

El objetivo matemático de un agente en el aprendizaje por refuerzo es encontrar una estrategia o política de comportamiento que maximice la suma total de recompensas que reciba con el transcurso del tiempo. Para

lograrlo, el sistema tiene que estar resolviendo continuamente el dilema fundamental entre la exploración (probar nuevas acciones para ver si son mejores) y la explotación (repetir las acciones que sabe que dan recompensas). Este paradigma, tal como lo argumentan Mnih et al. (2015), se ha revelado como la herramienta perfecta para resolver problemas complejos de control en tiempo real, con aplicaciones destacadas en el pilotaje de vehículos autónomos, la optimización de rutas logísticas en las cadenas globales de suministro, el diseño de estrategias de comercio financiero de alta frecuencia y el control automático de brazos robóticos en la industria manufacturera.

4.1.6 Retos y limitaciones del Machine Learning

Por muy potente que sea el aprendizaje automático y por muy de moda que esté actualmente, es importante que tengamos una actitud técnica madura para reconocer los importantes retos, sesgos y limitaciones de estos modelos. La primera gran limitación es que el modelo depende completamente de la calidad de los datos de entrenamiento. Si los datos de entrada están incompletos, desactualizados o incluyen sesgos históricos de discriminación social, el modelo aprenderá y amplificará esos errores de forma sistemática, fenómeno conocido en las ciencias de la computación como la regla “Garbage In, Garbage Out” (GIGO) (Barocas & al., 2023).

El segundo reto técnico de gran escala es el equilibrio entre el sobreajuste y el subajuste. Un modelo que es demasiado complejo puede sufrir sobreajuste, “aprendiendo” los datos de entrenamiento con tanta

precisión que pierde toda la capacidad de generalizar a nuevos datos en producción, mientras que el subajuste ocurre cuando el modelo es demasiado simple para capturar la estructura básica del problema. Por último, como indican Doshi-Vélez y Kim (2017), la falta de explicabilidad e interpretabilidad de muchos algoritmos avanzados plantea serios dilemas éticos y jurídicos en sectores altamente regulados como la banca y la salud, donde una empresa no puede limitar su respuesta a decir que “el algoritmo lo decidió así”, obligando a la industria a invertir activamente en el desarrollo de metodologías de IA Explicable (XAI).

5.1 Deep Learning y redes neuronales artificiales

5.1.1 Introducción al aprendizaje profundo

Deep Learning, o aprendizaje profundo, es un área específica de Machine Learning que ha revolucionado las habilidades de la informática en los últimos diez años. Los ingenieros tenían que invertir semanas o meses en la elaboración manual de las variables o características representativas (feature engineering) que los algoritmos tradicionales de aprendizaje automático necesitaban analizar. La habilidad del Deep Learning de ejecutar este procedimiento de extracción de características con total automatización y en orden jerárquico, a partir de datos no estructurados en su estado original (por ejemplo, señales de voz, textos masivos, imágenes o archivos de video), es la razón del salto revolucionario que ha dado (Goodfellow, Bengio, & Courville, 2016).

Esta capacidad de aprendizaje directo desde los datos en bruto ha sido la razón que ha propiciado el triunfo comercial de la inteligencia artificial contemporánea en campos donde los procedimientos tradicionales fracasaban constantemente. La arquitectura de sus modelos, que se componen de varias capas interconectadas de procesamiento matemático, es la causa por la cual este paradigma recibe el nombre de "profundo". El sistema va aprendiendo a representar los datos en niveles de abstracción superiores y más complejos, conforme la información se desplaza hacia abajo por estas capas. Las capas iniciales de una red, como indican LeCun et al. (2015), son capaces de detectar cambios de color o bordes simples en una imagen; las intermedias, por su parte, pueden identificar texturas o formas geométricas; y las finales tienen la capacidad de reconocer con una precisión sobrehumana objetos complejos, conceptos semánticos abstractos o rostros.

5.1.2 Inspiración biológica de las redes neuronales

Es fundamental estudiar la estructura biológica del cerebro humano, que es la fuente de inspiración inicial que originó esta disciplina para entender el diseño arquitectónico de los modelos de Deep Learning. El sistema nervioso central de los organismos funciona a través de billones de células especializadas llamadas neuronas. Estas se comunican entre ellas por medio de señales electroquímicas que circulan por el cuerpo celular, las dendritas y el axón, unidas a través de puntos de contacto conocidos como sinapsis. Warren McCulloch y Walter Pitts, quienes fueron científicos, sugirieron convertir este proceso biológico en un modelo matemático más sencillo durante las décadas de 1940 y 1950.

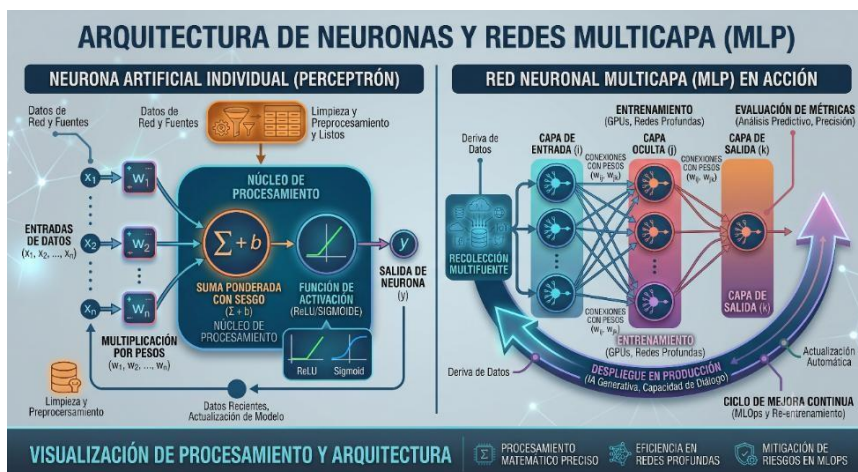
Posteriormente, Frank Rosenblatt desarrolló el Perceptrón (McCulloch & Pitts, 1943) (Rosenblatt, 1958).

No obstante, evitar caer en antropomorfismos simplistas y conservar una distancia analítica clara es esencial. A pesar de que la neurona artificial se asemeja conceptualmente a la recepción de varias entradas, su procesamiento ponderado y la emisión de una señal de salida, las Redes Neuronales Artificiales (ANN) actuales son estructuras fundamentalmente matemáticas fundamentadas en el álgebra lineal, el cálculo diferencial y la teoría probabilística. Su procedimiento de aprendizaje, que se basa en el ajuste iterativo de matrices numéricas a través de algoritmos de optimización, es muy distinto al funcionamiento complejo y apenas comprendido de la plasticidad del cerebro humano (Agrawal & al., 2018). Además, no tienen conciencia, emociones ni procesos metabólicos.

5.1.3 Componentes matemáticos de una red neuronal

El funcionamiento interno de una red neuronal artificial se basa en un grupo de elementos matemáticos meticulosamente coordinados que convierten vectores numéricos de entrada en pronósticos de salida. El Perceptrón o neurona artificial es la unidad elemental de procesamiento, y se encarga de recibir una serie de señales numéricas de entrada. Luego, cada señal se multiplica por un valor cuantitativo conocido como peso (w) y a continuación se le añade una constante denominada sesgo (b).

Figura N°4 Arquitectura de neuronas y redes multicapa



Nota. Fuente: Imagen creada con IA

Los pesos denotan la fuerza o relevancia de cada entrada en el modelo, y el sesgo hace posible que se mueva la función de decisión para modificar el aprendizaje de un modo flexible (Haykin, 2009). Después de sumar las entradas con ponderación, el resultado pasa por una función matemática no lineal llamada función de activación (por ejemplo, Sigmoide, Tanh o ReLU). Incluir funciones no lineales es un elemento fundamental; de lo contrario, una red neuronal de mil capas tendría la misma capacidad matemática que un modelo lineal sencillo y no podría resolver problemas complejos de la realidad. El algoritmo de Propagación Hacia Atrás (Backpropagation) es el que permite la enseñanza en red. Este realiza una estimación del error que hay entre el valor real y la predicción emitida, y a través de un cálculo de derivadas parciales (el gradiente) hace una corrección hacia atrás de todos los pesos de manera iterativa; así va disminuyendo paulatinamente el fallo global del sistema (Rumelhart & al, 1986).

5.1.4 Arquitecturas modernas de Deep Learning

La investigación científica reveló con el tiempo que una red neuronal genérica no era capaz de procesar todos los tipos de datos, lo cual motivó la creación de arquitecturas específicas en función del problema. Las Redes Neuronales Convolucionales (CNN) se crearon para trabajar con información visual y espacial, como imágenes y videos. Las CNN emplean convoluciones, que son operaciones matemáticas sofisticadas, para implementar filtros en las imágenes. Esto posibilita la identificación de rasgos visuales sin importar la posición o rotación de los elementos dentro del encuadre, lo cual ha transformado áreas como el diagnóstico médico mediante imágenes y la visión por computadora (Krizhevsky & al, 2017).

Por otra parte, se crearon las redes neuronales recurrentes (RNN) y las redes LSTM (Long Short-Term Memory) para manejar datos que dependen del tiempo y son secuenciales, como el texto escrito, las series financieras o las señales de audio. Estas redes tienen conexiones con memoria interna que posibilitan el procesamiento de la información en periodos sucesivos (Hochreiter & Schmidhuber, 1997). No obstante, la creación en 2017 de la arquitectura Transformer, que desechó el método recurrente y presentó el "mecanismo de atención", representó un progreso más inquietante para el futuro del Deep Learning. Esta innovación facilita que el modelo procese secuencias enteras de información en paralelo y, al mismo tiempo, analice las conexiones contextuales entre palabras distantes. Así se convierte en la base tecnológica para construir los grandes modelos lingüísticos contemporáneos (Vaswani & al., 2017).

6.1 Procesamiento del lenguaje natural (NLP) y modelos generativos

6.1.1 Definición y alcance del procesamiento del lenguaje natural

El Procesamiento del Lenguaje Natural (PLN o NLP, por sus siglas en inglés) es una subdisciplina clave en la intersección de la inteligencia artificial, la lingüística computacional y la ciencia de datos, cuyo objetivo principal es permitir que las computadoras entiendan, interpreten, procesen y generen lenguaje humano tanto en su forma escrita como oral. Desde la época de las computadoras de tarjetas perforadas, la comunicación con las máquinas ha sido una ardua tarea para los humanos, que debían aprender lenguajes de programación rígidos y estructurados. El NLP viene a darle la vuelta a esta historia al convertir a las lenguas naturales (como el español, inglés o mandarín) en la interfaz directa de comunicación entre el usuario y los sistemas informáticos (Jurafsky & Martin, 2021).

El campo técnico del NLP cubre dos grandes dominios conectados entre sí: la Comprensión del Lenguaje Natural (NLU – Natural Language Understanding) y la Generación del Lenguaje Natural (NLG – Natural Language Generation). NLU está centrada en resolver la complejidad, ambigüedad y variabilidad abrumadora del lenguaje humano, abordando tareas como la extracción de entidades nombradas, el análisis de sentimientos, la detección de intenciones, así como el procesamiento sintáctico y semántico. NLG, por su parte, estructura ideas y traduce las representaciones numéricas internas de la computadora en texto en lenguaje humano, coherente, fluido y contextualmente apropiado. Como

indican Manning y Schütze (1999), comprender el NLP se trata de superar el reto de la polisemia, el sarcasmo, las diferencias de dialecto, y la dependencia del contexto cultural, lo que representa la prueba suprema para la IA.

6.1.2 Modelos de lenguaje a gran escala (LLM)

Por completo, el desarrollo más reciente del NLP ha sido dominado por la consolidación de los Modelos de Lenguaje a Gran Escala (LLM). Un modelo de lenguaje es, en su forma más pura desde el punto de vista matemático, una distribución de probabilidad sobre secuencias de palabras; su objetivo fundamental es predecir cuál es la palabra o “token” más probable que debe venir a continuación dentro de un texto, dado todo el contexto anterior. El cambio cualitativo que originó los actuales LLMs fue la escala: al entrenar redes neuronales tipo Transformer con miles de millones de parámetros sobre porciones gigantescas de la información contenida en internet, los modelos empezaron a mostrar capacidades cognitivas emergentes de razonamiento, resumen, traducción y generación de código informático que no habían sido explícitamente programadas (Brown & al., 2020).

El ciclo de vida técnico de un LLM moderno se compone de varias etapas críticas. Primero, se hace un preentrenamiento masivo no supervisado, en el que la red procesa billones de palabras para aprender la estructura, la gramática y los conocimientos del mundo. A esta etapa le sigue un proceso de ajuste fino supervisado (Fine-Tuning) con datos curriculares curados por expertos, y finalmente se emplean técnicas de Aprendizaje por Refuerzo a partir de Retroalimentación Humana (RLHF) que hacen

que el comportamiento del modelo se alinee con los valores éticos humanos, la seguridad de la información y la utilidad para el usuario final (Ouyang & al, 2011). De esta manera los LLM se transforman en plataformas de procesamiento de conocimiento general al servicio de la sociedad y la industria.

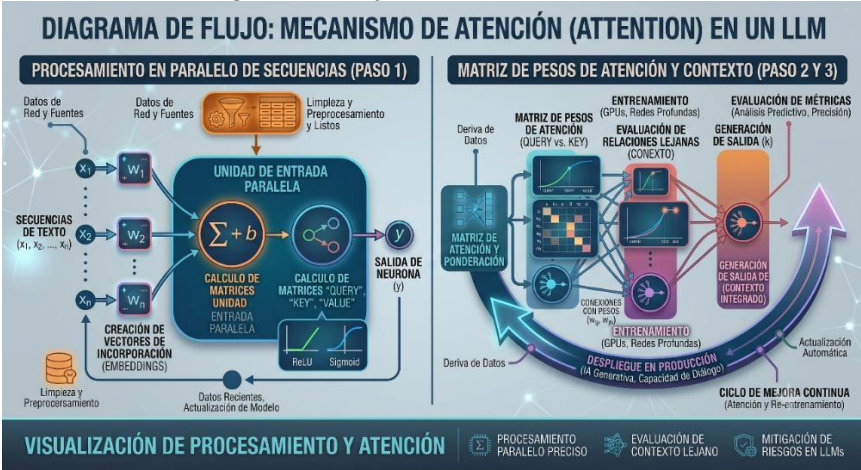
6.1.3 Modalidades de la IA Generativa

Si bien el procesamiento de texto mediante modelos LLM ha monopolizado la atención masiva, el ecosistema de IA Generativa involucra múltiples modalidades sensoriales e informáticas, las cuales están redefiniendo la producción cultural y comercial. La inteligencia artificial generativa multimodal emplea embeddings o representaciones vectoriales compartidas entre los distintos formatos, traduciendo los unos a otros de forma fluida, de tal forma que una instrucción escrita puede dar lugar a contenido en formato de imagen, sonido, código informático, video tridimensional, entre otros (Ramesh & al., 2022).

En el ámbito de la generación de imágenes, el desarrollo ha sido transformado por los Modelos de Difusión, que aprenden a reconstruir estructuras complejas mediante la destrucción progresiva del ruido gaussiano añadido a un archivo visual, sintetizando imágenes fotorrealistas extremas a partir de texto. Los algoritmos en el modo de audio no solo permiten la transcripción de voz a texto con alta fidelidad (speech-to-text), sino también la síntesis y clonación de voz humana con inflexiones emocionales naturales. Finalmente, en el caso de los vídeos, las redes generativas profundas combinan la coherencia temporal de los

fotogramas con la física simulada para crear secuencias cinematográficas completas a partir de descripciones textuales, lo que está transformando radicalmente el diseño, el entretenimiento y el marketing digital, como señalan Brooks et al (2024).

Figura N°5
Diagrama de flujo: Mecanismo de atención en un LLM



Nota. Fuente: Imagen creada con IA

6.1.4 Aplicaciones de NLP en negocios y Ciberseguridad

Las aplicaciones prácticas de las tecnologías de procesamiento del lenguaje natural y los modelos generativos en la vida corporativa y en la ciberseguridad actual son amplias e inmediatas. En el mundo empresarial, el NLP permite automatizar el análisis de miles de contratos legales, extraer los datos más importantes de facturas o informes financieros no estructurados, y diseñar sistemas de atención al cliente

disponibles permanentemente. Las arquitecturas de Generación Aumentada por Recuperación (RAG - Retrieval-Augmented Generation) se integran para conectar los LLM a las bases de conocimiento internas de la compañía, evitando así revelar datos confidenciales al exterior y asegurando respuestas estrictamente basadas en la documentación verificada de la empresa (Lewis & al, 2020).

En el terreno de la ciberseguridad, el NLP es un frente crítico tanto de defensa como de ataque. Desde el punto de vista defensivo, los modelos de procesamiento del lenguaje natural analizan continuamente el tráfico de correo electrónico institucional y los registros de chat corporativos para identificar campañas sofisticadas de ingeniería social, detectar la exfiltración inapropiada de secretos industriales o analizar firmas lingüísticas de malware. Desde el punto de vista de las amenazas, los delincuentes de la ciberdelincuencia usan los LLMs para automatizar la escritura de correos de spear-phishing altamente personalizados en varios idiomas, in errores gramaticales tradicionales, o para encontrar vulnerabilidades en el código fuente a través de auditorías automatizadas. Por eso, la aplicación del NLP en ciberseguridad requiere un seguimiento técnico permanente y la aplicación de estrictas políticas de gestión de la informática.

7.1 Ecosistema tecnológico de la inteligencia artificial

7.1.1 Los Datos como activo estratégico y combustible de la IA

En la economía digital actual se suele decir que “los datos son el nuevo petróleo”, pero dentro del ecosistema de la inteligencia artificial esta metáfora es incompleta si no se contempla la calidad y el tratamiento de esos datos. Los datos son el combustible principal que necesitan los algoritmos de Machine Learning y los modelos de Deep Learning para funcionar. A diferencia de los activos tangibles tradicionales, que se deprecian con el uso, los datos son un activo estratégico no rival que se vuelve más valioso cuanto más se integran, analizan y reutilizan en los modelos de decisión de la empresa (Provost & Fawcett, 2013).

Los datos sólo pueden aportar un verdadero valor estratégico para la organización, si se van más allá de la mera acumulación masiva de información (data hoarding) y se desarrollan políticas disciplinares de Gobernanza de Datos. La gobernanza de datos incluye la definición de estándares de calidad, la curaduría, el etiquetado adecuado, la gestión de metadatos y la aplicación de protocolos de privacidad y seguridad desde el diseño. Davenport y Westerman (2018) señalan que las corporaciones que consiguen consolidar una cultura verdaderamente orientada a datos consiguen convertir sus bases de datos en ventajas competitivas sostenibles, permitiendo que sus modelos de inteligencia artificial detecten tendencias del mercado, prevengan la fuga de clientes y optimicen operaciones en tiempo real con una precisión matemática imposible de alcanzar para la competencia tradicional.

7.1.2 Infraestructura Hardware y computacional

El vertiginoso avance de la inteligencia artificial de los últimos años no habría sido posible sin una transformación paralela de la arquitectura del hardware y de la capacidad de procesamiento informático. Entrenar modelos con miles de millones de parámetros implica realizar billones de operaciones matemáticas por segundo, lo que sobrepasa por mucho las capacidades de las procesadoras tradicionales (CPU). Esta limitación física impulsó la adopción masiva de las Unidades de Procesamiento Gráfico (GPU) y el posterior desarrollo de Unidades de Procesamiento Tensorial (TPU) e integrados ASIC diseñados específicamente para acelerar el cálculo matricial y el álgebra lineal subyacente al Deep Learning (Jouppi & al, 2017).

A este nivel de especialización de hardware se une la infraestructura del Cloud Computing y del Edge Computing. La computación en la nube abrió el acceso a supercomputadoras y clusters de GPUs al permitir que los startups y las pequeñas empresas pudieran alquilar potencia de cálculo bajo demanda, sin necesidad de realizar inversiones de capital millonarias en sus propios Data Centers. El Edge Computing o computación en el borde desplaza, complementariamente, la inferencia de los modelos de IA desde servidores centrales distantes directamente a los dispositivos finales (teléfonos móviles, cámaras de seguridad, sensores industriales IoT y vehículos autónomos). De acuerdo con Shi et al. (2016), esta arquitectura distribuida reduce la latencia de las respuestas, el consumo de ancho de banda y la privacidad de los datos procesando la información localmente en lugar de enviarla a la nube.

7.1.3 Principales actores y plataformas de IA

En el contexto mundial, el avance de la inteligencia artificial está dominado por una feroz rivalidad entre un puñado de gigantes tecnológicos, institutos de investigación académica y compañías especializadas que están a la cabeza en el desarrollo de los modelos base más sofisticados. Empresas como OpenAI, Google (DeepMind), Microsoft, Anthropic, Meta o Amazon concentran inversiones multimillonarias en el desarrollo de la IA, compitiendo por avanzar en la frontera de la IA mediante el desarrollo de arquitecturas multimodales, la optimización de los parámetros de sus modelos o su eficiencia energética (Mazzucato, 2021).

Esta carrera tecnológica no se limita a la capacidad pura de los modelos de lenguaje o visión, sino que se desarrolla en la creación de ecosistemas integrados donde los desarrolladores pueden utilizar las capacidades de IA mediante APIs, plataformas en la nube (como Microsoft Azure, Google Cloud Platform y AWS) y marcos de desarrollo de código abierto. El ejemplo de la estrategia de Meta al liberar modelos de la familia LLaMA ha impulsado el desarrollo de un vibrante ecosistema bajo licencia abierta, permitiendo que investigadores y empresas adapten modelos de lenguaje de alto rendimiento en sus propios servidores privados, sin necesidad de depender de APIs propietarias. Según el análisis de Lee (2018), esta dinámica del mercado crea un entorno de innovación acelerada, en el que la capacidad de integración y adopción de estas plataformas define la posición competitiva de las empresas en la economía global.

8.1 Beneficios, riesgos, ética y regulación de la IA

8.1.1 Beneficios organizacionales directos

El uso de tecnologías de inteligencia artificial dentro de las empresas tiene un efecto significativo sobre la eficiencia, la productividad y la capacidad de innovación de las empresas. El beneficio más inmediato es la automatización de los procesos repetitivos de gran volumen, dejando libres a los colaboradores humanos de tareas rutinarias de bajo valor agregado, para que concentren sus esfuerzos en la resolución de problemas creativos, la planificación estratégica y la atención personalizada a clientes de gran valor (Davenport & Westerman, Why So Many High-Profile Digital Transformations Fail, 2018).

Además de ahorrar costos operacionales y de aumentar la velocidad de ejecución, la IA mejora en gran medida la calidad de las decisiones empresariales con analítica prescriptiva y predictiva. Al procesar continuamente millones de variables, los modelos inteligentes permiten optimizar la gestión de las cadenas de suministro globales, predecir la demanda de productos con semanas de antelación, detectar fraudes financieros en milisegundos y personalizar la oferta comercial para cada cliente individual. Como afirman Brynjolfsson y McAfee (2014), las empresas que utilizan la IA como eje de su estrategia organizacional no solo obtienen mejoras incrementales en sus márgenes de ganancia, sino que cambian por completo las reglas de sus respectivas industrias.

8.1.2 Riesgos tecnológicos, desafíos e impacto laboral

Los extraordinarios beneficios organizacionales de la IA conllevan riesgos tecnológicos, dilemas operativos y profundas transformaciones del mercado laboral si se implementan de forma irreflexiva. Los sesgos algorítmicos son uno de los riesgos técnicos más críticos, ya que ocurren cuando los modelos aprenden y perpetúan patrones históricos de discriminación o desigualdad presentes en los datos de entrenamiento, lo que deriva en decisiones injustas en procesos de contratación laboral o concesión de créditos bancarios (O'Neil, 2016). A esto hay que sumar las alucinaciones en los modelos generativos donde el sistema sintetiza información falsa o ficticia expresada con absoluta solidez formal y la creciente dependencia tecnológica, que puede erosionar las capacidades analíticas críticas del personal humano.

En cuanto a su impacto social y laboral, la automatización basada en IA no solo está afectando los empleos rutinarios de trabajo manual, sino que está reorganizando de forma significativa las profesiones intelectuales de cuello blanco en áreas como el derecho, la contabilidad, la programación y la medicina. Si bien la historia económica muestra que las revoluciones tecnológicas destruyen puestos obsoletos mientras generan nuevas categorías laborales, la velocidad actual de esta transición plantea serios desafíos de reconversión laboral (reskilling y upskilling) y el riesgo de una polarización económica. Así lo advierte Acemoglu (2021) quienes los impactos de la mitigación requieren que las corporaciones y los Estados implementen políticas de una transición justa que den prioridad a la inteligencia aumentada y la capacitación continua de la fuerza laboral.

8.1.3 Marcos regulatorios internacionales

Gobiernos y organismos multilaterales han empezado a aprobar marcos regulatorios internacionales para traducir en leyes vinculantes los principios éticos y preocupaciones de privacidad que buscan gobernar el desarrollo de la inteligencia artificial. El Reglamento Europeo de Inteligencia Artificial (EU AI Act) es el hito legislativo más ambicioso con impacto global, que establece una regulación basada en un enfoque de gestión de riesgos para categorizar las aplicaciones de IA en cuatro niveles: riesgo inaceptable (prohibidas), alto riesgo (sujetas a estrictas auditorías y certificaciones), riesgo específico (sujetas a transparencia) y riesgo mínimo.

Los Principios de Inteligencia Artificial de la OCDE y la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO, por su parte, aportan orientaciones multilaterales para impulsar una IA centrada en el ser humano, la inclusión social y el desarrollo sostenible a nivel mundial. Estos marcos regulatorios internacionales están dando origen al denominado "Efecto Bruselas", como señalan Bradford y otros (2023), donde las multinacionales tecnológicas se ven obligadas a adaptar sus estándares globales a las exigencias regulatorias europeas para no perder acceso al mercado comunitario, sentando las bases de la gobernanza informática mundial en el siglo XXI.

CAPÍTULO II

FUNDAMENTOS DE LOS AGENTES INTELIGENTES

2.1 ¿Qué son los agentes de inteligencia artificial?

Los agentes de inteligencia artificial constituyen uno de los campos de mayor desarrollo dentro de la IA actual, ya que poseen la capacidad de percibir su entorno, procesar información y realizar acciones de forma autónoma o semiautónoma con el fin de cumplir objetivos específicos. Su desarrollo ha conducido a la aparición de sistemas cada vez más adaptativos, capaces de dialogar con los usuarios, las aplicaciones y los demás sistemas en los más dispares contextos.

2.1.1 Concepto y definición de los Agentes de Inteligencia Artificial

Afirma Bravo, Ch (2025) que la aparición de la Inteligencia Artificial se puede ya considerar como una de las transformaciones más significativas en la historia del ser humano. Sus capacidades y atributos continúan asombrando (o aterrorizando) tanto a los cercanos como a los lejanos, con avances cada vez más rápidos. En ese vasto campo de posibilidades, se sitúa un concepto que es relativamente nuevo: el agente de IA. Por esta razón, a continuación, examinaremos específicamente de qué se trata, cuáles son sus rasgos fundamentales y qué funciones pueden desempeñarle los seres humanos (considerando también el uso malintencionado que puede llevar a cabo el cibercrimen).

Los agentes de IA tienen la capacidad de desempeñar una variedad extensa de funciones, que van más allá del procesamiento del lenguaje natural. Estas incluyen resolver problemas, interactuar con entornos externos, tomar decisiones y llevar a cabo acciones. Los agentes de IA llevan a cabo tareas complejas en las aplicaciones empresariales, que abarcan el diseño de software, la automatización de TI, la creación de código y la asistencia conversacional. Hacen uso de métodos avanzados

de procesamiento de lenguaje natural, propios de los modelos de lenguaje grandes (LLM), para entender y reaccionar a las entradas del usuario de manera secuencial y decidir cuándo es necesario hacer uso de herramientas externas. (Gutowska, 2026)

En un sistema de agentes de inteligencia artificial, múltiples agentes pueden colaborar para administrar tareas complicadas de manera independiente, por ejemplo: solucionar dificultades en la facturación, ofrecer incentivos personalizados para mejorar el grado de satisfacción del cliente y gestionar modificaciones en las cuentas según el sitio web de Snowflake (2025). Para conseguirlo, examinan datos que se encuentran en diversos formatos y áreas de una compañía, como bases de datos, sitios web, wikis internas y correos electrónicos; con el fin de determinar la información más confiable y reciente. Luego, a través de algoritmos de análisis, detección de patrones y toma de decisiones, se utilizan para producir diferentes soluciones potenciales.

2.1.2 Características principales

Coloma, J.et.al (2020) afirma las dimensiones adicionales de autonomía, movilidad, la habilidad para interactuar sin depender de la presencia del usuario y la capacidad de razonamiento son las características que distinguen a los agentes de software de otras aplicaciones. Esto supone la habilidad de manejar información proveniente de contextos externos, como Internet, bases de datos y redes, dado que un conjunto de creencias, actitudes y conocimientos que son comprendidos por el agente del usuario. Entre las principales características se encuentran:

- **Autonomía:** La autonomía es la cualidad que más se menciona y la que menos se entiende. El término "autónomo" no implica que el agente pueda hacer lo que le plazca. Significa que finaliza un proceso de principio a fin, dentro de unas reglas que estableces tú mismo, sin necesidad de darle tu aprobación a cada paso.

La diferencia en términos prácticos es inmensa. Una inteligencia artificial generativa te asiste para redactar un correo, el cual tú revisas y envías. Un agente autónomo observa que un pedido ha estado detenido durante dos días, determina a quién notificar, elabora la notificación, la envía y registra el incidente. Pasas de "tengo una herramienta que me ayuda" a "tengo algo que me desocupa la mesa". Y esa diferencia no es de grado, sino de tipo: una sugiere, la otra lleva a cabo.

- **Reactividad:** Cuando un agente mantiene una interacción constante con su entorno y reacciona a las transformaciones que suceden en este (de manera oportuna para que su respuesta resulte útil), se le conoce como reactivo.
- **Proactividad:** Un agente se considera proactivo cuando puede crear y tratar de lograr metas, no solamente impulsado.
- **Competencia social:** Los agentes tienen la habilidad de interactuar con otros agentes (e incluso con humanos) mediante algún tipo de lenguaje de comunicación entre ellos, y quizás también colaborar con otros. En ocasiones, se refiere a ella como la capacidad comunicativa del agente. Esto es para que el agente

inteligente pueda acceder a la información necesaria desde diversas fuentes.

- **Capacidad de razonamiento:** en otras palabras, el agente inteligente es capaz de deducir y extrapolar basándose en los conocimientos y experiencias que tiene hasta ahora.

2.1.3 Componentes de un agente inteligente

Los agentes de IA interactúan de manera ágil con los sistemas digitales y toman decisiones inteligentes, lo que demanda una intervención mínima del ser humano. Pero, ¿cuál es el motivo por el que estos agentes son verdaderamente inteligentes? Los agentes de IA, en esencia, se fundamentan en un grupo de elementos interrelacionados que les posibilitan observar su entorno, procesar datos, decidir, cooperar, ejecutar acciones relevantes y adquirir conocimientos a partir de su experiencia.

Por un lado, los agentes reactivos son simples reflejos que reaccionan inmediatamente a los estímulos, en ocasiones mediante actuadores que les posibilitan la interacción con su ambiente. Los agentes reflexivos basados en modelos optimizan su proceso de toma de decisiones haciendo uso de un modelo interno del entorno. En el extremo opuesto, los agentes cognitivos proactivos tienen la capacidad de planificar a largo plazo y razonar de manera avanzada. Algunos agentes se enfocan en tareas concretas, mientras que otros están hechos para dirigir a otros agentes como una suerte de "conductor" en una orquestación de IA.

Con esa salvedad, estos son los elementos fundamentales de los agentes de inteligencia artificial, todos ellos son esenciales para la creación de sistemas inteligentes y adaptativos. (Stryker, 2026)

Según la Universidad Europea (2026) los agentes inteligentes se construyen sobre una serie de componentes que les posibilitan actuar, decidir y percibir. Para entender cómo funcionan y qué los distingue de otros sistemas de software, es fundamental entender estos elementos.

- **Sensores o entradas:** obtienen datos del medio ambiente, los cuales pueden ser en forma de señales físicas, texto o comandos de voz.
- **Sistema de procesamiento:** el corazón del agente. Examina los datos recibidos, emplea reglas, algoritmos o modelos y establece la acción más apropiada.
- **Base de conocimiento:** depósito de datos sobre el ambiente y las vivencias anteriores del agente. En sistemas sofisticados, se basa en técnicas de aprendizaje profundo y redes neuronales artificiales.
- **Actuadores o salidas:** los sistemas a través de los cuales el agente lleva a cabo sus acciones, ya sea moviendo un robot, mostrando una sugerencia o enviando un mensaje.
- **Función de aprendizaje:** está presente en los agentes más avanzados, adapta su comportamiento de acuerdo con los resultados conseguidos. Para lograr esto, normalmente se combina con grandes cantidades de datos que se manejan a través de plataformas de big data.



Nota. Fuente: Imagen creada con IA

2.1.4 Ciclo de percepción, razonamiento y acción

Los agentes de inteligencia artificial funcionan a través de un ciclo constante de percepción, razonamiento y acción que les permite interactuar de forma inteligente con su entorno. Este ciclo es la base de su funcionamiento, ya que permite captar información, analizar los datos obtenidos y ejecutar acciones orientadas al cumplimiento de objetivos específicos. Los programas tradicionales ejecutan instrucciones predeterminadas sin considerar cambios en el entorno, mientras que los agentes inteligentes, a diferencia de ellos, adaptan su comportamiento a partir de la información que procesan.

La percepción, que corresponde a la primera fase del ciclo, consiste en obtener información del entorno. Ello es realizado por el agente mediante la utilización de diferentes mecanismos de entrada como sensores físicos, cámaras, micrófonos, registros de eventos, bases de datos, servicios web, interfaces de programación de aplicaciones (API) o mensajes generados

por los usuarios. En esta etapa se pretende reunir información de interés acerca de la situación del entorno, descubrir alteraciones y advertir de circunstancias que necesiten una respuesta. La calidad de la percepción afecta directamente el desempeño del agente, pues las decisiones tomadas a partir de información incompleta, desactualizada o errónea pueden incidir en la efectividad de las acciones posteriores. (Government, 2026)

Una vez que se ha recogido la información, el agente empieza el proceso de razonamiento. En esta fase, examina los datos disponibles, interpreta el contexto, detecta patrones y evalúa las posibles alternativas para determinar la acción más adecuada. Dependiendo de la arquitectura del agente, este proceso puede basarse en reglas lógicas, algoritmos de búsqueda, modelos probabilísticos, técnicas de aprendizaje automático o modelos de lenguaje a gran escala (Large Language Models, LLM). En esta etapa, el agente también puede recurrir a datos guardados en su memoria, aplicar saberes que ya posea o recurrir a instrumentos externos, todo ello para profundizar su análisis y elevar la calidad de la decisión a tomar.

La acción es la última fase del ciclo. Una vez que el agente ha escogido la opción que más le convenga, ejecuta una acción que cambia el estado del entorno o entrega una respuesta al usuario. Las acciones pueden ser enviar una notificación, responder una consulta, recomendar una solución, bloquear un acceso no autorizado, ejecutar un comando, activar un dispositivo físico o iniciar un proceso automatizado. Una vez que se ejecuta la acción, el entorno cambia y el agente vuelve a percibir el nuevo

estado, reiniciando el ciclo continuamente. Este patrón cíclico permite que el agente se ajuste a situaciones dinámicas y responda de manera oportuna a los cambios en su entorno. (Lanham, 2025)

2.2. Tipología de agentes Inteligentes

Según Giacomo, B (2025) los cinco tipos principales de agentes de IA describen cinco formas básicas en las que los agentes toman decisiones sobre qué hacer: reaccionar a las entradas, mantener un estado interno, planificar para lograr objetivos, optimizar compensaciones y aprender de la experiencia. Este marco de trabajo se mantiene porque analiza el comportamiento de la toma de decisiones en vez de tecnologías concretas. Al centrarse en cómo un agente reacciona, razona, optimiza o se adapta (no en las herramientas que usa ni en los roles que desempeña), sigue aplicable a los sistemas modernos creados con modelos de lenguaje grandes, capas de orquestación y herramientas externas.

Podemos clasificar a los agentes de manera general, dependiendo de su capacidad para tomar decisiones y de cuánto conocimiento tienen sobre el mundo que les rodea. Existen agentes muy sencillos, que se parecen a autómatas reactivos, y otros más complejos que conservan una memoria del pasado o elaboran planes para el futuro.

2.2.1 Agentes reactivos

A veces se les llama Agentes Reactivos Simples a los más elementales. Operan de acuerdo con una lógica directa: "SI... "Entonces...". Su

conducta se basa únicamente en lo que perciben en ese momento, sin considerar la historia anterior. Considere el termostato de su hogar: Si la temperatura cae por debajo de 20 grados, entonces activa la calefacción. Si sube más allá de 22, la apaga. No evoca el frío que hacía hace una hora, simplemente responde a la estimulación presente. Con este tipo de agentes, se pueden llevar a cabo numerosas automatizaciones simples en las empresas: Si recibes un correo con el tema "Solicitud de presupuesto", entonces debes reenviarlo al departamento correspondiente a ventas. Son sencillos de poner en práctica, pero tienen limitaciones debido a que carecen de memoria y contexto. (Giacomo, 2025)

El funcionamiento de un agente reactivo se caracteriza por mantener una interacción continua con el entorno, respondiendo de manera inmediata a los cambios que detecta. Cada vez que recibe una nueva entrada de información, identifica el estado actual del entorno y determina cuál es la respuesta previamente establecida para esa situación. Una vez ejecutada la acción, el agente permanece atento a nuevos estímulos, repitiendo el mismo proceso de forma constante.

Este tipo de agentes no realiza procesos de análisis complejos ni construye una representación interna del entorno. Su objetivo es ofrecer una respuesta rápida y eficiente ante situaciones específicas, lo que permite reducir el tiempo de procesamiento y optimizar tareas que requieren decisiones inmediatas. Por ello, son ampliamente utilizados en sistemas de automatización, monitoreo y control, donde las condiciones de operación suelen ser claras y las respuestas están previamente definidas. (Coca, 2015)

2.2.2 Agentes basados en objetivos

Estos agentes son más activos afirma Zax, D (2026), no se limitan solo a reaccionar ante estímulos o a sostener un patrón actual, sino que tienen una meta concreta por cumplir (como "incrementar en un 10% el número de clientes potenciales calificados durante este mes" o "disminuir el tiempo promedio de respuesta a las solicitudes de soporte por debajo de las dos horas"). La necesidad de escoger las acciones que los acerquen más a su meta final es lo que impulsa sus decisiones.

Un agente de inteligencia artificial que utiliza un enfoque proactivo y dirigido a metas para resolver problemas y tomar decisiones es conocido como agente basado en objetivos según Zax, D (2026). Es un caso de IA agéntica, donde los sistemas de inteligencia artificial efectúan verdaderamente acciones por cuenta de los usuarios (a diferencia, por ejemplo, de un sencillo chatbot LLM para el servicio al cliente). Los agentes objetivos se ubican justo en el centro de la jerarquía de cinco niveles de complejidad. Son más complejos que los agentes reflejos simples (que están sujetos a reglas definidas de antemano) y que los agentes reflejos basados en modelos (que incorporan un modelo interno del mundo). Sin embargo, son menos complejos que los agentes de aprendizaje (que tienen la capacidad de ajustarse y progresar con el tiempo, generalmente a través del aprendizaje reforzado o del aprendizaje profundo) y los agentes basados en la utilidad (que pueden computar las compensaciones por medio de lo que se denomina función de utilidad).

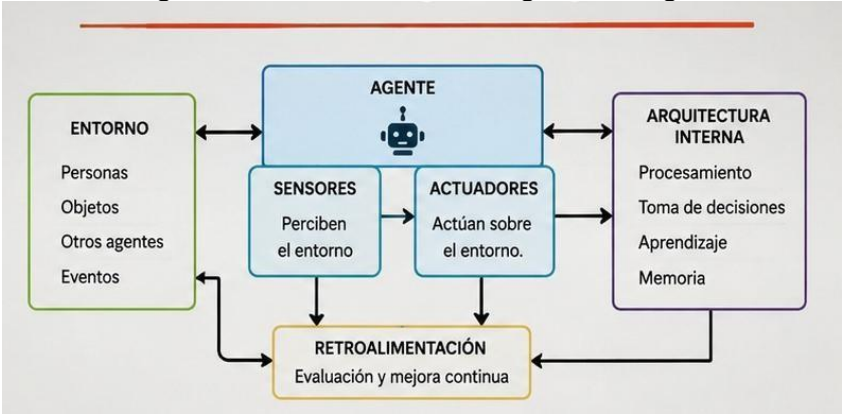
Así también Giacomo, B (2025) manifiesta que la automatización de almacenes, un campo de la robótica, es un ejemplo de agente basado en objetivos. Un robot de almacén que debe recoger un artículo específico de una estantería específica podría simplemente actuar ante los obstáculos inmediatos, como lo hace un robot aspirador al colisionar con las paredes (un "agente meramente reactivo"). No obstante, el plan más eficaz sería trazar una ruta que reduzca al mínimo las desviaciones y evite los obstáculos ya conocidos. A modo de ejemplo, en un ambiente cambiante como un almacén moderno, un módulo de planificación puede consultar su base de conocimientos, examinar la situación actual y proyectar estados futuros (por ejemplo, a partir del conocimiento sobre las acciones planeadas de otros robots). Todo esto con el objetivo de obtener el resultado esperado de manera eficaz.

2.2.3 Agentes basados en utilidad

Los agentes inteligentes que toman decisiones en función de una función de utilidad, la cual mide la deseabilidad de diversos estados, se conocen como agentes basados en utilidad. Estos agentes se proponen maximizar la utilidad general, teniendo en cuenta las incertidumbres y las compensaciones. Los agentes basados en utilidad, a diferencia de los que están basados en objetivos y se enfocan en el cumplimiento de metas concretas, buscan mejorar el rendimiento seleccionando acciones que produzcan la mayor utilidad. Los agentes basados en utilidad funcionan al valorar de manera constante la deseabilidad de diversas acciones o estados a través de una función de utilidad, que le asigna un valor numérico a cada posible resultado.

Los agentes fundamentados en la utilidad brindan un método eficaz para maximizar el rendimiento y la satisfacción en varios usos. Cuando analizan acciones fundamentadas en una función de utilidad, tienen la capacidad de tomar decisiones complejas que ofrezcan el máximo beneficio general. No obstante, el potencial de desalineación y la complejidad de los cálculos de utilidades son factores importantes que deben tratar los desarrolladores. Los agentes basados en utilidad tienen la capacidad de optimizar los resultados y la toma de decisiones en áreas como los sistemas de recomendación y el comercio financiero, si se diseñan e implementan con precisión. (Lanham, 2025)

Imagen N°7 Fundamentos de los Agentes Inteligentes



Nota. Fuente: Imagen creada con IA

Como menciona Belcic, I; Stryker, C (2026) una ecuación matemática que ilustra la manera en que el agente debe valorar el provecho de cualquier acción potencial que tenga la capacidad de llevar a cabo se llama función de utilidad. Fundamentalmente, es el conjunto de valores del agente y muestra cómo este ordena los factores importantes al tomar decisiones. Las funciones de utilidad otorgan a cada resultado de una

acción potencial un valor numérico, lo que permite medir las preferencias que el agente debe sostener. Los agentes basados en la utilidad emplean la función de servicios para discutir entornos complejos, ponderar los intercambios y optimizar los servicios que eligen.

La seguridad, la eficiencia, la distribución de recursos y los costos de oportunidad multiobjetivo son algunos de los factores que toma en cuenta una función de utilidad adecuada. La función de utilidad es el fundamento de un agente basado en la utilidad y es lo que los distingue de otras clases de agentes de inteligencia artificial.

2.2.4 Agentes autónomos

Los agentes de inteligencia artificial autónomos son programas informáticos diseñados para desempeñar tareas, tomar decisiones y comunicarse con otros sistemas sin la intervención continua de los seres humanos, con el fin de lograr objetivos concretos afirma Nuncio, R (2024). Estos agentes se valen de modelos de lenguaje extensos como motor de razonamiento para dividir problemas difíciles en subtarefas que pueden llevarse a cabo con la ayuda de herramientas externas. En el contexto del Instituto Tecnológico Europeo, caracterizamos esta tecnología como el progreso natural de la automatización robótica de procesos hacia la autonomía cognitiva total.

Según el Instituto Tecnológico Europeo (2026) los agentes de inteligencia artificial autónomos se distinguen de la automatización convencional en cuanto a su habilidad para razonar de manera iterativa y decidir con base en el contexto. Un flujo de trabajo estándar se basa en

una lógica lineal de "si esto, entonces aquello", mientras que un agente autónomo tiene la capacidad de analizar el resultado de una acción y determinar el siguiente paso con base en el objetivo final. Esta habilidad es esencial para la productividad de las empresas con IA, porque hace posible delegar responsabilidades complejas que antes necesitaban supervisión humana continua, como el manejo de atención al cliente de segundo nivel o el análisis prospectivo de riesgos financieros.

Aunque todos los agentes autónomos son, en términos técnicos, agentes de IA, no todos los agentes de IA tienen la autonomía. Los agentes de inteligencia artificial incluyen a aquellos de asistencia, como los copilotos, que necesitan la intervención humana para llevar a cabo diferentes actividades. A pesar de que los agentes autónomos y los agentes de IA de asistencia son capaces de tomar decisiones y aprender a partir de información nueva, únicamente los primeros pueden llevar a cabo una serie de tareas seguidas. Los agentes autónomos tienen la capacidad de realizar tareas y flujos de trabajo sin asistencia, mientras que los agentes de IA son más bien una herramienta fundamental que asiste a los trabajadores humanos a enfocarse en las cosas más relevantes.

Los agentes autónomos están diseñados para ser autosuficientes con escasa o ninguna intervención humana, mientras que los agentes de IA están creados para colaborar con la intervención de un ser humano. (García J. , 2026)

2.2.5 Agentes colaborativos o multiagente

Los agentes colaborativos o sistemas multiagente (Multi-Agent Systems, MAS) se refieren a un conjunto formado por dos o más agentes inteligentes que actúan de forma coordinada para resolver problemas, alcanzar un objetivo común o realizar tareas que serían difíciles o menos eficientes para un único agente. Cada agente tiene cierto grado de autonomía, puede percibir su entorno, tomar decisiones y comunicarse con los demás agentes, intercambiando información y coordinando acciones.

A diferencia de un agente individual, un sistema multiagente reparte la tarea entre varios agentes especializados. Cada uno puede tomar una función específica, ya sea para recopilar información, analizar datos, planificar acciones o supervisar resultados. Esta distribución de responsabilidades permite al sistema procesar la información de forma más eficiente, adaptarse a los cambios del entorno y responder con mayor rapidez a situaciones complejas. (Harkar, 2026)

Entonces se podría decir que los sistemas de agentes colaborativos o de multiagentes se basan en la interacción coordinada de varios agentes inteligentes, cada uno de los cuales desempeña un rol específico y contribuye al logro de un objetivo común Coloma, J.et.al (2020). Lo primero que ocurre es que cada agente obtiene información del entorno a través de distintas fuentes, tales como sensores, bases de datos o entradas del usuario. Luego, procesa esa información de forma autónoma, aplicando sus propias reglas o algoritmos de IA, y comunica los resultados al resto de agentes mediante los protocolos de comunicación. A partir de este intercambio de información, los agentes coordinan sus

acciones, distribuyen las tareas de acuerdo con sus capacidades y, cuando es necesario, negocian la mejor estrategia para evitar conflictos o duplicar esfuerzos. Por último, cada agente desempeña su tarea correspondiente y, tras obtener los resultados, el sistema evalúa el desempeño colectivo para introducir los ajustes necesarios, lo cual permite a los agentes adaptarse a los cambios en el entorno y mejorar continuamente su funcionamiento.

2.3. Arquitectura y funcionamiento de los Agentes IA

Entendemos por la arquitectura de agente inteligente la forma en que se organiza y estructura un agente para que pueda actuar de manera autónoma e inteligente. En otras palabras, es el conjunto de componentes que le permite percibir lo que ocurre en su entorno, procesar la información que recibe, tomar decisiones y ejecutar acciones orientadas a cumplir un objetivo específico. Esta arquitectura define cómo trabajan e interactúan entre sí elementos como la percepción, el razonamiento, el aprendizaje y la actuación, permitiendo que el agente responda de manera eficiente a diferentes situaciones e, incluso, adapte su comportamiento cuando el entorno cambia.

2.3.1 Sensores y actuadores

La incorporación de sensores y actuadores en aparatos inteligentes ha transformado los ámbitos de la robótica, la atención médica y el IoT, facilitando que se pase de la percepción a la acción. Esta interacción brinda flexibilidad, autonomía y capacidad de respuesta en tiempo real,

además de crear innovaciones como los dispositivos inteligentes portátiles, la medicina bioelectrónica y la robótica blanda. No obstante, continúan existiendo dificultades en cuanto a la interoperabilidad, la miniaturización y el control de la coordinación de señales en entornos dinámicos.

Según Wan et al. (2026), los sensores y actuadores son los componentes básicos que hacen posibles los dispositivos inteligentes. Los sensores recogen los estímulos o cambios del entorno y envían la información al sistema para que lo procese y pueda conocer el estado del ambiente. Los actuadores, por su parte, convierten las decisiones del sistema en respuestas físicas o digitales, lo cual permite al dispositivo interactuar de forma autónoma con su entorno. Los autores afirman que el verdadero potencial de los sistemas inteligentes está en la integración de ambos componentes mediante un ciclo cerrado de percepción-decisión-acción, donde la información captada por los sensores guía las acciones ejecutadas por los actuadores y éstas generan nuevos datos que nuevamente son percibidos por los sensores.

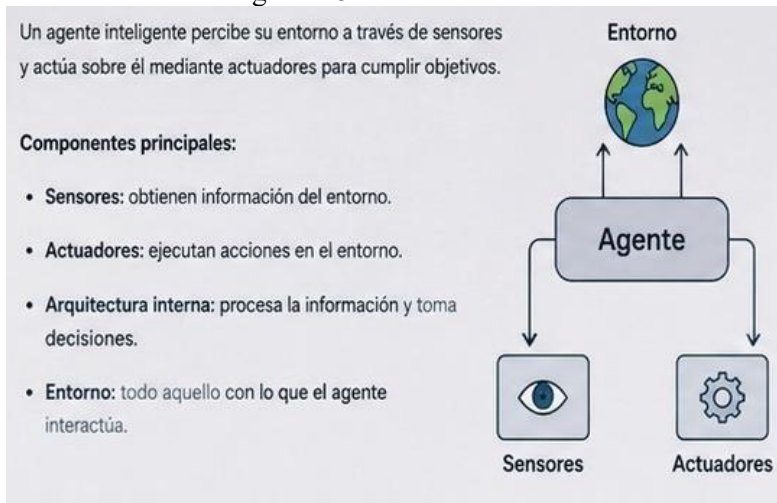
Este proceso de retroalimentación continua potencia la capacidad de respuesta, la precisión, la adaptabilidad y la autonomía de los dispositivos inteligentes, lo que favorece su aplicación en ámbitos como la robótica, la Internet de las Cosas (IoT), los dispositivos biomédicos y las interfaces humano-máquina.

2.3.2 Memoria a corto y largo plazo

Los agentes de IA, como los humanos, dependen de sistemas de memoria para procesar, almacenar y recuperar información. En IA, sin embargo,

la memoria suele estar dividida en memoria a corto plazo (MCP) y memoria a largo plazo (MLP), cada una con funciones distintas dentro del marco cognitivo del agente. Es fundamental entender las diferencias entre MCP y MLP para diseñar sistemas de IA que sean capaces de adaptarse, aprender y realizar tareas complejas. Este artículo explora las diferencias, funcionalidades y aplicaciones entre la MCP y la MLP en agentes de IA.

Imagen N°8 Sensores actuadores



Nota. Fuente: Imagen creada con IA

En los agentes de inteligencia artificial, la memoria a corto plazo se caracteriza por almacenar información de forma temporal, manteniéndola únicamente durante el tiempo necesario para realizar una tarea o interacción específica. Generalmente dura segundos o minutos, al cabo de los cuales se borra la información o se reemplaza por datos nuevos. Además, su capacidad es limitada, ya que sólo puede mantener una pequeña cantidad de información simultáneamente, por lo que el agente debe priorizar los datos más relevantes y descartar los que ya no

son útiles. Estas características hacen que la memoria a corto plazo funcione como un espacio de trabajo inmediato, en el cual el agente puede procesar la información más reciente y tomar decisiones de forma rápida y eficiente, especialmente en situaciones que requieren respuestas en tiempo real. (Coloma,J.et.al, 2020)

2.3.3 Herramientas y conectores

Según Arias, O (2025) las herramientas son un elemento fundamental en la estructura de los agentes de inteligencia artificial, ya que mejoran las habilidades del modelo de lenguaje, posibilitando que interactúe con recursos externos y realice acciones concretas para alcanzar las metas asignadas. A pesar de que la generación de respuestas y el razonamiento son funciones del modelo de lenguaje, las herramientas le permiten ejecutar tareas que necesitan acceso a información actualizada, interacción con otros sistemas o procesamiento especializado.

La operación de las herramientas se fundamenta en un mecanismo que permite al agente determinar cuándo no es posible resolver una tarea solo con la información del modelo y, por lo tanto, optar por utilizar una herramienta externa. Cuando la acción se ha llevado a cabo, el agente agrega el resultado al proceso de razonamiento, donde lo examina e integra con el contexto de la conversación para producir una respuesta más fundamentada y exacta.

Además, Coca, Y (2015) manifiesta el conector de agente de IA permite que los agentes de IA se conecten con un LLM, para que puedan ofrecer capacidades de interacción y razonamiento. Este conector está pensado para utilizarse con un subproceso ad hoc en un loop de retroalimentación, dando interacción con el usuario y selección de herramientas en forma

automática. Por ejemplo, utilice este conector para permitir que un agente de IA pueda seleccionar y ejecutar tareas de forma autónoma, dentro de subprocesos ad hoc, evaluando el contexto del proceso actual y determinando las tareas y herramientas pertinentes que debe usar. Puede utilizar el conector del agente de IA independientemente, aunque está pensado para utilizarse con un subproceso ad hoc a fin de definir las herramientas que puede utilizar un agente de IA.

2.3.4 Modelos de lenguaje (LLM)

Los Grandes Modelos de Lenguaje (LLM, por sus siglas en inglés) son sistemas de inteligencia artificial capacitados para procesar, comprender y producir texto con una fluidez sorprendente. Su funcionamiento es fácil de describir, pero increíblemente complejo de desarrollar: aprenden detectando patrones en cantidades enormes de texto. Si alguna vez has jugado a completar frases o has anticipado cuál sería la siguiente palabra en un mensaje de texto, ya sabes un poco de cómo funcionan. Sin embargo, los LLMs lo hacen a una escala gigantesca y con una precisión impresionante.

Los grandes modelos de lenguaje (LLM), como GPT-4, Claude o Gemini, se han convertido en herramientas imprescindibles en múltiples ámbitos, desde la educación hasta la automatización empresarial. Pero ¿cómo accedemos a estos modelos? Los podemos utilizar de dos maneras principales: a través de una aplicación o chatbot y mediante una API (Interfaz de Programación de Aplicaciones), que permite integrarlos a otros sistemas o productos. (Arias, 2025)

Como menciona Giacomo, B (2025) estos modelos han desarrollado una sorprendente habilidad para entender y producir lenguaje humano con fluidez y coherencia. Esto es crucial para los agentes de IA, ya que les permite comprender sus instrucciones expresadas en lenguaje natural (sin necesidad de comandos codificados), interactuar con textos (correos electrónicos, documentos, páginas web) de manera mucho más inteligente e incluso comunicarse de forma casi humana. Antes de estos LLM, hacer agentes que pudiesen manejar tareas basadas en lenguaje era mucho más difícil y necesitaba conocimientos súper especializados. Hoy esta capacidad se ha convertido en una especie de commodity tecnológico, un motor potente que puede integrarse en muchas aplicaciones, incluidos los agentes.

2.4. Plataformas y frameworks para la el desarrollo de agentes inteligentes

Las plataformas y frameworks, en el desarrollo de agentes inteligentes, son la base tecnológica que permite diseñar, implementar y desplegar sistemas con capacidad para percibir su entorno, razonar y actuar de forma autónoma. Estas herramientas ofrecen componentes predefinidos para controlar memoria, planificación, comunicación, aprendizaje y coordinación entre agentes, lo cual facilita la construcción de soluciones más eficientes y escalables. A continuación, se exponen las principales plataformas y frameworks que se utilizan en la actualidad, así como sus características, ventajas y aplicaciones en los distintos contextos de la inteligencia artificial.

- **ChatGPT:** OpenAI, una empresa dedicada a la investigación de la Inteligencia Artificial (IA), es la creadora de ChatGPT, una tecnología que permite el procesamiento del lenguaje natural. Esta herramienta de procesamiento de lenguaje es una de las primeras tecnologías de inteligencia artificial que emplea redes neuronales profundas para generar respuestas a frases y preguntas. Fue creada para aprender de enormes volúmenes de datos, lo que le posibilita generar respuestas más naturales y exactas a instrucciones y preguntas. La traducción automática, la recuperación de información y el entendimiento del lenguaje natural también se han beneficiado con esta tecnología.

Asimismo, ChatGPT posibilita que los usuarios generen textos con mayor personalización. Esto implica que los usuarios tienen la capacidad de generar contenido más adecuado para sus necesidades concretas y, además, que sea más pertinente para el destinatario final; esto se opone a los textos estandarizados disponibles en Google, por ejemplo, que no brindan un nivel elevado de personalización. Desde su lanzamiento como herramienta de acceso libre en diciembre de 2022, ha sido ampliamente adoptada por una gran cantidad de sectores y se ha integrado a numerosas aplicaciones variadas; imagino que seguirá expandiéndose en el futuro. (Anton, 2024)

- **Gemini :** Fraksa,N (2024) manifiesta que Gemini es el modelo de lenguaje a amplia escala (LLM, por sus siglas en inglés) que ha desarrollado Google. En un sentido más amplio, es una familia de modelos de IA multimodales que han sido creados para manejar una

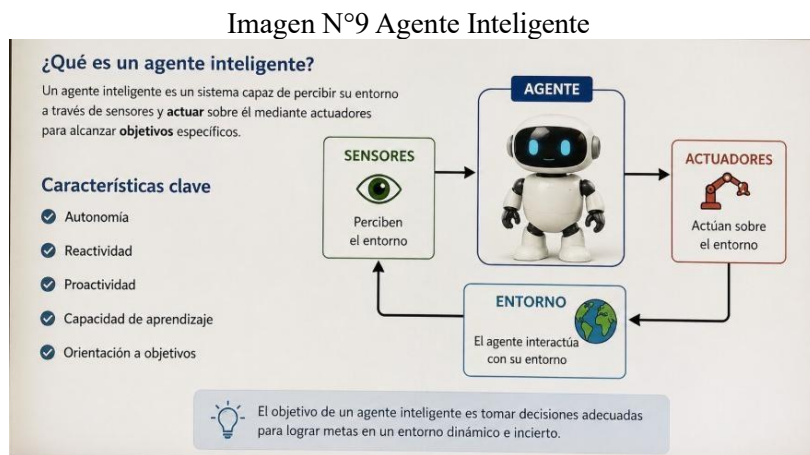
diversidad de tipos o modalidades de datos, como son el audio, las imágenes, el código de software, el texto y el video.

Gemini, de la misma manera que Claude de Anthropic, toma su nombre tanto del chatbot como de la familia LLM que lo soporta. Gemini también es el modelo que alimenta el chatbot de IA generativa (gen AI) de Google (antes conocido como Bard). Las aplicaciones Gemini, en sus versiones web y móviles, actúan como interfaz de chatbot para los modelos subyacentes.

- **Copilot:** Este es un asistente inteligente basado en IA generativa que ayuda a los usuarios a realizar diversas tareas mediante instrucciones expresadas en lenguaje natural. Este sistema, desarrollado por Microsoft y alimentado por modelos de lenguaje masivo (LLM), es capaz de entender el contexto de las solicitudes y producir respuestas coherentes, sugerencias, resúmenes, texto escrito, análisis de datos y fragmentos de código. Su objetivo es elevar la productividad y hacer más sencilla la comunicación entre los usuarios y los programas informáticos, disminuyendo el tiempo necesario para realizar tareas repetitivas o más complejas. (Pascal,B.et.al, 2025)

Copilot funciona mediante inteligencia artificial generativa y procesamiento del lenguaje natural. Cuando un usuario introduce una consulta o una instrucción, el sistema analiza su contenido y determina la intención de la petición. Luego, el modelo de lenguaje procesa la información teniendo en cuenta el contexto disponible, ya sea un documento, una hoja de cálculo, una conversación o un entorno de programación. De acuerdo con este análisis, produce una respuesta o propuesta, que puede incluir texto, código, gráficos, tablas, resúmenes o

recomendaciones. Por último, el usuario puede aceptar el resultado, modificarlo o solicitar nuevos ajustes, lo que establece un proceso de interacción continua que permite obtener respuestas cada vez más precisas y adaptadas a sus necesidades. (Pascal,B.et.al, 2025)



Nota. Fuente: Imagen creada con IA

- **LangChain:** De acuerdo con Bergmann, D; Stryker, C (2026) es un framework de orquestación “open source” para el desarrollo de aplicaciones con modelos de lenguaje grandes (LLM). Las herramientas y las API de LangChain, disponibles en bibliotecas basadas en Python y Javascript, simplifican el proceso de crear aplicaciones basadas en LLM, como chatbots y agentes de IA.

Lang Chain actúa como una interfaz genérica para prácticamente cualquier LLM, y ofrece un entorno de desarrollo centralizado para construir aplicaciones basadas en LLM y conectarlas con fuentes externas de datos y flujos de trabajo de software. El enfoque basado en módulos de LangChain permite a los desarrolladores y expertos en datos

comparar de forma dinámica diferentes instrucciones e incluso distintos modelos fundacionales con una necesidad mínima de reescribir código.

- **CrewAI:** Para Winland, V; Syed, M; Gutowska, A (2026) João Moura es el creador de crewAI, una estructura de orquestación multiagente de código abierto. Este marco, que se basa en Python, tiene la capacidad de orquestar agentes autónomos de inteligencia artificial (IA) e integrarlos como un equipo unitario para realizar tareas. CrewAI busca brindar un marco robusto para la automatización de los procesos de trabajo de varios agentes. El término "crew" hace referencia a los agentes de inteligencia artificial que colaboran entre sí para delegar tareas de manera autónoma y formular preguntas entre ellos, como lo haría un equipo de trabajo en la vida real. Cada equipo multiagente está formado por agentes de inteligencia artificial con roles complementarios que utilizan herramientas personalizadas y existentes para llevar a cabo un conjunto de tareas que se les han encomendado. Los modelos de lenguaje funcionan como un motor de razonamiento para los agentes al elegir una serie de acciones.² Es posible configurar a los agentes de CrewAI para emplear cualquier API o modelo de lenguaje extenso/grande (LLM) que sea de código abierto.
- **AutoGen:** Core es la capa esencial de AutoGen, la estructura elemental que hace posible el funcionamiento del marco. Según Microsoft, "la API Core lleva a cabo la transmisión de mensajes, los agentes controlados por eventos y los entornos de ejecución tanto locales como distribuidos". En otras palabras, posibilita que los agentes se comuniquen entre ellos, les permite activarse ante ciertos

eventos y les deja ejecutarse de manera local en su equipo o en múltiples servidores.

AutoGen es "extensible", lo que quiere decir que los usuarios tienen la posibilidad de añadir nuevas funcionalidades. Extensiones predeterminadas de AutoGen: Incluye elementos como LocalSearchTool, que posibilita la búsqueda en tu propio conjunto de archivos, y MultimodalWebSurfer, que facilita la navegación por internet. Los desarrolladores son también estimulados por Microsoft a crear sus propias extensiones. (Zax, 2026)

Mientras que AutoGen es flexible y caótico, CrewAI es estructurado y ordenado. La conversación es la base filosófica de AutoGen. Los agentes se reúnen en un grupo de chat y comienzan a conversar entre ellos para resolver la tarea. Lo innovador de AutoGen es que facilita la combinación de agentes de IA con seres humanos y con ejecutores de código real dentro del mismo chat. AutoGen generalmente tiene un agente especial conocido como UserProxy. Este agente funciona como tu representante. Si el otro agente (el Asistente) redacta un código en Python, el UserProxy tiene la capacidad de ejecutarlo en tu computadora (o en un entorno Docker) y devolverle al Asistente los resultados. (García J. , 2026)

2.5. Aspectos éticos y regulatorios

El rápido avance de la inteligencia artificial y el creciente uso de agentes inteligentes en diferentes ámbitos han planteado importantes desafíos en relación con la ética, la seguridad y el cumplimiento de la normativa. Estas tecnologías, aunque ofrecen grandes ventajas en cuanto a la

automatización, la eficiencia y el apoyo en la toma de decisiones, también plantean interrogantes respecto a la privacidad de los datos, la transparencia de los algoritmos, la responsabilidad de sus acciones y la protección de los derechos de las personas.

2.5.1 Transparencia algorítmica

Menciona en su artículo Abad, L (2026) la transparencia algorítmica es un principio esencial que asegura que los sistemas de inteligencia artificial operen de forma confiable, responsable y verificable. Consiste en aportar información suficiente sobre el funcionamiento de los algoritmos, los datos que utilizan y los criterios empleados para generar sus resultados, permitiendo comprender, en la medida de lo posible, cómo se toman las decisiones automatizadas. Este principio se propone lograr varios objetivos esenciales.

En primer lugar, facilita la comprobación de que el sistema funciona, ya que muchos modelos de inteligencia artificial se consideran “cajas negras”, porque es difícil entender los procesos internos que llevan a una determinada conclusión. La transparencia hace posible evaluar si el sistema está actuando como fue diseñado para alcanzar sus objetivos. En segundo lugar, propicia una mejora continua de los sistemas, ya que permite conocer los componentes, procesos y desempeño de los mismos, con lo cual se pueden identificar las oportunidades de optimización, corregir las deficiencias y comprobar si están cumpliendo adecuadamente con su finalidad.

2.5.2 Privacidad y protección de datos

La privacidad informativa o la privacidad de la información están vinculadas con la privacidad de los datos UNESCO (2026). En la época digital, el espacio físico, la autonomía individual y cómo una persona genera y emplea datos se empaquetan en tecnologías emergentes y se convierten en datos y espacios virtuales. Por ejemplo, antes de la llegada de la tecnología, uno podía ir a la biblioteca y consultar información cada vez que quisiera. Solamente los empleados con autoridad en la biblioteca podían acceder al historial o a la clase de información que el usuario estaba consultando.

Según la ISO/IEC (2023), se puede establecer que la seguridad de la información consiste en los procedimientos, las prácticas adecuadas y las metodologías que tienen como objetivo proteger tanto la información como los sistemas de información del acceso, uso, divulgación, interrupción, alteración o eliminación no autorizada. Por otro lado, a pesar de que no existe un acuerdo universal acerca de lo que es la ciberseguridad, podemos afirmar de manera sencilla que es una disciplina dedicada a resguardar los sistemas, las redes y la información informática contra ataques accidentales o maliciosos. Esto incluye la defensa ante el robo de identidad, ataques DoS (denegación de servicio), phishing, software malicioso y otras formas de riesgos a la seguridad informativa. (Astudillo, 2025)

2.5.3 Sesgos en IA

El sesgo se refiere a la tendencia sistemática que puede emerger en los modelos de inteligencia artificial, ya sea por un diseño sesgado o por datos de entrenamiento sesgados. El sesgo puede presentarse en varias

fases del proceso de construcción del algoritmo, desde que se define el problema hasta que se recopilan y preparan los datos. Si no se diseñan y supervisan de manera apropiada. Los sesgos pueden infiltrarse en los sistemas de inteligencia artificial mediante datos históricos que muestran prejuicios de las personas

Cuando nos referimos a sesgos, podemos mencionar dos categorías. Por un lado, puede haber una discriminación intencional o explícita (disparate treatment), y, por otro lado, puede haber una discriminación no intencional (disparate impact). Es un hecho que los sistemas de inteligencia artificial suelen replicar esta última, ya que los sesgos generalmente son el resultado de un tratamiento inadecuado de los datos con los que se ha entrenado el modelo. (Lopez & García, 2024)

2.5.4 Gobernanza de IA

La gobernanza de la IA se refiere a un conjunto de políticas, directrices, prácticas y normas que definen un marco administrativo para el uso de la tecnología y el aprovechamiento de sus ventajas comerciales, siempre manteniendo el respeto por la igualdad, los derechos humanos (incluidos los derechos del autor) y la transparencia.

La gobernanza de la inteligencia artificial facilita el control de los riesgos y el establecimiento de un uso tecnológico que tenga en cuenta consideraciones éticas, mantenga la integridad y promueva confianza entre los clientes, los usuarios, los empleados y cualquier otra parte interesada. Así, la gobernanza de la Inteligencia Artificial se puede definir brevemente como el poder que tiene una entidad para utilizar los

beneficios que brinda la tecnología de manera controlada. (ISO TOOLS, 2025)

CAPÍTULO III

AGENTES DE INTELIGENCIA ARTIFICIAL PARA LOS NEGOCIOS

3.1 Transformación digital e industria impulsada por agentes de IA

3.1.1 Contextualización de la transformación digital

Se puede apreciar que el concepto de transformación digital ha experimentado una profunda transformación cualitativa, al considerar la evolución de la tecnología dentro del entorno corporativo en los últimos tiempos. En las últimas dos décadas, la digitalización de las empresas se ha basado principalmente en digitalizar los procesos manuales y en recopilar grandes cantidades de datos en la nube. Sin embargo, este modelo tradicional ha generado un cuello de botella serio a nivel operativo, las entidades contemporáneas cuentan con más datos de los que su personal humano puede procesar y dependen de sistemas informáticos pasivos que necesitan de una supervisión y ejecución manual continua (Davenport & Ronanki, 2018).

En esta situación de exceso de información, la verdadera transformación digital de nueva generación se define por el paso de un software estático e interactivo a la utilización de agentes autónomos de inteligencia artificial. Un agente de inteligencia artificial es una entidad computacional dotada de capacidad cognitiva, diseñada para percibir su entorno digital, razonar sobre objetivos empresariales complejos, tomar decisiones y ejecutar secuencias de acciones de forma proactiva y con mínima supervisión humana (Wooldridge, 2020). Esto contrasta con una aplicación convencional o un tablero de control estático que solo muestra estadísticas pasadas. De tal forma, la transformación digital no solo deja de ser un simple proceso de digitalización de documentos, sino que se transforma en una automatización directa del raciocinio y la ejecución operativa.

3.1.2 El concepto de “Negocios Inteligentes”

La aparición de la tecnología de agentes cognitivos ha propiciado en la literatura gerencial actual la idea de “Negocios Inteligentes” (Smart Businesses). Una empresa inteligente no es simplemente una compañía que utiliza herramientas de marketing digital o computadoras sofisticadas; es, más bien, una organización cuyos procesos operativos están vinculados a sistemas de inteligencia artificial con la capacidad de aprender y perfeccionarse continuamente. En esas compañías los agentes de IA no operan como programas autónomos, sino como una capa de inteligencia común que acopla la gestión de la cadena de suministro con las ventas en tiempo real, la facturación empresarial y el servicio al cliente (Iansiti & Lakhani, 2020).

Los negocios inteligentes son estratégicamente importantes porque pueden tomar decisiones a la velocidad de los datos, no a la de la burocracia humana. Un sistema de agentes inteligentes revisa la oferta de los competidores, la demanda del mercado y los niveles de inventario en milisegundos, haciendo ajustes dinámicos sin intervención externa. Por ejemplo, en vez de esperar que un comité de gerentes revise los informes de ventas al finalizar el mes para hacer cambios en los precios del catálogo. Las empresas que logran hacer la transición a los modelos de negocio inteligentes, afirman Porter y Heppelmann (2015), no solo reducen los costos fijos operativos a niveles antes inimaginables, sino que adquieren una flexibilidad estratégica que les permite establecer nuevas reglas del juego en sus respectivas industrias.

3.1.3 La evolución industrial: De la industria 4.0 a la industria 5.0

Comprender cómo se van incorporando los agentes de IA en el ámbito productivo, requiere analizar el rápido cambio de paradigma que está ocurriendo entre la Industria 4.0 y la nueva Industria 5.0. La Revolución Industrial Cuarta (Industria 4.0) se caracterizó principalmente por la interconexión digital de las fábricas mediante el Internet de las Cosas (IoT), el análisis de Big Data y la automatización mediante robots rígidos en entornos industriales (Lasi & al, 2014). Aunque este enfoque aumentó enormemente la capacidad de producción a gran escala, a menudo careció de flexibilidad en términos humanos, de sostenibilidad ambiental y de capacidad de respuesta ante grandes interrupciones en las cadenas mundiales de suministro.

Ante estas limitaciones, la Industria 5.0 emerge como un paradigma avanzado y complementario, que vuelve a colocar al ser humano en el centro del proceso productivo, promoviendo una visión centrada en la persona, resiliente y sostenible. Los agentes de inteligencia artificial en este nuevo sistema tienen un rol fundamental como colaboradores cognitivos (cobots y agentes lógicos) en estrecha colaboración con operarios e ingenieros. Xu et al. (2021) señala que la Industria 4.0 pretendía reemplazar el trabajo humano por la automatización a ciegas, mientras que la Industria 5.0 utiliza agentes de inteligencia artificial para potenciar la creatividad y el proceso de toma de decisiones humanas, lo que permite personalizar productos de alta calidad a gran escala, sin que esto afecte a la eficiencia industrial.

3.1.4 Automatización empresarial de siguiente generación

Lo que la industria tecnológica denomina Hiperautomatización o Automatización de Siguiete Generación ha surgido de la convergencia entre los agentes de IA y la automatización de los procesos empresariales. En el pasado, la automatización corporativa se apoyaba en herramientas de Automatización Robótica de Procesos (RPA) capaces de copiar y pegar información entre pantallas, guiándose por reglas mecánicas fijas e inmutables. Sin embargo, el RPA tradicional tenía una rigidez absoluta, por ejemplo, si un cliente enviaba un correo con un lenguaje inesperado o si el formato de una factura cambiaba ligeramente, el bot fallaba y necesitaba la intervención de un programador (van der Aalst & al, 2018).

La Hiperautomatización, impulsada por agentes de Inteligencia Artificial (IA), se enfrenta a esta vulnerabilidad integrando capacidades cognitivas de razonamiento probabilístico, procesamiento de lenguaje natural y visión computacional dentro del flujo de trabajo. Un agente de IA de última generación será capaz de leer un contrato escrito en lenguaje natural, extraer sus cláusulas financieras más importantes, comparar las cifras con las políticas internas de la empresa, identificar eventuales fraudes y requerir autorizaciones humanas únicamente cuando detecte un nivel de riesgo inusual. Agostinelli et al (2021) señalan que este desarrollo tecnológico permite a las empresas automatizar procesos complejos “de punto a punto” que previamente demandaban semanas de tareas administrativas, transformando así a los departamentos de operaciones en máquinas de ejecución ultrarrápidas e infalibles.



Nota. Fuente: Imagen creada con IA

3.2 Agentes de IA en marketing, ventas y comercialización

3.2.1 Segmentación inteligente de audiencias

La clásica segmentación de mercados basada en clasificaciones estáticas como la ubicación geográfica, el género o la edad, ya no sirve para predecir el comportamiento real de los consumidores en el ámbito actual del marketing y la comercialización. Los consumidores de hoy interactúan de forma caótica a través de un abanico de canales digitales, dejando huellas de datos complejas que cambian en cuestión de horas. Gracias a los agentes de inteligencia artificial (Wedel & Kannan, 2016), el uso de la segmentación conductual y psicográfica dinámica en tiempo real ha transformado esta disciplina.

Los sistemas de inteligencia artificial pueden analizar simultáneamente miles de variables de comportamiento, entre ellas el historial de clics, los patrones de navegación en internet, el tiempo dedicado a un catálogo, las interacciones en redes sociales y la sensibilidad a los cambios de precios.

Se consigue a través de agentes que alimentan modelos de agrupamiento (clustering) y algoritmos de aprendizaje no supervisado. El agente de inteligencia artificial clasifica a los clientes en micro segmentos muy homogéneos que se van renovando cada vez que el usuario interactúa. Esta clasificación inteligente permite a los gerentes de marketing dejar atrás las estrategias generales y masivas, y dirigirse a cada segmento con mensajes comerciales adecuados a sus actuales intenciones de compra y necesidades psicológicas.

3.2.2 Hyper-Personalización de campañas de marketing

La segmentación dinámica es directamente proporcional a la posibilidad de desarrollar campañas de marketing basadas en el paradigma de la hiperpersonalización. La personalización tradicional consistía simplemente en colocar el nombre del cliente en la parte superior de un correo masivo. Por el contrario, la hiperpersonalización impulsada por agentes de IA genera contenido comercial único, contextualizado y dinámico para cada individuo en el momento en que está más dispuesto a realizar una compra (Kumar & al, 2019).

Un agente de inteligencia artificial especializado en ventas puede observar el comportamiento de un cliente en una plataforma web y en cuestión de segundos crear una propuesta comercial personalizada, elegir las imágenes promocionales más adecuadas a su perfil visual, adaptar el tono comunicativo y establecer cuál es el canal más adecuado para hacer llegar la comunicación (WhatsApp, correo electrónico o notificación push). Los descubrimientos de Overgoor et al. (2019) indican que las

compañías que utilizan la hiperpersonalización con sistemas de IA generativa y agentes de recomendación experimentan un aumento de más del 30 % en sus tasas de conversión y una mejora importante en el valor del ciclo de vida del cliente (Customer Lifetime Value - CLV), lo que permite optimizar el retorno de la inversión publicitaria.

3.2.3 Asistentes y Chatbots comerciales avanzados

Los antiguos chatbots basados en rígidos árboles de decisión han evolucionado a sofisticados agentes conversacionales de negocios, lo cual ha causado una profunda transformación en el canal de atención comercial y ventas directas. Los chatbots de primera generación provocaban una frustración enorme en los usuarios porque no entendían frases complejas que no estuvieran en sus scripts predefinidos. Por el contrario, los agentes comerciales actuales se basan en arquitecturas de generación aumentada por recuperación (RAG) y modelos de lenguaje a gran escala (LLM), lo que les permite mantener conversaciones empáticas, fluidas y contextualmente coherentes en lenguaje natural.

Estos agentes de inteligencia artificial no sólo contestan las preguntas más frecuentes, sino que además pueden guiar a todo el embudo de ventas al cliente: evalúan la intención de compra del usuario, hacen comparaciones de productos en tiempo real, manejan objeciones comerciales complejas, aplican descuentos autorizados, procesan el pago dentro del chat de manera directa y segura. Luo et al. (2019) señalan que tener asistentes de inteligencia artificial listos para resolver preguntas difíciles, todos los días del año y a cualquier hora, elimina las esperas que

antes hacían abandonar los carritos de compra, lo cual provoca un incremento significativo de las ventas directas en servicios financieros y en plataformas de comercio electrónico.

Figura N° 11 Flujo de Generación aumentada por recuperación



Nota. Fuente: Imagen creada con IA

3.2.4 Motores y sistemas de recomendación inteligentes

Es fundamental examinar el efecto de los motores inteligentes de recomendación para finalizar este apartado de marketing, ya que son una de las aplicaciones más rentables que la inteligencia artificial ha tenido en la economía digital actual. Se ha podido comprobar que una parte importante de los ingresos de las principales plataformas, como Netflix, Amazon y Spotify, procede directamente de los productos o contenidos que les sugiere su algoritmo de IA. Los actuales sistemas de recomendación han superado a los modelos tradicionales de filtrado

colaborativo y se han transformado en agentes de recomendación conscientes del contexto (Ricci & al, 2015).

Estos agentes de inteligencia artificial están permanentemente analizando los datos sobre las preferencias pasadas del usuario, pero adaptan sus recomendaciones en función del contexto inmediato: la hora del día, el dispositivo que está utilizando, su ubicación geográfica y su estado emocional deducido a partir de sus búsquedas más recientes. También, mediante técnicas de aprendizaje por refuerzo, el agente aprende por su cuenta qué sugerencias consiguen un clic o una venta y cuáles son ignoradas, optimizando continuamente sus políticas de recomendación. Los motores de recomendación basados en agentes de inteligencia artificial (IA) no sólo facilitan la búsqueda de productos en amplios catálogos, sino que también operan como vendedores cruzados inteligentes (cross-selling y up-selling), incrementando sistemáticamente el valor medio de compra en plataformas de comercio electrónico.

3.3 Agentes de IA para la gestión operativa corporativa

3.3.1 Gestión documental e inteligencia de datos no estructurados

En la empresa moderna, más del 80% de la información empresarial valiosa se encuentra almacenada en datos no estructurados: contratos PDF, facturas escaneadas, correos electrónicos, auditorías e informes técnicos de mantenimiento. En el pasado, el procesamiento de esta enorme cantidad de documentos obligaba a que equipos enteros de personal administrativo leyera manualmente cada archivo para ingresar

datos importantes en los sistemas de planificación de recursos empresariales (ERP). No sólo este modelo manual era muy caro y lento, sino que generaba un gran número de errores en la escritura que afectaban la contabilidad y el rendimiento empresarial (Davenport & Ronanki, 2018).

La incorporación de agentes de inteligencia artificial especializados en gestión documental ha transformado profundamente este flujo de trabajo, otorgando a los sistemas la capacidad de leer y comprender semánticamente. Al incorporar modelos de Procesamiento del Lenguaje Natural (NLP), visión computacional para el reconocimiento óptico de caracteres (OCR) y técnicas de Generación Aumentada por Recuperación (RAG), estos agentes tienen la capacidad de leer, clasificar y extraer información crucial de miles de documentos en cuestión de segundos. Un agente de IA documental puede leer un contrato complejo de cien páginas, localizar las cláusulas de penalización por mora, comprobar la validez de las firmas digitales y registrar los compromisos financieros, todo ello de forma autónoma (Lewis & al, 2020).

La verdadera ventaja operativa de esta gestión documental inteligente, está en la capacidad de los agentes para razonar sobre el contexto y detectar anomalías, sin necesidad de reglas fijas. El agente, en vez de detenerse ante las variaciones de formato de un archivo, emplea modelos generativos para entender la intención del texto y validar los datos contra las bases de conocimiento internas de la compañía. Según el análisis de Baidoo-Anu y Ansah (2023), esta automatización cognitiva de datos no estructurados reduce los tiempos de auditoría en más de un 90%,

requiriendo la intervención de un abogado o auditor humano únicamente cuando identifica cláusulas atípicas o niveles de riesgo financiero fuera de los parámetros autorizados.

3.3.2 Agentes de IA en la gestión del talento humano

Los departamentos de Recursos Humanos de las empresas modernas han descubierto que los agentes de inteligencia artificial son aliados estratégicos para optimizar el ciclo de vida del colaborador dentro de la organización. En la primera etapa de atracción y selección de personal, los analistas de talento humano pasaban fatigas revisando a mano cientos de hojas de vida, lo que además de consumirles tiempo, los conducía a tomar decisiones sesgadas de manera inconsciente. Los cazatalentos autónomos son capaces de analizar los perfiles de los postulantes, evaluar sus proyectos previos y hacer evaluaciones iniciales de manera conversacional, calificando la idoneidad técnica de cada candidato con una precisión constante (Tambe & al, 2019).

Los agentes de IA no solo aceleran la contratación inicial, sino que también son fundamentales para personalizar los planes de formación y desarrollo continuo del personal existente. Cuando el empleado firma su contrato, agentes tutores virtuales diseñan planes de entrenamiento ajustados a sus vacíos específicos de conocimiento, guiándolo a través del proceso de inducción a través de interfaces conversacionales interactivas disponibles 24 horas al día. Este acompañamiento continuo acelera la curva de aprendizaje del nuevo personal, y garantiza que las

competencias del equipo se mantengan alineadas con las demandas cambiantes del mercado tecnológico y comercial.

En el terreno del clima organizacional y la retención del talento crítico, los agentes analíticos brindan capacidades de diagnóstico predictivo sin precedentes. Al analizar de manera anónima y agregada las interacciones en los canales de comunicación interna, los patrones de asistencia y las evaluaciones de desempeño, los agentes inteligentes pueden medir el índice de satisfacción laboral e identificar signos tempranos de agotamiento profesional (burnout). Wang et al. (2019) señala que este monitoreo preventivo permite a la gerencia de gestión humana intervenir a tiempo con políticas de incentivos y flexibilización laboral, reduciendo drásticamente las tasas de rotación no deseada de personal calificado.

3.3.3 Automatización de la gestión financiera y contable

La gestión financiera de una empresa requiere una precisión matemática absoluta, una transparencia total y el cumplimiento estricto de normativas tributarias y contables locales e internacionales. Los agentes de IA están tomando el control de tareas operativas de alta complejidad en este departamento, como la conciliación bancaria automatizada, la administración de cuentas por pagar y por cobrar, y la creación de presupuestos predictivos en tiempo real. Los contables solían comparar manualmente los estados bancarios con las facturas emitidas y recibidas,

buscando discrepancias de centavos; hoy, los agentes de IA pueden hacer esta comparación de forma instantánea y continua.

En el área del control de tesorería y gestión de riesgos corporativos los agentes inteligentes se comportan como detectores continuos de anomalías y fraudes financieros. Al observar cada movimiento de dinero que entra y que sale de la compañía, un agente financiero puede reconocer patrones raros de gastos como por ejemplo gastos fuera de lo normal a proveedores que se acaba de registrar o transferencias realizadas en horarios fuera de lo laboral para interrumpir la transacción de manera proactiva. El sistema, de forma inmediata, envía una alerta al auditor o al director financiero indicando la irregularidad detectada, acompañada de un informe detallado de los factores de riesgo identificados (Buczak & Guven, 2015).

Esta inteligente automatización financiera transforma el rol de los departamentos contables, pasando de ser unidades dedicadas al registro histórico de transacciones a centros analíticos estratégicos. Los agentes de IA producen simulaciones del flujo de caja futuro en función de variables macroeconómicas, tasas de interés e inflación, lo cual permite a la tesorería corporativa optimizar la colocación de los excedentes de capital o anticipar necesidades de financiamiento con meses de antelación. Como señala Goodell (2020), la inclusión de agentes de IA en la gestión financiera, además de disminuir los costos operacionales de la administración contable, resguarda el patrimonio de la compañía ante eventualidades económicas y ataques de ciberfraude.

Figura N° 12 Flujo de auditoría y evolución



Nota. Fuente: Imagen creada con IA

3.3.4 Inteligencia de Negocios (Business Intelligence) asistida por agentes

La Inteligencia de Negocios tradicional dependía en gran medida de analistas de datos especializados que debían escribir complejas consultas de bases de datos (SQL) y construir tableros estáticos (dashboards) para que la alta dirección pudiera revisar las métricas de la empresa al final de cada periodo. Pero en la velocidad del negocio actual, esperar días o semanas por un reporte de datos, es tomar decisiones con datos del pasado. La Inteligencia de Negocios asistida por agentes de IA está revolucionando esta dinámica al democratizar el acceso a la información y permitir la exploración analítica en tiempo real a través del lenguaje natural (Sharda & al, 2017).

Hoy en día, un gerente general o director de operaciones no requiere saber programar ni pedir informes al área de tecnologías de información,

simplemente debe hacer una pregunta directa en lenguaje cotidiano al agente de IA institucional a través de su teléfono o computadora. El agente traduce automáticamente la pregunta en una serie de instrucciones lógicas que se aplican a las bases de datos corporativas, examina millones de registros de ventas en cuestión de segundos y responde con un resumen explicativo, tablas comparativas y gráficos dinámicos que se adaptan a la pregunta.

Esta inmediatez en la consulta analítica permite transformar los datos masivos de la empresa en decisiones ejecutivas oportunas, permitiendo corregir el rumbo del negocio antes de que un problema afecte los resultados financieros trimestrales. Los agentes de BI no sólo responden preguntas pasivas, sino que, además, ejecutan algoritmos de minería de datos en segundo plano para descubrir correlaciones ocultas en las ventas o patrones de consumo que los analistas humanos no habían notado. Según Davenport y Bean (2021), el uso de Business Intelligence asistido por agentes conversacionales ayuda a las empresas a alcanzar una cultura verdaderamente basada en datos, acelerando el ciclo de toma de decisiones en todos los niveles de la organización.

3.4 Agentes de IA en el ecosistema de microempresas y PYMES

3.4.1 Democratización de la tecnología para pequeños negocios

En el pasado, las tecnologías de la información más avanzadas eran accesibles casi únicamente para las grandes corporaciones multinacionales que disponían de presupuestos suficientemente cuantiosos como para contratar consultoras de software y construir

infraestructuras físicas de altísima capacidad. Las PYMES y los emprendimientos locales debían trabajar con herramientas rudimentarias y procesos manuales por falta de capital y personal técnico especializado. Los agentes de inteligencia artificial surgidos en la nube bajo modelos de software han roto definitivamente esta brecha histórica, dando inicio a una era sin precedentes de democratización tecnológica (Iansiti & Lakhani, 2020).

Hoy en día, una microempresa con pocos empleados puede tener exactamente el mismo nivel de poder analítico y automatización cognitiva que una gran multinacional, pagando solo por suscripciones mensuales ajustadas al uso de agentes de IA en la nube. Esta accesibilidad económica y técnica permite que un pequeño negocio local pueda automatizar su contabilidad, gestionar de forma inteligente sus inventarios y diseñar campañas publicitarias de gran impacto sin necesidad de contratar ingenieros informáticos ni agencias especializadas (Nimje, 2024).

Este acceso equitativo a herramientas inteligentes redefine la dinámica competitiva en el mercado global, posibilitando que los pequeños emprendimientos compitan en agilidad y calidad de servicio con las corporaciones establecidas. Las PYMES que integran agentes de IA consiguen montar flujos de trabajo ultra eficientes, que compensan su menor escala de capital con una velocidad de ejecución sobrehumana. Según Prinz et al. (2021), la democratización de la IA no solo potencia la supervivencia de las pequeñas empresas, sino que sirve como impulsor

de innovación y desarrollo económico en los mercados locales y regionales.

3.4.2 Automatización integral de procesos de bajo costo

Para una pequeña empresa o microemprendimiento el recurso más escaso y crítico no suele ser solo el dinero, sino el tiempo de los fundadores y empleados clave, que deben dividir sus jornadas diarias entre atender clientes, llevar la contabilidad, contestar mensajes, empaquetar productos y negociar con proveedores. Esta sobrecarga de tareas operativas rutinarias, consume la energía intelectual del equipo e impide a los dueños de PYMES concentrarse en la planificación estratégica y en el crecimiento comercial de su negocio (Brynjolfsson & McAfee, 2014).

Los agentes de IA de bajo coste, que se integran en plataformas de desarrollo sin código (no-code) o de bajo código (low-code), ponen solución a esta problemática asumiendo de forma autónoma las tareas administrativas mecánicas del día a día. Una microempresa, sin necesidad de conocimientos de programación y mediante sencillas configuraciones visuales, puede desplegar agentes que agenden automáticamente citas con clientes, emitan facturas electrónicas inmediatamente después de efectuada una venta, envíen recordatorios de cobro y actualicen los niveles de stock en tiempo real cuando se agota un producto (van der Aalst & al, 2018).

Los emprendedores, al delegar estas tareas repetitivas en agentes digitales que operan continuamente y no se equivocan por agotamiento,

recuperan valiosas horas semanales de trabajo. El tiempo que se libera se redirige directamente a actividades de alto valor añadido como la innovación en el diseño de nuevos productos, la búsqueda de alianzas comerciales estratégicas y la atención presencial personalizada a los clientes más importantes. Según Agostinelli et al. (2021), la hiperautomatización de bajo costo permite a las PYMES escalar sus operaciones y multiplicar el volumen de facturación sin necesidad de aumentar prematuramente sus costos fijos operacionales.

3.4.3 Comercio electrónico inteligente para PYMES

El comercio electrónico ha sido un canal de ventas fundamental para miles de pequeñas y medianas empresas que desean vender sus productos más allá de las fronteras geográficas de su localidad. Pero es muy difícil competir en el entorno digital contra estas grandes plataformas de e-commerce que tienen recomendaciones personalizadas en tiempo real y servicio al cliente 24 horas. Los agentes de inteligencia artificial especializados en comercio electrónico, ofrecen a las PYMES la oportunidad de convertir sus sencillas tiendas virtuales en plataformas de venta altamente inteligentes y competitivas.

Un agente de IA para e-commerce se comporta como un asesor de ventas virtual muy capacitado que conversa con el usuario a través de la página web o las redes sociales de la empresa. El agente mediante fluidas conversaciones en lenguaje natural comprende las necesidades del cliente, analiza sus preferencias y presupuesto, y le sugiere los productos de catálogo que mejor se ajusten a sus requerimientos específicos.

Además, el sistema es capaz de realizar de manera autónoma el seguimiento de carritos de compra abandonados, enviando ofertas personalizadas con la intención de motivar al cliente a completar su transacción (Treiblmaier & Sillaber, 2021).

Para una pequeña o mediana empresa, tener agentes de ventas digitales que puedan atender a cientos de compradores al mismo tiempo con el mismo nivel de amabilidad, velocidad y efectividad, se traduce en un incremento directo en las tasas de conversión y en la fidelización del cliente. Esta atención omnicanal y continua elimina las barreras temporales y espaciales en las ventas en Internet, lo que permite a las empresas locales consolidar una sólida presencia digital. En definitiva, la incorporación de agentes inteligentes en el comercio electrónico permite a las PYMES desarrollar modelos de negocio resilientes, sostenibles e integrados en la economía digital global.

3.5 Agentes de IA en la toma de decisiones estratégicas

3.5.1 Analítica predictiva y prescriptiva

En el entorno corporativo tradicional, los gerentes tomaban sus decisiones sobre la base de un enfoque eminentemente reactivo, en el cual los comités ejecutivos examinaban reportes contables y financieros que mostraban solamente lo que había ocurrido hace semanas o meses atrás. Este modelo de gestión histórica presentaba severas limitaciones en mercados dinámicos, obligando a que los líderes de negocios actuaran cuando los problemas ya habían afectado la rentabilidad de la organización. La inclusión de agentes de inteligencia artificial ha

cambiado radicalmente esta dinámica, introduciendo un paradigma analítico proactivo capaz de procesar volúmenes masivos de datos en tiempo real, anticipando escenarios futuros y guiando la estrategia corporativa con una precisión matemática imposible de alcanzar por los métodos analíticos del pasado (Fitzgerald & al, 2014).

La evolución desde la analítica predictiva a la analítica prescriptiva es el salto tecnológico más cuantitativo dentro de la gestión ejecutiva moderna. La analítica predictiva, a través de algoritmos de Machine Learning, estima los eventos o patrones de demanda que se producirán en el futuro, mientras que la analítica prescriptiva evalúa miles de variables interconectadas de forma autónoma y recomienda a la alta gerencia la acción concreta y óptima para capitalizar una oportunidad o mitigar un riesgo operativo. Un agente prescriptivo no sólo alerta sobre un problema que se avecina, sino que simula los diferentes caminos que se pueden seguir, calcula las probabilidades de éxito que tiene cada alternativa y formula una hoja de ruta detallada para que la dirección tome las decisiones empresariales (Sharda & al, 2017).

Figura N° 13 Madurez analítica en la toma de decisiones

Los cuadros de mando o dashboards empresariales de primera generación están comenzando a presentar signos claros de obsolescencia frente a las necesidades de la economía digital actual. Estos tableros de control solían diseñarse como gráficos estáticos que acumulaban una abrumadora cantidad de métricas y de indicadores clave de rendimiento (KPIs), obligando a los ejecutivos a dedicar horas valiosas a interpretar gráficos complejos para tratar de entender las causas de un problema operativo. Las nuevas interfaces inteligentes superan definitivamente esta saturación de datos al convertirse en entornos interactivos impulsados por agentes de IA que conversan directamente con los usuarios de negocio en lenguaje natural (Few, 2013).

Esa es la principal diferencia de estos tableros inteligentes: pueden pasar de ser meros mostradores de datos a analistas cognitivos activos dentro de la organización. Un agente de IA integrado en la interfaz, monitorizando continuamente el flujo de datos corporativos, es capaz de detectar anomalías estadísticas, tendencias emergentes o desviaciones en los costos operativos sin requerir una consulta previa por parte del usuario. En lugar de ofrecer frías tablas numéricas, la interfaz inteligente lanza alertas contextuales explicativas que detallan con precisión de dónde surge el problema y proponen a la gerencia las medidas de corrección más adecuadas (Davenport & Bean, 2021).

Esa evolución hacia las interfaces conversacionales inteligentes, para la alta dirección de las corporaciones actuales, democratiza el acceso a la Inteligencia de Negocios y acelera los tiempos de respuesta de forma notable. Un director general ya no necesita la mediación de un equipo de analistas para comprender el estado de su empresa. Ahora basta con

plantear una pregunta directa en lenguaje cotidiano a la interfaz para obtener explicaciones profundas, gráficos adaptados y proyecciones inmediatas. Así, como argumentan Provost y Fawcett (2013), los dashboards asistidos por IA dejan de ser herramientas pasivas de reporte para convertirse en asesores estratégicos en tiempo real que simplifican la complejidad informativa de la empresa moderna.

3.5.3 Sistemas expertos modernos y soporte decisiones

Los sistemas de soporte a decisiones (DSS) están experimentando una profunda reinención conceptual gracias a la convergencia entre las bases de conocimiento no estructuradas de las empresas y los modelos de lenguaje a gran escala basados en arquitecturas RAG (Retrieval-Augmented Generation). Los primeros sistemas expertos de finales del siglo pasado se basaban en árboles de reglas rígidas codificados manualmente, que requerían de costosas remodelaciones técnicas cada vez que se modificaba una ley o un procedimiento interno. Los sistemas expertos actuales, en cambio, actúan como avanzados consultores virtuales que pueden leer, entender y razonar sobre miles de documentos corporativos en formato de texto libre (Power, 2002).

Esta habilidad para procesar e interpretar información no estructurada es de gran valor estratégico para sectores con alta complejidad regulatoria. Por ejemplo, cuando una institución financiera debe decidir sobre la concesión de préstamos corporativos multimillonarios, el sistema experto actual no solamente se centra en el análisis de las cifras contables del solicitante, sino que además cruza la operación con la legislación bancaria vigente, evalúa el historial de cumplimiento tributario del

cliente y emite un informe riguroso donde detalla las razones normativas por las cuales recomienda o no la autorización de la transacción (Turban & Volonino, 2011).

Estos sistemas de soporte a decisiones permiten a las organizaciones estandarizar la calidad de sus decisiones estratégicas y minimizar el impacto del error humano en procesos críticos. Más que sustituir el criterio humano, estos sistemas expertos operan bajo esquemas de colaboración donde el agente de IA lleva a cabo la investigación documental y el análisis comparativo pesado, dejando la decisión final en manos del ejecutivo responsable. De acuerdo con Sharda et al. (2017), este modelo de decisiones aumentadas aporta un nivel de rigor analítico permanente a lo largo de toda la organización, protegiendo a la empresa frente a posibles contingencias legales y mejorando su eficiencia operativa.

3.6 Modelos de Negocio Emergentes Basados en Agentes de IA

3.6.1 AI as a service (AIaaS)

La rápida expansión de los agentes de inteligencia artificial ha alterado profundamente las estructuras de costos de la industria tecnológica, impulsando la consolidación del modelo de negocio de Inteligencia Artificial como Servicio (AI como Servicio - AIaaS). En el paradigma de

desarrollo de software anterior, si una organización quería incorporar capacidades analíticas avanzadas, tenía que realizar importantes inversiones de capital para comprar licencias propietarias, adquirir la infraestructura física de servidores y contratar equipos especializados para mantener el sistema. El modelo AIaaS elimina definitivamente estas barreras financieras permitiendo que las empresas consuman capacidades cognitivas avanzadas bajo demanda a través de suscripciones en la nube o pagos por uso de APIs (Armbrust & al, 2010).

Para las empresas proveedoras de tecnología, el modelo AIaaS se erige como un canal de expansión comercial sin precedentes en la economía digital. Un solo agente de IA que procese documentos, traduzca idiomas o evalúe riesgos crediticios puede venderse simultáneamente a miles de corporaciones en distintas partes del mundo, con un costo marginal de distribución cercano a cero. Esta arquitectura en la nube permite actualizar constantemente los modelos, con el objetivo de que todos los clientes tengan las últimas mejoras algorítmicas, sin que esto afecte sus flujos operativos (Iansiti & Lakhani, 2020).

Para las pequeñas, medianas y grandes empresas usuarias, el modelo AIaaS les permite pasar de un gasto de capital fijo e inflexible (CapEx) a un gasto operativo variable y escalable (OpEx). Esta flexibilidad financiera permite que un negocio aumente o disminuya su uso de agentes inteligentes según la estacionalidad de sus ventas o sus cambiantes necesidades operativas, pagando solamente por el poder de cálculo que realmente utiliza. Brynjolfsson y McAfee (2014) plantean que la adopción de AIaaS democratiza la innovación tecnológica a nivel

mundial, posibilitando que organizaciones de cualquier tamaño compitan en igualdad de condiciones operativas frente a los líderes consolidados de la industria.

3.6.2 Agentes como asistentes y empleados corporativos virtuales

Una de las tendencias de negocios más disruptivas en la era digital es el cambio desde las aplicaciones de software convencionales hacia el concepto de fuerza laboral sintética o empleados virtuales especializados (digital workers). Un software clásico es una herramienta pasiva a la que el trabajador humano debe recurrir para hacer un cálculo o guardar un registro; un empleado virtual constituido como un agente de IA es una entidad autónoma que tiene un rol corporativo definido, un correo institucional, acceso seguro a los sistemas de la compañía y objetivos operativos concretos que debe cumplir de forma independiente (Davenport & Miller, 2022).

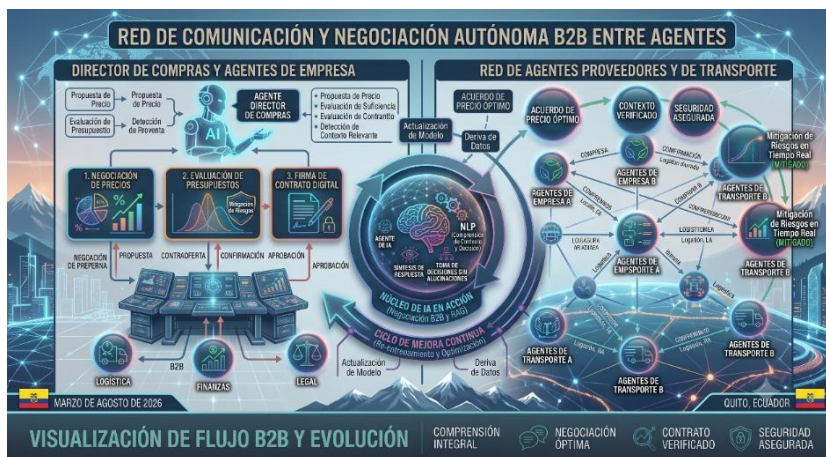
En las corporaciones contemporáneas, estos trabajadores telemáticos están adquiriendo funciones oficiales en sectores tales como la verificación contable, la administración de compras y el soporte técnico al cliente. Un agente contratado como “analista de compras virtual”, por ejemplo, podría monitorear constantemente las solicitudes de suministros provenientes de las distintas sucursales de la empresa, enviar correos electrónicos solicitando cotizaciones a varios proveedores externos, evaluar las propuestas considerando tanto el costo como el tiempo de entrega, y redactar finalmente la orden de compra lista para la autorización del gerente de área (Wooldridge, 2020).

Esta integración de empleados sintéticos reconfigura la división del trabajo dentro de la organización, permitiendo que las tareas administrativas repetitivas de alto volumen sean gestionadas por agentes digitales que operan las 24 horas del día, sin errores por cansancio. Esta delegación de la carga operativa libera al personal humano de la rutina burocrática, permitiéndole reorientar sus esfuerzos intelectuales hacia la creatividad, la innovación de productos, el liderazgo de equipos y la construcción de relaciones comerciales de alto valor. Como señalan Brynjolfsson y Hitt (2000), la formación de equipos híbridos donde humanos y empleados virtuales trabajen conjuntamente de forma sistemática se traduce en un incremento sustancial de la productividad y la rentabilidad del negocio.

3.6.3 Ecosistemas inteligentes y redes de agentes interconectados

El horizonte más avanzado de la transformación digital empresarial es el de pasar del uso de agentes de IA aislados dentro de una única compañía a la construcción de ecosistemas de agentes interconectados (Multi-Agent Systems). En estos entornos de red mundial, los agentes de inteligencia artificial de distintas empresas, proveedores, bancos e instituciones públicas se hablan, negocian y hacen negocios directamente entre ellos a través de protocolos digitales estandarizados, sin necesidad de que un humano intermediario valide cada paso (Wooldridge, 2020).

Figura N°14 Flujo B2B entre agentes



Nota. Fuente: Imagen creada con IA

Imaginemos un proceso automatizado logístico e industrial en el que el agente de IA de una planta de ensamblaje detecta que se agotará el inventario de un componente electrónico en las próximas 48 horas. Enseguida, el agente hace una solicitud a la red y se conecta a un intercambio automatizado con los agentes de IA de diferentes fabricantes y empresas de transporte internacional; evalúa los costos, revisa las disponibilidades, negocia los plazos de entrega, firma digitalmente el contrato por medio de claves criptográficas y efectúa el pago correspondiente. En milisegundos y de forma totalmente autónoma se realiza todo este complejo proceso comercial B2B (Tapscott & Tapscott, 2016).

La consolidación de estas redes de agentes interconectados promete llevar a su mínima expresión los costos de transacción y los tiempos burocráticos que históricamente han frenado el comercio internacional. Al eliminar las demoras operativas generadas por la gestión manual de correos, facturas y aprobaciones, la cadena de valor global alcanza un nivel de fluidez y adaptación instantánea sin precedentes. Tal y como

concluyen Iansiti y Lakhani (2020), la habilidad para integrar a las empresas dentro de estos ecosistemas de agentes autónomos será el factor clave que defina la relevancia comercial y la competitividad de las organizaciones en la economía digital del futuro.

3.7 Indicadores de impacto empresarial y retorno de inversión

3.7.1 Medición de impacto en la productividad operativa

Para que cualquier estrategia o proyecto de adopción de agentes de inteligencia artificial pueda considerarse legítimamente exitoso dentro del entorno corporativo contemporáneo, es totalmente indispensable dejar de lado la fascinación por la novedad tecnológica y evaluar con rigor numérico el impacto real de estos agentes en la productividad de las operaciones. En las ciencias administrativas contemporáneas, la evaluación del impacto operativo se orienta hacia la medición cuantitativa de la eficiencia temporal y la aceleración de los ciclos laborales, contrastando indicadores precisos entre el tiempo que demandaba a un equipo humano realizar un proceso manual tradicional, frente al tiempo que requiere el mismo flujo de trabajo cuando es gestionado por un sistema de agentes inteligentes (Kaplan & Norton, 1996).

En sectores con grandes volúmenes de transacciones no estructuradas (procesamiento de reclamos en compañías de seguros, evaluación de solicitudes de crédito bancario o procesamiento de facturas en cadenas de suministro), las métricas operativas muestran verdaderas reducciones de tiempo. Los procesos complejos que antes necesitaban semanas de

trámites burocráticos hechos a mano en papel, ahora se hacen en minutos de forma autónoma gracias a los agentes de IA. Esta compresión del tiempo de respuesta no solo hace más ágil al negocio, sino que también incrementa la capacidad de procesamiento de la compañía sin incurrir en aumentos proporcionales de costos fijos (Davenport & Ronanki, 2018).

De igual manera, se mide estrictamente la productividad operativa por la tasa de reducción de errores humanos y reprocesos. En la organización tradicional la fatiga del personal que efectúa tareas repetitivas de digitación o de revisión documental produce un porcentaje inevitable de fallas, que conllevan a costosos reprocesos administrativos o a sanciones regulatorias. Los agentes de inteligencia artificial trabajan con un nivel de consistencia y precisión constante, las 24 horas del día, eliminando los errores de digitación o de interpretación en la gestión contable. Según el análisis de Brynjolfsson y McAfee (2014), la combinación de una velocidad de ejecución superior a la humana y la eliminación casi total de los errores mecánicos genera una ventaja competitiva estructural, que redefine los estándares de productividad en la industria.

3.7.2 Incremento de la rentabilidad financiera

En la iniciativa privada, el objetivo prioritario de toda estrategia de transformación digital es la creación de valor económico sostenible y el incremento de la rentabilidad financiera para los accionistas y propietarios del negocio. La incorporación estratégica de agentes de inteligencia artificial incide directamente y positivamente en el estado de resultados de las corporaciones a través de dos mecanismos financieros simultáneos: la reducción sostenida de los costos operativos fijos y la

generación activa de nuevas fuentes de ingresos comerciales (Porter, 1985).

En el terreno de la optimización de costos, la automatización cognitiva basada en agentes IA permite a las empresas lograr economías de escala de forma mucho más ágil. Las compañías pueden aumentar exponencialmente su volumen de clientes o de transacciones diarias, al delegar la carga administrativa pesada a empleados virtuales sintéticos, sin tener que expandir de forma lineal su plantilla de personal administrativo o contratar espacio físico de oficina adicional. Esta contención de costos fijos eleva inmediatamente el margen operativo (EBITDA) y fortalece la estructura financiera de la organización frente a posibles contracciones o crisis del mercado (Iansiti & Lakhani, 2020).

En cuanto al crecimiento de los ingresos, los agentes de IA impulsan las ventas ejecutando campañas comerciales hiper-personalizadas, implementando motores de recomendación cruzada en tiempo real y poniendo a disposición asistentes conversacionales 24/7, que evitan la pérdida de prospectos comerciales durante las horas no laborables. Una respuesta ágil a las inquietudes del comprador y una sugerencia puntual de productos complementarios incrementan de manera regular el promedio de compra por ticket y reducen las tasas de abandono de clientes. Así, la sinergia entre el control riguroso de los gastos operativos y el incremento en las ventas brutas genera un impacto multiplicador sobre la rentabilidad financiera neta del negocio, según explican Chintagunta et al. (2016).

3.7.3 Metodologías para el cálculo del retorno de inversión (ROI) en proyectos de IA

Para determinar la viabilidad financiera y la justificación económica de un proyecto de inversión en agentes de IA, es necesario emplear metodologías de cálculo de Retorno de Inversión (ROI) que vayan más allá de las fórmulas contables tradicionales. El cálculo tradicional del ROI debe ser adaptado analíticamente para que capte de forma precisa tanto los ahorros cuantitativos directos a corto plazo como los beneficios cualitativos e intangibles a medio y largo plazo (Remenyi & al, 2007).

Los analistas financieros corporativos deberán costear integralmente la variable de Inversión Total, incluyendo no solo el valor explícito de las licencias de software o el pago por llamadas a APIs en la nube, sino todos los costos ocultos asociados a la implementación. Entre estos gastos destacan los de preparación, limpieza y gobernanza de datos, la integración técnica del sistema con la infraestructura legada de la empresa, las horas de capacitación del personal humano y los honorarios de auditorías de ciberseguridad e impacto ético. Ignorar cualquiera de estos factores distorsionaría el análisis financiero y llevaría a expectativas falsas de retorno para la junta directiva (Brynjolfsson & Hitt, 2000).

Por otra parte, para cuantificar los Beneficios Netos es necesario contabilizar el valor monetario de las horas-hombre liberadas en tareas administrativas, la reducción en los costos de adquisición de clientes (CAC), las pérdidas evitadas por la prevención de fraudes contables y el

incremento neto de las ventas atribuible a la IA. Sharda et al. (2017) encontraron que los proyectos de agentes de IA bien diseñados y alineados con la estrategia del negocio tienden a tener un Payback menor a 12 meses. Esa pronta amortización confirma que la inteligencia artificial no es un gasto tecnológico de más, sino una de las decisiones financieras más rentables y estratégicas en la economía empresarial contemporánea.

CAPÍTULO IV

AGENTES DE INTELIGENCIA ARTIFICIAL EN LA CIBERSEGURIDAD

4.1. Introducción a la Ciberseguridad

El internet tiene el potencial de ser la innovación más relevante del siglo XXI, y tiene un impacto en la vida cotidiana de las personas en más

maneras de las que piensan. El internet también ha propiciado la aparición de nuevas modalidades delictivas, las cuales son denominadas delitos cibernéticos. Aunque Internet se transformó en "el espacio" donde las personas pasan su tiempo y socializan, también brinda grandes oportunidades a los ciberdelincuentes. Los criminales se percataron de que, dado que todo ha pasado a ser digital, también debían adaptarse y hacerse digitales. No les costó entender que el mundo entero emplea los servicios de Internet para actividades como pedir comida, hacer compras, realizar operaciones bancarias y más. Además, todos ellos tienen algo en común: realizar transacciones financieras mediante medios digitales.

4.1.1 Conceptos fundamentales de Ciberseguridad

Afirmar en su libro Kiser, Q (2020) en las últimas décadas, la dependencia del mundo de la tecnología es mayor que nunca. Esto ha provocado un incremento significativo en la generación de datos digitales. Las computadoras guardan datos de personas, empresas, agencias gubernamentales y otras entidades, los cuales se trasladan a otras computadoras mediante una red todos los días. Una brecha en el sistema que brinda acceso a un atacante es capaz de tener efectos muy significativos. La divulgación de información personal puede impactar negativamente en la reputación de una empresa y provocar que los clientes y los socios pierdan confianza.

La ciberseguridad, que también se conoce como seguridad informática, es el proceso de garantizar el ICA de la información, donde I representa Integridad, C Confidencialidad y A Disponibilidad. La ciberseguridad incluye un grupo de herramientas significativas, métodos para

administrar riesgos, tecnologías, capacitación y enfoques para salvaguardar la información, los dispositivos, las redes y los programas contra accesos no autorizados y ataques.

Aporta también Walker, B (2019) que la ciberseguridad está relacionada con el ciberespacio. El término "ciberespacio" hace referencia a un espacio interactivo que incluye redes digitales, las cuales recogen, guardan y procesan información para facilitar diversas maneras de comunicarse. Así, el ciberespacio abarca Internet y varios sistemas que respaldan a los servicios, la infraestructura y los usuarios de Internet. La ciberseguridad tiene gran relevancia, aunque su consecución es bastante complicada. Existen numerosas dinámicas en juego que dificultan la naturaleza de la seguridad cibernética, pero es importante que hagamos un esfuerzo por conservar el equilibrio ante un ecosistema cibernético en permanente cambio. Lo que hagamos hoy determinará el futuro de la seguridad cibernética.

4.1.2 Amenazas digitales actuales

Entender los delitos que pueden ser perpetrados por un individuo a través de plataformas digitales, tales como el Internet y las computadoras, también es crucial. Estos son los distintos tipos de delitos informáticos que la gente comete hoy en día.

Una de las amenazas digitales presentes en la actualidad es el ransomware. Este tipo de malware cifra toda la información de su

computadora y le impide acceder a ella. El atacante que utilizó el ransomware dejará una aplicación para que le pague; después de eso, los archivos se descifrarán automáticamente. En 2017, un ransomware llamado WannaCry logró infectar a todos los dispositivos con Windows.

Clickjacking también está presente. Este es un ataque que engaña al usuario para que haga clic en un botón o enlace que parece auténtico, pero contiene scripts dañinos insertados. Este libro únicamente ha abarcado una pequeña parte de la terminología de ciberseguridad, pero esto funcionará como plataforma para un entendimiento más profundo. (Kiser, 2020)

Una de las amenazas digitales más críticas a las que se enfrentan tanto organizaciones como usuarios es el robo y la filtración de datos, debido al impacto que esto tiene en la confidencialidad, la privacidad y la continuidad de las operaciones. El robo de datos se produce cuando un atacante logra acceder sin autorización a información sensible, con el objetivo de usarla, venderla o explotarla en busca de beneficios económicos o estratégicos. La filtración de datos, o data breach, sucede cuando la información se expone, se revela o se transmite a personas no autorizadas, ya sea de manera intencional o accidental.

La manipulación de información y la desinformación son también una amenaza digital que consiste en la creación, alteración o difusión deliberada de información falsa, engañosa o fuera de contexto con el propósito de influir en las opiniones, decisiones o comportamientos de las personas. Estas prácticas tienen como objetivo generar confusión, desacreditar a personas o instituciones, influir en procesos políticos o sociales, dañar la reputación de organizaciones o promover intereses

económicos, ideológicos o estratégicos. (Acosta, Benavides, & Garcia, 2020)

4.1.3 Ciberdelincuencia y riesgos organizacionales

Según Cavada, J (2020) el término ciberdelincuencia no cuenta con una definición jurídica única y universalmente aceptada, debido a las diferencias existentes entre las legislaciones nacionales y los instrumentos internacionales. No obstante, existe un amplio consenso en la comunidad académica y entre los organismos especializados en que el ciberdelincuencia comprende el conjunto de actividades delictivas que utilizan las tecnologías de la información y las comunicaciones (TIC) como medio para cometer un delito o que tienen como objetivo sistemas informáticos, redes, dispositivos o la información que estos almacenan, procesan o transmiten. En la literatura especializada también se emplean términos como delincuencia informática, delitos informáticos, criminalidad informática o ciberdelincuencia, los cuales, aunque presentan ligeras diferencias según el contexto jurídico o doctrinal, hacen referencia a conductas ilícitas relacionadas con el uso de las tecnologías digitales. En este sentido, el concepto de ciberdelincuencia engloba tanto los delitos tradicionales que han evolucionado hacia el entorno digital, como el fraude electrónico, el robo de identidad o la estafa mediante phishing, como aquellos delitos que solo pueden existir gracias al desarrollo de las tecnologías de la información, entre ellos el acceso ilícito a sistemas, la distribución de malware, los ataques de denegación de servicio (DDoS), el ransomware y otras formas de ciberdelincuencia dirigidas contra infraestructuras y activos digitales.

Figura N°15 Agentes de IA en Ciberseguridad



Nota. Fuente: Imagen creada con IA

Por otro lado, Caamaño, E; Gil, R (2024) enfatiza en el avance de las tecnologías de la información ha impulsado la transformación digital de las organizaciones, lo que ha mejorado la productividad, la innovación y la calidad de los servicios prestados. Sin embargo, este desarrollo ha incrementado también la exposición al cibercrimen, creando riesgos que pueden poner en jaque la continuidad de las operaciones, la integridad de los activos de información y la confianza de clientes y demás partes interesadas. Las organizaciones están expuestas a amenazas como el fraude, el robo de información, la corrupción, el soborno y diferentes delitos informáticos, los cuales pueden generar pérdidas económicas, daños en su reputación, e incluso afectar la estabilidad de sectores estratégicos. En este sentido, la adopción de medidas de ciberseguridad y una gestión adecuada de los riesgos son fundamentales para evitar la explotación de vulnerabilidades, salvaguardar la infraestructura tecnológica y asegurar la resiliencia ante las crecientes amenazas del entorno digital.

4.2. Principios de seguridad de la información

Conforme el ámbito empresarial continúa expandiéndose, las amenazas se vuelven más complejas y los costos de corrección aumentan aún más. La ciberseguridad se trata de salvaguardar la confidencialidad, integridad y disponibilidad de los datos. No es sencillo proteger dispositivos y redes, dado que se requiere protección frente a una variedad de ataques distintos, algunos de los cuales no eres consciente.

4.2.1 Confidencialidad

Es un campo de la informática que se centra en ofrecer herramientas técnicas que aseguren la confidencialidad, la integridad y la disponibilidad de los datos. La confidencialidad asegura que solo quienes están debidamente autorizados puedan acceder a la información o consultarla; la integridad, que esta no ha sido alterada sin autorización, y la disponibilidad garantiza (aunque suene redundante) que siempre esté disponible cuando se necesite. La información no es segura si uno de estos elementos falla.

Implementar medidas de confidencialidad es esencial para asegurar que la información solo esté al alcance de individuos autorizados, disminuyendo así el peligro de accesos no permitidos y filtraciones de datos. A tal efecto, las entidades implementan controles de acceso que comprenden procedimientos de autenticación para confirmar la identidad de los usuarios antes de concederles el acceso a información sensible. Además, utilizan métodos de cifrado que salvaguardan los datos tanto cuando están en tránsito como cuando están almacenados, evitando que

terceros puedan descifrarlos sin autorización. Asimismo, se utilizan técnicas de enmascaramiento y anonimización de datos, que ocultan o alteran la información confidencial para prevenir la exposición de datos delicados, aun cuando estos sean empleados para análisis, pruebas o intercambio de información. (Astudillo, 2025)

4.2.2 Integridad

Para Kiser, Q (2020) la integridad de los datos es un principio de la seguridad de la información que garantiza que la información permanezca completa, exacta, coherente y libre de modificaciones no autorizadas durante todo su ciclo de vida. Esto implica que los datos se mantengan intactos entre diferentes actualizaciones o consultas, asegurando que cualquier cambio realizado sea válido, autorizado y correctamente registrado. La integridad se establece desde la fase de diseño de la base de datos mediante la aplicación de reglas, restricciones y procedimientos estandarizados, y se preserva a través de mecanismos de comprobación de errores y procesos de validación que detectan inconsistencias o alteraciones.

Además, para fortalecer este principio, las organizaciones implementan protocolos de autenticación, políticas de seguridad como aquellas relacionadas con la creación y gestión de contraseñas y sistemas de control de acceso que asignan permisos específicos a los usuarios, definiendo quién puede acceder, modificar o gestionar determinada información. Asimismo, es importante mantener bloqueadas o eliminar

las cuentas que ya no se utilizan, reduciendo así el riesgo de accesos no autorizados y contribuyendo a la protección de la integridad de los datos.

El portal web Fortinet (2026) menciona que, con frecuencia, se compromete deliberadamente la integridad. Un atacante tiene la posibilidad de ignorar un sistema de detección de intrusos (IDS, por sus siglas en inglés), modificar las configuraciones de los archivos para conceder acceso no autorizado o alterar los registros que el sistema guarda con el fin de encubrir el ataque y comprometer la seguridad informativa. La integridad también puede ser vulnerada de manera accidental. Es posible que alguien, por descuido, cometa un error de otro tipo o introduzca el código equivocado. Asimismo, si los procedimientos, políticas y protecciones de seguridad de la compañía son insuficientes, es posible que la integridad se vea comprometida sin que nadie dentro de la organización tenga responsabilidad.

4.2.3 Disponibilidad

Según Kiser, Q (2020) el término "disponibilidad de información", como se mencionó anteriormente, corresponde a la rama del saber y la práctica informática dentro de la seguridad de la información que permite acceder a los datos de individuos y entidades con los que es posible trabajar. La seguridad de la información se fundamenta en la confidencialidad, la integridad, la autenticación y la disponibilidad; estos elementos son los pilares sobre los que se sustentan casi todas las herramientas y soluciones digitales utilizadas diariamente por las empresas.

Podríamos describir la disponibilidad de manera más técnica: significa que la información debe estar disponible para aquellos que quieran acceder a ella, incluidas las aplicaciones, las operaciones y los usuarios. Este acceso debe llevarse a cabo únicamente para las personas que estén autorizadas en cada situación.

En áreas de negocios con un flujo de información elevado, mejorar el rendimiento de sus equipos y colaboradores, además de los procesos, está estrechamente relacionado con posibilitar una circulación continua de datos entre ambos lados, para que cada integrante de la organización tenga acceso a ellos oportunamente y pueda realizar su trabajo eficazmente. En este contexto, si la información no está disponible, puede obstaculizarse el flujo de trabajo y perjudicar algunos servicios que la organización brinda a su público. (Arias, 2025)

4.3 Aplicación de IA en Ciberseguridad

El uso de la inteligencia artificial (IA) en la ciberseguridad ha cambiado cómo las entidades detectan, evitan y reaccionan ante los peligros informáticos. La IA, debido a su habilidad de examinar grandes cantidades de datos, reconocer patrones de conducta y aprender constantemente, hace posible que se fortalezcan los sistemas de protección ante ataques cada vez más complejos. En este segmento se exponen las aplicaciones más relevantes de la inteligencia artificial (IA) en el campo de la ciberseguridad, enfatizando su papel en la detección de riesgos.

4.3.1 Detección de anomalías

La identificación de anomalías mediante inteligencia artificial es un proceso donde un modelo de IA examina una serie de datos y señala los registros que son considerados atípicos en relación a una base de referencia, la cual refleja una conducta normal y funciona como el punto inicial para la comparación. La referencia esperada para un conjunto de datos se define durante el entrenamiento del modelo, fusionando los datos históricos con las metas del proyecto y las previsiones del sector.

Aunque la detección de anomalías puede llevarse a cabo mediante el análisis de datos convencional, se fundamenta en reglas que se establecen manualmente. El alcance específico y estático de dichas reglas genera restricciones que pueden ser superadas gracias a la habilidad del modelo de inteligencia artificial de desarrollarse y ajustarse con el tiempo. En las situaciones en que el sistema procesa datos semi supervisados y no supervisados, también es posible detectar anomalías de inteligencia artificial sin la necesidad de contar con datos de referencia. (Chen, 2025)

4.3.2 Respuesta automatizada

Manifiesta Arias, O (2025) que esta función posibilita realizar acciones inmediatamente ante incidentes de seguridad, sin tener que depender completamente del ser humano. Los sistemas de inteligencia artificial pueden detectar conductas anómalas y poner en marcha respuestas previamente establecidas para contener o reducir una amenaza en cuestión de segundos, gracias al análisis permanente de eventos, registros y alertas que provienen de diversas fuentes. Esta habilidad acorta

notablemente el tiempo de respuesta, disminuye la repercusión de los ataques y ayuda a preservar la continuidad operativa.

Figura N° 16 el cerebro de la Ciberseguridad



Nota. Fuente: Imagen creada con IA

Además, Coloma,J.et.al (2020) aporta que los sistemas inteligentes, a diferencia de los métodos tradicionales, donde los analistas de seguridad tienen que examinar manualmente cada alerta antes de actuar, emplean algoritmos de aprendizaje automático y modelos de análisis comportamental para establecer el riesgo asociado a un evento y escoger la respuesta más apropiada. Estas medidas pueden abarcar la desconexión automática de un dispositivo comprometido de la red, el bloqueo de direcciones IP maliciosas, la desactivación temporal de cuentas sospechosas, la detención de procesos considerados amenazantes, la actualización de las reglas de seguridad en cortafuegos o el envío inmediato de alertas para que el personal especializado prosiga con la investigación.

4.3.3 Inteligencia de amenazas

También conocida como inteligencia de amenazas cibernéticas (CTI), es información exhaustiva y ejecutable sobre las amenazas a la ciberseguridad. La inteligencia de amenazas asiste a los equipos de seguridad en la adopción de una perspectiva más proactiva para identificar, mitigar y prevenir ataques cibernéticos. La inteligencia de amenazas implica más que simplemente datos sin procesar acerca de las mismas. Se trata de información sobre amenazas que se ha correlacionado y analizado con el fin de dotar a los expertos en seguridad de un entendimiento detallado acerca de las potenciales amenazas que sus organizaciones pueden afrontar, incluyendo la forma de detenerlas. (Kosinski, 2026)

El proceso de inteligencia de amenazas es ininterrumpido y abarca la recolección, el procesamiento, el análisis y la divulgación de información vinculada a riesgos cibernéticos potenciales. Este proceso facilita la transformación de datos que provienen de fuentes variadas en información útil para respaldar la toma de decisiones estratégicas y operativas relacionadas con seguridad. Asimismo, posibilita que se reconozcan las vulnerabilidades que pueden ser aprovechadas, que se haga un seguimiento de lo que realizan los actores maliciosos y que se actualicen continuamente las medidas de protección en función del cambio en el panorama de amenazas. Así, las organizaciones tienen la posibilidad de potenciar sus políticas de ciberseguridad, perfeccionar su gestión del riesgo y aumentar su capacidad para defender sus activos cruciales frente a ataques cada vez más complicados. (Coca, 2015)

4.4. Agentes IA para la protección de infraestructuras

La defensa de las infraestructuras tecnológicas es uno de los retos más importantes que enfrenta la ciberseguridad. Esto se debe a la complejidad creciente de los entornos digitales y al incremento de los peligros que se dirigen contra redes, sistemas y aparatos conectados. En este escenario, los agentes de inteligencia artificial asumen un rol esencial al implementar aptitudes de vigilancia constante, evaluación instantánea y reacción automatizada ante potenciales sucesos de seguridad. Estos agentes, debido a su habilidad para aprender de la conducta de los sistemas, detectar anomalías y ajustarse a nuevas situaciones de ataque, mejoran la defensa de las infraestructuras críticas y ayudan a asegurar que la información esté disponible, íntegra y confidencial. Las aplicaciones más importantes de los agentes de inteligencia artificial en diversos tipos de infraestructuras tecnológicas y su contribución al fortalecimiento de la ciberseguridad se examinan en las siguientes secciones.

4.4.1 Redes empresariales

Las redes empresariales son la base que permite la comunicación e intercambio de información entre los usuarios, aplicaciones, servidores y dispositivos que conforman una organización. Son uno de los principales objetivos de los ciberdelincuentes, ya que concentran una gran cantidad de datos y servicios críticos. En este sentido, los agentes de inteligencia artificial son fundamentales para reforzar la seguridad de estas redes, ya que supervisan continuamente el tráfico, analizan el comportamiento de los dispositivos e identifican de forma temprana actividades sospechosas.

Los agentes de IA, a diferencia de los mecanismos tradicionales de seguridad que se basan en reglas predefinidas y en firmas de amenazas conocidas, utilizan técnicas de aprendizaje automático para identificar patrones de comportamiento normales en la red y detectar desviaciones que pueden indicar la presencia de ataques, accesos no autorizados o movimientos laterales por parte de un atacante. Esto les permite reconocer tanto amenazas conocidas como ataques novedosos o de día cero (Zero-Day), que normalmente los sistemas convencionales no serían capaces de identificar. (Kiser, 2020)

Los agentes inteligentes, además de detectar incidentes, pueden tomar medidas automáticas para mitigar la amenaza, como aislar equipos comprometidos, bloquear conexiones sospechosas, restringir el acceso de usuarios con comportamientos anómalos o actualizar dinámicamente las políticas de seguridad de la red. Estas capacidades reducen considerablemente el tiempo de respuesta ante un incidente y restringen la propagación del ataque hacia otros sistemas de la organización.

4.4.2 Sistemas en la nube

La computación en la nube es una tendencia novedosa para entidades que están tras cualidades de calidad asociadas con este modelo, como por ejemplo aplicaciones que necesitan un gran nivel de disponibilidad a un precio reducido y aprovechar las características de arquitecturas enfocadas en servicios. El Cloud Computing es una modalidad que posibilita brindar servicios mediante la web y otras redes de

comunicación, incluyendo una variedad de recursos vinculados con las Tecnologías de la Información y Comunicación (TIC). Estos recursos, que incluyen la capacidad para procesar información, el almacenamiento, las plataformas informáticas y aplicaciones, etc., están distantes del usuario o "en la nube", y se les paga dependiendo de cómo los utilice este. (Parrales, 2018)

Los agentes de inteligencia artificial tienen un rol significativo en el resguardo de los sistemas en la nube a través de la vigilancia constante del comportamiento de recursos y servicios que se despliegan en este contexto. Con el estudio de patrones de actividad, eventos y registros, estos agentes tienen la capacidad de identificar irregularidades que podrían señalar la existencia de vulnerabilidades o amenazas, incluso antes de que causen un efecto significativo. Además, son capaces de detectar configuraciones que no son seguras, supervisar el empleo de privilegios de acceso y comprobar que las políticas de seguridad definidas por la organización se cumplan.

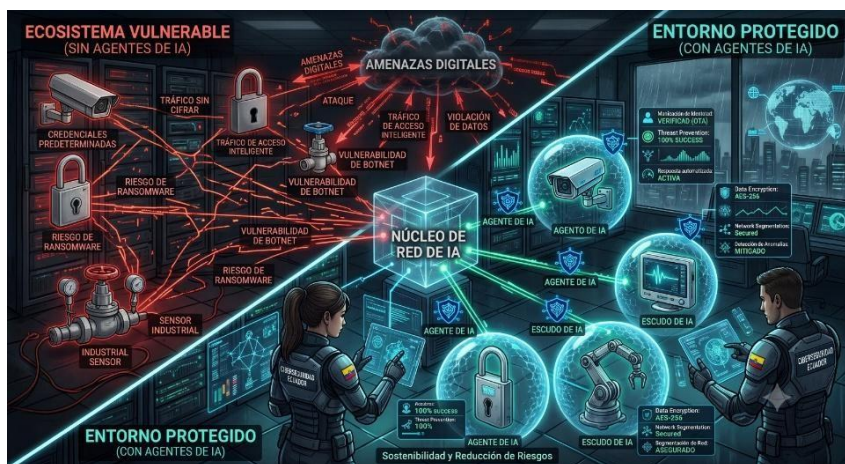
En consecuencia, la incorporación de agentes de inteligencia artificial en los sistemas en la nube no solo mejora la capacidad para prevenir y detectar incidentes de seguridad, sino que también optimiza la administración de los recursos tecnológicos, favorece la continuidad de los servicios y contribuye a garantizar la confidencialidad, integridad y disponibilidad de la información almacenada y procesada en estos entornos. (Lanham, 2025)

4.4.3 Dispositivos IoT

Kevin Ashton, un pionero de la tecnología británico, fue el primero en emplear el término «Internet de las Cosas (IoT)» en 1999. De acuerdo con Ashton, el Internet de las Cosas es un sistema que conecta objetos físicos a la red mundial por medio de sensores. El Internet de las Cosas (IoT) abarca dispositivos inteligentes que se comunican con otros equipos, objetos, entornos e infraestructuras. Esta nueva tecnología ha cambiado radicalmente la vida de las personas, pues les permite vivir y trabajar con mayor inteligencia, además de adquirir un control absoluto sobre sus vidas. El IoT no solo brinda dispositivos inteligentes para la automatización de las casas, sino que también es fundamental para las empresas. (Joyanes, 2018)

El Internet de las Cosas (IoT) se compone de una red de equipos de medición, aparatos mecánicos y objetos que poseen un identificador único (UID) Inbamani,A.et.al (2025). Puede transferir información por medio de una red sin la intervención del ser humano. Hoy en día, el IoT se ha vuelto un aparato de telecomunicaciones muy relevante y es cada vez más conocido entre los expertos y los científicos. Debido a sus técnicas interdisciplinarias. El Internet de las Cosas es una plataforma que une el mundo físico y el ciberespacio. Simultáneamente, en un análisis exhaustivo, se evalúan las redes y el diseño del Internet de las cosas (IoT) para aplicaciones económicas masivas. Se han requerido implementaciones para garantizar áreas de servicio con calidad satisfactoria, cumplimiento eficaz y operaciones. Numerosos inventos de tecnologías sin cables llevan a que se fabriquen más dispositivos IoT.

Figura N° 17 Amenazas digitales



Nota. Fuente: Imagen creada con IA

Uno de los grandes retos de la Internet de las Cosas es la gestión y protección de un ecosistema que está formado por millones de dispositivos heterogéneos que funcionan a la vez y producen un flujo continuo de información. La variedad de fabricantes, sistemas operativos, protocolos de comunicación y niveles de potencia computacional dificulta la implementación de mecanismos de seguridad uniformes. Además, la incorporación continua de nuevos dispositivos amplía la superficie de ataque, por lo que resulta indispensable mantener un monitoreo permanente, una adecuada gestión del ciclo de vida de los equipos y la actualización constante de sus configuraciones y firmware. Estas medidas permiten robustecer la resiliencia de las infraestructuras basadas en IoT y garantizar el funcionamiento seguro, fiable y eficiente de los servicios que dependen de esta tecnología. (Joyanes, 2018)

4.5 Agentes para detección y respuesta a incidentes

La TDR (detección y respuesta a amenazas) es un término que se refiere a los métodos y herramientas que las organizaciones emplean para identificar, analizar e intentar minimizar las amenazas de ciberseguridad. Integra métodos de detección avanzados, soluciones de seguridad integradas y capacidades de respuesta automatizada para asistir a las organizaciones en la adaptación a un panorama de amenazas en constante cambio y en la disminución del riesgo.

TDR asiste a los grupos de seguridad para que contengan los incidentes con rapidez y recuperen los sistemas con una interrupción mínima. Las organizaciones requieren tácticas proactivas para identificar acciones maliciosas antes de que generen perjuicios, ya que amenazas como el ransomware, el phishing y las vulnerabilidades de día cero se vuelven más comunes y avanzadas. (Krantz & Jonker, 2026)

4.5.1 SIEM inteligente

SIEM Desde que surgieron, los sistemas han cambiado la visión de la ciberseguridad, proporcionando un método distinto para integrar información de seguridad diseminada en una sola entidad coherente. Estas soluciones, al incorporar el Aprendizaje Automático (AA) y la Inteligencia Artificial (IA), son capaces no solo de normalizar y procesar enormes volúmenes de datos, sino también de examinar irregularidades y patrones que podrían señalar un suceso de seguridad. La recopilación de datos es uno de los procesos básicos en la inteligencia artificial basada en SIEM. Se trata de reunir información de seguridad proveniente de distintas fuentes, como servidores, dispositivos de red, bases de datos,

aplicaciones y demás. Se ha reunido una amplia variedad de datos que comprende registros, datos de eventos, inteligencia sobre amenazas y otros tipos de información vinculada con la seguridad. (Acosta, Benavides, & Garcia, 2020)

Para García, J (2026) la característica más sobresaliente de la inteligencia artificial SIEM La mayor ventaja que tienen estos sistemas es la posibilidad de automatizar estos procesos claves de normalización y agregación de datos. Estos sistemas son capaces de examinar los datos con mayor rapidez al emplear la IA y el aprendizaje automático, organizando, uniendo y normalizando de manera inteligente los datos de seguridad. Esta automatización disminuye de manera notable el tiempo y la energía que normalmente se necesitaban para estas actividades, lo que posibilita que los equipos de seguridad se enfoquen en elementos más estratégicos de la ciberseguridad. La IA cuando se normalizan y añaden los datos. El SIEM emplea algoritmos de inteligencia artificial para optimizar la identificación de amenazas. Estos algoritmos han sido entrenados para identificar las cualidades de amenazas ya conocidas y descubrir nuevas amenazas en desarrollo a través del análisis de patrones conductuales. En un escenario de amenazas que están en constante cambio, esta capacidad es esencial.

4.5.2 SOC automatizado

Según Peña, J (2021) la automatización de SOC tiene la capacidad de acelerar los procesos de SOC al asumir algunas funciones. La IA es, por ejemplo, muy apta para recolectar información de seguridad proveniente

de diversas fuentes, realizar análisis avanzados de datos y detectar potenciales problemas a partir de amenazas y anomalías ya conocidas. De esta manera, la IA lucha contra el exceso de alertas y posibilita que los analistas enfoquen su atención en los problemas reales. La automatización de SOC tiene la capacidad de ayudar a resolver problemas después de haber sido detectados. Los analistas tienen la capacidad de elaborar libros de estrategias y runbooks para ciertas acciones o tareas de corrección, los cuales pueden ser ejecutados automáticamente para llevar a cabo la tarea con rapidez y a gran escala.

Actualmente, las plataformas de inteligencia artificial están evolucionando hacia una mayor autonomía, permitiendo que los modelos de aprendizaje automático clasifiquen alertas e incluso ejecuten respuestas con una intervención humana mínima. Este proceso se apoya en la orquestación, que integra distintas herramientas de seguridad mediante API; la automatización, que realiza tareas predefinidas según condiciones específicas; los manuales o playbooks, que establecen las acciones para responder a incidentes; y los flujos de trabajo, que organizan de forma secuencial todo el proceso, desde la detección de una amenaza hasta su respuesta. (Peña, 2021)

Los Centros de Operaciones de Seguridad (SOC) enfrentan diversos desafíos derivados de la creciente complejidad del panorama de las ciberamenazas. Entre ellos destacan la sobrecarga de alertas, donde el alto volumen de eventos y falsos positivos dificulta la identificación de incidentes reales; la escasez de profesionales especializados en

ciberseguridad; la gestión de entornos híbridos y descentralizados que amplían la superficie de ataque; y la falta de integración entre múltiples herramientas de seguridad, lo que retrasa la respuesta ante incidentes. En este contexto, la inteligencia artificial se convierte en un apoyo estratégico al automatizar tareas, filtrar información y priorizar amenazas, fortaleciendo el trabajo de los analistas sin sustituir su criterio, ya que la protección más efectiva surge de la combinación entre la experiencia humana y las capacidades tecnológicas. (Caamaño & Gil, 2024)

4.5.3 Threat Hunting

El Threat Hunting es un enfoque proactivo de la ciberseguridad que busca detectar amenazas que logran eludir las herramientas de detección automatizada. El Threat Hunting, en lugar de aguardar alertas, implica la búsqueda activa de comportamientos sospechosos, señales débiles o anomalías que puedan sugerir que un atacante ya está dentro del entorno.

El Threat Hunting parte de la premisa de que las brechas de seguridad pueden suceder y suceden, aun en entornos con una buena defensa. Los cazadores examinan la telemetría mediante servicios, identidades, redes y puntos finales en la nube para detectar actividades maliciosas que se entrelazan con las operaciones normales. Esto frecuentemente comprende la identificación de abuso de credenciales, métodos de subsistencia en la tierra, desplazamiento lateral y mecanismos persistentes que no generan advertencias convencionales. (Trend, 2026)

Los atacantes están logrando evitar la detección cada vez mejor. Se combinan con actividades legítimas usando métodos como el abuso de credenciales, las herramientas para acceder a distancia y los malware sin archivos. La mayoría de las veces, estos métodos ignoran los sistemas de seguridad convencionales. Puede haber vacíos si se confía solamente en las alertas. El Threat Hunting se ocupa de esto al detectar amenazas con anticipación, frecuentemente antes de que surjan señales evidentes. El Threat Hunting, a diferencia de las estrategias reactivas que se basan en alertas, se centra en técnicas sigilosas y duraderas que tienen el potencial de no activar las defensas automatizadas. (Shinde, 2026)

4.6. IA ofensiva y riesgos emergentes

La inteligencia artificial ha tenido un impacto importante en el campo de la ciberseguridad, al tiempo que ha reforzado las capacidades defensivas, también se ha utilizado para objetivos ofensivos. Los delincuentes de la red utilizan estas tecnologías para automatizar sus ataques, crear amenazas más sofisticadas y mejorar su capacidad de esquivar los sistemas de protección tradicionales. Como resultado, se han generado nuevos riesgos que ponen a prueba a las organizaciones y requieren estrategias de seguridad más sofisticadas. Por ello, es fundamental comprender las principales aplicaciones ofensivas de la IA y los riesgos emergentes a ella asociados, para anticipar las amenazas y fortalecer las capacidades de prevención y respuesta.

Figura N° 18 IA en la ciberseguridad



Nota. Fuente: Imagen creada con IA

4.6.1 Deepfakes

Un deepfake, también conocido como ultrafalso, es un video extremadamente realista que ha sido manipulado digitalmente para mostrar a personas diciendo o haciendo cosas que en realidad nunca ocurrieron. En un nuevo contexto en el que se agravan los peligros de las suplantaciones de identidad, estas representaciones sintéticas creadas a través de técnicas informáticas fundamentadas en inteligencia artificial (IA) constituyen grandes peligros para la privacidad. Detectar si las declaraciones o apariciones públicas de personajes influyentes son el resultado de representaciones ficticias o se corresponden con la realidad es cada vez más difícil debido a la complejidad de los métodos para crear deepfakes. (García F. , 2021)

La tecnología deepfake utiliza inteligencia artificial avanzada mediante redes generativas antagónicas (GANs), que incluyen dos algoritmos

esenciales: el discriminador y el generador. El generador comienza el procedimiento al crear el contenido falso inicial basado en un resultado que se desea, empleando un conjunto de datos de entrenamiento. El discriminador, por su parte, examina la veracidad de este contenido e indica áreas que podrían mejorarse. Este proceso iterativo posibilita que el generador perfeccione su habilidad para crear contenido más realista, mientras que el discriminador se vuelve mejor en detectar las diferencias a corregir.

Las GANs son esenciales para la generación de deepfakes, porque examinan patrones en imágenes auténticas y los reproducen en contenido creado. Los sistemas GAN analizan diferentes perspectivas de las imágenes objetivo para obtener detalles completos en las fotografías. En los videos, examinan no solo las perspectivas, sino también la conducta, el desplazamiento y los patrones de habla. Para mejorar el realismo del producto final, este análisis multifacético pasa por múltiples ciclos a través del discriminador. (Kiser, 2020)

4.6.2 Phishing inteligente

En su artículo (Schmitt & Flechais, 2024) afirma que es una técnica de phishing avanzada que emplea inteligencia artificial (IA) con el fin de elaborar ataques más personalizados, persuasivos y difíciles de identificar. El phishing inteligente difiere del phishing tradicional, que consiste en enviar mensajes genéricos a un gran número de personas, porque analiza información pública o filtrada sobre la víctima, como su puesto, empresa, contactos o actividad en redes sociales, para elaborar correos electrónicos, mensajes o sitios web muy creíbles. La IA también

puede adaptar automáticamente el contenido, imitar el estilo de escritura de personas conocidas e incluso traducir mensajes a otros idiomas, lo que aumenta las posibilidades de que la víctima revele información confidencial o descargue archivos maliciosos.

El phishing inteligente es una amenaza emergente en el ámbito de la ciberseguridad, ya que combina técnicas de ingeniería social con algoritmos de inteligencia artificial, reduciendo los errores que tradicionalmente permitían identificar estos fraudes. Por ello, además de poner en marcha soluciones tecnológicas de detección, es fundamental reforzar la formación y concienciación de los usuarios para que sean capaces de identificar este tipo de ataques y evitar el compromiso de la información y de los sistemas.

El phishing inteligente difiere del phishing tradicional en su capacidad de adaptarse a la víctima en tiempo real. Los atacantes pueden valerse de la inteligencia artificial para procesar grandes cantidades de información de redes sociales, sitios web de empresas o filtraciones de datos, y así confeccionar mensajes muy personalizados que encajen a la perfección con la persona o la organización a la que van dirigidos. Los modelos de IA generativa también pueden escribir correos electrónicos sin errores gramaticales, imitar el estilo de escritura de directivos o compañeros de trabajo e incluso generar conversaciones creíbles a través de chatbots, lo cual aumenta mucho la probabilidad de éxito del ataque y dificulta su detección por parte de los usuarios y de los sistemas tradicionales de filtrado. (Shamir, 2024)

4.6.3 Ingeniería social automatizada

La ingeniería social automatizada es una técnica de ataque que utiliza inteligencia artificial y herramientas de automatización para manipular a las personas y obtener información confidencial, credenciales de acceso o inducir las a realizar acciones que comprometan la seguridad de una organización. A diferencia de la ingeniería social convencional, que implica la interacción manual del atacante, la automatización permite recolectar datos sobre las víctimas, elaborar mensajes personalizados, sostener conversaciones mediante chatbots o asistentes virtuales y llevar a cabo campañas masivas con escasa participación humana. Como resultado, los ataques son más rápidos, más convincentes y más difíciles de identificar.

La IA generativa también ha ampliado el alcance de la ingeniería social automatizada, posibilitando la creación de correos electrónicos, mensajes instantáneos, publicaciones en redes sociales, e incluso contenido de voz o vídeo que se hace pasar por personas reales. Estas capacidades hacen posible llevar a cabo fraudes como el spear phishing (phishing dirigido), el CEO Fraud (fraude del director ejecutivo) y la suplantación de identidad a través de deepfakes. Ante este escenario las organizaciones deben combinar controles tecnológicos, robusta autenticación y permanentes programas de concientización, a fin de fortalecer la capacidad de los usuarios para reconocer y prevenir este tipo de ataques. (Schmitt & Flechais, 2024)

4.7. Gestión de riesgos y cumplimiento normativo

La gestión de riesgos y el cumplimiento normativo son elementos fundamentales para fortalecer la seguridad y la gobernanza dentro de una organización. La gestión de riesgos comprende un conjunto de actividades coordinadas orientadas a identificar, evaluar, tratar y controlar los riesgos que pueden afectar el logro de los objetivos organizacionales, con el propósito de reducir o mitigar sus posibles impactos negativos. Por otra parte, el cumplimiento normativo (compliance), por su parte, consiste en garantizar la adhesión a las leyes, regulaciones, estándares y principios éticos aplicables a la organización, promoviendo una cultura de responsabilidad y buenas prácticas que debe ser compartida por todos los colaboradores.

Figura N° 19 Deepfakes



Nota. Fuente: Imagen generada con IA

4.7.1 ISO 27001

La norma ISO/IEC 27001 (2022) es la más reconocida a nivel global para los sistemas de gestión de seguridad de la información (SGSI). Esta norma establece las condiciones que un SGSI tiene que satisfacer.

La norma ISO/IEC 27001 ofrece a las compañías, sin importar su tamaño o sector, directrices para implementar, instaurar, sostener y optimizar constantemente un sistema de administración de seguridad de la información.

Cumplir con la ISO/IEC 27001 significa que una empresa u organización ha implementado un sistema para administrar los riesgos asociados a la seguridad de los datos que tiene o gestiona, y que dicho sistema cumple con todos los principios y buenas prácticas establecidos en esta norma internacional.

Dado el incremento de la ciberdelincuencia y la aparición continua de nuevas amenazas, es posible que manejar los riesgos cibernéticos parezca complicado o incluso inalcanzable. La ISO/IEC 27001 asiste a las entidades en la identificación y el tratamiento proactivo de los puntos vulnerables, así como en la toma de conciencia acerca de estos riesgos.

La ISO/IEC 27001 (ISO, 2022) fomenta una perspectiva holística en lo que respecta a la seguridad de la información, que incluye a las personas, las políticas y la tecnología. Un sistema de gestión de la seguridad de la información que se implemente siguiendo esta norma es una herramienta fundamental para gestionar riesgos, fortalecer la resiliencia cibernética y garantizar la excelencia en las operaciones.

4.7.2 NIST Cybersecurity Framework

El marco de ciberseguridad del NIST (2026) (NIST CSF) proporciona orientación integral y mejores prácticas que las organizaciones del sector privado pueden seguir para mejorar la seguridad de la información y la gestión de riesgos de ciberseguridad. El Instituto Nacional de Estándares y Tecnología (NIST) de EE. UU. es una agencia no reguladora que promueve la innovación mediante el avance de la ciencia de medición, los estándares y la tecnología.

El CSF del NIST es lo suficientemente flexible como para integrarse con los procesos de seguridad existentes dentro de cualquier organización, en cualquier industria. Proporciona un excelente punto de partida para implementar la seguridad de la información y la gestión de riesgos de ciberseguridad en prácticamente cualquier organización del sector privado en Estados Unidos.

El NIST Cybersecurity Framework (CSF) se estructura en funciones, categorías, subcategorías y referencias informativas que proporcionan un enfoque organizado para gestionar los riesgos de ciberseguridad. Sus funciones representan actividades continuas e interrelacionadas que fortalecen la postura de seguridad de una organización, mientras que las categorías y subcategorías detallan acciones específicas para su implementación. Tradicionalmente, el marco comprende cinco funciones principales: Identificar, orientada al conocimiento de los activos, el entorno y los riesgos; Proteger, enfocada en la implementación de controles y salvaguardas para reducir el impacto de las amenazas; Detectar, destinada al monitoreo continuo y la identificación oportuna de incidentes; Responder, que establece los procesos para analizar, contener y mitigar los ataques; y Recuperar, dirigida a restablecer las operaciones,

fortalecer la resiliencia y garantizar la continuidad del negocio tras un incidente de ciberseguridad. (IBM, 2026)

4.7.3 Protección de datos personales

Afirma la revista digital de la Universidad San Francisco de Quito (2026) que, en la actualidad, los progresos tecnológicos posibilitan que tanto las entidades públicas como las privadas utilicen datos personales para llevar a cabo sus distintas actividades, en una escala sin precedentes. La ambición humana de llegar a un nivel más alto de control y conocimiento sobre la información privada de las personas surgió hace siglos. Por lo tanto, la defensa de los datos se analiza desde la necesidad humana de salvaguardar áreas de libertad personal. La perspectiva está orientada a salvaguardar los datos de individuos, o sea, personas naturales en particular que se ven cosificadas debido al mal uso de la tecnología.

La ausencia de control y, en consecuencia, el uso inadecuado de los datos personales es perjudicial para los derechos esenciales de las personas. Para comprender la relevancia del derecho a la protección de datos, es necesario inicialmente aproximarse a las nociones de derecho a la privacidad y derecho a la intimidad. El derecho a la intimidad, entendido como el derecho del individuo a la soledad y como una facultad destinada a proteger un espacio específico de manera exclusiva.

Según Caamaño, E; Gil, R (2024) la protección de datos personales resulta de especial importancia en el ámbito judicial, donde se administra gran cantidad de información sensible y confidencial relacionada con

personas que participan de procesos judiciales, tales como datos de identificación, antecedentes, pruebas, expedientes, información financiera y otros registros de carácter personal. En este sentido, se deben adoptar las medidas técnicas, organizativas y legales necesarias para garantizar la confidencialidad, integridad y disponibilidad de la información, evitando accesos no autorizados, alteraciones, pérdidas o divulgaciones indebidas. Una correcta gestión de estos datos fortalece la confianza en las instituciones de justicia, protege los derechos de las personas y ayuda al cumplimiento de los principios de seguridad y tratamiento responsable de la información recogidos en la normativa vigente.

Cumplir con las normas y regulaciones legales es otro elemento relevante de la ciberseguridad. Numerosas industrias poseen requisitos particulares relacionados con la garantía de información que deben ser satisfechos, y si no se cumplen, las organizaciones podrían afrontar sanciones y multas de gran envergadura. La Ley Orgánica de Protección de Datos Personales (LOPD) empezó a aplicarse en Ecuador el 26 de mayo del 2021. Esta ley regula cómo las empresas públicas y privadas, además de los profesionales independientes, manejan la información personal de aquellos con los que interactúan, tales como clientes, proveedores o colaboradores, con el objetivo de asegurar el derecho constitucional ecuatoriano a la protección de datos personales. (Astudillo, 2025)

BIBLIOGRAFIA

- Inbamani,A.et.al. (2025). Blockchain in healthcare and IoT: A systematic literature review. *Science Direct*.
- Abad, L. (2026). Transparencia algorítmica:enfoques y desafíos para el Estado democrático . *Revista Española de la Transparencia*.
- Acemoglu, D. (2021). *Harms of AI*. National Bureau of Economic Research.
- Acosta, M., Benavides, M., & Garcia, N. (2020). Delitos informáticos: Impunidad organizacional y su complejidad en el mundo de los negocios. *Redalyc*.
- Aggarwal, C. (2015). *Data Mining*. Springer.
- Agostinelli, S., & al, e. (2021). Research challenges for intelligent robotic process automation. *Computers in Industry*.
- Agrawal, A., & al., e. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press.
- Anton, G. (2024). *ChatGPT para profesores*. Malaga: Independiente.
- Arias, O. (2025). *La Inteligencia Artificial y su impacto en la sociedad*. Caracas: Autonomo.
- Armbrust, M., & al, e. (2010). A view of cloud computing. *Communications of the ACM*, 50-58.
- Astudillo, K. (2025). *Como iniciarte en Ciberseguridd*. Independiente.
- AutomationAnywhere*. (4 de Octubre de 2024). Obtenido de <https://www.automationanywhere.com/la/company/blog/automa>

tion-ai/exploring-ai-agents-types-capabilities-and-real-world-applications

- Baidoo-Anu, D., & Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*.
- Barocas, S., & al., e. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press.
- Belcic, I., & Stryker, C. (10 de Mayo de 2026). IBM. Obtenido de <https://www.ibm.com/es-es/think/topics/utility-based-agent>
- Bergmann, D., & Stryker, C. (2026). *IBM Think*. Obtenido de <https://www.ibm.com/es-es/think/topics/langchain>
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Bommasani, R., & al., e. (2021). On the Opportunities and Risks of Foundation Models. *Machine Learning*.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Bradford, A., & al, e. (2023). *The Brussels Effect: How the European Union rules the digital world*. . Oxford University Press.
- Bravo, C. (12 de Junio de 2025). eset. Obtenido de <https://www.eset.com/latam/blog/cultura-y-seguridad-digital/que-son-agentes-ia/>
- Brooks, T., & al., e. (2024). *Video generation models as world simulators*. OpenAI Research.
- Brown, T., & al., e. (2020). Language Models are Few-Shot Learners. *Computation and Language*.
- Brynjolfsson, E., & Hitt, L. (2000). Beyond Computation: Information Technology, Organizational Transformation and Business Performance. *Journal of Economics Perspectives*, 23-48.
- Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.
- Buczak, A., & Guven, E. (2015). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Communications Surveys & Tutorials*, 1153-1176.
- Caamaño, E., & Gil, R. (2024). Prevención de riesgos por ciberseguridad desde la auditoria forense: conjugando el talento humano organizacional. *Redalyc*.
- Cavada, J. (2020). *Biblioteca del congreso Nacional de Chile*. Obtenido de <https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/102>

21/29012/2/Definicion_y_regulacion_de_cibercrimen_y_delito_informatico_JPC_edit.pdf

- Chen, M. (26 de Junio de 2025). *Oracle*. Obtenido de <https://www.oracle.com/latam/artificial-intelligence/anomaly-detection/>
- Chingaturta, P. e. (2016). Editorial—Marketing Science and Big Data. *Marketing Science*, 341-342.
- Coca, Y. (2015). Agentes inteligentes. Aplicación a la realidad virtual. *Redalyc*.
- Cockburn, I., Henderson, R., & Stern, S. (2018). *The Impact of Artificial Intelligence on Innovation*. National Bureau of Economic Research. doi:10.3386/w24449
- Coloma, J. et al. (2020). Inteligencia artificial, sistemas inteligentes, agentes inteligentes. *Dialnet*.
- Crevier, D. (1993). *Ai: The Tumultuous History Of The Search For Artificial Intelligence*. Basic Books.
- Davenport, T., & Bean, R. (2021). *NewVantage Partners Releases 2021 Big Data and AI Executive Survey*. NewVantage Partners.
- Davenport, T., & Miller, S. (2022). *Working with AI: Real Stories of Human-Machine Collaboration*. MIT Press.
- Davenport, T., & Ronanki, R. (February de 2018). Artificial Intelligence for the Real World. *Harvard Business Review*.
- Davenport, T., & Westerman, G. (2018). *Why So Many High-Profile Digital Transformations Fail*. Harvard Business Review.
- Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *arXiv: Machine Learning*.
- Europea, U. (10 de Marzo de 2026). *Universidad Europea*. Obtenido de <https://universidadeuropea.com/blog/agentes-inteligentes/#agentes-inteligentes-componentes>
- Faraksa, N. (2024). *Google Gemini Unleashed*. New york: Independiente.
- Few, S. (2013). *Information Dashboard Design: Displaying Data for At-a-glance Monitoring*. Analytics Press.
- Fitzgerald, M., & al, e. (2014). Embracing Digital Technology: A New Strategic Imperative. *MIT Sloan Management Review*, 1-12.
- Fortinet. (21 de Febrero de 2026). Obtenido de <https://www.fortinet.com/lat/resources/cyberglossary/cia-triad>
- García, F. (2021). «Deepfakes» el próximo reto en la detección de noticias falsas. *Dialnet*.
- García, J. (2026). *Agentes IA Guia para principiantes*. Autonomo.

- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media.
- Giacomo, B. (2025). *Agentes de IA para empresas: Cómo crear un equipo de empleados digitales con inteligencia artificial para acelerar el marketing, triplicar las ventas y automatizar tu negocio*. Bruno Editore.
- Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence*, 1-48.
- Goodell, J. (2020). COVID-19 and finance: Agendas for future research. *Finance Research Letters*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Goverment, A. (24 de Marzo de 2026). *Digital Gov*. Obtenido de https://www.digital.gov.au/policy/ai/agent-ai-addendum-background?utm_source=chatgpt.com
- Gutowska, A. (2026). *IBM*. Obtenido de <https://www.ibm.com/es-es/think/topics/ai-agents>
- Harker, S. (2026). *IBM*. Obtenido de <https://www.ibm.com/mx-es/think/topics/multi-agent-collaboration>
- Hastie, T., & al., e. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer.
- Haykin, S. (2009). *Neural networks and learning machines. 3rd. Edition*. Pearson.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 1735-1780.
- Iansiti, M., & Lakhani, K. (2020). *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World*. Harvard Business School.
- IBM. (2026). Obtenido de <https://www.ibm.com/mx-es/think/topics/nist>
- Instituto Tecnológico Europeo. (2026). Obtenido de <https://institutotecnologicoeuropeo.com/agentes-de-ia-autonomos-la-revolucion-de-la-productividad/>
- ISO. (2022). *ISO/IEC*. Obtenido de <https://www.iso.org/es/norma/27001>
- ISO TOOLS. (18 de Mayo de 2025). Obtenido de https://isotools.org/2024/11/26/gobernanza-de-la-ia-concepto-tipos-y-marcos-normativos/#Que_es_la_Gobernanza_de_la_IA
- James, G., & al., e. (2021). *An Introduction to Statistical Learning: with Applications in R*. Springer.

- Jordan, M., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 255-260.
- Jouppi, N., & al, e. (2017). In-datacenter performance analysis of a tensor processing unit. *2017 ACM/IEEE 44th Annual International Symposium on Computer Architecture (ISCA)*. Toronto: IEEE.
- Joyanes, L. (2018). *Industria 4.0 La cuarta revolucion industrial*. Mexico: Alfaomega.
- Jurafsky, D., & Martin, J. (2021). *Speech and language processing (3rd ed. draft)*. Pearson.
- Kaplan, R., & Norton, D. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business School Press.
- Kiser, Q. (2020). *CIBER Seguridad*. Independiente.
- Kosinski, M. (Mayo de 2026). *IBM*. Obtenido de <https://www.ibm.com/es-es/think/topics/threat-intelligence>
- Krantz, T., & Jonker, A. (Marz de 2026). *IBM*. Obtenido de <https://www.ibm.com/es-es/think/topics/threat-detection-and-response>
- Kreuzberger, D., & al., e. (223). Machine Learning Operations (MLOps): Overview, Definition, and Architecture. *IEEE Access*, 31866-31879.
- Krizhevsky, A., & al, e. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM* (págs. 84-90). Association for Computing Machinery.
- Kumar, V., & al, e. (2019). Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing. *California Management Review*.
- Lanham, M. (2025). *Agentes de IA en accion*. Nueva York: Manning Publications.
- Lasi, H., & al, e. (2014). Industry 4.0. *Business & Information Systems Engineering*, 239-242.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 436-444. doi:10.1038/nature14539
- Lee, K. (2018). *AI Superpowers: China, Silicon Valley, and the New World Order*. Harper Business.
- Lewis, P., & al, e. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems* (págs. 9459-9474). Curran Associates Inc.
- Lopez, F., & García. (2024). ia y sesgos: una visión alternativa expresada desde la ética y el derecho. *Dialnet*.

- Luo, X., & al, e. (2019). Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science*.
- Manning, C., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
- Mazzucato, M. (2021). *Mission Economy: A Moonshot Guide to Changing Capitalism*. HarperCollins Publishers.
- McCarthy, J., & al., e. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence: : August 31, 1955. *AI Magazine*, 12-14.
- McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* , 115-133.
- Mitchell, T. (1997). *Machine Learning*. McGraw-Hill.
- Mnih, V., & al., e. (2015). Human-level control through deep reinforcement learning. *Nature*, 529-533.
- Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Nilsson, N. (2014). *Principles of Artificial Intelligence*. Morgan Kaufmann.
- Nimje, P. (2024). The Rise of Low-Code/No-Code Development Platforms. *International Journal of Advanced Research in Science Communication and Technology*.
- Nuncio, R. (2024). *Una historia de la Inteligencia Artificial*. Ciudad de Mexico: AdAstra.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing.
- Ouyang, L., & al, e. (2011). Training language models to follow instructions with human feedback. *NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems* (págs. 27730-27744). Curran Associates Inc.
- Overgoor, G., & al, e. (2019). Letting the Computers Take Over: Using AI to Solve Marketing Problems. *California Management Review*.
- Parrales, J. (2018). Los sistemas de información en la nube y su utilidad en las organizaciones comerciales de la ciudad de Manta, Manabí-Ecuador. *Maskana-Ciencia de la Educacion*.
- Pascal,B.et.al. (2025). *Agentic Artificial Intelligence*. Madrid: Autonomo.

- Peña, J. (Junio de 2021). *Repositorio Universitat Oberta Catalunya*.
Obtenido de
<https://openaccess.uoc.edu/server/api/core/bitstreams/3530dc34-2ad5-41ce-b344-c888c21a9b31/content>
- Poole, D., & Mackworth, A. (2017). *Artificial Intelligence Foundations of Computational Agents*. Cambridge University Press.
- Porter, M. (1985). *The Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press.
- Porter, M., & Heppelmann, J. (Octubre de 2015). How Smart, Connected Products Are Transforming Companies. *Harvard Business Review*.
- Power, D. (2002). *Decision Support Systems: Concepts and Resources for Managers*. Quorum Books Westport.
- Prinz, N., & al., e. (2021). Low-Code Development Platforms – A Literature Review. *AMCIS 2021, Digital Innovation and Entrepreneurship, Virtual Conference*. Association for Information Systems (AIS).
- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*.
- Ramesh, A., & al., e. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv: Computer Vision and Pattern Recognition*.
- Remenyi, D., & al., e. (2007). *The Effective Measurement and Management of ICT Costs and Benefits*. CIMA Publishing.
- Ricci, F., & al, e. (2015). *Recommender Systems Handbook*. Springer.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*,, 386-408.
- Rumelhart, D., & al, e. (1986). Learning representations by back-propagating errors. *Nature*, 533-536.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach (4th ed.)*. Pearson.
- Schmitt, M., & Flechais, I. (2024). Digital deception: generative artificial intelligence in social engineering and phishing. *Springer Nature*.
- Shamir, L. (2024). Analysis and prevention of AI-based phishing email attacks. *arxiv*.
- Sharda, R., & al, e. (2017). *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*. Pearson.
- Shi, W., & al, e. (2016). Edge Computing: Vision and Challenges. *IEEE Internet of Things Journal*, 637-646.

- Shinde, J. (2026). Introduction to Threat Hunting Guide. *Scrib. snowflake*. (2025). *snowflake*. Obtenido de <https://www.snowflake.com/es/fundamentals/what-are-ai-agents-understanding-their-role-and-impact/>
- Stryker, C. (2026). *IBM*. Obtenido de <https://www.ibm.com/es-es/think/topics/components-of-ai-agents>
- Sutton, R., & Barto, A. (2018). *Reinforcement learning: An introduction (2nd ed.)*. MIT Press.
- Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, 751-752.
- Tambe, P., & al, e. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *California Management Review*.
- Tapscott, D., & Tapscott, A. (2016). *Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World*. Penguin Publishing Group.
- Treiblmaier, H., & Sillaber, C. (2021). The impact of blockchain on e-commerce: A framework for salient research topics. *Electronic Commerce Research and Applications*.
- Trend, M. (13 de Julio de 2026). *Trend*. Obtenido de https://www.trendmicro.com/es_es/what-is/threat-detection/threat-hunting.html
- Turban, E., & Volonino, L. (2011). *Information Technology for Management: Improving Strategic and Operational Performance*. Wiley.
- Turing, A. (1936). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 230-265.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 433-460.
- UNESCO. (2026). *UNESCO*. Obtenido de <https://www.unesco.org/mil4teachers/es/module8>
- USFQ. (2026). ¿Vida privada o muerte a la privacidad?: protección de datos personales en la relación empresa-cliente en Ecuador. *Revistas USFQ*.
- van der Aalst, W., & al, e. (2018). Robotic Process Automation. *Business & Information Systems Engineering*, 269-272.
- Vaswani, A., & al., e. (2017). Attention is all you need. *NIPS'17: Proceedings of the 31st International Conference on Neural*

- Information Processing Systems* (págs. 6000-6010). Advances in Neural Information Processing Systems.
- Walker, B. (2019). *Ciber Seguridad*. Independiente.
- Wan,Z.et.al. (2025). Sensor-actuator integration in intelligent devices: from functional synergy to emerging application. *ScienceDirect*.
- Wang, S., & al, e. (2019). Blockchain-Enabled Smart Contracts: Architecture, Applications, and Future Trends. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2266-2277.
- Wedel, M., & Kannan, P. (2016). Marketing Analytics for Data-Rich Environments. *Journal of Marketing*.
- Winland, V., Syed, M., & Gutowska, A. (2026). IBM. Obtenido de <https://www.ibm.com/mx-es/think/topics/crew-ai>
- Wooldridge, M. (2020). *An introduction to multiagent systems (2nd ed.)*. John Wiley & Sons.
- Xu, X., & al, e. (2021). Industry 4.0 and Industry 5.0—Inception, conception and perception. *Journal of Manufacturing Systems*, 530-535.
- Zax, D. (12 de Junio de 2026). IBM. Obtenido de <https://www.ibm.com/es-es/think/topics/goal-based-agent>

ANEXO 1

Revisión de pares ciegos.



EDITORIAL M.S.

Av. Ernest Rutherford y John Dalton. Tlf: +593 960677758
Ambato – Ecuador



INFORME DE EVALUACIÓN DE REVISORES CIEGOS

Datos de la obra	
Nombre de la Obra	AGENTES DE INTELIGENCIA ARTIFICIAL PARA LOS NEGOCIOS Y LA CIBERSEGURIDAD
Fecha de Evaluación	19 de agosto del 2026

Datos del revisor		
Nombre	MARCO ANTONIO	
Apellidos	SALAZAR CASTILLO	Cédula: 1850423011
Grado académico	Maestría X	Doctorado
Título pregrado	INGENIERO/A EN CIENCIAS DE LA COMPUTACION	
Área de posgrado	MASTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL	

Contenido del texto		
N°	Descripción	Observaciones
1	Título, prólogo e introducción (claridad y estructura)	Se da a entender de manera clara y concisa los temas abordados en el documento.
2	Relevancia, Originalidad Revisión de literatura	Documento que muestra texto muy explícito en el área abordada, mostrando una recopilación de información apropiada.
3	Estructura metodológica. Diseño experimental	Explicada de forma clara y concisa.
4	Resultados, Discusión , Conclusiones.	Recopilación de datos obtenidos concretos, mostrando aportes de vital importancia a la investigación realizada por parte de los autores.
5	Referencia (variedad y claridad	Muy bien establecidas de acuerdo a los lineamientos establecidos.

Dictamen			
Descripción	Sí	No	Observaciones
Publicable	X		
Publicar con correcciones			
No publicar			

Atentamente,



Verificar documento en Firmado:
Código de verificación por:
MARCO ANTONIO
SALAZAR CASTILLO

Ing. MARCO ANTONIO SALAZAR CASTILLO Mag.
Revisor

INFORME DE EVALUACIÓN DE REVISORES CIEGOS

Datos de la obra	
Nombre de la Obra	AGENTES DE INTELIGENCIA ARTIFICIAL PARA LOS NEGOCIOS Y LA CIBERSEGURIDAD
Fecha de Evaluación	Agosto 20 /2026

Datos del revisor		
Nombre	Roberto Rolando	
Apellidos	López López	Cédula: 1802301562
Grado académico	Maestría	Doctorado X
Título pregrado	Ingeniero en sistemas	
Area de posgrado	Master en Administración de empresas. PhD. en educación	

Contenido del texto		
Nº	Descripción	Observaciones
1	Título, prólogo e introducción (claridad y estructura)	Se da a entender de manera clara y concisa los temas abordados en el documento.
2	Relevancia Originalidad Revisión de literatura	Documento que muestra texto muy explícito en el área abordada, mostrando una recopilación de información apropiada.
3	Estructura metodológica Diseño experimental	Explicada de forma clara y concisa.
4	Resultados Discusión Conclusiones	Recopilación de datos obtenidos concretos, mostrando aportes de vital importancia a la investigación realizada por parte de los autores.
5	Referencia (variedad y claridad)	Muy bien establecidas de acuerdo a los lineamientos establecidos.

Dictamen			
Descripción	Sí	No	Observaciones
Publicable	X		
Publicar con correcciones			
No publicar			

Atentamente,



PhD. López López Roberto Rolando
Revisor

SOBRE EL LIBRO

La inteligencia artificial y la ciberseguridad están transformando la manera en que las organizaciones operan, toman decisiones y protegen sus activos más valiosos: la información y la confianza. Este libro explora el potencial de los agentes de inteligencia artificial como aliados estratégicos en el mundo empresarial y como guardianes digitales en un entorno cada vez más amenazado.

A través de fundamentos sólidos, aplicaciones prácticas, casos reales y estrategias innovadoras, los autores demuestran cómo IA IA puede optimizar procesos, anticipar riesgos, mejorar la toma de decisiones y construir organizaciones más resilientes, competitivas y sostenibles.

Una obra esencial para líderes, profesionales, estudiantes y emprendedores que buscan comprender y aprovechar las tecnologías emergentes para impulsar sus negocios y proteger su futuro.

¿QUÉ ENCONTRARÁS EN ESTE LIBRO?



Fundamentos de la inteligencia artificial y su impacto en los negocios.



Agentes de IA aplicados a procesos empresariales clave.



Estrategias y mejores prácticas de ciberseguridad empresarial.



Casos prácticos, tendencias y herramientas para un entorno digital seguro y competitivo.

LOS AUTORES



Gustavo Eduardo Fernández Villacrés
Freddy Patricio Baño Naranjo
Wilfrido Amilcar Trujillo Quinto
José Luis Vásconez Salazar

Profesionales y académicos comprometidos con la innovación, la tecnología y el desarrollo empresarial, con amplia experiencia en inteligencia artificial, ciberseguridad, transformación digital y gestión estratégica.

“

La inteligencia artificial no reemplaza a las personas; potencia su capacidad de decidir, innovar y proteger lo que realmente importa: su negocio, su información y su futuro.

”



ISBN 978-9907-0-1798-4



DOI 10.64584/aian1456

2026
EDICIÓN



INNOVACIÓN
SEGURIDAD
FUTURO