

# CONCEPTOS ACTUALES EN INTELIGENCIA ARTIFICIAL Y SU APLICACIÓN A LA ATENCIÓN EN SALUD

John Jaime Sprockel Díaz, MD



---

*Colombia · México · Panamá · Estados Unidos*

Los autores y colaboradores presentan temas de actualidad en los cuales los procedimientos y la dosificación de los medicamentos están tomados de las recomendaciones actuales que aparecen en la literatura universal. Por lo tanto, ante los posibles errores humanos o cambios en la medicina, ni los autores ni los colaboradores ni cualquier otra persona que haya participado en la preparación de esta obra garantiza que la información contenida en ella sea precisa o completa, y tampoco son responsables de los posibles errores u omisiones de resultados con la información obtenida. Sería recomendable recurrir a otras fuentes de información para tener certeza de que la misma en este escrito es precisa.

Esto es de particular importancia en relación a los fármacos nuevos o de uso no frecuente. Sería recomendable también consultar a las empresas farmacéuticas para conseguir información adicional si es necesario.

**DECLARACIÓN SOBRE LA IA GENERATIVA Y LAS TECNOLOGÍAS ASISTIDAS POR IA EN EL PROCESO DE ESCRITURA.** Durante la preparación de esta obra, el autor usó ChatGPT como ayuda para la redacción y edición del manuscrito en busca de claridad y concisión. Posterior al uso de esta herramienta, el autor revisó y editó cuidadosamente el contenido según fuese necesario y asume toda la responsabilidad por el contenido de la publicación.

**CONCEPTOS ACTUALES EN INTELIGENCIA  
ARTIFICIAL Y SU APLICACIÓN A LA ATENCIÓN  
EN SALUD**

© 2026 GRUPO DISTRIBUNA

ISBN impreso: 978-628-7673-70-0

ISBN digital: 978-628-7673-71-7

**Autor**

John Jaime Sprockel Díaz, MD

**Corrección de estilo**

Felipe Gaitán

Andrés Mantilla Meluk

**Coordinación editorial**

Andrés Mantilla Meluk

Laura Rubiano Velasco

**Diseño y diagramación**

Marcela Torres Caballero

Impreso en Bogotá, Colombia

*Printed in Bogotá, Colombia*

**GRUPO DISTRIBUNA**

Celulares: +57 321 2097895 - 310 8739208

E-mail: [suanny@libreriamedica.com](mailto:suanny@libreriamedica.com)

[servicioalcliente@libreriamedica.com](mailto:servicioalcliente@libreriamedica.com)

Ciudad de México - CDMX, México: PERNAMBUCO

853, Colonia, Lindavista Norte

Teléfono: +52 55 4544 5702

E-mail: [distribunamexico@libreriamedica.com](mailto:distribunamexico@libreriamedica.com)

**HECHO DEPÓSITO LEGAL**

Prohibida la reproducción parcial o total del material editorial o gráfico de esta publicación sin previa autorización escrita del editor. El esfuerzo y entrega de médicos colegas hicieron posible terminar este proyecto. Fotocopiarlo es una forma de irrespetarse e irrespetar el trabajo y dignidad de los autores.

Gracias por su apoyo al adquirir un original.

LA EDITORIAL

Para nosotros es muy importante su opinión acerca de esta obra. Escribanos:  
**[opinioneditorial@libreriamedica.com](mailto:opinioneditorial@libreriamedica.com)**



Consulte el  
catálogo de  
publicaciones  
on-line

**[www.libreriamedica.com](http://www.libreriamedica.com)**

**[www.ebookmedico.com](http://www.ebookmedico.com)**



**¿POR QUÉ PUBLICAMOS?**

Nos motiva construir contenidos, información y conocimiento con excelencia y responsabilidad social. Exhortamos a nuestros lectores a aceptar el desafío de hacer de este cúmulo de valiosa información, experiencia, evidencia e investigación, plasmado en nuestros libros o procesos académicos facilitados, un elemento de impacto en el entorno social y asistencial donde cada uno se encuentre, y de esta manera poder brindar a la comunidad mayores y mejores posibilidades de calidad de vida.

# DEDICATORIA



A Álvaro Sprockel, mi padre. Aunque no alcanzó a ver estos logros profesionales y académicos, sé que, desde donde esté, comparte con orgullo cada paso de este camino que él ayudó a forjar. Su ejemplo y enseñanzas son una presencia constante que me inspiran a seguir adelante con la misma pasión y entrega que él me inculcó.

A mi familia, por su amor incondicional, su apoyo constante y por ser el motor que impulsa mis sueños.

A los profesionales de la salud que, día a día, buscan mejorar la vida de sus pacientes y se atreven a abrazar la innovación en su práctica.

Y a todos aquellos que creen en el poder del conocimiento y la tecnología como herramientas para transformar el cuidado de la salud. Que este libro inspire el camino hacia un futuro en el que la inteligencia artificial sea una aliada en nuestro compromiso con la vida.



# AUTOR



## **John Jaime Sprockel Díaz, MD**

Médico cirujano, Universidad Nacional de Colombia

Médico internista, Universidad El Bosque

Magíster en Ingeniería de Sistemas y Computación, Pontificia Universidad Javeriana

*Magister of Business Administration*, IEBS Biztech School

Unidad de Cuidados Intensivos, Subred Integrada de Servicios de Salud del Sur - Hospital El Tunal

Servicio de Medicina Interna, Hospital San José de Bogotá

Profesor investigador asociado, Fundación Universitaria de Ciencias de la Salud - FUCS:

- Programa de Semiología, pregrado de Medicina
- Rotación de Pisos, posgrado de Medicina Interna
- Curso electivo, Inteligencia Artificial en Medicina, de la especialidad y maestría de Epidemiología Clínica
- Miembro, Instituto de Investigaciones



# CONTENIDO



|                 |      |
|-----------------|------|
| Prefacio        | XIII |
| Agradecimientos | XV   |

## SECCIÓN I. INTELIGENCIA ARTIFICIAL GENERAL

|    |                                                                                                    |    |
|----|----------------------------------------------------------------------------------------------------|----|
| 1. | Definición de inteligencia artificial                                                              | 3  |
| 2. | Aprendizaje en inteligencia artificial                                                             | 7  |
| 3. | Gradiente descendente y algoritmo de retropropagación                                              | 17 |
| 4. | Procesamiento del lenguaje natural                                                                 | 29 |
| 5. | Tokenización y <i>embeddings</i> en procesamiento del lenguaje natural                             | 37 |
| 6. | <i>Transformers</i>                                                                                | 47 |
| 7. | Mecanismos de atención                                                                             | 59 |
| 8. | Explicación de <i>consultas</i> , <i>claves</i> y <i>valores</i> en los modelos <i>transformer</i> | 67 |
| 9. | Redes neuronales y <i>transformers</i> basadas en grafos                                           | 71 |



|     |                                                                                |    |
|-----|--------------------------------------------------------------------------------|----|
| 10. | Aprendizaje por refuerzo                                                       | 85 |
| 11. | Redes neuronales líquidas: un salto hacia el futuro del aprendizaje automático | 97 |

## SECCIÓN II. INTELIGENCIA ARTIFICIAL GENERATIVA

---

|     |                                                                                                        |     |
|-----|--------------------------------------------------------------------------------------------------------|-----|
| 12. | Modelos de lenguaje de gran tamaño                                                                     | 107 |
| 13. | Fases de entrenamiento de los modelos de lenguaje de gran tamaño                                       | 113 |
| 14. | Ajuste fino en los modelos de lenguaje de gran tamaño                                                  | 123 |
| 15. | Aprendizaje por refuerzo a partir de comentarios humanos para la alineación de los modelos de lenguaje | 131 |
| 16. | Ventana de contexto                                                                                    | 137 |
| 17. | Cuantización de modelos de lenguaje                                                                    | 143 |
| 18. | Aprendizaje en contexto                                                                                | 151 |
| 19. | Ingeniería de instrucciones                                                                            | 157 |
| 20. | Generación aumentada por recuperación                                                                  | 167 |
| 21. | <i>Transformers</i> visuales                                                                           | 177 |
| 22. | Modelos de difusión                                                                                    | 183 |
| 23. | DreamBooth                                                                                             | 191 |
| 24. | Ajuste fino eficiente de parámetros y adaptación de bajo rango para modelos de lenguaje de gran tamaño | 195 |
| 25. | <i>Embeddings</i> de inversión textual                                                                 | 203 |
| 26. | <i>Habla a texto</i> mediante inteligencia artificial                                                  | 207 |
| 27. | Whisper y SeamlessM4T: los modelos estados del arte en tareas de reconocimiento automático del habla   | 219 |



|            |                                     |            |
|------------|-------------------------------------|------------|
| <b>28.</b> | <b>Modelos de visión y lenguaje</b> | <b>227</b> |
| <b>29.</b> | <b>Mamba</b>                        | <b>243</b> |

### **SECCIÓN III. APLICACIONES DE LA INTELIGENCIA ARTIFICIAL EN MEDICINA**

---

|            |                                                                                                                                                                     |            |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>30.</b> | <b>AlphaFold 2: la revolución en la predicción de la estructura de proteínas</b>                                                                                    | <b>255</b> |
| <b>31.</b> | <b>Uso de ChatGPT en la escritura de artículos científicos</b>                                                                                                      | <b>261</b> |
| <b>32.</b> | <b>Uso de inteligencia artificial en revisiones sistemáticas</b>                                                                                                    | <b>267</b> |
| <b>33.</b> | <b>Usos de la inteligencia artificial en la educación médica</b>                                                                                                    | <b>275</b> |
| <b>34.</b> | <b>Respuestas de <i>chatbots</i> de inteligencia artificial a consultas de pacientes</b>                                                                            | <b>291</b> |
| <b>35.</b> | <b><i>Chatbots</i> basados en inteligencia artificial, en su interacción con los médicos: respuestas a preguntas, evaluación de casos y generación de resúmenes</b> | <b>301</b> |
| <b>36.</b> | <b>Sistemas computacionales de asistencia para la anamnesis de la historia clínica</b>                                                                              | <b>315</b> |
| <b>37.</b> | <b>Datos multimodales en la atención en salud</b>                                                                                                                   | <b>325</b> |
| <b>38.</b> | <b>Segmentación de imágenes en medicina</b>                                                                                                                         | <b>335</b> |
| <b>39.</b> | <b>Datos sintéticos en medicina</b>                                                                                                                                 | <b>345</b> |
| <b>40.</b> | <b>Modelos de visión y lenguaje aplicados en múltiples tareas de la medicina</b>                                                                                    | <b>355</b> |
| <b>41.</b> | <b>Modelos de visión y lenguaje aplicados a patología computacional</b>                                                                                             | <b>371</b> |
| <b>42.</b> | <b>Gemelos digitales en la atención en salud</b>                                                                                                                    | <b>397</b> |
| <b>43.</b> | <b>Los GPT personalizados en medicina y en la atención en salud</b>                                                                                                 | <b>409</b> |



|     |                                              |     |
|-----|----------------------------------------------|-----|
| 44. | Robótica y medicina                          | 423 |
| 45. | Navegación quirúrgica asistida por ordenador | 437 |

SECCIÓN IV. ASPECTOS TÉCNICOS DE LA  
INTELIGENCIA ARTIFICIAL Y PLATAFORMAS

---

|     |                                                                                               |     |
|-----|-----------------------------------------------------------------------------------------------|-----|
| 46. | ¿Por qué la inteligencia artificial depende de las GPU?                                       | 451 |
| 47. | Generación de código con modelos enormes del lenguaje ( <i>Text-to-Code AI Models</i> )       | 455 |
| 48. | Agentes basados en modelos enormes del lenguaje (LLM)                                         | 465 |
| 49. | Métricas para la evaluación de modelos enormes del lenguaje                                   | 477 |
| 50. | Modelos humanos, automáticos y mixtos para la evaluación de los modelos enormes del lenguaje  | 489 |
| 51. | Puntos de referencia ( <i>benchmarks</i> ) para la evaluación de modelos enormes del lenguaje | 497 |
| 52. | <i>Mixture of Experts</i>                                                                     | 507 |
| 53. | Simuladores virtuales del mundo                                                               | 519 |
| 54. | Integración de los modelos enormes del lenguaje en robótica                                   | 523 |
| 55. | OpenAI y ChatGPT                                                                              | 535 |
| 56. | Hugging Face                                                                                  | 543 |
| 57. | Scale AI                                                                                      | 547 |
| 58. | CivitAi                                                                                       | 551 |



## SECCIÓN V. FUTURO, ÉTICA Y LEGAL

---

|            |                                                                                                                                             |            |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>59.</b> | Aspectos éticos generales de la interacción de la inteligencia artificial con el sistema de atención en salud                               | <b>557</b> |
| <b>60.</b> | Primera ley para la regulación de la inteligencia artificial por la Eurocámara                                                              | <b>565</b> |
| <b>61.</b> | Referencias más recientes al riesgo de la inteligencia artificial                                                                           | <b>571</b> |
| <b>62.</b> | Aplicación de medidas de seguridad ( <i>guardrails</i> ) y violaciones de seguridad ( <i>jailbreaking</i> ) en modelos enormes del lenguaje | <b>583</b> |
| <b>63.</b> | <i>Software</i> de código abierto e inteligencia artificial                                                                                 | <b>593</b> |
| <b>64.</b> | Delegación del consentimiento informado a la inteligencia artificial conversacional: avances y desafíos éticos                              | <b>607</b> |
|            | Índice analítico                                                                                                                            | <b>615</b> |



# PREFACIO



Mucho ha pasado desde la publicación del primer libro que escribí acerca de la inteligencia artificial (IA) aplicada a la medicina en 2020. El avance en esta tecnología ha sido impresionante y ha transformado aspectos tanto técnicos como sociales de maneras que pocos habríamos anticipado. En los últimos años, los desarrollos en torno a la arquitectura de los *transformers* han dado pie a innovaciones como los modelos de difusión, los modelos enormes del lenguaje y los modelos multimodales. Estos avances han provocado una disrupción profunda en nuestra sociedad, que no solo ha tocado la investigación y la práctica médica, sino también la vida diaria y la manera como concebimos la tecnología. Este escenario de cambio acelerado me impulsó a actualizar y enriquecer el contenido de aquella primera obra, no solo abordando los aspectos básicos, sino también profundizando en el contexto técnico detrás de estos avances y su aplicación específica en la medicina y otras áreas de la salud.

Además, los recientes premios Nobel otorgados en 2024 a los científicos Demis Hassabis y John Jumper (química), y a Geoffrey Hinton y John Hopfield (física), resaltan la importancia creciente de la IA en el ámbito científico y nos ofrecen una señal de su impacto futuro en nuestras vidas. Estos logros demuestran el papel esencial que desem-



peñará la inteligencia artificial, desde la investigación biomédica hasta la interacción cotidiana, y reafirmará la necesidad de entender estos desarrollos para anticipar sus efectos y aprovechar su potencial.

La inteligencia artificial ha dejado de ser un tema de ciencia ficción para convertirse en un recurso concreto y de uso cada vez más amplio en medicina. Con su capacidad de análisis masivo de datos y sus aplicaciones en diagnóstico, predicción y tratamiento, la IA representa una herramienta fundamental para el profesional de la salud. Este libro pretende ser una guía tanto para aquellos que desean comprender los fundamentos técnicos de estos sistemas, como para los que buscan una perspectiva detallada sobre su aplicabilidad en la atención médica y las ciencias de la salud.

La escritura de este libro ha sido un proceso de descubrimiento continuo, en el que el apoyo de herramientas basadas en inteligencia artificial, como ChatGPT, ha demostrado sus posibilidades de manera sorprendente. ChatGPT no solo facilitó la redacción, para lograr una agilidad y consistencia en el estilo de cada capítulo, sino que también permitió un control preciso de la estructura del contenido y, en algunos casos, la generación de imágenes de apoyo. Estas experiencias con IA en la elaboración de la obra ilustran de manera tangible cómo la tecnología puede mejorar significativamente la productividad y calidad en procesos de escritura extensivos.

A medida que avanzamos hacia una era en la que la IA se convierte en una parte fundamental de nuestra vida, los invito a explorar las páginas de este libro, con la expectativa de que encuentren en él una comprensión más clara de las bases y el futuro de la inteligencia artificial en la medicina y otras áreas de la salud. Este libro ha sido escrito para ser una herramienta útil y accesible para profesionales de la salud y cualquier lector interesado en conocer la tecnología que transformará nuestro futuro.

# AGRADECIMIENTOS



A mi familia, por ser el pilar de mi vida y el apoyo constante en cada paso de este proyecto. Gracias por su paciencia y comprensión frente a las ausencias que este libro demandó, y por compartir conmigo la emoción de cada pequeño logro. A Hellen, Mateo y Valery, mis grandes amores y la razón por la que encuentro la motivación para levantarme cada día; son ustedes quienes llenan mi vida de propósito y alegría. A mis padres, Álvaro y Dioselina, por su confianza, por el sustento y el ejemplo que me han dado. Sus enseñanzas y valores me han acompañado en cada momento de este camino.

A la Fundación Universitaria de Ciencias de la Salud y al Hospital San José, mi segundo hogar profesional, por brindarme un espacio invaluable de aprendizaje y crecimiento continuo. Gracias por ser el laboratorio donde he podido desarrollar y perfeccionar mis habilidades investigativas, docentes y de desarrollo profesional. Este camino, que en sus inicios parecía poco convencional en nuestro entorno, se perfila ahora como la base para un futuro en el que la inteligencia artificial en salud será una herramienta habitual en la formación médica y en la práctica clínica.

A la Subred Integrada de Servicios de Salud del Sur - Hospital El Tunal, y en especial al doctor Jhon Parra, por su confianza y apoyo constante.



Su compromiso con el avance de la investigación en nuestro entorno y su creencia en las posibilidades que esta ofrece han sido fundamentales. A los médicos que participaron en este proceso de investigación, por su colaboración y dedicación. Su trabajo ha sido una pieza esencial para hacer realidad este proyecto.

Sin todos ustedes, este libro no habría sido posible.



**SECCIÓN I**

**INTELIGENCIA  
ARTIFICIAL GENERAL**



# DEFINICIÓN DE INTELIGENCIA ARTIFICIAL

John Sprockel, MD



La inteligencia artificial (IA) se define como la ciencia y la ingeniería de crear máquinas inteligentes. Este término fue acuñado por John McCarthy en 1955. Las definiciones modernas abarcan desde la capacidad de una computadora para realizar tareas humanas hasta la creación de sistemas que emulan procesos intelectuales humanos, como el razonamiento, el aprendizaje y el reconocimiento de patrones. La IA general busca construir máquinas con capacidad intelectual indistinguible de la humana. El aprendizaje automático, una rama de la IA, permite a las computadoras mejorar su rendimiento a partir de datos sin programación explícita.

La primera referencia del término *inteligencia artificial* (IA) se atribuye al profesor emérito de Stanford, John McCarthy. En 1955, el matemático norteamericano definió dicho concepto como «la ciencia y la ingeniería de hacer máquinas inteligentes». Esta denominación fue adoptada luego por la Conferencia de Dartmouth durante el verano de 1956, en la Universidad Dartmouth College, en Hanover (Nuevo Hampshire,



Estados Unidos). Tal evento ha sido considerado como el punto germinal de esta área del saber.

En la actualidad, no existe consenso unificado acerca de cómo debe definirse este término. La Enciclopedia Británica define la IA como «la capacidad de una computadora digital o un robot controlado por computadora para realizar tareas comúnmente asociadas con seres inteligentes». El concepto se emplea con frecuencia en el proyecto de desarrollar sistemas dotados de los procesos intelectuales característicos de los humanos, como la capacidad de razonar, descubrir significado, generalizar o aprender de experiencias pasadas (1).

Por su parte, el Diccionario de Cambridge plantea las siguientes definiciones (2):

1. El uso o estudio de sistemas informáticos o máquinas que tienen algunas de las cualidades del cerebro humano, como la capacidad de interpretar y producir lenguaje de una manera que parece humana; reconocer o crear imágenes; resolver problemas, y aprender de datos que se les suministran.
2. Un sistema informático o máquina particular que tiene algunas de las cualidades del cerebro humano, como la capacidad de interpretar y producir lenguaje de una manera que parece humana; reconocer o crear imágenes; resolver problemas, y aprender de los datos proporcionados a dicho sistema.
3. El uso de programas informáticos que tienen algunas de las cualidades de la mente humana, como la capacidad de comprender el lenguaje, reconocer imágenes y aprender de la experiencia.
4. El estudio de cómo producir computadoras que tengan algunas de las cualidades de la mente humana, como la capacidad de comprender el lenguaje, reconocer imágenes, resolver problemas y aprender.
5. Tecnología informática que permite hacer algo de una manera similar a como lo haría un humano.

Estas definiciones reflejan diversas perspectivas sobre la IA, y sus aplicaciones se centran en su relación con la cognición humana y los objetivos de su estudio. Comparten la idea central de imitar cualidades humanas como el aprendizaje, el reconocimiento de patrones y la



resolución de problemas. La diferencia principal radica en su enfoque: mientras algunas definiciones enfatizan la creación y uso de sistemas computacionales específicos, otras se concentran en la disciplina académica o tecnológica en general. Además, algunas definiciones destacan la capacidad de los sistemas de IA para procesar y aprender de los datos; otras resaltan la simulación de procesos cognitivos humanos.

Otras definiciones para este término incluyen:

- La teoría y el desarrollo de sistemas informáticos capaces de realizar tareas que suelen requerir inteligencia humana, como la percepción visual, el reconocimiento del habla, la toma de decisiones y la traducción entre idiomas (3).
- Disciplina científica que se ocupa de crear programas informáticos, los cuales ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico (4).
- La exhibición de la inteligencia humana por parte de las máquinas, cuando estas pueden pensar y actuar como los humanos (5).

Un término con una importancia especial en la actualidad es el de *inteligencia artificial general* (*artificial general intelligence*, AGI), la cual había sido llamada hace algún tiempo *IA fuerte* (*strong AI*), y tiene como objetivo construir una máquina cuya capacidad intelectual general sea indistinguible de la de un ser humano (1).

Asimismo, es necesario hacer una distinción importante con respecto al término *aprendizaje de máquina* o *aprendizaje automático* (*machine learning*, ML). Se trata de una rama de la IA enfocada en desarrollar algoritmos y modelos que permiten a las computadoras aprender y mejorar su desempeño a partir de datos, sin ser programadas de forma explícita para cada tarea específica. Es una disciplina que se ocupa de la implementación de programas informáticos capaces de aprender de manera autónoma (6). A través del ML, las máquinas pueden identificar patrones y tomar decisiones con una mínima intervención humana, lo cual es fundamental al realizar tareas complejas de IA, como el reconocimiento de voz o de imágenes. Esta capacidad de aprender de los datos diferencia al ML de la programación tradicional y lo ubica como un pilar central en el desarrollo de sistemas inteligentes.



Otra definición relevante es la de *inteligencia artificial generativa*, la cual constituye un subcampo de la IA y se centra en el desarrollo de algoritmos y modelos capaces de generar datos, como imágenes, texto, sonido, video, código, datos sintéticos o cualquier otro tipo de información [7].

## REFERENCIAS

---

1. Copeland BJ. Artificial intelligence [Internet]. Encyclopedia Britannica. 2024 [citado el 12 de febrero de 2024]. Disponible en: <https://www.britannica.com/technology/artificial-intelligence>
2. Cambridge Dictionary. Artificial intelligence [Internet]. [citado el 12 de febrero de 2024]. Disponible en: <https://dictionary.cambridge.org/dictionary/english/artificial-intelligence>
3. Oxford Reference. Artificial intelligence [Internet]. [citado el 12 de febrero de 2024]. Disponible en: <https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095426960>
4. Diccionario de la Real Academia de la Lengua Española, edición del tricentenario. Inteligencia [Internet]. 2023 [citado el 12 de febrero de 2024]. Disponible en: <https://dle.rae.es/inteligencia>
5. Modak S, Mokarizadeh H, Karbassiyazdi E, et al. Chapter 10: The AI-assisted removal and sensor-based detection of contaminants in the aquatic environment. En: Asadnia M, Razmjou A, Beheshti A. Artificial Intelligence and Data Science in Environmental Sensing (A volume in Cognitive Data Science in Sustainable Computing). Londres: Academic Press (2ª edición); 2022. p. 211-244.
6. Hosch WL. Machine learning: artificial intelligence [Internet]. Encyclopedia Britannica. 2024 [citado el 12 de febrero de 2024]. Disponible en: <https://www.britannica.com/technology/machine-learning>
7. Kanbach DK, Heiduk L, Blueher G, et al. The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective. Rev Manag Sci. 2024;18:1189-1220.

# APRENDIZAJE EN INTELIGENCIA ARTIFICIAL

John Sprockel, MD



El aprendizaje en la inteligencia artificial (IA) se define como el proceso mediante el cual sistemas, máquinas y algoritmos mejoran su rendimiento con el tiempo, gracias a la exposición a datos, la experiencia y la interacción con el mundo real. El aprendizaje automático (ML) se sitúa como una rama esencial de la IA, la cual permite a las máquinas aprender de los datos sin estar explícitamente programadas para tareas específicas.

En los siguientes párrafos se desglosa el aprendizaje en la IA en categorías como *aprendizaje supervisado*, *no supervisado*, *semisupervisado*, *autosupervisado* y *por refuerzo*. Se detallan sus métodos, aplicaciones y ejemplos. Además, se introduce el aprendizaje por representación y el *transfer learning*, para lo cual se explica cómo estos transforman datos brutos en información útil y aprovechan conocimientos previos, respectivamente. El texto también aborda técnicas avanzadas como el aprendizaje incremental, multiinstancia, multitarea, federado, de reforzamiento inverso y activo, y se expande el entendimiento de las metodologías de aprendizaje automático, así como sus aplicaciones prácticas en diversas áreas.



El aprendizaje en el área de la inteligencia artificial (IA) se refiere al proceso mediante el cual las máquinas, los sistemas o algoritmos mejoran su rendimiento en tareas específicas con el tiempo, gracias a la exposición a datos, la experiencia y la interacción con el mundo real. Este proceso es fundamental para el desarrollo de sistemas de IA, ya que les permite adaptarse, aprender de sus errores y tomar decisiones más informadas.

Existen diferentes métodos de aprendizaje en IA, cada uno con sus propias aplicaciones y técnicas específicas. A través del aprendizaje, los sistemas de IA pueden identificar patrones, hacer predicciones, reconocer relaciones y adaptarse a nuevos escenarios, lo que les permite realizar tareas complejas que van desde el reconocimiento de imágenes hasta la toma autónoma de decisiones.

El *aprendizaje automático* o *de máquina* (*machine learning*, ML) es una rama crucial de la IA, enfocada en la creación de algoritmos y modelos que habilitan a las máquinas para aprender y mejorar su desempeño a partir de los datos, sin necesidad de estar explícitamente programadas para cada tarea específica. Esta disciplina se dedica al desarrollo de un *software* capaz de aprender de manera autónoma, y ello permite a las máquinas identificar patrones y tomar decisiones con una mínima intervención humana.

El ML es fundamental para desarrollar tareas complejas asociadas a la IA, tales como el reconocimiento de voz o imágenes. A diferencia de los enfoques de programación tradicionales, en los que las instrucciones se definen de forma manual, el ML se basa en la capacidad de las máquinas para aprender de los datos, por lo cual se convierte en un elemento esencial en el avance y el desarrollo de sistemas inteligentes.

El aprendizaje en IA se puede dividir en varios tipos según la forma en que se presentan los datos durante el proceso de entrenamiento. Cada uno tiene sus propias características y aplicaciones que son explicados a continuación:

- *Aprendizaje supervisado*: los modelos se entrenan utilizando un conjunto de datos etiquetados, es decir, cada entrada de datos viene con la salida correcta. El objetivo es que el modelo aprenda a predecir la salida a partir de nuevas entradas, después de haber analizado los



ejemplos de entrenamiento. Se usa principalmente para tareas de clasificación y regresión. Entre los ejemplos se incluyen el reconocimiento de dígitos escritos a mano, la detección de *spam* en correos electrónicos y la predicción de precios de viviendas.

- *Aprendizaje no supervisado*: en el aprendizaje no supervisado, los modelos se entrenan con datos que no están etiquetados, lo que significa que no se proporcionan respuestas correctas. El objetivo aquí es explorar la estructura de los datos para identificar patrones. Este tipo de aprendizaje es útil en tareas de agrupamiento, reducción de dimensionalidad y detección de anomalías. Los ejemplos comunes incluyen la segmentación de clientes en *marketing* y la compresión de datos.
- *Aprendizaje semisupervisado*: el aprendizaje semisupervisado se encuentra entre el aprendizaje supervisado y no supervisado. Utiliza tanto datos etiquetados como no etiquetados para el entrenamiento. Esta técnica es particularmente útil cuando las etiquetas de los datos son difíciles de obtener, lo que a menudo ocurre en situaciones del mundo real. Se aplica en casos como la clasificación de texto y la identificación de imágenes, escenarios en los que etiquetar grandes conjuntos de datos puede ser impracticable.
- *Aprendizaje autosupervisado*: es un subtipo de aprendizaje no supervisado en el que los datos se utilizan para generar sus propias supervisiones. En otras palabras, el modelo se entrena para predecir una parte de los datos a partir de otras partes de los mismos datos. Esto frecuentemente se hace mediante la creación de tareas auxiliares, como predecir la siguiente palabra en una oración o el futuro cuadro en un video. Se utiliza en el procesamiento del lenguaje natural y en el desarrollo de modelos de visión por computadora, entre otros campos.

## APRENDIZAJE POR REFUERZO

---

Un caso especial lo constituye el *aprendizaje por refuerzo*. Se trata de un tipo de aprendizaje de máquina en el que un agente aprende a tomar decisiones mediante la experimentación en un entorno y recibe recom-



penas o penalizaciones por sus acciones. A diferencia de otros tipos de aprendizaje automático, el aprendizaje por refuerzo no se basa en datos de entrenamiento etiquetados de forma previa o en la detección de patrones en conjuntos de datos. En cambio, se centra en encontrar una estrategia, llamada *política*, que le permita al agente maximizar la suma de recompensas que recibe a lo largo del tiempo.

## Técnicas de aprendizaje por refuerzo

### *Métodos basados en valor*

En estos métodos, como *Q-learning* y *Deep Q-Networks* (DQN), el agente aprende a evaluar el valor de cada estado, o la combinación estado-acción, para maximizar la recompensa total. Los valores se actualizan de forma iterativa, a medida que el agente explora el entorno.

### *Métodos basados en políticas*

Estos métodos, como *Policy Gradients*, se enfocan de manera directa en aprender la política óptima sin necesidad de evaluar el valor de cada acción. Esto es especialmente útil en espacios de acción continua o cuando las políticas son más naturales de expresar que los valores.

### *Métodos actor-crítico*

Combinan los métodos basados en valor y en políticas. El *actor* aprende a elegir acciones basándose en una política, mientras que el *crítico* evalúa la acción tomada por el actor, lo cual ayuda a mejorar la política.

### *Aprendizaje por refuerzo multiagente*

Los múltiples agentes interactúan en el mismo entorno. Cada agente puede aprender de sus acciones y de los otros agentes.

## Casos en los que se usa el aprendizaje por refuerzo

- *Juegos*: desde juegos simples, como el tres en raya, hasta juegos complejos, como *Go* y ajedrez, o incluso videojuegos más sofisticados.



- *Robótica*: para enseñar a los robots a realizar tareas complejas como caminar, recoger y colocar objetos, o navegar a través de entornos.
- *Sistemas de recomendación*: en los cuales un agente podría aprender a hacer recomendaciones personalizadas para maximizar algún objetivo, como el tiempo de visualización o la satisfacción del usuario.
- *Control automático*: como en la regulación de la temperatura en sistemas de calefacción, ventilación y aire acondicionado (*heating, ventilation and air conditioning*, HVAC), o en la conducción autónoma.
- *Optimización de carteras de inversión*: en la que el objetivo es maximizar el retorno de la inversión bajo cierto nivel de riesgo.

El aprendizaje por refuerzo sigue siendo un área de investigación muy activa y se aplica en una creciente variedad de campos, lo cual demuestra su versatilidad y potencial para resolver problemas complejos de toma de decisiones.

## APRENDIZAJE POR REPRESENTACIÓN

---

El *aprendizaje por representación*, también conocido como *representational learning* o *feature learning*, es una técnica en el campo del aprendizaje automático que se enfoca en la identificación y extracción automática de representaciones útiles o características relevantes de los datos de entrada. Esta técnica permite que los algoritmos de aprendizaje automático descubran por sí mismos las representaciones necesarias para la detección de patrones o para la realización de tareas específicas, en lugar de depender de la ingeniería manual de características realizada por humanos.

El objetivo del aprendizaje por representación es transformar los datos de entrada crudos en una forma que haga más efectivo y eficiente el proceso de aprendizaje. Estas nuevas representaciones de los datos a menudo revelan información importante y esencial sobre las estructuras subyacentes o las relaciones intrínsecas dentro de los datos.



## Técnicas de aprendizaje por representación

### *Autoencoders*

Son redes neuronales utilizadas para aprender representaciones codificadas de los datos de entrada, al tratar de reconstruir la entrada original a partir de la representación codificada.

### *Redes neuronales convolucionales (CNN)*

Especializadas en procesar datos con una estructura de cuadrícula, como imágenes, en las que aprenden representaciones jerárquicas de las características visuales.

### *Redes neuronales recurrentes (RNN) y transformers*

Utilizadas para procesar secuencias de datos, como texto o series temporales, en las cuales aprenden representaciones contextuales de los elementos en la secuencia.

## Aplicaciones del aprendizaje por representación

- *Reconocimiento de imágenes y visión por computadora:* en los que se aprenden características visuales como bordes, formas y texturas, que son cruciales para tareas como la identificación de objetos.
- *Procesamiento del lenguaje natural (natural language processing, NLP):* para aprender representaciones de palabras o frases que capturan el significado semántico y las relaciones sintácticas.
- *Análisis de series temporales:* en los cuales se identifican patrones subyacentes y características temporales en los datos.

El aprendizaje por representación ha sido fundamental para el éxito de muchas aplicaciones modernas de IA, pues permite que los sistemas aprendan de forma automática las características más efectivas para una tarea dada, lo que resulta en mejoras significativas en el rendimiento y la generalización de los modelos de aprendizaje automático.



## APRENDIZAJE POR TRANSFERENCIA

---

El *transfer learning* o *aprendizaje por transferencia* es una técnica en el campo del aprendizaje automático que consiste en aprovechar el conocimiento adquirido durante el entrenamiento de un modelo en una tarea particular con el objetivo de mejorar o acelerar el aprendizaje en una segunda tarea relacionada, pero diferente. La idea es que, en lugar de comenzar el proceso de aprendizaje desde cero para cada nueva tarea, se pueda transferir parte del conocimiento previamente adquirido, lo que resulta útil cuando se dispone de una cantidad limitada de datos para la nueva tarea.

### Características del *transfer learning*

#### *Reutilización de modelos*

Permite reutilizar modelos previamente entrenados en una tarea, ajustándolos para aplicarlos a una tarea nueva, pero similar.

#### *Ahorro de tiempo y recursos*

Reduce de manera significativa el tiempo y los recursos computacionales necesarios para entrenar modelos profundos desde cero.

#### *Mejora del rendimiento*

Puede mejorar el rendimiento de los modelos en la nueva tarea, especialmente cuando los datos de entrenamiento son escasos.

### Técnicas comunes de *transfer learning*

#### *Fine-tuning*

Consiste en tomar un modelo entrenado en una tarea y ajustar (o *afinar*) sus parámetros con un nuevo conjunto de datos correspondiente a una tarea diferente. A menudo, esto involucra modificar solo las últimas capas del modelo.



### ***Extracción de características***

Utiliza las representaciones aprendidas (características) por un modelo preentrenado y las aplica a una nueva tarea, a menudo agregando nuevas capas específicas para la nueva tarea y entrenando solo esas capas.

### **Casos de aplicación del *transfer learning***

- *Visión por computadora*: transferir conocimientos de un modelo entrenado en un gran conjunto de datos de imágenes, como *ImageNet*, para tareas específicas como la clasificación de imágenes en dominios diferentes o más reducidos.
- *NLP*: utilizar modelos preentrenados en grandes *corpus* de texto para mejorar tareas como la traducción automática, el análisis de sentimientos o la clasificación de texto.
- *Reconocimiento de voz*: aplicar el conocimiento aprendido de modelos entrenados en grandes conjuntos de datos de voz para mejorar la precisión en sistemas de reconocimiento de voz, en dominios específicos o dialectos.

El *transfer learning* ha revolucionado la forma en que se desarrollan y se aplican los modelos de aprendizaje automático, lo cual permite que, incluso con limitadas cantidades de datos, los modelos puedan alcanzar un alto rendimiento en diversas tareas.

Además de los tipos de aprendizaje en máquina ya mencionados, existen otros enfoques y metodologías en el amplio espectro del aprendizaje automático y la IA.

## **APRENDIZAJE INCREMENTAL (*ONLINE LEARNING*)**

---

Este tipo de aprendizaje se refiere a la actualización continua del modelo, a medida que recibe nuevos datos, en lugar de entrenar un modelo con un conjunto fijo de datos. Es particularmente útil para sistemas que necesitan adaptarse a cambios en tiempo real o a flujos de datos continuos.



## **APRENDIZAJE MULTIINSTANCIA**

---

En este marco de trabajo, la formación se realiza sobre bolsas de instancias en lugar de instancias individuales. Cada bolsa contiene múltiples instancias y se etiqueta positiva si contiene al menos una instancia positiva. Se utiliza, por ejemplo, en tareas de clasificación de documentos o en la detección de objetos en imágenes.

## **APRENDIZAJE MULTITAREA (*MULTI-TASK LEARNING*)**

---

Este enfoque implica entrenar un modelo en varias tareas relacionadas de forma simultánea, utilizando una estructura compartida. La idea es que el modelo pueda aprender información útil de cada tarea, lo cual puede ayudar a mejorar su rendimiento en las otras tareas.

## **APRENDIZAJE FEDERADO (*FEDERATED LEARNING*)**

---

Se entrena un modelo a través de múltiples dispositivos descentralizados o servidores que poseen datos locales, sin intercambiarlos. Esto es útil para mejorar la privacidad y seguridad de los datos y se usa en aplicaciones móviles, entre otras herramientas.

## **APRENDIZAJE DE REFORZAMIENTO INVERSO (*INVERSE REINFORCEMENT LEARNING*)**

---

En lugar de aprender a partir de recompensas asignadas a acciones, este modelo intenta entender o inferir la función de recompensa que un agente parece estar siguiendo, basándose en sus comportamientos observados.



## APRENDIZAJE ACTIVO

---

Este método selecciona de forma activa los ejemplos más informativos o útiles para ser etiquetados y usados en el entrenamiento. Es particularmente valioso cuando etiquetar nuevos datos es costoso o requiere de expertos humanos.

## CONCLUSIÓN

---

El aprendizaje en la IA es un campo vibrante y en constante evolución que abarca una amplia gama de técnicas y metodologías. Desde el aprendizaje supervisado hasta enfoques más avanzados como el aprendizaje federado, cada método tiene su propio conjunto de ventajas y aplicaciones prácticas en el mundo real. A medida que continuamos explorando y desarrollando estas tecnologías, es esencial mantenernos informados y comprender las capacidades y limitaciones de cada enfoque. Al hacerlo, podemos aprovechar al máximo el potencial de la IA para resolver problemas complejos, mejorar la toma de decisiones y contribuir al progreso en diversas áreas de la sociedad y la industria.

## LECTURA RECOMENDADA

---

- Datos.gob.es. ¿Cómo aprenden las máquinas? Machine Learning y sus diferentes tipos [Internet]. Ministerio para la Transformación Digital y de la Función Pública (España). 2020 [citado el 1 de junio de 2024]. Disponible en: <https://datos.gob.es/es/blog/como-aprenden-las-maquinas-machine-learning-y-sus-diferentes-tipos>

# GRADIENTE DESCENDENTE Y ALGORITMO DE RETROPROPAGACIÓN

John Sprockel, MD



El gradiente descendente es un algoritmo de optimización utilizado para minimizar una función objetivo que, con frecuencia, es una función de costo o pérdida en el contexto del aprendizaje automático. Este gradiente opera de manera iterativa, ajustando los parámetros del modelo, como los pesos en una red neuronal, en la dirección opuesta del gradiente de la función de costo con respecto a esos parámetros. Mientras tanto, el algoritmo de retropropagación es una técnica específica usada para entrenar redes neuronales. Se trata de una aplicación del gradiente descendente que se utiliza para calcular con eficiencia los gradientes de la función de costo con respecto a cada peso en la red, mediante la propagación del error desde la salida de la red hacia atrás, a través de sus capas. Dicho proceso permite ajustar los pesos de la red, de tal manera que el error de predicción se reduce y ello contribuye a que la red aprenda de los datos de entrada. En el presente capítulo se expondrán los componentes y parámetros críticos de la retropropagación, las variaciones del gradiente descendente, como el descenso del gradiente estocástico (*stochastic gradient descent*, SGD), y el impacto de la arquitectura de la red en el aprendizaje.



El *gradiente descendente* y el *algoritmo de retropropagación* son dos conceptos fundamentales en el aprendizaje automático y la inteligencia artificial (IA), especialmente en el entrenamiento de redes neuronales. Estos forman la columna vertebral del entrenamiento de las redes neuronales modernas y permiten avances significativos en campos como el reconocimiento de voz, la visión por computadora y el procesamiento del lenguaje natural.

El *gradiente descendente* es un algoritmo de optimización utilizado para minimizar alguna función objetivo, a menudo relacionada con el error o la pérdida en el aprendizaje automático. La *función objetivo* o *función de costo* cuantifica el error entre las salidas predichas y las salidas reales. Así, el gradiente descendente actualiza los parámetros del modelo, que suelen ser los pesos de una red neuronal, y los mueve en la dirección que reduce el error. En este contexto, el gradiente se refiere a la derivada de la función de costo respecto a los parámetros del modelo, lo cual indica la dirección y la rapidez con la que el costo varía con respecto a los cambios en los parámetros.

Por su parte, el *algoritmo de retropropagación* (o *backpropagation*) es una técnica específica usada para entrenar redes neuronales y una aplicación particular del gradiente descendente. Este algoritmo permite que la información del error (la diferencia entre la salida deseada y la salida actual de la red) se propague hacia atrás, a través de la red, de modo que se puedan actualizar los pesos de forma eficiente. Ello se logra aplicando la regla de la cadena del cálculo para estimar los gradientes de la función de costo con respecto a cada peso, empezando por las últimas capas y moviéndose hacia las primeras.

Aunque el concepto de *gradiente* se ha conocido desde el siglo XIX en el contexto del cálculo, su aplicación en el aprendizaje automático comenzó a ganar popularidad en las décadas de 1950 y 1960. No obstante, debido a limitaciones computacionales y teóricas, su uso no fue tan extendido hasta el resurgimiento de las redes neuronales en los últimos decenios. Así, la idea de retropropagación ha estado presente desde los años 1960, pero fue en 1986, con la publicación de un artículo seminal desarrollado por David E. Rumelhart, Geoffrey E. Hinton, y Ronald J. Williams, que su aplicación en redes neuronales despegó.



Este trabajo demostró cómo la retropropagación podría utilizarse para entrenar redes con múltiples capas ocultas, con lo cual se superaron de manera significativa las limitaciones de las redes más simples de aquella época.

A continuación, se presenta un pseudocódigo simplificado para el algoritmo de retropropagación utilizado en redes neuronales:

```
Inicializar los pesos de la red de forma aleatoria.
Mientras el criterio de parada no se cumpla:
    Para cada ejemplo de entrenamiento (x, y):
        # Propagación hacia adelante
        Para cada capa l de la red:
            Calcular la activación de los nodos de la capa
            l usando las activaciones de la capa anterior
            (o la entrada x para la primera capa).

            # Retropropagación del error
            Calcular el gradiente del error con respecto
            a la salida (diferencia entre la salida
            predicha y la real).

        Para cada capa l, comenzando por la última y
        retrocediendo hasta la primera:
            Calcular el gradiente del error con respecto
            a los pesos de la capa l.
            Actualizar los pesos de la capa l restando
            un porcentaje (tasa de aprendizaje) del
            gradiente calculado.

Evaluar el desempeño de la red en el conjunto de
validación.
Si se cumple el criterio de parada (por ejemplo, la
convergencia del error), detener el entrenamiento.
```

Este es un esquema básico, en una implementación real se tendrían que considerar aspectos como la elección de la función de activación, el cálculo de los gradientes (que depende de la función de costo y la



función de activación utilizadas) y la definición exacta del criterio de parada. La retropropagación es el corazón del entrenamiento de redes neuronales y les permite aprender de los errores y mejorar las predicciones de forma iterativa.

En el algoritmo de retropropagación, se tienen en cuenta varios parámetros clave para el entrenamiento efectivo de redes neuronales. Estos parámetros incluyen:

- *Pesos y sesgos (biases)*: estos son los principales parámetros ajustados durante el entrenamiento. Los pesos controlan la fuerza de la conexión entre neuronas en capas adyacentes, y los sesgos permiten ajustar la salida de una neurona independientemente de sus entradas. La retropropagación calcula cómo el error cambia con respecto a estos parámetros y los actualiza para minimizar el error.
- *Tasa de aprendizaje (learning rate)*: este parámetro controla la velocidad a la que se actualizan los pesos y los sesgos. Una tasa de aprendizaje demasiado alta puede hacer que el aprendizaje sea inestable, mientras que una tasa demasiado baja puede lograr que el aprendizaje sea muy lento, y es posible que se quede atascado en mínimos locales. La elección de una tasa de aprendizaje adecuada es crucial para un buen entrenamiento.
- *Función de activación*: determina la salida de una neurona, dadas ciertas entradas. Las funciones de activación comunes incluyen la sigmoide, la tangente hiperbólica y la unidad lineal rectificada (*rectified lineal unit*, ReLU). Estas funciones introducen no linealidades en el modelo, lo que es esencial para aprender patrones complejos.
- *Función de costo (o pérdida)*: mide qué tan bien el modelo está realizando una tarea dada. La retropropagación utiliza el gradiente de esta función para actualizar los pesos y sesgos. Las funciones de costo comunes incluyen el error cuadrático medio (*mean squared error*, MSE) para la regresión y la entropía cruzada para la clasificación.
- *Momentum*: es un parámetro que ayuda a acelerar el descenso del gradiente en la dirección correcta y a suavizar las actualizaciones, lo que puede llevar a una convergencia más rápida y evitar quedarse atascado en mínimos locales.



- *Regularización*: ayuda a prevenir el sobreajuste, al agregar un término de penalización a la función de costo. Las técnicas de regularización comunes incluyen L1 (Lasso), L2 (Ridge) y *dropout*.
- *Épocas (epochs)*: se refiere al número de veces que el algoritmo de aprendizaje pasa por el conjunto completo de datos durante el entrenamiento. Una época completa implica cada muestra del conjunto de datos pasando una vez hacia adelante y hacia atrás, a través de la red.
- *Tamaño del lote (batch size)*: se refiere al número de muestras de entrenamiento consideradas antes de hacer una actualización de los pesos. El tamaño del lote puede variar desde 1, que corresponde al descenso del gradiente estocástico (*stochastic gradient descent*, SGD), hasta el tamaño total del conjunto de datos, que corresponde al gradiente descendente por lote.

Estos parámetros son fundamentales para el algoritmo de retropropagación y deben ajustarse de forma cuidadosa para optimizar el rendimiento de la red neuronal. La selección y el ajuste de estos parámetros, conocido como *afinamiento de hiperparámetros*, es una parte crucial del proceso de entrenamiento de la red.

## VARIACIONES DEL GRADIENTE DESCENDENTE

---

El algoritmo del gradiente descendente es fundamental en el aprendizaje automático para optimizar las funciones de costo. Sin embargo, su implementación básica puede ser lenta o ineficaz para grandes volúmenes de datos, o en situaciones en las que el paisaje del error es complejo. Por ello, se han desarrollado variantes del algoritmo.

### SGD

El SGD actualiza los parámetros para cada muestra de entrenamiento o para pequeños grupos de ellos, en lugar de hacerlo para el conjunto completo de datos.



## **Ventajas**

Ayuda a acelerar el proceso de entrenamiento y puede llevar a una convergencia más rápida. Además, la naturaleza estocástica del SGD contribuye a evitar mínimos locales en la función de costo.

## **Desafíos**

La trayectoria de convergencia puede ser *ruidosa*, debido a la actualización frecuente, y puede requerir una disminución cuidadosa de la tasa de aprendizaje.

## **Gradiente descendente por minilotes**

Es una mezcla entre el enfoque de lote completo y el SGD, en la cual la actualización de los parámetros se realiza para cada subconjunto del conjunto de entrenamiento (minilote).

## **Ventajas**

Combina la estabilidad del gradiente descendente de lote completo con la eficiencia y la capacidad de escape de mínimos locales del SGD.

## **Desafíos**

Requiere la elección de un tamaño adecuado para el minilote, lo cual puede influir de forma significativa en el rendimiento del entrenamiento.

## **Métodos adaptativos**

### ***Estimación adaptativa de momentos (adaptive moment estimation, ADAM)***

Combina ideas de *momentum* y tasas de aprendizaje adaptativas. Mantiene estimaciones separadas de los momentos del gradiente (primero y segundo) para cada parámetro.

- **Ventajas:** suele ser efectivo en la práctica y requiere menos ajustes manuales de la tasa de aprendizaje.



- *Desafíos*: puede ser sensible a los valores iniciales de los hiperparámetros, y a veces puede conducir a una convergencia inestable.

### ***Propagación de la media cuadrática (root mean squared propagation, RMSprop)***

Este método modifica la tasa de aprendizaje para cada parámetro, dividiendo la tasa de aprendizaje por una media móvil del cuadrado de los gradientes.

- *Ventajas*: permite que el aprendizaje sea más adaptable a cada parámetro y es especialmente útil en problemas con gradientes muy variables.
- *Desafíos*: al igual que con ADAM, la configuración inicial puede afectar el rendimiento del modelo.

### ***Algoritmo de gradiente adaptativo (adaptive gradient algorithm, AdaGrad)***

Ajusta la tasa de aprendizaje de cada parámetro, basándose en los gradientes pasados, y permite que los parámetros con gradientes frecuentes tengan tasas de aprendizaje decrecientes.

- *Ventajas*: es particularmente bueno para tratar con datos esparcidos y en parámetros que se actualizan con menos frecuencia.
- *Desafíos*: el ajuste continuo puede hacer que la tasa de aprendizaje se reduzca demasiado, lo que ralentiza el entrenamiento.

## **IMPACTO DE LA ARQUITECTURA DE LA RED EN EL APRENDIZAJE AUTOMÁTICO**

---

La arquitectura de una red neuronal, específicamente su profundidad (número de capas) y su ancho (número de unidades o neuronas en cada capa), desempeña un papel crucial en su capacidad para aprender y generalizar, a partir de los datos de entrenamiento. Estos factores impactan de forma directa en el proceso de retropropagación y, en consecuencia, en el aprendizaje global del modelo.



## Profundidad de la red (número de capas)

Las redes más profundas pueden contener funciones con mayor complejidad y capturar relaciones más abstractas en los datos. Sin embargo, a medida que la red se profundiza, el algoritmo de retropropagación puede enfrentarse al problema de los gradientes que desvanecen o explotan, lo que hace que el entrenamiento sea menos efectivo.

Para problemas simples, una arquitectura menos profunda puede ser suficiente y, a su vez, más eficiente. En cambio, problemas complejos, como el reconocimiento de imágenes o el procesamiento del lenguaje natural, suelen requerir redes más profundas para capturar la jerarquía de características y relaciones.

## Ancho de la red (número de neuronas por capa)

Una mayor cantidad de neuronas permite que la red capture más características y patrones distintos dentro de los datos. Sin embargo, un exceso puede llevar a un sobreajuste, en especial si no se dispone de suficientes datos de entrenamiento.

La elección del ancho adecuado depende de la complejidad del conjunto de datos y del problema. Experimentar con diferentes configuraciones y aplicar técnicas de regularización puede ayudar a encontrar el equilibrio correcto entre capacidad y generalización.

## IMPACTO EN LA RETROPROPAGACIÓN

---

En redes muy profundas, el gradiente de la función de pérdida puede disminuir de forma exponencial a medida que se retropropaga, lo que dificulta la actualización de los pesos en las primeras capas. Esto se puede mitigar con técnicas como la inicialización adecuada de pesos, funciones de activación no saturantes y el uso de capas de normalización.

De modo similar, los gradientes pueden crecer de forma exponencial en redes profundas, lo cual causa actualizaciones de pesos inestables.



Se pueden utilizar técnicas como el *gradient clipping* para prevenir este fenómeno.

## **CONSEJOS PARA ELEGIR LA ARQUITECTURA ADECUADA**

- *Comenzar con modelos simples.* Iniciar con arquitecturas menos complejas y aumentar la profundidad o el ancho de manera gradual, con base en la necesidad y los resultados obtenidos.
- *Usar arquitecturas probadas.* En ciertos tipos de problemas, considerar el uso de arquitecturas que han demostrado su efectividad, como las redes neuronales convolucionales (CNN), para visión por computadora, o las redes neuronales recurrentes (RNN), para secuencias temporales.
- *Validación cruzada.* Usar técnicas de validación cruzada para evaluar el rendimiento de diferentes arquitecturas y evitar el sobreajuste.
- *Regularización y técnicas de optimización.* Aplicar técnicas como el *dropout*, la normalización por lotes y los ajustes en la tasa de aprendizaje para mejorar el entrenamiento y la generalización del modelo.

Cada una de estas variaciones responde a desafíos específicos en el entrenamiento de modelos de aprendizaje automático y pueden ser más adecuadas para ciertos tipos de datos o problemas. La elección entre ellas depende de las características particulares del problema, la disponibilidad de datos y la complejidad del modelo.

Se ha demostrado de forma amplia la versatilidad del descenso por gradiente en el campo del aprendizaje automático, al adaptarse a las arquitecturas de redes neuronales cada vez más complejas. Esta versatilidad se refleja en cómo el algoritmo no solo ajusta los pesos de las neuronas individuales, sino también elementos más complejos como los filtros en las CNN y la matriz de atención en los *transformers*, lo cual permite una diversa gama de aplicaciones y mejoras en el rendimiento.

Así, en las CNN, el descenso por gradiente desempeña un papel crucial, al ajustar los filtros (o núcleos) que son utilizados para extraer características de las imágenes. Estos filtros son esenciales para



capturar patrones visuales como bordes, texturas y formas. A través del proceso de retropropagación, el descenso por gradiente ajusta sistemáticamente los valores de estos filtros para minimizar el error de predicción y permite que la red aprenda a identificar las características relevantes, a fin de realizar tareas como la clasificación de imágenes o el reconocimiento facial.

En cuanto a los *transformers*, que han revolucionado el procesamiento del lenguaje natural, el descenso por gradiente se usa para ajustar la matriz de atención. Los *transformers* utilizan mecanismos de atención que ponderan la importancia relativa de diferentes palabras en una oración o distintos componentes en una secuencia de datos. Al entrenar, el descenso por gradiente optimiza estas matrices de atención, lo cual permite que el modelo determine de manera más efectiva qué partes de los datos son relevantes para la tarea en cuestión, como la traducción de idiomas, la generación de texto o la comprensión del lenguaje.

La habilidad del descenso por gradiente para adaptarse y ajustar diversos parámetros y estructuras dentro de modelos complejos ha sido fundamental en los avances de IA. Además, su capacidad para trabajar con diferentes arquitecturas y sus componentes, desde neuronas individuales hasta filtros convolucionales y matrices de atención, demuestra su flexibilidad y eficacia. Esta adaptabilidad asegura que el descenso por gradiente continúe siendo una herramienta valiosa en el desarrollo de modelos de IA más avanzados y especializados.

En síntesis, el gradiente descendente y el algoritmo de retropropagación son pilares fundamentales en la teoría y la práctica del aprendizaje automático y la IA, en especial en el entrenamiento de redes neuronales. A través de la aplicación de estos métodos, podemos afinar modelos de IA para lograr un rendimiento óptimo en una amplia gama de tareas, desde la visión por computadora hasta el procesamiento del lenguaje natural. La comprensión profunda de estos algoritmos y su implementación adecuada es crucial para el desarrollo de sistemas de IA efectivos y eficientes.



## LECTURAS RECOMENDADAS

---

- Khanacademy.org. Descenso de gradiente [Internet]. Khan Academy. [citado el 1 de junio de 2024]. Disponible en: <https://es.khanacademy.org/math/multivariable-calculus/applications-of-multivariable-derivatives/optimizing-multivariable-functions/a/what-is-gradient-descent>
- Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323(6088):533-6.



# PROCESAMIENTO DEL LENGUAJE NATURAL

John Sprockel, MD



El procesamiento del lenguaje natural (*natural language processing*, NLP) es una rama de la inteligencia artificial (IA) que facilita la interacción entre humanos y computadoras mediante el lenguaje natural, abordando tareas como la tokenización, el etiquetado de partes del discurso (*POS tagging*), el análisis sintáctico y el reconocimiento de entidades. Las técnicas fundamentales incluyen el análisis morfológico y el modelado de temas, mientras que sus aplicaciones varían desde asistentes virtuales hasta análisis de sentimientos. El NLP enfrenta retos como la ambigüedad del lenguaje y la variedad lingüística. Los modelos de *transformers* y lenguaje de gran tamaño (*large language model*, LLM) han revolucionado el NLP, pues mejoran la comprensión y generación de texto, aunque plantean desafíos éticos y sociales.

El procesamiento del lenguaje natural (*natural language processing*, NLP) es una rama de la inteligencia artificial (IA) que se enfoca en la interacción entre computadoras y humanos, a través del lenguaje natu-



ral. El objetivo del NLP es leer, descifrar, entender y construir sentido de los idiomas humanos de una manera valiosa. En los últimos años, el NLP ha encontrado su camino en diversas aplicaciones diarias como los asistentes virtuales, la traducción automática, el análisis de sentimientos y más.

En el corazón del NLP se encuentra la capacidad de las máquinas para procesar y analizar grandes cantidades de datos de lenguaje natural. Esto involucra una serie de tareas, que incluyen el etiquetado de partes del discurso, el reconocimiento de entidades nombradas, la extracción de relaciones y la comprensión del lenguaje, entre otras. El desarrollo de tecnologías de NLP se ha acelerado gracias a los avances en aprendizaje automático y aprendizaje profundo, lo que permite a las máquinas entender mejor la complejidad y los matices del lenguaje humano. Ello abre nuevas posibilidades para aplicaciones y servicios que pueden mejorar la forma en que vivimos y trabajamos.

Las técnicas fundamentales del NLP incluyen una variedad de métodos y estrategias diseñados para permitir que las máquinas entiendan y manipulen el lenguaje humano. Estas técnicas son cruciales para desarrollar aplicaciones efectivas de NLP.

## TÉCNICAS FUNDAMENTALES DEL NLP

---

- *Tokenización.* Es el proceso de dividir el texto en unidades más pequeñas, llamadas tókenes, que pueden ser palabras, frases o símbolos. La tokenización es el primer paso en muchas tareas del NLP, ya que prepara los datos para un análisis más profundo.
- *Análisis morfológico.* Se refiere al proceso de analizar las raíces de las palabras y sus inflexiones. Esto incluye la identificación de verbos, sustantivos, adjetivos y otros componentes, así como sus tiempos, géneros y números.
- *Etiquetado de partes del discurso (POS tagging).* Implica asignar etiquetas a cada palabra de un texto (sustantivos, verbos y adjetivos, entre otros), con base en su definición y su contexto.



- *Análisis sintáctico (parsing)*. Consiste en analizar la estructura gramatical de las oraciones, identificando sujeto, predicado y otros elementos, para entender cómo se organizan las palabras en una oración.
- *Reconocimiento de entidades nombradas (named entity recognition, NER)*. Se utiliza para identificar y clasificar elementos claves en el texto, como nombres de personas, organizaciones, ubicaciones, fechas y cantidades.
- *Extracción de relaciones*. Consiste en identificar y extraer relaciones entre entidades nombradas, como vínculos de parentesco, organizacionales y geográficos, entre otros.
- *Análisis de sentimientos*. Es el proceso que consiste en determinar la actitud o emoción del hablante o escritor en el texto, ya sea positiva, negativa o neutral.
- *Desambiguación del sentido de las palabras (word sense disambiguation, WSD)*. Implica determinar el significado de una palabra con base en su contexto, en especial en palabras que tienen múltiples significados.
- *Modelado de temas*. Se utiliza para descubrir los temas subyacentes en grandes volúmenes de texto, agrupando palabras y documentos, de manera que reflejen diferentes temas o asuntos.
- *Traducción automática*. Consiste en convertir texto de un idioma a otro, manteniendo el significado y la intención originales tanto como sea posible.

Estas técnicas son las piedras angulares del NLP y se combinan de diversas maneras para construir aplicaciones complejas, como asistentes virtuales, sistemas de recomendación, herramientas de resumen automático y más. La evolución continua de tales técnicas, impulsada por el aprendizaje automático y el aprendizaje profundo, está abriendo nuevas posibilidades y desafíos en el campo del NLP.

Así, existen varias tareas clave que ayudan a las máquinas a entender, interpretar y generar lenguaje humano. Estas tareas incluyen:

- *Extracción de relación*: detectar y extraer relaciones entre entidades nombradas, como relaciones laborales, ubicaciones y eventos, entre otras.



- *WSD*: resolver la ambigüedad de palabras que tienen múltiples significados según el contexto.
- *Análisis de sentimientos*: evaluar la actitud, la opinión o los sentimientos expresados en un fragmento de texto.
- *Clasificación de texto*: asignar categorías o etiquetas predefinidas a un texto, sobre la base de su contenido.
- *Resumen automático*: generar una versión más corta de un texto que retenga los puntos clave e información importante.
- *Generación de lenguaje natural* (*natural language generation, NLG*): producir texto comprensible a partir de datos estructurados o de una entrada específica.
- *Transcripción o speech-to-text*: implica convertir la palabra hablada en texto escrito. Esta tarea es fundamental en muchas aplicaciones del NLP y tiene una amplia gama de usos, que van desde la transcripción de audio y video en tiempo real hasta la creación de documentos escritos a partir de grabaciones de voz.
- *Traducción automática*: convertir texto de un idioma a otro conservando el significado original tanto como sea posible.
- *Detección de plagio*: identificar si un texto ha sido copiado de otra fuente.
- *Pregunta y respuesta* (*question and answer, Q&A*): construir sistemas capaces de responder preguntas formuladas en lenguaje natural.
- *Comprensión de lectura automática*: permitir que las máquinas lean, entiendan y respondan preguntas sobre un texto dado.

Las aplicaciones del NLP son diversas y se extienden a múltiples sectores, con lo cual impactan de manera significativa la forma como interactuamos con la tecnología y como se procesa la información. Entre algunas de las aplicaciones más destacadas se hallan:

- *Asistentes virtuales*: como Siri, Alexa y Google Assistant, que utilizan el NLP para entender y responder a comandos de voz y preguntas en lenguaje natural.
- *Traducción automática*: herramientas como Google Translate usan el NLP para convertir texto o voz de un idioma a otro, lo cual facilita la comunicación global.



- *Análisis de sentimientos*: diversas empresas usan NLP para analizar opiniones y sentimientos en reseñas de productos, publicaciones en redes sociales y otros textos, a fin de entender la percepción del consumidor.
- *Sistemas de recomendación*: plataformas como Netflix y Amazon utilizan el NLP para analizar las preferencias y comentarios de los usuarios, con lo cual mejoran sus recomendaciones de productos o contenido.
- *Detección de spam y filtrado de contenido*: los servicios de correo electrónico y las plataformas de redes sociales emplean el NLP para identificar y filtrar mensajes no deseados o contenido inapropiado.
- *Herramientas de accesibilidad*: el NLP permite la creación de herramientas que ayudan a personas con discapacidades, como lectores de pantalla para usuarios con problemas de visión o sistemas de comunicación mejorados para aquellos con dificultades del habla.
- *Chatbots y servicios de atención al cliente*: múltiples empresas usan *chatbots* basados en NLP para proporcionar respuestas automáticas a preguntas frecuentes de los clientes. Esto mejora la eficiencia del servicio.
- *Análisis legal y de documentos*: el NLP se utiliza para analizar grandes volúmenes de documentos legales, para lo cual se extrae información relevante, y esto ayuda en la toma de decisiones.
- *Búsqueda y extracción de información*: el NLP facilita la búsqueda de información específica en grandes bases de datos de texto, lo cual permite a los usuarios encontrar de forma rápida lo que necesitan.
- *Resumen automático de textos*: algunas herramientas de NLP generan resúmenes concisos de artículos, informes y otros documentos, lo que ahorra tiempo y facilita la comprensión de grandes volúmenes de información.

## RETOS EN EL NLP

---

El NLP representa un reto significativo por varias razones.



- *Ambigüedad del lenguaje.* La ambigüedad es inherente al lenguaje humano. Una misma palabra o frase puede tener múltiples acepciones según el contexto. Por tanto, desambiguar estos significados es un desafío constante en el NLP.
- *Variedad lingüística.* Existen numerosos idiomas, dialectos y jergas, cada uno con sus propias reglas gramaticales, sintácticas y de vocabulario. Además, el lenguaje evoluciona de forma constante con la introducción de nuevos términos y modismos.
- *Contexto y conocimiento del mundo.* A menudo, la comprensión del lenguaje depende del conocimiento del mundo y del contexto, que puede ser implícito y no estar expresado de forma directa en el texto o el habla. Integrar esta comprensión contextual y el conocimiento general en los sistemas de NLP es complejo.
- *Sarcasmo e ironía.* La detección de tonos como el sarcasmo y la ironía es particularmente desafiante, ya que suele requerir una comprensión profunda del contexto, la cultura y las intenciones del hablante, que no siempre son claras dentro del texto.
- *Procesamiento de grandes volúmenes de datos.* Con frecuencia, el NLP implica analizar grandes conjuntos de datos textuales, lo que requiere recursos computacionales significativos y algoritmos eficientes para procesar y extraer información útil.
- *Integración de múltiples disciplinas.* El NLP es un campo interdisciplinario que abarca lingüística, psicología, informática y ciencia de datos. Fusionar conocimientos de estas áreas para desarrollar sistemas efectivos de NLP es un desafío.
- *Subjetividad y opiniones.* En tareas como el análisis de sentimientos, entender la subjetividad y las opiniones humanas expresadas en el texto requiere no solo el análisis del lenguaje, sino también una comprensión de las emociones y puntos de vista humanos.

Estos retos hacen que el NLP sea un área de estudio y desarrollo intenso, así como en constante evolución, con investigadores y profesionales trabajando para superar estas dificultades y mejorar la capacidad de las máquinas tanto para entender como para procesar el lenguaje humano de manera efectiva.