**Artificial Intelligence: A Modern Approach**

Fourth Edition

**Stuart Russell
Peter Norvig**

Pearson, 2020

por: JESÚS LÓPEZ*

175-188

El libro *Artificial Intelligence: A Modern Approach*. Fourth Edition, de Stuart Russell y Peter Norvig, está conformado por 27 capítulos y 1.119 páginas dedicadas al campo de la Inteligencia Artificial (IA). En el libro se presenta una visión completa y desde una perspectiva moderna de la IA que facilita su definición, la comprensión de su campo, y la presentación de diferentes enfoques y técnicas para el desarrollo de sistemas inteligentes. Con este propósito se abordan temas tales como:

- Definiciones de IA y consecuencias.
- Los agentes inteligentes como modelos de sistemas que aprecian su entorno operan sobre él y aprenden de esas experiencias.
- Técnicas de búsqueda y optimización de soluciones a problemas complejos.
- Tipos de conocimiento que se pueden utilizar y representar en el diseño de sistema inteligentes.
- Enfoques de aprendizaje automático: supervisado, no supervisado y por refuerzo.
- Técnicas de aprendizaje profundo: redes neuronales profundas y redes neuronales recurrentes.
- Enfoques de razonamiento: lógico y probabilístico.
- Técnicas de planificación para el desarrollo de acciones y el logro de objetivos.

* El autor de estas tres reseñas bibliográficas es Profesor-investigador del Área de Teoría y Método de la Planificación del Centro de Estudios del Desarrollo, Cendes-UCV
Correo-e: camiseba@gmail.com - <https://orcid.org/0000-0003-2633-7835>

- Técnicas de percepción para procesar información sensorial que requiere la creación de representaciones del mundo.
- Retos y desafíos de la robótica en la mejora y desarrollo de sistemas inteligentes.

La idea central que integra el contenido del texto es el diseño de agentes inteligentes, de allí que los autores definan a la IA: «como el estudio de los agentes que reciben percepciones del entorno y realizan acciones» (p. vii). Esta conceptualización la inscriben en el trasfondo cultural en el que la IA se ha desarrollado, por lo cual dedican la parte inicial del libro a la revisión de sus fundamentos, historia, estado del arte y sus riesgos y beneficios.

A los fines de resaltar las ideas centrales de esta parte del texto, se ha elaborado una síntesis descriptiva.

Cuadro 1

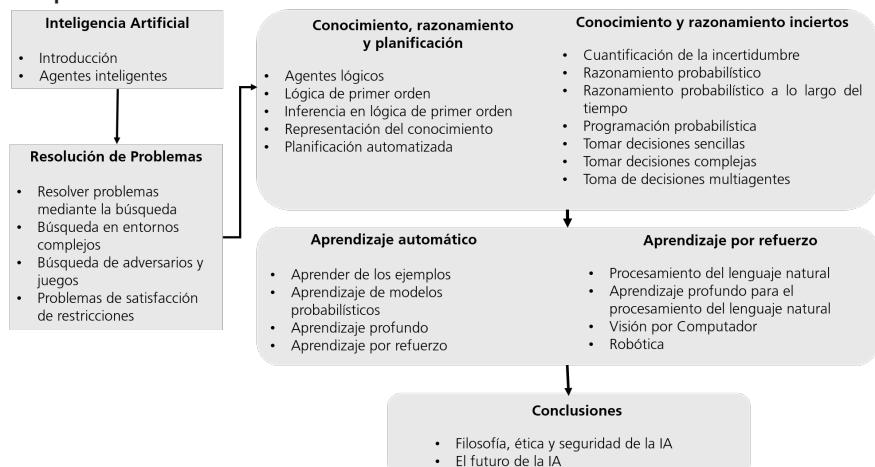
Síntesis descriptiva

Fundamentos	Énfasis en la definición y alcance de la IA: clásicas y actuales, así como en las tareas que pueden asumir los sistemas inteligentes Concepto de agentes inteligentes: unidad básica de la IA que integra entorno, decisiones y acción Arquitectura de un agente: percepción, conocimiento, inferencia y actuación
Historia	Pioneros e hitos: orígenes década de los 50 hasta el pasado reciente Inviernos de la IA: desilusión, reducción de fondos y razones Renacimiento de la IA: avances en hardware, algoritmos y disponibilidad de datos
Estado del arte	Aprendizaje automático: técnicas y aplicaciones en diversos campos Aprendizaje profundo: reconocimiento de imágenes, procesamiento de lenguaje natural y generación de contenido Visión por computadora: técnicas y aplicaciones de reconocimiento y seguimiento de objetos, reconstrucción 3D Procesamiento de lenguaje natural: traducción automática, análisis de sentimientos y generación de textos
Riesgos	Sesgos en los algoritmos: pueden mantener y profundizar desigualdades sociales Pérdida de privacidad: recopilación de una gran cantidad de datos personales Desempleo: automatización de labores realizadas en la actualidad por humanos Autonomía de las máquinas: implicaciones éticas
Beneficios	Mejora de la calidad de vida: medicina, transporte, educación y energía Incremento de la productividad: automatización de actividades repetitivas Nuevas oportunidades de negocio: innovación y diferenciación de la competencia

Fuente: elaboración propia.

A partir de la idea central señalada y del trasfondo cultural en cuestión los autores desarrollan el hilo conductor del texto en siete (7) partes, a saber: a) resolución de los problemas; b) conocimiento, razonamiento y planificación; c) conocimiento y razonamiento inciertos; d) aprendizaje automático; e) aprendizaje por refuerzo; y, f) conclusiones. A los efectos de tener una visión de conjunto del libro se ha elaborado un esquema que integra su contenido por partes y capítulos.

Cuadro 2

Composición del texto

Fuente: elaboración propia.

En cada uno de los veintisiete (27) capítulos del libro (que conforman las siete (7) partes) se presentan los aspectos conceptuales, las metodologías y algoritmos que se pueden utilizar en cada caso, para luego realizar un resumen sobre el tema abordado y recomendar las notas bibliográficas e históricas pertinentes.

Entre los planteamientos y argumentaciones de este texto destacan, entre otros:

- La IA se ocupa principalmente de la acción racional y la conciben como la ciencia del diseño de agentes aproximadamente racionales.
- En su desarrollo teórico y metodológico, el campo de la IA «pasó de la lógica booleana al razonamiento probabilístico, y del conocimiento artesanal al aprendizaje automático a partir de datos» (p.35). Por otra parte, ha logrado una integración efectiva con otras disciplinas.

- Resolver los riesgos y consecuencias éticas de la IA (por su posible evolución impredecible) requiere un cambio en la concepción que tenemos de ella.
- El diseño del programa agente depende de la información, del proceso de decisión y de las particularidades del contexto en cuestión. El cual puede mejorar su desempeño a través del aprendizaje.
- Este diseño va desde agentes de decisiones basados en el conocimiento hasta agentes de aprendizaje profundo, para lo cual se dispone de tecnologías componentes a partir de las cuales se ensamblan los diseños, tales como: razonamiento lógico, probabilístico o neuronal, algoritmos de aprendizaje, etc.
- Existe una variedad de algoritmos de búsqueda que un agente puede utilizar para seleccionar una secuencia de acción en diversos entornos. Incluso en entornos parcialmente observables, no deterministas y desconocidos.
- A partir del análisis de una variedad de juegos se precisa cómo el agente puede actuar en un entorno adverso y se establece el cómo este puede jugar bien en la práctica.
- A un agente inteligente basado en el conocimiento se le define una lógica sobre la cual pueda razonar sobre el mundo y pueda adoptar buenas decisiones.
- El lenguaje de representación del conocimiento es más poderoso que la lógica proposicional. Mientras que la última se centra en los hechos, la lógica de primer orden refiere a objetos y relaciones. En el proceso de inferencia se utilizan una serie de algoritmos.
- La teoría de la probabilidad y, en especial, las redes bayesianas son la base adecuada para el razonamiento incierto, mejorando así las decisiones de un agente estrictamente lógico.
- Es determinante el combinar la teoría de la probabilidad, la teoría de la utilidad y un sistema de decisiones para lograr un agente aproximadamente racional. Incluso cuando los resultados de una acción son inciertos y requieren para llegar a la recompensa de muchas acciones se deben utilizar modelos de transición a partir de cadenas de Markov.

- La teoría de juegos permite describir y sistematizar la interacción multiagentes a la hora de decidir en una situación de interdependencia estratégica.
- El aprendizaje supervisado depende de la naturaleza del agente, del componente a mejorar y de la retroalimentación que se encuentre disponible.
- El tipo de aprendizaje depende de los métodos utilizados en la retroalimentación (redes bayesianas, modelos de Markov, etc.).
- El aprendizaje profundo funciona bien, entre otros casos, para el reconocimiento visual de objetos, el reconocimiento de voz, etc. Las redes recurrentes para el procesamiento de secuencias, por ejemplo, traducciones.
- A partir del aprendizaje por refuerzo un agente puede ser competente en un entorno desconocido, dadas sus percepciones y recompensas esporádicas.
- La robótica trata de agentes incorporados al mundo físico que incluso lo pueden cambiar o transformar. Los algoritmos de filtrado probabilísticos como los de Kalman son útiles para la percepción de los robots.
- La inteligencia artificial es una tecnología poderosa y quienes trabajan con ella tienen el imperativo ético de reducir los efectos secundarios no deseados que implica su utilización en situaciones concretas.
- En relación con el futuro de la IA, los autores señalan que, a pesar de los avances que se han registrado en tan corto tiempo, ellos se acogen a la frase del ensayo de Alan Turing (1950) sobre la maquinaria informática y la inteligencia, donde señala: «*Solo podemos ver a una corta distancia, pero podemos ver que queda mucho por hacer*».

Bajo nuestra perspectiva, es de suma relevancia el comprender el marco de trabajo, las teorías y métodos de la IA debido a su impacto en los procesos de planificación, en el diseño de estrategias, en la prospectiva y la construcción de escenarios, en los procesos de negociación, en la adopción de decisiones etc. Y, muy en especial, en su aplicabilidad a los fines de la construcción de futuros sostenibles en democracia. Por tanto, se hace necesario el disponerlo en nuestros centros de investigación, de enseñanza y observatorios de plataformas para: el desarrollo de modelos IA, la automatización de procesos robóticos, el procesamiento de lenguaje natural, la visión por computadora, el desarrollo de chatbots o de asistentes virtuales.