



# Inteligencia Artificial y Problemas Sociales

Wulfrano Arturo Luna Ramírez  
*Coordinador*

# Inteligencia Artificial y Problemas Sociales

Wulfrano Arturo Luna Ramírez  
*Coordinador*



Dr. José Antonio De los Reyes Heredia  
*Rector General*

Dra. Norma Rondero López  
*Secretaria General*

UNIDAD CUAJIMALPA  
Mtro. Octavio Mercado González  
*Rector*

Dr. Gerardo Francisco Kloss Fernández del Castillo  
*Secretario*

Dra. Gloria Angélica Martínez de la Peña  
*Directora de la División de Ciencias de la Comunicación y Diseño*

Mtra. Silvia Gabriela García Martínez  
*Secretaria Académica de la División de Ciencias de la Comunicación y Diseño*

*Miembros del Consejo Editorial DCCD*

Dr. Manuel Rodríguez Viqueira  
Mtra. Lorena Alejandra Guerrero Morán  
Dr. Noé Abraham González Nieto  
Mtro. Francisco Mata Rosas

*Miembros del Comité Editorial DCCD*

Dr. César Augusto Rodríguez Cano  
Dr. Rodrigo Martínez Martínez  
Mtro. Alejandro Rodea Chávez  
Dr. Mario Alberto Morales Domínguez  
Dr. Joaquín Sergio Zepeda Hernández  
Mtro. José Alfredo Andrade García

# Inteligencia Artificial y Problemas Sociales

Wulfrano Arturo Luna Ramírez  
*Coordinador*



**Q334.7** Inteligencia artificial y problemas sociales [recurso electrónico] /  
**I58** Wulfrano Arturo Luna Ramírez, coordinador.-- Ciudad de México  
**2024:** : UAM, Unidad Cuajimalpa, División de Ciencias de la  
Comunicación y Diseño, 2024.

Datos electrónicos (1 archivo pdf : 2.3 MB)

ISBN (versión impresa) : 978-607-28-3181-0

ISBN (versión digital) : 978-607-28-3182-7

1. Inteligencia artificial 2. Inteligencia artificial -- Aspectos  
morales y éticos. 3. Inteligencia artificial -- Aspectos sociales. 3.  
Innovaciones tecnológicas

I. Luna Ramírez, Wulfrano Arturo, coord.

Clasificación Dewey: **006.304 I61 2024**

---

Inteligencia Artificial y Problemas sociales / Wulfrano Arturo Luna Ramírez |  
Primera edición, 2024.

D.R. © Universidad Autónoma Metropolitana  
Unidad Cuajimalpa  
División de Ciencias de la Comunicación y Diseño  
Avenida Vasco de Quiroga #4871, Colonia Santa Fe Cuajimalpa,  
Alcaldía Cuajimalpa, C.P. 05348, Ciudad de México.

*Diseño Editorial:* Javier Nieves Martínez

*Cuidado de la edición:* Miguel Ángel Hernández Acosta

*Diseño de portada:* Javier Nieves Martínez

<http://www.cua.uam.mx/publicaciones-electronicas/>

Prohibida la reproducción parcial o total de este libro por cualquier medio sin la autorización por escrito de la Universidad Autónoma Metropolitana, el editor o el autor.

Este libro fue arbitrado y dictaminado positivamente por tres dictaminadores, bajo el sistema doble ciego. Ha sido valorado positivamente y liberado para su publicación tanto por el Comité Editorial, como por el Consejo Editorial de la División de Ciencias de la Comunicación y Diseño, Universidad Autónoma Metropolitana, Unidad Cuajimalpa.

ISBN (versión impresa) : 978-607-28-3181-0

DOI: <https://doi.org/10.24275/9786072831827>

Derechos reservados © 2024 | Impreso en México

# Índice

<b>Introducción</b> WULFRANO ARTURO LUNA RAMÍREZ	9
<b>Inteligencia Artificial responsable</b> CHRISTIAN LEMAÎTRE LEÓN	19
<b>Tecnologías Inmersivas, computación espacial y cognitiva, Gemelos Digitales y la capacitación del futuro</b> ANASOL PEÑA-RIOS	51
<b>Tecnología, capital y persona: entre Prometeo y Polifemo. Notas críticas desde el marxismo</b> ENRIQUE G. GALLEGOS	91
<b>La evaluación tecnológica como un vínculo entre la política y la tecnología</b> CAMILO BECERRIL RAMÍREZ	109
<b>Aplicaciones de IA: perspectivas de análisis para la discusión</b> KARINA GUTIÉRREZ FRAGOSO, ELÍAS RUIZ HERNÁNDEZ Y GERMAN CUAYA SIMBRO	121

<b>Respuestas emocionales en la interacción con dispositivos tecnológicos</b>	<b>157</b>
GLORIA ADRIANA MENDOZA FRANCO	
<b>Interfaces cerebro computadora con perspectivas a su aplicación en robots de servicio doméstico</b>	<b>193</b>
MONTSERRAT ALVARADO GONZÁLEZ	
<b>Sistemas Multi-Agente y Modelación Basada en Agentes: una herramienta interdisciplinaria</b>	<b>213</b>
WULFRANO ARTURO LUNA RAMÍREZ	

# Introducción

WULFRANO ARTURO LUNA RAMÍREZ

## **Seminario Inteligencia Artificial y Problemas Sociales**

El Seminario de Discusión Académica Inteligencia Artificial y Problemas Sociales fue un espacio de reflexión sobre los alcances de la Inteligencia Artificial (IA) y su impacto en las sociedades actuales. A su vez, permitió determinar aquellas herramientas que puede aportar esta disciplina al entendimiento y a la modelación de los problemas sociales observados en nuestros entornos. El interés de este espacio de discusión radica en explorar cuáles son las capacidades de la IA para plantear opciones metodológicas y herramientas para analizar y explicar las problemáticas en cuestión. Así, se buscó explicar conceptos, aplicaciones y casos de estudio como paso previo para la formulación de propuestas que atiendan o coadyuven a encontrar soluciones.

El Seminario se llevó a cabo en varias sesiones presenciales, de septiembre a noviembre de 2019, en dos universidades: la Universidad Autónoma Metropolitana, Unidad Cuajimalpa (UAM-C) y la Universidad Veracruzana (UV). El propósito de este seminario

fue adentrarse en la discusión del papel que juegan las teorías y técnicas de la IA en el análisis y propuestas de solución a la problemática social, con especial énfasis en los casos de México y Latinoamérica.

El Seminario contó con la participación de profesores de los departamentos de Teoría y Procesos del Diseño, Tecnologías y Sistemas de Información de la UAM-C, y de las Facultades de Ciencias Agrícolas, y Contaduría y Administración de la UV. En él participaron investigadores de distintas instituciones, tanto nacionales como del extranjero, cuyo quehacer académico y profesional se ejerce alrededor de distintas áreas del saber, principalmente relacionadas con la computación, el diseño y la filosofía.

### **La interdisciplina como base, la complejidad como horizonte**

El Seminario se enmarca en los esfuerzos interdisciplinarios que fundamentan la actividad académica desarrollada en la UAM-C, donde éstos son la constante desde su fundación y se reflejan en proyectos como la Maestría en Diseño, Información y Comunicación.

Hacia dentro del Departamento de Tecnologías de la Información se buscó remarcar y promover la vocación interdisciplinaria de la IA a la luz de la influencia mutua entre subdisciplinas de las Ciencias de la Computación y las Tecnologías de la Información y la Comunicación. En tanto que, hacia afuera de este departamento, se valoró el potencial de la IA para el estudio y la generación de soluciones a las problemáticas sociales, así como sus vínculos con las ambientales. Esto, a la par de promover el encuentro de otros puntos de vista que enriquecen, a su vez, los enfoques y perspectivas de la IA. En este contexto de convergencia interdisciplinaria cabe recordar las palabras de un estudioso de los Sistemas Complejos y la propia interdisciplina, el Dr. Rolando García, quien dice que hay dos momentos donde ésta se logra: el primero se da cuando se plantea un objeto de estudio y el segundo cuando

ese objeto de estudio planteado resulta ser un sistema complejo (García 2006).

Además, de acuerdo con la visión de García, un ejercicio muy productivo e interesante es que los especialistas planteen preguntas, que desde su disciplina no pueden contestar, a colegas de otras formaciones para así ejercer de facto la indagación interdisciplinaria. De esta manera, al ser los problemas sociales de clara naturaleza compleja, es completamente relevante y pertinente estudiarlos a raíz del intercambio ya planteado, que intenta beneficiarse de la confluencia de distintas disciplinas presentes en la División de Ciencias de la Comunicación y Diseño (DCCD) y en la UAM-C en su conjunto.

### **Problemas “locales” y el estudio de problemas sociales**

Ahora bien, el contexto actual nos impele a mirar la amplísima lista de pendientes por resolver en nuestro entorno inmediato (en la sociedad en la que estamos inmersos), que configura un sistema complejo e, incluso, una armazón de ellos.

Además, hay que señalar que la resolución de problemas (a veces tan desdeñada por instituciones, investigadores y las mismas autoridades que deberían encargarse de ello), sean éstos nacionales o regionales, no implica privarse de la participación en las discusiones internacionales ya sean académicas o de otra índole. Es decir, la discusión/solución de los problemas denominados “locales” es en sí un referente fundamental para resolver problemas similares en otras latitudes.

En otras palabras, los problemas locales nunca son estrictamente tales: si representan un problema, su solución como tal plantea una superación epistemológica o práctica que puede ser aplicable en otros problemas de índole similar, una vez ajustados a sus contextos. Es esta una parte fundamental de la misión de nuestra Universidad y, en particular, de la Unidad Cuajimalpa. Al ser la DCCD un pilar importante de ella, debería buscarse la participación en su consecución. Este Seminario planteó un esfuerzo en esa dirección.

Como ya se dijo, dado su propio carácter inter y multidisciplinario, la IA podría orientar algunos de estos esfuerzos de solución. Por lo tanto, en este Seminario se pretendió que los distintos participantes e invitados discutieran desde sus respectivas disciplinas la importancia e impacto de las ciencias y tecnologías relacionadas con la IA en nuestras sociedades. Por otro lado, también se discurió sobre la importancia y aporte que las Ciencias Sociales y las Humanidades pueden realizar para el desarrollo de la propia IA, por ejemplo, para incluir de manera efectiva al usuario final (individual o colectivo) en los procesos de investigación, desarrollo y evaluación de las soluciones propuestas. Una avenida que es interesante explorar bajo esta perspectiva se encuentra en los Sistemas Multi-Agente y la Modelación Basada en Agentes. Estas dos áreas de la IA se benefician de la interacción con los conocimientos y objetos de estudio provenientes de las Ciencias Sociales y las Humanidades. Este intercambio ha dado lugar en otras latitudes a la creación de nuevos enfoques como las Humanidades Digitales o, bien, las Ciencias Sociales Computacionales. Lo anterior evidencia el potencial de las convergencias interdisciplinarias en torno a objetos de estudio comunes.

### **Temas discutidos en el seminario**

El Seminario se centró en la discusión en algunos temas como guía, más que como estructura rígida, en torno de los cuales se desarrollaron las exposiciones de los especialistas que en él confluieron: orígenes y ética de la investigación en IA; Educación y Plataformas Digitales y la aplicación de la IA a éstas; análisis de expresiones sociales y medios digitales; uso de tecnologías inteligentes y su relación con el individuo y la sociedad; las aplicaciones de la IA a la agronomía y sistemas de producción y distribución alimentaria, y la Modelación Basada en Agentes.

Dos temas resultaron fundamentales por su transversalidad: la ética y la crítica a los métodos tecno-científicos actuales. Dentro del primer punto se puede decir que la IA plantea desafíos

éticos que impactan en las sociedades y les exigen a los científicos reconocer su responsabilidad en los alcances de sus investigaciones. La investigación científica, en tanto institución social, está sujeta a presiones internas y externas en su desarrollo y posibilidades. Vale la pena plantear algunas preguntas a este respecto: ¿seguiremos como instituciones de investigación y como sociedad supeditados a los “desarrollos” tecnocientíficos que sólo actúan en beneficio de los grandes corporativos de procesamiento de datos, y sus intereses comerciales y militares? ¿Cómo afrontan los investigadores en IA tales retos? ¿Cómo se puede plantear que la IA, su enseñanza y su desarrollo en tanto ciencia y tecnología, se alinee en el plano internacional con los Objetivos de Desarrollo Sostenible de las Naciones Unidas (Gamez 2015) y, más importante aún, a las necesidades nacionales, si se consideran, por ejemplo, los Programas Nacionales Estratégicos definidos por el Plan Nacional de Desarrollo del gobierno mexicano actual (Pronaces 2024)? Lo anterior demarca, por un lado, los temas relacionados con el ejercicio de la ciencia entendida como derecho humano y el acceso universal al conocimiento y, por otro, la prevención, atención y solución de problemáticas nacionales, agrupados en torno a temas como agentes tóxico y procesos contaminantes, agua, salud, seguridad humana, cultura, sistemas socio-ecológicos, educación, soberanía alimentaria, energía y cambio climático, y vivienda. Temas todos donde la IA puede plantear soluciones, enfoques de investigación y desarrollo pertinentes. De esta forma, los autores que aquí confluyen hacen énfasis en diversos aspectos relacionados con estos programas de manera directa o indirecta.

Por otro lado, entre los problemas éticos fundamentales que se plantean en torno a la IA está el de la definición de lo que es “bueno” y lo que es “malo” para establecer una guía de acción y valoración de las acciones. Además, dicho actuar implica tanto la responsabilidad como la libertad (es decir, la libre decisión de actuar) y una concepción de obligatoriedad moral como

conducta debida (como naturaleza y fundamentos del actuar “como se debe”), ya sea individual o colectiva. Ahora bien, siendo las máquinas (y, por ende, sus programas de control y operación) las encargadas de realizar ciertas acciones, la cuestión se complica aún más: ¿quién define tal conducta debida y a qué intereses sirve? ¿La llamada “economía de la atención” basada en la persuasión automatizada (que se aprovecha de las “debilidades psicológicas humanas” para inducir comportamientos atentando contra la libertad personal) es algo imparabile y por ello debemos resignarnos a aceptarlo? ¿Qué opciones se pueden plantear?

Respecto del segundo punto, algunos autores marxistas nos dicen que si bien los métodos de explotación generados por la moral capitalista (que tiene como sumo bien la plusvalía) dieron origen a una explotación brutal, fue en el siglo XX cuando se dio pleno paso a métodos científicos y racionalizados de producción como los del trabajo en cadena que divide la operación productiva en múltiples partes, lo que provoca que el individuo quede atomizado al margen del resultado completo. De esta forma, se genera una impersonalización del trabajo, y se convierte en una operación repetitiva, monótona, mecánica y agobiante. Diríase como colofón a lo planteado que, en cierta forma, la tecnología y las máquinas degradan al hombre cuando son incluidas en el trabajo en cadena descrito antes y bajo el supuesto de la máxima ganancia como principio de operación y organización.

### **Organización del libro**

El capítulo 1 da inicio al recorrido temático del libro; versa sobre una perspectiva histórica del área de la IA que parte de una breve mirada de sus antecedentes históricos y discurre críticamente su desarrollo actual. Dentro de esta reflexión, a cargo del Dr. Christian Lemaître, se plantea una discusión pertinente y necesaria que, como ya se dijo, es transversal a todos los temas y líneas de discusión: la ética en la IA y sus tecnologías

derivadas (para ello considera como punto central la responsabilidad de todos los involucrados). Así, se traza la necesidad de una IA responsable.

En el capítulo 2 la Dra. Anasol Peña-Rios discurre, de una manera muy vasta, sobre la intersección entre Tecnologías Inmersivas, la computación espacial, los dispositivos prendables o vestibles (conocidos en inglés como *wearables*, un artefacto que se puede implementar como parte de la vestimenta o como accesorio de un atuendo, de ahí su asimilación a prenda/prendable o vestible), los Gemelos Digitales y la IA. Además, se muestra cómo esta convergencia hace posible que un empleado sea entrenado y asesorado vía remota en una representación digital de su entorno y materia de trabajo, y las posibles aplicaciones de estas tecnologías en otros dominios.

Por su parte, el capítulo 3, a cargo del Dr. Enrique Gallegos, recoge el intercambio entre IA y las Ciencias Sociales en uno de los momentos más álgidos dentro de la discusión del Seminario al plantear una crítica a la tecnología. El autor infiere sobre la relación entre capital, tecnología y persona a partir de sus notas críticas desde el marxismo.

El Mtro. Camilo Becerril, en el capítulo 4, muestra que la evaluación tecnológica puede hacer el vínculo para conectar la actividad política o democrática con el fenómeno de la tecnología.

El capítulo 5 muestra ciertas perspectivas de análisis para discutir las aplicaciones de la IA y sus diversas ramas (como la ciencia de datos) en distintos campos como el agro y la industria. A cargo de los Dres. Karina Gutiérrez, Elías Ruiz y German Cuaya, el texto discute la IA en una perspectiva industrial que reconfigura los espacios de trabajo y producción. Además, se mencionan las implicaciones sociales y éticas de la IA en la medicina.

Apartir del campo emocional, el capítulo 6 aborda las respuestas de este tipo en la interacción con dispositivos tecnológicos. La Mtra. Gloria Mendoza presenta una discusión sobre las

interfaces humano-computadora y la integración de las emociones en su diseño y sus posibles aplicaciones, así como en los robots sociales.

En el capítulo 7, la Dra. Montserrat Alvarado da continuidad a la discusión de las interfaces humano-computadora, en un área de especialización que hasta hace poco estaba en el dominio exclusivo de la ciencia ficción: las interfaces cerebro-robot, en este caso, para controlar un robot de servicio doméstico.

El volumen se cierra con el capítulo 8, en el que quien estas líneas escribe hace un recorrido por una rama de la IA que se ha postulado como unificadora del campo, que toma como elemento central un modelo teórico y de Ingeniería de Software: los agentes artificiales. También se hace mención sobre la Modelación Basada en Agentes como herramienta de utilidad tanto para las ciencias exactas como para las sociales y humanas.

### **Comentarios finales**

Este volumen recoge en sus capítulos los planteamientos de diversos especialistas en torno a la aplicación de la IA a los problemas sociales y la repercusión de éstos, a su vez, en la configuración de la disciplina.

El enfoque interdisciplinario de la discusión y los temas tratados durante el Seminario y los textos aquí presentados ponen de manifiesto cómo la ética y otras disciplinas humanas dejan su impronta en la IA de manera transversal.

Por un lado, se reconoce la importancia de alcanzar el progreso de las condiciones sociales y materiales mediante la generación de métodos, mecanismos, máquinas (sistemas y dispositivos, diríamos ahora, incluyendo algoritmos y robots) que ayuden al ser humano en sus tareas diarias, que eleven la productividad, mejoren la comunicación, asistan en los diagnósticos médicos, las cirugías, y un largo etcétera. Se señala también que con el arribo de la IA no sólo se han acelerado y perfeccionado su uso, sino que incluso abarca las cuestiones estéticas (es decir, el arte

no sólo asistido, sino producido por máquinas). Pero se advierte de no caer en una postura facilista del progreso material basado en la tecnificación, que lleva a una visión deformada que adjudica una supuesta “neutralidad” de la ciencia/tecnología que deja de dar cabida al dicho común: “Yo hago la herramienta/proceso/desarrollo para que sea la mejor; si se usa para construir o destruir, eso ya no me compete”.

Además de pedestre, la postura mencionada antes raya en el cinismo si se consideran ciertos problemas actuales: discriminación automatizada, vigilancia extrema y acumulación de capital y poder, entre otros temas discutidos en los capítulos aquí presentados. Ante esto, cabe plantear algunas preguntas: ¿son los científicos/tecnólogos de la IA (en tanto diseñadores de sistemas, algoritmos, dispositivos, protocolos, etcétera) los responsables de que el dispositivo/máquina/programa realice la conducta debida (ética-moral)? ¿Lo son los programadores/construtores/fabricantes? ¿O son sus empleadores (dueños de firmas/compañías, directivos, jefes), en tanto directores de lo que ha de ser fabricado/programado a través de sus políticas institucionales comerciales y líneas de producción los responsables? ¿Lo son los revendedores de tales tecnologías? ¿O, a fin de cuentas, lo son los usuarios (en tanto disparadores finales, o al menos intermediarios) de las acciones de los dispositivos/programas?

Al ser este el tema central de las discusiones mostradas en este volumen es previsible que las siguientes ediciones de este Seminario de Discusión Académica sobre la Inteligencia Artificial y problemas sociales (y las discusiones del área en el conjunto de la comunidad académica y la sociedad toda) se extiendan en distintas direcciones y se especialicen de diversas formas, pero sean tamizadas, indirecta o directamente, por la cuestión ético-moral.

### **Agradecimiento**

Mi agradecimiento a la Mtra. Lucila Mercado Colín y el Dr. Gustavo Ortiz Hernández por su contribución a la organización del

Seminario de Discusión Académica Inteligencia Artificial y Problemas Sociales. Agradezco también al Mtro. Octavio Mercado González por la convocatoria, y al personal de la División de Ciencias de la Comunicación y Diseño por el apoyo durante el evento del cual se origina este trabajo.

Finalmente, este texto no hubiera sido posible sin la ayuda del equipo del Centro de Escritura y Argumentación y especialmente de la Mtra. Valeria Amanda Benítez Rosete, quienes aportaron su valioso conocimiento a la revisión del manuscrito, aunado al apoyo de Lizbeth Verónica Ochoa Fonseca.

## Referencias

- GAMEZ, María José (2015). “Objetivos y metas de desarrollo sostenible, Desarrollo Sostenible”. Acceso el 10 de enero de 2024. <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>.
- GARCÍA, Rolando (2006). *Sistemas Complejos. Conceptos, método y fundamentación epistemológica de la investigación interdisciplinaria*. Barcelona: Gedisa.
- PRONACES (2024). “Programas Nacionales Estratégicos”. Consejo Nacional de Humanidades, Ciencias y Tecnologías. Acceso el 9 de enero de 2024. <https://conahcyt.mx/pronaces/>.

# Inteligencia Artificial responsable

CHRISTIAN LEMAÎTRE LEÓN

## Resumen

La computación electrónica ha tomado un papel relevante dentro de la sociedad actual. Su origen se encuentra en el ámbito militar, del cual poco a poco emigró hacia propuestas comerciales. Su desarrollo se debe a la confluencia de distintas ciencias como la cibernética, la teoría de general de sistemas y la teoría de la información, además de los avances e innovaciones en el campo de la electrónica que dieron cabida a la aparición de lo que se conocería más tarde como Inteligencia Artificial. De ésta, si bien nació formalmente con ese nombre en la década de 1950, sus raíces intelectuales y tecnológicas pueden rastrearse hasta la Grecia o China clásicas.

El área ha evolucionado en distintos periodos de auge y declive principalmente por cuestiones relacionadas con intentos de obtención de financiamiento basados en promesas que después tuvieron que ser reevaluadas y adaptadas. Tal es el caso de las actuales propuestas como el “cómputo en la nube”, las redes

profundas y los métodos de mercadotecnia digital basada en la recopilación y análisis de grandes cantidades de datos. Esta disciplina plantea múltiples áreas de aplicación que, aunadas a una mayor tecnificación de distintos aspectos de la vida contemporánea, hace que renazca el interés por sus consecuencias para la sociedad. Esta discusión se ha presentado en distintos momentos en la historia de la ciencia, y uno de los puntos álgidos de ella fue la que se presentó en el área de la física a partir del Proyecto Manhattan y la creación de la bomba atómica, que dividió a la comunidad científica de manera profunda.

De esta forma, cabe cuestionarse sobre la responsabilidad social por los malos usos y consecuencias negativas del desarrollo de esta ciencia y sus tecnologías y aplicaciones derivadas. La tesis que aquí se presenta considera que los científicos implicados, al ser parte de la sociedad, no pueden sustraerse de dicha responsabilidad y, por el contrario, conscientes de ésta, deberían pugnar por un desarrollo del área en beneficio de la humanidad por sobre los modelos de negocio imperantes que atentan contra ella.

**Palabras clave:** historia de la Inteligencia Artificial, ética en la Inteligencia Artificial, cómputo en la nube, redes sociales, cibernética.

## Introducción

A lo largo de los últimos ochenta años, la computación electrónica ha tomado cada vez más un papel protagónico en todos los ámbitos de la sociedad contemporánea. Esta relación se inició en secreto en torno a proyectos relacionados con la Segunda Guerra Mundial, en Alemania, Inglaterra y Estados Unidos. Después, en la posguerra, la siguiente generación de computadoras electrónicas se desarrolló sobre todo en centros de investigación; aunado a ello aparecieron, de forma modesta, las primeras computadoras electrónicas comerciales. Para la década de 1950 existían grandes empresas de computación electromecánicas, basadas en máquinas tabuladoras cada vez más versátiles. Estas empresas fueron

las principales introductoras de las nuevas computadoras electrónicas. En paralelo aparecieron muchos otros desarrollos científicos y tecnológicos en torno a la naciente ingeniería electrónica, además de que se dio el descubrimiento del transistor, sin el cual la historia del mundo actual sería muy diferente.

En el campo más teórico aparecieron propuestas integradoras como la cibernética, de Norbert Wiener y Arturo Rosenblueth, o la teoría general de sistemas de Ludwig von Bertalanffy. Dentro de estas grandes corrientes destacaron, por un lado, la teoría matemática de la información de Claude Shannon; el primer modelo matemático del funcionamiento de una neurona, de Warren McCulloch y Walter Pitts, y la aparición de una propuesta de Alan Turing sobre cómo poder programar una computadora para emular la inteligencia humana junto con una prueba para determinar si ese objetivo ha sido alcanzado (conocida como la “prueba de Turing”).

Podemos resumir que en esos años apareció como objeto de estudio científico multidisciplinario la indagación del funcionamiento del cerebro, la cibernética, el modelo de McCulloch-Pitts y los planteamientos de Turing se centraron desde ópticas complementarias en el objetivo de entender cómo funciona el cerebro a partir de modelos matemáticos y computacionales. Estas nuevas vías de investigación científica establecieron las bases de muchos de los temas que nos ocupan hoy.

En 1956 un grupo de investigadores, encabezados por Claude Shannon, John McCarthy y Marvin Minsky, realizó un seminario que dio nacimiento al área de la Inteligencia Artificial (IA). Si bien el desarrollo posterior de esta disciplina pareció ser autónomo del contexto científico y tecnológico de las décadas 1940 y 1950, no podemos entender los desarrollos sorprendentes de aplicaciones de la IA si no tomamos en cuenta el conjunto de estas teorías desarrolladas en los inicios de la posguerra.

Hoy la IA, y en particular una de sus ramas de investigación (la de las redes neuronales profundas y los algoritmos asociados de

entrenamiento de estas estructuras matemáticas, lo que comúnmente se conoce como “aprendizaje profundo”), ha saltado a la palestra de las tecnologías más disruptivas de nuestro entorno económico y social.

En las siguientes secciones voy a exponer mi visión de cómo esta tecnología ha logrado llegar a donde está y cuáles son los problemas que se han detectado con su uso masivo. Esto con el objetivo de plantear algunas interrogantes mucho más serias sobre los retos que enfrentamos como especie en el planeta Tierra y los enfoques que deberían privar en el uso, en ese contexto, de las tecnologías de la computación, incluyendo la IA.

### **Breve recuento histórico de la Inteligencia Artificial**

La IA como disciplina se nutre de dos ideas centrales: *a*) la creación de autómatas antropomorfos o zoomorfos, es decir, artefactos que toman como inspiración modelos a imagen y semejanza del ser humano o de otros animales buscando emularlos de manera artificial (esta idea está presente prácticamente en todas las culturas); y *b*) el desarrollo de maquinaria capaz de realizar manipulaciones con información simbólica incluyendo la numérica, idea que da origen a los desarrollos de la historia de la computación. Estas dos líneas que desembocan en la IA pueden trazarse hasta las civilizaciones tan antiguas como la egipcia, la griega o la china, en donde se construyeron artefactos con cierta movilidad que imitaban figuras humanas o animales, así como dispositivos musicales que entonaban melodías sin intervención humana directa.

Por ejemplo, la cultura griega estaba permeada por la idea de que era posible crear seres semejantes a los humanos capaces de tener comportamientos tan complejos como ellos. En la mitología griega se establecía que Hefesto (dios de los herreros y de los artesanos) había construido dos ayudantes (las Kourai Khryseai) que realizaban de manera autónoma las más diversas tareas. En la vida real hubo varios artesanos capaces de construir artefactos

diversos con movimientos y aparente autonomía. El más conocido es Herón de Alejandría, quien construyó un buen número de ingeniosos aparatos. Sin embargo, quizás el hallazgo más enigmático y sorprendente es el mecanismo de Anticitera, artefacto descubierto a principios del siglo XX y analizado en la década de 1970 por Derek John de Solla Price, quien concluyó que este objeto era un complejo sistema de relojería astronómica, forjado en bronce, que poseía una precisión muy grande en el diseño de los múltiples engranes que lo conformaban. Una verdadera computadora analógica. Si bien existen algunas referencias en la literatura romana sobre artefactos de mecánica de precisión griegos, éste de Anticitera es el único que se ha encontrado hasta el momento. Las implicaciones de su descubrimiento son importantes pues demostraría un dominio de la mecánica de precisión que sólo podría compararse con los niveles alcanzados por la industria relojera del siglo XVIII y, en el campo de la astronomía, con las construcciones de relojes astronómicos de gran tamaño del siglo XIV.

Posteriormente, en las civilizaciones China e Islámica, encontramos artesanos de gran destreza que construyeron diversos tipos de autómatas. A finales del siglo XVII e inicios del XVIII en Europa se perfecciona la industria relojera de precisión y fue entonces cuando algunos de los maestros relojeros se convirtieron en creadores de diversos sofisticados autómatas capaces de mostrar destrezas sólo realizadas por seres humanos. Es así como aparecen muñecos autómatas capaces de ejecutar un fragmento de piezas musicales tocando pianos, clavecines o flautas, así como escribas capaces de plasmar con su mano mecánica textos en letra manuscrita de gran fineza en su trazo. Tales autómatas contruidos con complejos mecanismos de relojería y de caja musical constituían verdaderas computadoras analógicas de gran belleza y de una ingeniería que envidiarían los laboratorios de robótica actuales.

Estos autómatas (del siglo XVIII en adelante) eran piezas de entretenimiento de las cortes de toda Europa. Fue entonces, en

la segunda mitad de ese siglo, cuando apareció uno muy peculiar que mostró cómo la ingeniería podía ser utilizada desde el momento de su propio diseño como un engaño. Se trató del falso autómatas de Wolfgang von Kempelen conocido como el “turco” o el “ajedrecista”, debido a que era un muñeco vestido como turco y quien estaba situado frente a una caja donde reposaba un tablero de ajedrez con todas sus piezas. Este autómatas era capaz de jugar partidas completas de ajedrez a un alto nivel de desempeño, y causó gran interés en las cortes europeas más importantes en donde en más de una ocasión derrotó a los monarcas que osaban retarlo.

Sin embargo, ese “ajedrecista” era un timo, pues se trataba de una caja dentro de la cual podía instalarse un ajedrecista de baja estatura quien era el que en realidad jugaba y movía las manos del autómatas y las piezas del tablero. Sin duda que el mecanismo en su conjunto era una muestra de gran ingenio, pero no se presentaba como acto de magia, sino como un verdadero autómatas. En los términos actuales diríamos que era una violación ética del trabajo de los maestros creadores de autómatas. Por otra parte, no deja de sorprender cómo las personas de las cortes de toda Europa no dudaban de que fuera posible hacer máquinas capaces de practicar un juego tan difícil como el ajedrez. Al parecer era algo que se inscribía en el imaginario colectivo como una posibilidad real. Debido a que los autómatas en sí eran mecanismos muy complejos, se establecía un obstáculo para la comprensión de su funcionamiento para la gran mayoría de personas. Esa opacidad permitía todo tipo de interpretaciones e historias en torno a estos artilugios, algo similar a lo que hoy sucede con muchas aplicaciones de la IA.

Otra vertiente de historias del uso maquiavélico de la tecnología aparece a principios de la industria fílmica con la película de *Metrópolis*, de Fritz Lang. En ella, un robot humanoide femenino suplanta la personalidad de María, la protagonista de la historia, y crea el caos bajo las órdenes de su creador. Este robot, no

es autónomo, pero es capaz de emular perfectamente a María y confundir a todo el pueblo trabajador que veía en la verdadera a su líder. Estamos hablando de 1927.

Dejemos por el momento esta línea de desarrollo de la prehistoria de la IA y retomemos el hilo de la segunda línea del desarrollo de mecanismos capaces de manipular información simbólica y especialmente información numérica.

La necesidad de contar con herramientas auxiliares para realizar los cálculos surge con la aparición de los primeros sistemas numéricos. Por ejemplo, en la civilización sumeria aparece por primera vez la escritura y la aritmética, y se crearon escuelas de escribas que formaban a jóvenes en el arte de escribir y hacer cuentas. Para poder realizar estas últimas se crearon tablas de multiplicación para su sistema numérico de base 60, lo que aceleraba la ejecución de todos los cálculos necesarios para el mercado de bienes y la contabilidad del Estado. Posteriormente, en las diversas civilizaciones aparecieron diferentes tipos de ábacos. Sin embargo, el paso decisivo en la mecanización de los cálculos se dio en el siglo XVII con el invento de dos aparatos de cálculo que permitían realizar automáticamente sumas y restas. La primera fue construida por Wilhelm Schickard en 1623 y la segunda por Blaise Pascal en 1642. La primera no tuvo mayor impacto en la construcción de otras máquinas, mientras que la de Pascal contó con una difusión e influencia mayor. El siguiente avance significativo se debió al invento de Gottfried Leibniz, medio siglo más tarde, pues introdujo un ingenioso mecanismo para llevar la cuenta de las sumas sucesivas y poder así mecanizar las multiplicaciones.

Una componente esencial de las máquinas de procesamiento de la información surge a lo largo del siglo XVIII en la nascente industria textil cuando, en 1801, aparece el Telar de Jacquard. Esta máquina por ser automática podía acoplarse a una fuente central de energía para poder operar a gran escala, lo que dio nacimiento a la gran industria en los albores de la revolución

industrial. El elemento central de este telar era el uso de tarjetas perforadas que permitía codificar los patrones de diseño de una tela en una serie de patrones de agujeros perforados en las tarjetas. Podemos describir el propósito de este telar como el de un traductor mecánico de la información codificada en las tarjetas perforadas a los diseños de los patrones de colores (patrones de información) que se plasmaban en las telas. El origen de este proceso de traducción fue inventado en 1725 por Basile Bouchon quien se inspiró en el funcionamiento de los organillos musicales.

El siguiente componente indispensable en nuestra historia es la aparición de una metodología de simplificación sucesiva de tareas de cálculo complejas capaz de realizar con gran exactitud cálculos de funciones tan complicadas como las trigonométricas o logarítmicas a partir de operaciones elementales de sumas y restas. Este método fue desarrollado por Gaspard de Prony quien fue encargado durante la revolución francesa de calcular las tablas de funciones trigonométricas y de logaritmos con una gran precisión. Esta proeza la pudo realizar en tan solo dos años inspirándose en el método de la división del trabajo descrita por Adam Smith en su obra *La riqueza de las naciones* de 1776. La orquestación de esa monumental labor se basó en una organización del trabajo en tres niveles. El primero fue la responsabilidad de matemáticos de alto nivel encargados de describir la manera de descomponer las operaciones más complejas en sucesiones de operaciones simples de suma y resta. El segundo era responsabilidad de personas con entrenamiento matemático quienes debían preparar las secuencias de operaciones a ser calculadas. Finalmente, el tercer nivel era responsabilidad de un grupo más amplio de personas con un entrenamiento básico en aritmética, quienes eran responsables de ejecutar las sumas y restas que habían sido preparadas por el equipo del nivel medio. Este fue el primer ejemplo de la realización de una tarea a gran escala basada en la división del trabajo intelectual y no manual como el estudiado por Adam Smith.

Las máquinas de cálculo, la codificación de información en tarjetas perforadas y la división del trabajo para calcular los valores de una función cualquiera a base de sumas y restas fueron las tres fuentes principales de inspiración de Charles Babbage quien a partir de 1833 dedicó todo su esfuerzo y fortuna en diseñar una máquina capaz de calcular cualquier función analítica por medios mecánicos, la máquina analítica. Si bien no logró construirla, sus diagramas de diseño permitieron construir (más de un siglo después) un ejemplar que funcionaba exactamente como fue planeado. En esta tarea intelectual, Babbage tuvo un valioso apoyo en Ada Lovelace con quien discutió el tema a lo largo de años. Fruto de esto desarrolló ideas de gran originalidad como, por ejemplo, la propuesta de generar música con la máquina analítica, la cual era un artefacto diseñado para realizar cálculos numéricos que podría generar patrones musicales en lugar de patrones numéricos. Ésta es una idea antecedente directa de los planteamientos de la IA de mediados del siglo XX.

Ahora bien, la primera mitad del siglo XX se vio marcada por dos grandes momentos para la historia de la computación y de la IA. El primero es la publicación en 1926 de la propuesta de Alan Turing para definir lo que es una función calculable. En ese trabajo el británico propuso un método constructivista con la definición de una máquina teórica conocida como la “Máquina de Turing”. Esta idea proporcionó un resultado teórico que establecía el fundamento principal de la Ciencia de la Computación: toda función calculable puede ser programada en una Máquina de Turing.

El segundo momento fue marcado por la creación de las válvulas termoiónicas o bulbos, que iniciaron el desarrollo de la electrónica. Además, durante la Segunda Guerra Mundial, surgieron tres esfuerzos por construir computadoras electrónicas: la Z4, de Konrad Zuse en Alemania; la Colossus, de Tommy Flowers en Inglaterra, y la ENIAC, de Ecker y Mauchly.

La década de 1940 merece una mención especial, pues además de los desarrollos de los primeros prototipos de computadoras

electrónicas, emergieron otros avances teóricos fundamentales para entender las influencias directas del nacimiento de la IA. En 1943 se publicó un artículo de Warren McCulloch y Walter Pitts quienes presentaron un modelo matemático del funcionamiento de una neurona. Este trabajo es considerado como el originador de lo que hoy se conocen como redes neuronales artificiales. Asimismo, en 1948 se publicó el artículo de Claude Shannon “Una teoría matemática de la comunicación”, que consolidó el marco teórico de la pujante industria de las telecomunicaciones. A lo largo de los años de la posguerra, Norbert Wiener junto con Arturo Rosenblueth desarrollan una propuesta teórica general que denominan “cibernética” y que buscaba englobar una teoría de la comunicación y el control en los animales y las máquinas. El libro de Norbert Wiener, *Cibernética*, se publicó en 1948.

Finalmente, en 1950 Alan Turing publicó su muy influyente artículo “Maquinaria computacional e inteligencia”, en el que discute sobre las capacidades de las nuevas máquinas computadoras electrónicas y sus aplicaciones más allá de efectuar cálculos a alta velocidad. Turing las presenta como máquinas en las que se pueden codificar operaciones simbólicas de muy diverso tipo. Su propuesta más novedosa es quizá la de desarrollar programas de cómputo que permitan a las máquinas ir adquiriendo conocimiento conforme se van utilizando, simulando así lo que es el proceso de aprendizaje en los niños. Turing discute en general las posibilidades de que las computadoras puedan simular procesos cognitivos que se dan en el ser humano y que consideramos una muestra de inteligencia. Para evaluar esos futuros desarrollos propone una prueba experimental, conocida como prueba de Turing, que permitiría determinar si se podía considerar a una máquina (programa) como inteligente.

### **La cuestión ética de la ciencia**

No podemos dejar de lado un problema mayor que surgió en el mundo, y en particular en el mundo científico, con las detonaciones

nucleares sobre las ciudades de Hiroshima y Nagasaki en 1945 con las que Estados Unidos pretendía terminar la guerra contra Japón. Esas explosiones develaron de manera trágica el proyecto estadounidense de fabricar bombas atómicas gracias a la ayuda de un grupo de investigadores en Física quienes participaron en el “Proyecto Manhattan”. Lo anterior daba cuenta de un ejemplo aterrador del poder destructor del uso del conocimiento científico, lo cual escindió a la comunidad de físicos, entre quienes apoyaban la decisión del presidente de Estados Unidos y de su estado mayor de lanzar esas bombas sobre la población civil y quienes la rechazaban por razones éticas y humanitarias.

Ésta fue la primera vez que se dio una discusión a fondo sobre la responsabilidad social de los científicos. Si bien las investigaciones son resultado de decisiones personales, ellos no podían escudarse en la idea de una ciencia neutra, y trasladar a la sociedad los problemas que puedan causar sus aplicaciones o su desarrollo por parte de ellos. En ese entendido, para muchos quedó claro que los investigadores científicos deben hacerse responsables de lo que otros hagan con sus descubrimientos y luchar por que esas aplicaciones negativas se prohíban. En otras palabras, hay una responsabilidad ética inherente en todo desarrollo de nuevo conocimiento.

La discusión sobre la ética en la investigación y desarrollo de la física nuclear se acrecentó cuando científicos de la talla de Edward Teller impulsaron la construcción de la bomba de hidrógeno, lo que inició la carrera armamentista con la URSS en las décadas de 1950, 1960 y 1980. A esta línea belicista se opusieron científicos como Albert Einstein, Bertrand Russell, Józef Rotblat y Norbert Wiener. Este último, poco después de haber publicado el libro de *Cibernética*, escribió otro dirigido a un público más amplio y sin grandes conocimientos de matemáticas para introducir las propuestas esenciales de su teoría. Este libro *Cibernética y sociedad (The Human Use of Human Beings: Cybernetics and Society)* se convirtió en un referente para los

que pregonaban un uso pacífico y positivo de la tecnología en beneficio de la sociedad.

### **El surgimiento de la Inteligencia Artificial**

Al inicio de la década de 1950, la cibernética fue tomando cada vez más fuerza como marco general teórico que abarcaba muchos de los grandes avances como la teoría del control, la teoría de la información, la computación y la automatización industrial. Sin embargo, en Estados Unidos, las posiciones pacifistas de Wiener y su crítica a los planes belicistas de la carrera armamentista le hicieron perder apoyos financieros del gobierno y muchos científicos e ingenieros buscaron desprenderse de la influencia de la cibernética. Por otro lado, a nivel mundial, la popularidad de la disciplina no dejaba de crecer, al grado que la URSS decidió adoptarla como la ciencia oficial del régimen, lo que la volvió aún más sospechosa para el gobierno norteamericano.

Es en este contexto que en 1956 John McCarthy, Marvin Minsky, Claude Shannon y Nathaniel Rochester propusieron organizar un seminario en la universidad de Dartmouth para discutir sobre los proyectos que empezaban a desarrollarse en diversos laboratorios de diseño de programas que simulaban procesos de deducción en el ser humano. El evento académico estaría dirigido a 10 investigadores durante dos meses. Lo que trabajarían serían experimentos de un tipo de programación que manipulaba información simbólica no numérica. El plan que tenían era pedir financiamiento para llevar a cabo ese seminario a la Fundación Rockefeller. Con el fin de vencer las reticencias que habían encontrado por parte de responsables de esta entidad, decidieron utilizar un nombre muy llamativo que McCarthy había empezado a utilizar en sus cursos en Dartmouth: el seminario sería sobre “Inteligencia Artificial”. La propuesta fue aceptada y el término *Inteligencia Artificial* fue adoptado. La campaña de mercadotecnia fue todo un éxito. La propuesta definía el propósito del seminario de la siguiente forma: “se postula que cada aspecto

del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, describirse con tanta precisión que se puede programar una computadora para simularlo. [...] Creemos que se puede hacer un avance significativo en uno o más de estos problemas si un grupo cuidadosamente seleccionado de científicos trabaja en ello juntos durante un verano” (McCarthy, Minsky, Rochester y Shannon 1955 [traducción propia]).

El avance que obtuvieron fue de hecho la fundación de una nueva área de investigación que se derivaba directamente de la computación.

Un último comentario sobre el grupo promotor de este seminario de Dartmouth: estaba claro que no deseaban relacionarse para nada con el campo de la cibernética ni con Norbert Wiener, aunque este último veía esos experimentos como parte de la corriente natural de la cibernética. Otra diferencia crucial con la visión de Wiener fue la ausencia total de cualquier referencia a las perspectivas éticas que esta nueva disciplina centrada en simular las funciones cognitivas podría tener sobre los seres humanos.

### **Auge y desarrollo de la Inteligencia Artificial**

En la década de 1960 el campo de la IA tuvo un crecimiento importante, y los principales grupos de investigación (como el Instituto de Tecnología de Massachusetts [MIT], el Stanford Research Institute [SRI], la Universidad de Stanford, la Carnegie Tech y la Rand Corporation) se beneficiaron del apoyo económico proveniente de la Agencia de Proyectos de Investigación Avanzados de Defensa (DARPA, por sus siglas en inglés) del ejército norteamericano. Los proyectos de esos años se centraban en temas como la demostración de teoremas, los juegos como el ajedrez y las damas, el surgimiento del Aprendizaje Automático, el Procesamiento del Lenguaje Natural, y la resolución general de problemas. Al tipo de programación utilizado se le denominó programación heurística, es decir, una programación que no busca seguir un algoritmo que lleva a una solución exacta, sino

que proporciona soluciones más o menos aproximadas la mayoría de las veces. La pertinencia de una heurística en la solución de un problema concreto reside en proponer, la mayoría de las veces, soluciones adecuadas a la situación en la que se aplica. Hacia mediados de la década se emprendieron los primeros proyectos de investigación sobre robótica en Estados Unidos y Escocia. El proyecto de mayor envergadura fue la construcción del robot Shakey del SRI financiado por DARPA que mostró la factibilidad de crear un artefacto móvil del tamaño de una pequeña nevera montado sobre unas ruedas accionadas por motores de paso, con una cámara de televisión, un telémetro, y comunicada por ondas de radio con una computadora central. Shakey era capaz de deambular por el laboratorio y los corredores sin tropezar, siguiendo un plan que iba construyendo en función del objetivo a alcanzar y los obstáculos que detectaba en el camino. Éste fue el primer ejemplo de la integración de las capacidades de visión, de representación interna del entorno físico del robot, de solución teórica del camino a recorrer y de ejecución del plan obtenido a través del comando de los motores asociados a las ruedas.

Eran años de crecimiento y euforia de los pioneros de la IA. Cada uno a su estilo se lanzaba a vaticinar grandes resultados en pocos años. Fue el caso de Herbert Simon, quien en 1956 afirmó que en 10 años un programa de ajedrez le ganaría al campeón mundial. Pasaron 40 años para que eso sucediera, hubo que esperar que el poder de cómputo disponible para ejecutar esos programas creciera exponencialmente como consecuencia de la “Ley de Moore” (perspicaz observación de Gordon Moore, de Intel, en la década de 1960, sobre que el número de transistores de los circuitos integrados para una misma superficie se estaban duplicando más o menos cada dos años).

Una característica que encontramos a lo largo de la historia de la IA es que desde sus inicios, con la propuesta a la Fundación Rockefeller, el campo ha estado marcado por una grandilocuencia, más cercana a la mercadotecnia que de un planteamiento

innovador de la ciencia. Existe una línea muy fina que separa la perspectiva basada en una conjetura científica o tecnológica de las promesas basadas en un optimismo exagerado que busca impresionar para conseguir fondos para continuar las investigaciones de los grandes laboratorios. Como veremos más adelante, esa euforia mercadológica la estamos viviendo nuevamente.

Esta euforia de los principales centros de investigación de principios de la década de 1970 se vio frenada drásticamente cuando los gobiernos del Reino Unido, primero, y luego de Estados Unidos decidieron reducir de manera drástica los fondos asignados a los proyectos de robots por falta de resultados prácticos. A este periodo se le conoce como el primer invierno de la IA. Eso no quiere decir que se dejó de trabajar en investigación en muchos otros temas, simplemente que se redujeron los presupuestos asignados a todo lo que fuera investigación en IA.

Un resultado interesante de esta reducción de la exagerada presencia de la publicidad de la IA capitaneada por los grandes laboratorios de Estados Unidos y Reino Unido fue que otra corriente que había sido “despreciada” por la hegemónica empezó a tener mayor presencia y a obtener resultados de interés. La corriente conexionista, heredera del modelo matemático del funcionamiento de una neurona de McCulloch Pitts y posteriormente del Perceptrón de Frank Rosenblatt en la década de 1950, desarrolló los primeros modelos de lo que ahora conocemos como redes neuronales artificiales (pero eran modelos aún muy sencillos si se les compara con lo que se utilizan en la actualidad).

La década de 1980 vio surgir además de estos trabajos sobre redes neuronales, una nueva ola de entusiasmo en torno a la IA simbólica. Esta vez en torno al diseño y construcción de arquitecturas de computadoras que deberían procesar más eficientemente los programas de IA en los dos lenguajes especiales más populares en ese entonces: Lisp y Prolog.

El esfuerzo oficial que impulsó esa nueva oleada de interés y de financiamiento a los proyectos de IA fue el programa nacional

del Japón para la construcción de las “computadoras de la quinta generación”. En éste, el gobierno japonés conjuntaba a las principales industrias de cómputo de ese país y a la academia, no sólo de Japón, sino de muchas partes del mundo y en especial de Europa. La idea era usar una nueva arquitectura de computadora mediante circuitos de muy alto nivel de integración o VLSI (del inglés *Very Large Scale Integration*) que utilizaría al lenguaje de programación Lógica Prolog como lenguaje de máquina. El objetivo era hacer máquinas que fuesen muy eficientes corriendo programas de IA como los sistemas expertos en boga en esos años, a la par que realizaba computación en paralelo.

El anuncio de este programa de la quinta generación provocó un cisma a nivel mundial, tanto en Estados Unidos como en Europa, en donde se lanzaron poco después enormes proyectos en torno al desarrollo de la investigación aplicada en IA y en computación en general. Además, se construyeron nuevas arquitecturas de máquinas que habían estado desarrollándose a nivel de prototipos académicos. Fue el caso de las máquinas Lisp, que dieron pie a una serie de nuevas empresas de cómputo en Estados Unidos.

Durante un poco más de 10 años la IA fue nuevamente la promesa tecnológica del futuro, y los gobiernos y empresas invirtieron grandes sumas de dinero en esos proyectos. Para mediados de la década de 1990, los proyectos de máquinas Prolog y máquinas Lisp fueron abandonados por no cumplir con sus promesas y porque la aplicación continua de la Ley de Moore hizo que los procesadores de arquitectura tradicional (llamada arquitectura von Neumann) sobrepasaran en velocidad y reducción de costos las complejas arquitecturas Lisp y Prolog. Así inició un segundo invierno de la IA, en el que otra vez el exceso de euforia y las promesas no pudieron cumplirse, y los detentores de los fondos de investigación e innovación decidieron invertir en otros proyectos tecnológicos.

Si bien fue muy sonado el fracaso industrial y político de estos proyectos de nuevas arquitecturas, queda por escribirse la

historia de los beneficios que provocó. Lo anterior sobre todo en los miles de ingenieros jóvenes de Japón y Estados Unidos, así como en el papel que jugaron en los desarrollos que vinieron después.

En efecto, la década de 1990 estuvo marcada por un fenómeno mayúsculo, la aparición y la diseminación del uso de Internet gracias a la World Wide Web. De nuevo, la observación de Moore se cumplió y el poder de cómputo disponible fue ya suficientemente grande para resolver muchas tareas que eran imposibles de realizar una década antes. Lo más importante es que la inteligencia y el tenaz trabajo de los ingenieros en microelectrónica han sido constantes y han permitido que la famosa “ley” siga cumpliéndose, con algunas modificaciones, hasta nuestros días.

La vieja promesa de que un programa de ajedrez le ganaría al campeón mundial se hizo realidad cuando la computadora Deep Blue le ganó a Gary Kasparov en febrero de 1996. Este evento, que conmocionó al mundo, puede verse como el punto culminante de la IA simbólica. Lo que hizo ese programa no fue nada radicalmente novedoso, simplemente utilizó al máximo y de manera muy hábil el poder de cómputo y de almacenamiento de la super computadora de IBM. Fue una proeza del uso de la “fuerza bruta”. No fue un avance de la programación, fue un avance de la electrónica.

### **El cómputo en la nube**

El desarrollo de la microelectrónica, que sigue obstinadamente la Ley de Moore, así como el crecimiento de Internet basado en otro crecimiento exponencial (el de la capacidad de transmisión de información de las fibras ópticas), con rapidez dibujó una infraestructura mundial en donde era posible procesar cada vez más información y transmitirla a cualquier parte del mundo, o casi. Estas condiciones objetivas de capacidades físicas impulsaron un movimiento de concentración, no sólo de poder de cómputo, sino de poder económico y político.

La historia de la computación, desde el punto de vista de arquitectura de sistemas, ha sido una que se ha movido de arquitecturas centralizadas a arquitecturas descentralizadas. Un ir y venir en una especie de movimiento espiral. A finales de la década de 1990 y principios de este siglo se fue dando un nuevo momento de centralización dentro de la descentralización: nacieron los primeros centros que concentraban ciento y miles de procesadores con capacidades crecientes de memoria. Nacían los “Data Centers” y la posibilidad de que los dueños de esas infraestructuras ofrecieran nuevos servicios. Era el cimiento del “cómputo en la nube”.

De nuevo la microelectrónica, y nuestra vieja amiga, la Ley de Moore, venían al apoyo del desarrollo de nuevas tecnologías y negocios. El cómputo en la nube no ha dejado de crecer y de sofisticarse, al mismo tiempo que ha asegurado el crecimiento de un pequeño número de empresas de cómputo gigantes, verdaderos oligopolios que dominan las interacciones de miles de millones de personas a través de la World Wide Web.

Por primera vez, el común de los mortales pudo acceder a información multimedia dispuesta en cualquier sitio del mundo y pudo subir información a sitios especializados o simplemente intercambiar textos, fotos o videos con otras personas. Fue así como la cantidad de información que circula por Internet explotó. Un evento clave para desencadenar el proceso de concentración de la información digital en Internet fue el ataque a las Torres Gemelas de Nueva York el 11 de septiembre de 2001. A partir de ese momento los servicios de inteligencia de los países desarrollados, pero muy especialmente el de Estados Unidos, empezaron a poner en marcha una búsqueda de terroristas a partir de la información en Internet. Años más tarde, en 2013 con la difusión de documentos hecha por Edward Snowden, se supo la envergadura mundial de la infraestructura de vigilancia puesta en marcha por los servicios de inteligencia norteamericana. Se conoció, por ejemplo, la existencia del sistema “TreasureMap”

que tenía por objetivo realizar un mapa de todo el internet, de todas las computadoras conectadas, en todas partes y todo el tiempo. Ese tipo de sistema de vigilancia debió estar soportado por enormes centros de datos secretos.

Si bien las tecnologías utilizadas por los servicios de inteligencia no eran conocidas forzosamente por la industria de las tecnologías de la información, la acumulación creciente de datos circulando en la red a los cuales algunas empresas empezaban a tener acceso provocó los mismos efectos que en el mundo secreto de la inteligencia: la tentación de guardar toda la información a la que pudiesen acceder y almacenar en centros de datos cada vez más grandes. El cómputo en la nube había nacido.

Claro que una cosa era crear esa enorme infraestructura de cómputo en los centros de datos y otra era definir los nuevos modelos de negocio que podrían explotar la infraestructura y los datos almacenados. En sí, el desarrollo del cómputo en la nube es una historia fascinante, por desgracia no es el tema de este capítulo. Así, nos limitaremos a establecer que en la actualidad es el modelo de negocio de cómputo preponderante. Lo que nos interesa es cómo esa enorme concentración de información aparentemente anodina, proveniente de los intercambios de los usuarios de redes sociales y otras aplicaciones, motivó el desarrollo de tecnologías para su explotación y para desarrollar nuevos modelos de negocio que han permitido el crecimiento de las gigantescas empresas tecnológicas que hoy conocemos.

### **El nuevo auge de la Inteligencia Artificial**

En el mundo académico de la IA, y en particular en las aplicaciones de Procesamiento Automático de Lenguaje Natural, muy pronto se dio la necesidad de contar con colecciones de palabras organizadas semánticamente. Para esto era necesario agrupar las palabras sinónimas o con significados cercanos, o bien relacionar conceptos generales con algunos más específicos (hiperónimo/hipónimo), o de un concepto general y las partes que

lo componen (holónimo/merónimo), etcétera. Estas colecciones eran necesarias para poder hacer programas que pudieran manejar textos en lenguaje natural de manera útil; ya sea, por ejemplo, para reconocer órdenes o generar respuestas de manera automática. Sin embargo, su generación es sumamente laboriosa: requiere que seres humanos vayan etiquetando cada palabra con los conjuntos que les corresponden. Una dificultad adicional es la de manejar conceptos abstractos como *libertad*, *deseo*, *tristeza*, etcétera, que implican relaciones semánticas múltiples y frecuentemente dependientes del contexto, del idioma y la cultura al que se refiere el texto en el que aparecen. A este ejercicio de asociación de conceptos semánticos a las palabras se les conoce como “etiquetado”. Es una etapa obligada para desarrollar cualquier colección de datos que pretendamos utilizar por un programa de computación más o menos sofisticado.

La primera gran colección de palabras etiquetadas fue desarrollada originalmente en 1985 por G. A. Miller y C. Fellbaum, y se denominó WordNet. Ésta dio origen a una asociación mundial de colecciones de palabras etiquetadas en muchos idiomas y sigue siendo una referencia clave.

En 2006 Fei-Fei Li, investigadora de la universidad de Stanford, inició el desarrollo de un proyecto de colección de imágenes etiquetadas, para ello se inspiró en las redes de conceptos desarrolladas para WordNet. Así nació ImageNet, una colección de imágenes etiquetadas que no ha dejado de crecer desde entonces.

WordNet e ImageNet son los dos grandes insumos iniciales que permitieron el auge de las técnicas de IA que en la actualidad son predominantes en el mundo: las “redes neuronales profundas” y los algoritmos de “aprendizaje profundo”. Estos métodos se fueron desarrollando conceptualmente en la década de 1980 y las primeras implementaciones que mostraban resultados interesantes surgieron una década después. Sin embargo, se trataba de algoritmos que requerían un poder de cómputo muy alto que no estaba disponible aún. Hubo que esperar cerca de 20 años

para que los efectos de la Ley de Moore permitieran acceder a un poder de cómputo tan poderoso como para hacer realidad el uso práctico de estos algoritmos.

Entre 2011 y 2015 aparecieron diferentes programas de reconocimiento de imágenes que obtenían resultados cada vez más sorprendentes en las competencias de reconocimiento de imágenes que se realizan anualmente en los congresos más importantes del área. A partir de 2015, las grandes empresas tecnológicas empezaron a crear equipos de desarrollo de productos basados en esta tecnología que pronto iban a inundar el mercado. Fue el inicio del nuevo “boom” de la IA. A diferencia de los anteriores, los millones de dólares involucrados para la investigación y el desarrollo de aplicaciones excedió en varios órdenes de magnitud lo que la comunidad de IA había conocido en las décadas anteriores.

En paralelo se desencadenó una campaña de mercadotecnia y relaciones públicas sin precedentes, haciendo que el término de “IA” sustituyera los términos de “redes” y “aprendizaje profundos”, lo que borró para el gran público todas las otras áreas de desarrollo de investigación y desarrollo de IA. Apareció otra vez esa euforia que se vivió en la década de 1960, pero en esta ocasión con miles de millones de dólares de las grandes empresas por delante. Al igual que en otros momentos de la historia de la tecnología, aparecieron las versiones de los iluminados que presagiaban todo tipo de calamidades y el eventual dominio de los robots sobre los seres humanos. A estos augurios se les puede aplicar la misma metáfora que se extendió en la década de 1960 con la primera ola de euforia de la IA, es decir que todas las promesas y expectativas que se estaban generando en torno a las aplicaciones de la IA y la robótica era similar a señalar que en la carrera por llegar a la Luna habíamos logrado un primer gran paso: ¡subirnos a la copa de un árbol!

Por muy sorprendentes que sean los resultados alcanzados hoy, los algoritmos de IA están muy lejos de lo que hace nuestro

cerebro. Sin duda, en las décadas por venir habrá desarrollos importantes, pero esperemos que estén enfocados en lograr el bien común y no la multiplicación de las riquezas de unos pocos, que es lo que realmente está promoviendo todo este ajetreo alrededor de la IA.

### **Las alertas éticas en torno de la IA**

Los primeros resultados sorprendentes de las capacidades de reconocimiento de patrones (como imágenes de gatos, perros o textos de tweets) usando redes neuronales profundas desencadenó, principalmente en el medio tecnológico de Estados Unidos, una eclosión de empresas de base tecnológica, las famosas *startups*. Éstas introducían una gran variedad de aplicaciones de redes neuronales a situaciones concretas de toma de decisión en los ámbitos empresarial o gubernamental.

Los efectos negativos no tardaron en aparecer. En 2016 Cathy O'Neil publicó su libro *Armas de destrucción matemática*, el cual lanzó la alerta sobre los problemas del uso sin control de algoritmos de aprendizaje y de redes neuronales entrenados sobre conjuntos de datos que no tenían ningún control de calidad. Además, documentó varios casos que se volvieron paradigmáticos, como los sistemas de evaluación de la calidad del trabajo de los profesores de primaria y secundaria que tenían errores conceptuales y afectaron a muchos profesores en diversos estados de la Unión Americana. Otro tipo de aplicación fue la de los sistemas de selección de candidatos a ser contratados por cualquier empresa, o bien el caso del sistema que aconsejaba a los jueces si un detenido podía seguir su juicio estando en casa o representaba un peligro de reincidencia y había que encarcelarlo mientras se desarrollaba el proceso.

O'Neil, explicó al mundo de manera muy clara los problemas con estas aplicaciones que entrenan sus redes neuronales sobre datos plagados de sesgos y que dañaban a ciertas comunidades o perfiles de personas. Asimismo, mostró que los algoritmos

utilizados tanto en la arquitectura de las redes neuronales como en los procesos de entrenamiento no se podían conocer cubriendo con un manto de opacidad todos los procesos de decisión y recomendaciones de esos sistemas. Eran todas aplicaciones que mostraban faltas de ética, en su diseño, en su programación y en su utilización.

Desde esa fecha han aparecido múltiples denuncias de sistemas con esas mismas características, opacidad de los algoritmos y en los métodos de construcción de los conjuntos de datos de entrenamiento de las redes neuronales profundas que utilizan. La consecuencia directa ha sido una creciente pérdida de confianza en estos sistemas. Bajo el criterio del secreto industrial, las empresas se escudan en una política de opacidad en sistemas que afectan la vida de personas que ven vulnerados sus derechos humanos más básicos. O'Neil muestra cómo esa práctica comercial depredadora se escuda en argumentos de que los sistemas están basados en investigaciones científicas, que son de alto nivel, pero que son muy complejas para que el común de los mortales las entienda, de ahí el título de su libro: *Armas de destrucción matemática*.

Una de las principales conclusiones que se desprenden de lo expuesto por Cathy O'Neil es que bajo el escudo de que se están utilizando productos que vienen directamente de la ciencia, y por lo tanto son infalibles, se está desarrollando una práctica empresarial, corrupta y falta de ética. Otro evento perturbador en esa euforia del uso de la IA en aplicaciones prácticas ha sido sin duda el papel que están jugando las redes sociales. En la última década, el uso de éstas tuvo un incremento enorme. En ese sentido, Facebook y YouTube fueron las más grandes al rebasar los dos mil millones de usuarios cada una. No debemos olvidar que se trata de empresas, y que muchas de ellas cotizan sus acciones en la Bolsa de Valores y, por lo tanto, tienen que rendirle cuentas a sus accionistas representados en primer término por los grandes fondos de inversión que dominan el mundo financiero. En otras

palabras, su objetivo primero es ser rentables, y entre más lo sean, mejor. Una simple consulta de los resultados financieros de estas empresas arroja resultados impresionantes con crecimientos muy por arriba de otras ramas industriales.

Este éxito financiero se debe a que pusieron a punto un modelo de negocio, cada vez más sofisticado, de ventas por publicidad. Ese modelo reposa en el hecho del acceso que tienen estas empresas a información privada de sus usuarios: todo lo que escriben, las fotos y videos que comparten, los sitios que consultan en la Web, el tiempo que pasan en cada página, sus desplazamientos geográficos, etcétera. Con toda esta información actualizada en tiempo real, estas empresas elaboran perfiles muy detallados de los usuarios que les permiten vender el tiempo frente a la pantalla de estos usuarios a las empresas de publicidad. De esta forma los mensajes publicitarios pueden estar mejor focalizados en los valores de los parámetros obtenidos del perfil de cada uno. Todo esto, evidentemente, sin que los usuarios se percaten de que están siendo vigilados y manipulados por estas empresas. Pero para que este modelo funcione es necesario que los usuarios pasen el mayor tiempo posible frente a la pantalla de la plataforma. Si esto sucede la empresa podrá aumentar el número de franjas de tiempo que podrán subastar ante los anunciantes.

Aquí es donde entra en juego una batería de técnicas desarrolladas por las empresas tecnológicas para lograr aumentar lo más posible el tiempo de uso de sus plataformas: las técnicas de diseño adictivo que se han denunciado de manera creciente por los efectos nocivos en la salud emocional de los usuarios y las usuarias más vulnerables. Entre estas técnicas adictivas, podemos listar el botón de “me gusta”, el scroll infinito, o mostrar preferentemente las noticias o los mensajes más negativos y controversiales. Por ejemplo, Frances Haugen, quien publicó miles de documentos internos de Facebook a finales de 2021, afirmó en su comparecencia ante una comisión de congreso norteamericano que la red social conocía perfectamente los efectos nocivos de

estas técnicas en Instagram, las cuales causaban serios problemas emocionales (sobre todo en jóvenes adolescentes), y que, a pesar de saberlo, prefirió continuar con esas prácticas adictivas en beneficio de sus ganancias.

### **La quimera del tecnosolucionismo**

Ante las múltiples críticas a los efectos perversos del uso masivo de aplicaciones de inteligencia artificial como los ya mencionados, una de las respuestas más socorridas es que los problemas detectados y denunciados no son responsabilidad de la tecnología ni de la ciencia en las que estas aplicaciones se basan. El problema, según esta visión, reside en la aplicación errónea de esa tecnología, porque la ciencia y la tecnología son neutras, no son ni buenas ni malas. Es decir, lo que no ven los críticos son todas las aplicaciones positivas de estas tecnologías para el bien de la humanidad, por lo que si se quiere avanzar en la solución de los problemas que aquejan a la humanidad sólo lo podemos hacer con la ayuda de la ciencia y la tecnología. Los que suelen defender esta posición, o alguna variante de ella, pasan de inmediato a dar ejemplos de las promesas de la ciencia y la tecnología para acabar con el hambre en el mundo, o con enfermedades y epidemias que azotan grandes regiones de África, Asia y Latinoamérica. Esta posición es la que se conoce como *Tecnosolucionismo*. Así, en la actualidad la Inteligencia Artificial es presentada como la gran promesa que va a resolver los problemas de la educación, de la salud y del cambio climático, entre muchos otros que enfrentamos. Para apoyar esta visión se enumeran un sinnúmero de proyectos en curso que trabajan en esos temas.

Sin embargo, como todas las actividades humanas, la ciencia y la tecnología están sujetas a ser evaluadas socialmente, son *sistemas de acciones intencionales* de los diferentes agentes participantes en su desarrollo. Estos agentes, como lo expresa, León Olivé, buscan ciertos fines de forma deliberada, en función de determinados intereses, para lo cual ponen en juego creencias,

conocimientos, valores y normas susceptibles de una evaluación moral de la sociedad. Por eso la ciencia y la tecnología no pueden considerarse neutras desde el punto de vista valorativo.

Tomemos el ejemplo de la educación. Desde la década de 1960 se han encadenado los proyectos de aplicación de la computación y de la inteligencia artificial a la educación. Algunos de ellos han sido interesantes, como la creación de lenguajes de programación sencillos enfocados a los niños (como Logo o Scratch); el desarrollo de videos accesibles desde internet con material de cursos como los de la Academia Khan, o los complejos sistemas de tutores inteligentes comerciales que se utilizan en algunas costosas escuelas privadas en varias partes del mundo. Estas aplicaciones se desarrollan en contextos sociales, económicos precisos, e incluso muchos de ellos se han desarrollado en países del primer mundo. Todas estas soluciones tienen como premisa que los niños y las niñas deben tener acceso a equipo de cómputo en la escuela, y que debe estar conectado a internet. Éste es un primer escollo para los sistemas educativos nacionales de nuestras naciones del Sur Global. El espejismo de resolver con tecnología los problemas de la educación en nuestros países es muy fuerte y empuja a los gobiernos a buscar una cobertura universal de internet para todas las escuelas públicas, por lo que caen en una demencial carrera contra el constante surgimiento de nuevas tecnologías de telecomunicaciones que van haciendo obsoletas las anteriores. Cuando ya se está alcanzando una mayor cobertura con una tecnología, digamos la 4G, el propio gobierno decide pasar a fomentar la instalación de la red 5G que implica un cambio profundo en la infraestructura necesaria.

Uno puede preguntarse si las ofertas de tecnología educativa forman parte de una estrategia comercial de los monopolios de la industria de telecomunicaciones en esa carrera sin fin de consumismo extremo del último grito de la moda tecnológica, o bien de una estrategia de crear usuarios de las aplicaciones de cómputo que mantengan funcionando la bonanza de las

empresas tecnológicas de los países desarrollados. Muchas aplicaciones existentes o que podríamos concebir son herramientas interesantes en los procesos de enseñanza-aprendizaje y que pueden ser de utilidad para nuestra niñez y nuestra juventud, pero después de más de cincuenta años de promesas de solución de los problemas de la educación en el mundo lo que queda claro es que los sistemas educativos, en la mayoría de los países, van en retroceso. Tomemos el ejemplo del proyecto de *One Laptop per Child*, del MIT, que impulsó el diseño y construcción de computadoras de muy bajo costo para ser utilizadas en entornos escolares de todo el mundo, y en particular del Sur Global. Se promovió como la solución a los problemas de la educación y de la brecha digital existente en las comunidades pobres de todo el mundo. El líder del proyecto, Nicholas Negroponte, afirmaba en 2006 que el objetivo último de este proyecto era terminar con la pobreza en el mundo a través de la educación. A pesar de algunos éxitos locales y temporales, y de haber repartido cerca de 3 millones de computadoras en el mundo, en 2014 la asociación *One Laptop per Child* (OLPC), responsable de la construcción y disseminación de estas computadoras, cerró sus operaciones.

Hoy, una enorme cantidad de niños, niñas y jóvenes tienen acceso a teléfonos móviles conectados a internet lo cual, en lugar de ser una ventaja para el avance en el nivel educativo, se ha transformado en un elemento de enajenación y adicción de muchos de ellos. En otras palabras, la argumentación tecnosolucionista en torno al uso de la computación en las escuelas se transformó en un argumento mercadológico que benefició al modelo consumista impulsado por las grandes economías de empresas y gobiernos para entrenar usuarios de su tecnología.

La quimera del tecnosolucionismo ha provocado que los gobiernos tiendan a desentenderse de una de sus funciones sustantivas: asegurar la formación sólida de sus habitantes que les permita desarrollarse a plenitud con calidad de vida. Para ello hay que invertir seriamente en educación (empezando por salarios dignos

a los profesores), condiciones de trabajo que les permita mantenerse actualizados e invertir en buena infraestructura (incluyendo computadoras con programas didácticos diseñados en el país y adaptados a las circunstancias regionales). En esas circunstancias, las aplicaciones de inteligencia artificial en la educación podrán jugar un papel positivo con desarrollos más modestos, pero mejor adaptados a las circunstancias, lo que permitirá el despliegue de pequeñas y medianas empresas regionales constructoras de aplicaciones, sin necesidad de depender de las infraestructuras de cómputo en la nube de las empresas transnacionales. Esto sería un ejemplo de desarrollo tecnológico responsable.

## Conclusiones

Lo expuesto no representa más que una muy pequeña muestra del tipo de aplicaciones que ya existen, y que han revelado problemáticas desde el punto de vista ético. La efervescencia actual en torno a estas tecnologías de la IA de redes neuronales profundas y sus métodos de entrenamiento están generando miles de nuevas aplicaciones cada año, sin ningún control sobre los posibles malos usos y efectos sobre la población que ignora todo sobre la vigilancia a la que está sometida (no sólo por sus gobiernos, sino también por las multinacionales que violan constantemente su intimidad y la manipula para mantener un modo de consumismo extremo basado en prácticas adictivas).

A partir de estas muestras, podemos sacar algunas conclusiones importantes. La primera es que la mistificación de la IA y de sus algoritmos llenos de ciencia y matemática no es más que una inmensa cortina de humo para no mostrar que las empresas han montado modelos de negocio depredadores a sabiendas de los efectos nocivos que muchas de sus aplicaciones causan en ciertos sectores. Un puñado de empresas muy poderosas forman parte del oscuro sistema financiero que gobierna la economía mundial. Esta economía digital ha crecido en un vacío normativo en todo el mundo. Así, una de las primeras tareas es pugnar por

leyes nacionales e internacionales que permitan acabar con los estragos de esta economía del consumismo y de la manipulación de voluntades.

Como necesidad primera en esta normatividad deben prohibirse los modelos de negocio que basan su crecimiento en la captura de información privada de las personas sin su permiso o sin su conocimiento real, así como prohibirse el comercio de dicha información y de sus actividades. La privacidad no es una mercancía. Así, un primer paso podría ser la prohibición mundial del uso de las *cookies*.

Asimismo, las empresas oligopólicas del mundo digital deben romperse en tantas empresas independientes como sea necesario y deben integrarse a modelos de negocio de economía circular enfocadas en apoyar la consecución de los objetivos de desarrollo sostenible de la ONU.

El sentido común más elemental nos dice que no es posible crecer indefinidamente en un mundo finito. El cómputo en la nube y la proliferación de aplicaciones de Inteligencia Artificial cada vez más poderosas están acelerando el incremento desenfrenado de la economía digital del mundo desarrollado. Por ello, debe terminar el crecimiento indefinido de centros de datos. Además, se deben establecer normas internacionales para un uso sostenible y mejor distribuido de los existentes para el beneficio común de la humanidad. Es momento de emprender un desarrollo sostenible y descentralizado de las tecnologías asociadas al cómputo en la nube que permitan un control social colectivo de los servicios de comunicación de comunidades autónomas organizadas.

Por último, es importante señalar que la investigación en IA debe enfocarse en apoyar el bienestar del ser humano, de sus comunidades y del planeta. Hay mucho por hacer y el uso excesivo de la “fuerza bruta” de los grandes centros de cómputo no ha permitido el desarrollo de ideas más originales enfocadas en el bien común.

**Nota:** todos los conceptos utilizados en este trabajo pueden ser consultados directamente en Wikipedia para tener una explicación funcional basada en fuentes de confianza. Los avances en este campo crecen a velocidades enormes. Si se desea tener información más detallada sobre muchos de los temas mencionados aquí y otros que no lo fueron pero que son relevantes, sugiero la lectura de los libros que se listan a continuación:

- AMES, Morgan (2019). *The Charisma Machine: The Life, Death, and Legacy of One Laptop per Child*. Massachusetts: The MIT Press.
- CRAWFORD, Kate (2021). *Atlas of AI: power, politics, and the planetary costs of Artificial Intelligence*. New Haven: Yale University Press.
- MCCARTHY, John, Marvin Minsky, Nathaniel Rochester Claude Shannon (1955). *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Acceso el 25 de junio de 2022. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>.
- O'NEIL, Cathy (2018). *Armas de destrucción Matemática: como el Big Data aumenta la desigualdad y amenaza la democracia*. Madrid: Capitan Swing.
- OLIVÉ, León (2003). "Ética aplicada a las ciencias naturales y la tecnología". En *Cuestiones éticas de la ciencia y la tecnología en el siglo XXI*, editado por Andoni Ibarra y León Olivé Unzueta, 181-223. Madrid: Biblioteca Nueva.
- VÉLIZ, Carissa (2021). *Privacidad es poder. Datos, vigilancia y libertad en la era digital*. Barcelona: Debate.
- ZUBOFF, Shoshana (2020). *La era del capitalismo de la vigilancia: La lucha por un futuro humano frente a las nuevas fronteras del poder*. Barcelona: Paidós.

## Sobre el autor

### Dr. Christian Lemaître León

Investigador adscrito al Departamento de Tecnologías de la Información de la División de Ciencias de la Comunicación y Diseño,

Universidad Autónoma Cuajimalpa. Doctor en Informática por la Universidad de París 6. Físico por la Universidad Nacional Autónoma de México (UNAM). Ha realizado labores de docencia, investigación, desarrollo de innovaciones tecnológicas y consultoría en distintas instituciones como el Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS) de la UNAM, la Facultad de Ciencias de la UNAM, el Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM) campus Morelos, la Fundación Arturo Rosenblueth, la Universidad Veracruzana, el Laboratorio Nacional de Informática Avanzada (LANIA) y la Unidad Cuajimalpa de la Universidad Autónoma Metropolitana. Sus áreas de interés incluyen Inteligencia Artificial distribuida, sistemas multiagente, instituciones y organizaciones multiagente, planificación y aprendizaje, creatividad, urbanismo de sistemas de información y sistemas interactivos. De 2005 a 2009 fue jefe del Departamento de Tecnologías de la Información de la Unidad Cuajimalpa de la UAM, y de 2009 a 2013 fue director de la División de Ciencias de la Comunicación y Diseño de la UAM Cuajimalpa.

# Tecnologías Inmersivas, computación espacial y cognitiva, Gemelos Digitales y la capacitación del futuro

ANASOL PEÑA-RIOS

## Resumen

Las Tecnologías Inmersivas y la computación espacial tienen el potencial de cambiar la manera como nos comunicamos e interactuamos entre nosotros y con nuestro ambiente, y transformar a la sociedad más allá de la automatización de tareas y la optimización de procesos. Al combinarlas con mecanismos de computación cognitiva se pueden producir entornos adaptativos (conscientes del contexto) que brinden un enfoque centrado en el ser humano, además de que aumenten los sentidos para el pensamiento multidimensional y la resolución de problemas. Su uso también puede conducir a una transferencia de conocimiento mucho más rápida y una comprensión más profunda de diferentes procesos, lo que mejorará el aprendizaje basado en experiencia. Este documento introduce estas tecnologías, así como las tendencias actuales y futuras, y su posible impacto en la sociedad.

**Palabras clave:** realidad virtual, realidad aumentada, realidad mixta, computación espacial, Gemelos Digitales, capacitación,

transferencia de conocimiento, aprendizaje por experiencia, entrenamiento inmersivo, computación cognitiva.

## Introducción

La forma como nos comunicamos y colaboramos hoy difiere drásticamente de cualquier cosa que hubiéramos imaginado, y el futuro es quizás incluso menos predecible. Factores como la movilidad y la migración, el cambio demográfico y los eventos globales también modifican la naturaleza y la calidad de las interacciones humanas.

Las Tecnologías Inmersivas que incluyen la realidad mixta, la realidad aumentada y realidad virtual (RM, RA y RV, respectivamente), tienen como objetivo difuminar las líneas entre el mundo físico y el digital para crear nuevas experiencias de usuario. La realidad virtual se imaginó por primera vez en la ciencia ficción y surgió en la vida real a través de un sistema inmersivo para la visualización de películas creado en la década de 1950. El primer dispositivo de realidad aumentada fue creado en la siguiente década por investigadores de la Universidad de Harvard. El término *realidad virtual* fue acuñado en 1987 cuando la compañía VPL Research creó el primer hardware para uso comercial. El término *realidad aumentada* fue acuñado en 1990 por Boeing cuando a dos de sus investigadores se les pidió que idearan una alternativa a los costosos diagramas que se usaban para guiar a los trabajadores en la planta de producción. Después de esto hubo un decremento en la investigación y comercialización de hardware para aplicaciones inmersivas, sin embargo, tuvo un repunte en 2010 cuando Palmer Luckey, con 18 años, creó el prototipo de visor Oculus Rift VR, reavivando el interés por la realidad virtual. A partir de este momento, se presentó un crecimiento acelerado en las Tecnologías Inmersivas, financiado principalmente por los llamados gigantes tecnológicos. En 2014, Google presentó sus lentes Glass de realidad aumentada y en 2016 Microsoft presentó sus lentes HoloLens de realidad mixta. En el mismo año,

Facebook adquirió la compañía Oculus VR, apostando su estrategia comercial al llamado Metaverso, incluso cambiando su nombre oficialmente a Meta.

Las aplicaciones comerciales para la realidad virtual (RV) y aumentada (RA) están incrementando, alterando fundamentalmente la forma como las personas interactúan entre sí y con sus entornos. Las Tecnologías Inmersivas podrían convertirse en la plataforma de comunicación de próxima generación transformando la manera como trabajamos, accedemos a servicios e información, nos entretenemos y nos comunicamos con los demás —si bien aún existen barreras tecnológicas, necesidad de contenido más variado e inclusivo, más investigación acerca de los efectos fisiológicos a largo plazo y acceso limitado a estas tecnologías debido a la brecha digital—. Sin embargo, en la actualidad ya hay varias tendencias que muestran nuevas formas de juegos inmersivos y experiencias que colocan a las audiencias en medio de los eventos, así como arte inmersivo que es creado pensando en cómo lo experimentará el espectador.

Antes del lanzamiento comercial a gran escala relativamente reciente de los lentes de realidad virtual dirigidos a jugadores de videojuegos o *gamers*, esta tecnología se había utilizado durante mucho tiempo para aplicaciones médicas, incluida la capacitación en atención médica (Ruthenbeck y Reynolds 2015) y la rehabilitación física (Elor y Kurniawan 2020; Matamala-Gomez, Slater y Sanchez-Vives 2022). Lo anterior gracias a que los simuladores de cirugía inmersiva brindan un entorno libre de riesgos para practicar procedimientos que salvan vidas. Además, la realidad virtual también se ha utilizado para tratar pacientes con depresión, ansiedad y víctimas que sufren de trastorno de estrés postraumático (James *et al.* 2003). Los estudios han demostrado que los tratamientos de realidad virtual también pueden ayudar a tratar la paranoia y la adicción a las drogas, así como el control del dolor (Ioannou *et al.* 2020). Estos efectos paliativos pueden ser uno de los mayores beneficios de la realidad

virtual, ya que la reducción del dolor puede aumentar las tasas de supervivencia y mejorar la calidad de vida de las personas. La realidad virtual también puede ser una herramienta poderosa para cambiar el comportamiento y generar empatía. Investigadores han creado diferentes experiencias inmersivas que retratan en primera persona la experiencia de personas que sufren de prejuicios raciales implícitos (Banakou *et al.* 2020, Kishore *et al.* 2022).

A estos escenarios debemos sumar situaciones fuera de nuestro control como, por ejemplo, la necesidad de implementar mecanismos de teletrabajo debido a la crisis sanitaria de COVID-19, que forzó a muchas empresas a repensar cómo facilitar las interacciones entre sus empleados, las cuales fueron limitadas durante este tiempo, lo que creó muchos retos para la comunicación, la colaboración y la innovación. Así, la pandemia aceleró la adopción de Tecnologías Inmersivas, especialmente en el entretenimiento mediante eventos virtuales, y en la educación y la capacitación.

El objetivo de este texto es introducir al lector las tecnologías que habilitan las experiencias inmersivas, así como ilustrar cómo podrían ser usadas en el contexto de educación y capacitación. Esto con el objetivo de cerrar la brecha entre teoría y práctica mediante casos de uso reales. Finalmente, se presenta una reflexión acerca de sus beneficios y posibles riesgos tanto en este contexto como en la sociedad en general.

### **Tecnologías Inmersivas y computación espacial**

Hemos mencionado que las Tecnologías Inmersivas tienen el potencial de transformar la manera en que nos comunicamos y colaboramos, pero ¿qué son estas tecnologías?

- La realidad virtual (RV) es el término utilizado para describir un entorno tridimensional generado por computadora con el que una persona puede interactuar y explorar. Generalmente presenta contenido de alta definición y coloca al usuario en

él para estimular sus sentidos creando la ilusión de realidad. Los videojuegos son un buen ejemplo de realidad virtual.

- La realidad aumentada (RA) superpone información digital en el mundo real, lo que permite a los usuarios interactuar con el mundo real mientras miran una capa adicional de información, generalmente a través de un teléfono móvil. Un ejemplo de esto son los filtros de las redes sociales.
- La realidad mixta (RM) permite la interacción entre elementos reales y virtuales en tiempo real como si existieran en el mismo espacio y tiempo. Por ejemplo, contenido holográfico para entretenimiento y educación.

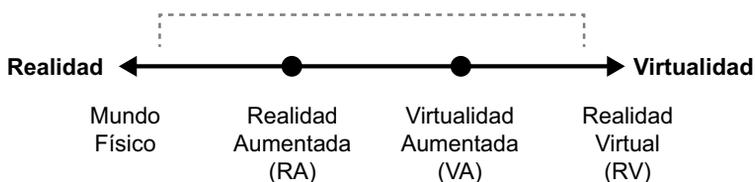


Figura 1. Continuo Realidad-Virtualidad. Elaboración propia a partir de Milgram y Kishino (1994).

Un término que engloba todas estas tecnologías es *realidad extendida* (o XR por sus siglas en inglés *eXTended Reality*), aunque estas tecnologías también pueden denominarse Tecnologías Inmersivas. Actualmente, la mayoría de las aplicaciones XR se centran en el contenido visual, pero también pueden incluir estímulos hápticos, auditivos y olfativos, entre otros.

Milgram y Kishino (1994) propusieron un continuo para representar los diferentes grados de separación entre la completa virtualidad y la realidad (Figura 1). La realidad mixta (RM) es un espectro entre estos extremos, conectando entornos totalmente físicos a entornos totalmente virtuales. La realidad virtual (RV) (Figura 2b) está en un extremo y describe un entorno compuesto

únicamente por objetos virtuales. En el otro extremo está la pura realidad física (Figura 2a).

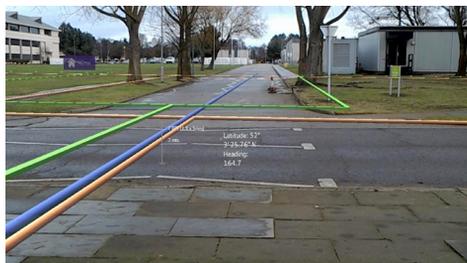
La *realidad aumentada* (RA) (Figura 2b) difiere de la *realidad virtual* en que superpone contenido digital en el mundo físico, mientras que la RV está completamente aislada del mundo físico, presentando un mundo virtual 3D sintético. Dada la presentación simultánea de elementos físicos y virtuales, las aplicaciones de RA funcionan mejor en situaciones donde es necesario interactuar con elementos físicos del mundo cotidiano. Por lo tanto, la RA complementa la realidad en lugar de reemplazarla por completo. En contraste, la *virtualidad aumentada* (VA) describe un entorno principalmente virtual, que se mejora al agregar objetos físicos y datos del mundo físico (en la Figura 2 se muestran varios ejemplos al respecto).

La *computación espacial* es un nuevo paradigma de la informática, basado en formas naturales de interacción en la convergencia de los mundos físico y virtual. Puede ser definido como “la interacción humana con una máquina en la que la máquina retiene y manipula referentes a objetos y espacios reales” (Greenwold 2003). La computación espacial se diferencia de otros campos relacionados (por ejemplo, el modelado 3D y el diseño digital) en que requiere que las formas y los espacios con los que trata pre-existan y tengan valencia en el mundo real (Greenwold 2003). Es decir, no es suficiente que la pantalla se use para representar un espacio virtual; debe estar relacionada de manera significativa con un lugar en el mundo físico.

Debido a que las aplicaciones de realidad extendida modelan y aumentan espacios físicos, son interfaces ideales para la aplicación de este paradigma. Para ello requieren de un mecanismo que rastree y controle los movimientos e interacciones de los objetos mientras una persona navega por el mundo digital o físico, en referencia a la práctica de usar acciones físicas (por ejemplo, interacciones controladas por gestos, movimientos de la cabeza y el cuerpo, o por procesamiento de lenguaje natural)



a) Realidad física



b) Realidad aumentada



c) Realidad virtual

Figura 2. Ejemplos de Tecnologías Inmersivas. Elaboración propia (Peña-Rios *et al.* 2018; Peña-Rios 2022).

como entradas para sistemas de medios digitales interactivos (Lathan y Ling 2020). Por lo tanto, aportan un enfoque centrado en el ser humano a las interfaces de usuario. De igual manera, la respuesta que dan estos sistemas de medios digitales al procesamiento de datos ya no puede estar contenida en una sola salida (por ejemplo, una pantalla), sino que deberían ocupar el espacio físico (en tercera dimensión) que nos rodea, manifestándose a través de comunicación multimodal (por ejemplo, mediante visualización de datos o video en el espacio físico, retroalimentación auditiva y/o salidas hápticas). Esto mezcla la línea entre el mundo físico y el mundo digital o simulado, también crea una sensación de integración entre ambos y, por lo tanto, una mayor inmersión en la experiencia. Esto es más difícil de lo que parece, ya que nuestros sentidos y cerebros evolucionan para brindarnos una experiencia finamente sincronizada y mediada.

Dependiendo de la tecnología usada se pueden alcanzar diferentes grados de *inmersión espacial* (o *presencia*) (Figura 3). La *presencia* se puede definir como la experiencia de estar en un entorno creado artificialmente que es lo suficientemente realista como para hacer que los participantes se olviden de que se encuentran en un entorno sintético (Dede 1995, Slater y Wilbur 1997, Ijsselsteijn 2005).

### **Gemelos Digitales y la computación cognitiva**

Como ya vimos en la sección anterior, la computación espacial propone computación híbrida real-virtual que borre las barreras entre el mundo físico y el virtual. Este paradigma está cercanamente ligado con el concepto de *Gemelos Digitales*.

Los *Gemelos Digitales* (Figura 4) podrían concebirse mejor como un concepto amplio, en lugar de una tecnología específica o un conjunto de tecnologías que tienen como objetivo reproducir sus contrapartes del mundo real de la manera más realista posible. En esencia, podrían considerarse como una “réplica digital de una entidad física viva o no viva” (El Saddik 2018).

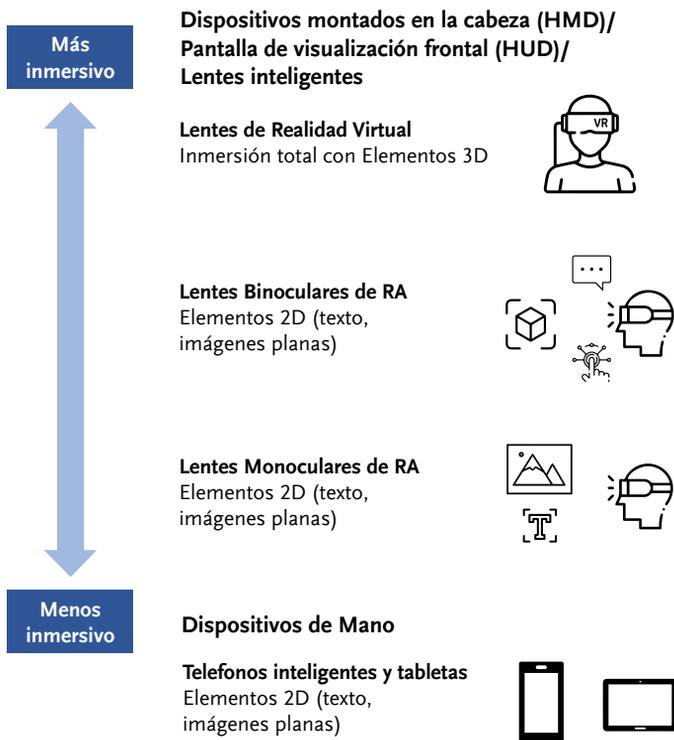


Figura 3. Grado de inmersión en relación con el hardware usado (Peña-Rios, 2022). Elaboración propia.

Su implementación generalmente combina diferentes tecnologías para crear modelos de simulación digital que aprenden y se actualizan a medida que cambian sus contrapartes físicas, sincronizando su estado, condición de trabajo o posición casi en tiempo real (Peña-Rios 2022). A diferencia de un modelo tradicional, que sólo puede brindar una instantánea del comportamiento en un momento específico, un Gemelo Digital describe con precisión el cambio a lo largo del tiempo de un activo o proceso físico de la vida real. Esta conexión se basa en “mediciones de datos masivas, acumulativas, casi en tiempo real y del mundo real en una

variedad de dimensiones” (Parrott y Warshaw 2017). Este puente de transmisión de datos bidireccional entre el mundo físico y el virtual permite que la entidad virtual exista simultáneamente (y sincronizada) con su contraparte física (El Saddik 2018).

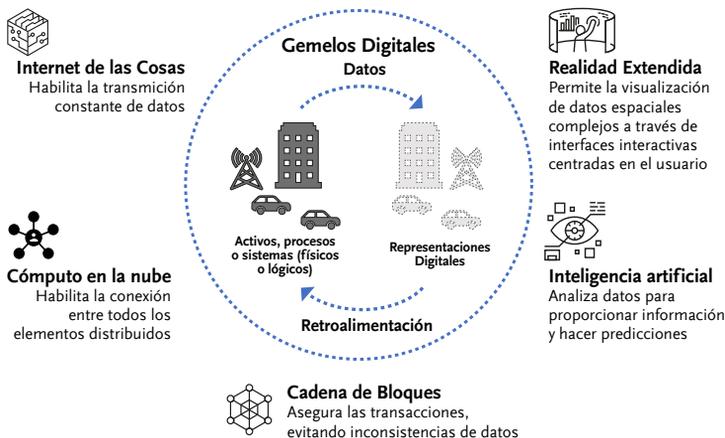


Figura 4. Gemelos Digitales y tecnologías relacionadas (Peña-Rios 2022).

Elaboración propia.

Al permitir que los usuarios visualicen sistemas y procesos complejos de forma remota es posible ejecutar de manera rentable simulaciones hipotéticas y tomar instantáneas digitales de situaciones de interés que pueden usarse para optimizar la operación y el mantenimiento de los activos y procesos físicos subyacentes. La expectativa es que el despliegue generalizado de estos gemelos se traducirá en una reducción de costos operativos, una evaluación de riesgos acelerada y una mejor colaboración en equipo (Peña-Rios 2022).

La computación espacial propone hacer Gemelos Digitales no sólo de objetos, sino de personas y ubicaciones, utilizando GPS, lidar (acrónimo del inglés LiDAR, *Light Detection and Ranging* o

*Laser Imaging Detection and Ranging*), video y otras tecnologías de geolocalización para crear un mapa digital de una habitación, un edificio o una ciudad. Los algoritmos de software integran este mapa digital con datos de sensores y representaciones digitales de objetos y personas para crear un mundo digital que puede ser observado, cuantificado y manipulado y que también puede manipular el mundo real (Lathan y Ling 2020).

La computación espacial combinada con mecanismos de computación cognitiva (por ejemplo, el Aprendizaje Automático, la visión por computadora, el Procesamiento Natural del Lenguaje) permitirá desarrollar sistemas expertos que pueden proporcionar recomendaciones relevantes para la toma de decisiones en función del contexto de la persona que está solicitando, además de que difuminará no sólo los límites entre el mundo virtual y físico, sino también los límites entre humanos y máquinas, lo que llevará “las interacciones hombre-máquina y máquina-máquina a nuevos niveles de eficiencia en muchos ámbitos de la vida, entre ellos la industria, la atención médica, el transporte y el hogar” (Lathan y Ling 2020).

La industria ya ha adoptado la integración de sensores dedicados, Gemelos Digitales, RA y el Internet de las cosas para optimizar la productividad y está empezando a adoptar la computación espacial. Por ejemplo, en el lugar de trabajo del futuro, un empleado podría ponerse auriculares de realidad aumentada para ver una imagen holográfica proyectada que muestra no sólo las instrucciones de reparación, sino también un mapa espacial de los componentes de la máquina. De esta manera los trabajadores pueden ser guiados a través y alrededor de la máquina para repararla de la manera más eficiente posible, con lo que existiría una reducción del tiempo de inactividad y sus costos. En otro ejemplo, un técnico podría utilizar una interfaz de realidad virtual para visualizar un Gemelo Digital de un sitio físico remoto y poder dirigir varios robots mientras construyen productos en una fábrica. Aquí los algoritmos de computación espacial podrían

ayudar a optimizar la seguridad, eficiencia y calidad del trabajo mejorando, por ejemplo, la coordinación de los robots y selección de las tareas que se les asignen (Peña-Rios 2022).

### **Humanos aumentados y la capacitación del futuro**

A medida que las tecnologías antes mencionadas se integren a la sociedad y sus actividades, y conecten personas, productos, procesos y plataformas, se requerirán nuevos paradigmas para respaldar la transformación digital.

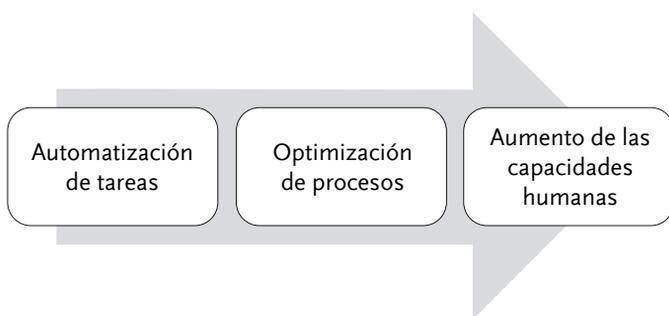


Figura 5. Transformación digital (Peña-Rios 2022). Elaboración propia.

En este sentido, las Tecnologías Inmersivas y la computación espacial tienen el potencial de transformar a la sociedad más allá de la automatización de tareas y la optimización de procesos, al brindar un enfoque centrado en el ser humano, y *aumentar* los sentidos para el pensamiento multidimensional y la resolución de problemas (Figura 5). Al combinarlas con mecanismos de cómputo cognitivo, se pueden producir entornos adaptativos (conscientes del contexto) para un soporte avanzado de decisiones. Estas soluciones podrían conducir a una transferencia de conocimiento mucho más rápida y una comprensión más profunda de diferentes procesos utilizando mecanismos de aprendizaje activo.

El constructivismo es una teoría del aprendizaje que afirma que los individuos construyen su conocimiento basándose en

sus experiencias y conocimientos previos. El cognitivismo está estrechamente relacionado con el constructivismo y establece que todo conocimiento se basa en uno previo. Según esta teoría, los individuos crean activamente su conocimiento, moldeando su perspectiva de la realidad. Uno de los principios clave del constructivismo es que el aprendizaje debe estar centrado en el estudiante. Así, el enfoque del aprendizaje cambia del docente a los estudiantes, quienes son vistos como agentes activos en su proceso de aprendizaje, trayendo sus propias experiencias y perspectivas únicas al salón de clases, para dar forma a su comprensión de los conceptos que se les enseñan. Este enfoque enfatiza el aspecto social del aprendizaje, ya que los estudiantes interactúan entre sí y construyen su conocimiento juntos, mejorando y diversificando el potencial de aprendizaje de toda la clase.

Hay tres etapas de adquisición de conocimiento: introductorio, avanzado y experto (Jonassen 1991). Los entornos de aprendizaje orientados al constructivismo son más efectivos para la adquisición de conocimientos avanzados, lo que permite a las personas crear conocimientos basados en la experiencia (Jonassen 1991, Ertmer y Newby 2013). Los enfoques más objetivistas respaldan mejor la adquisición inicial de conocimientos. Sin embargo, la transición a enfoques constructivistas a medida que los alumnos adquieren más conocimientos les proporciona las capacidades para abordar problemas complejos (Jonassen 1991, Ertmer y Newby 2013).

Las Tecnologías Inmersivas están orientadas a enfoques constructivistas, ya que permiten que los alumnos experimenten contenidos de manera inmersiva y personalizada. Sin embargo, también existen limitaciones y desafíos asociados con su uso en la educación y capacitación. Por ejemplo, un desafío es el costo de desarrollar contenido educativo que esté diseñado para alcanzar objetivos de aprendizaje específicos, ya que requiere experiencia y recursos especializados. Otro desafío es crear experiencias prácticas de aprendizaje que fomenten la participación, la colaboración y la creatividad. Lograr esto requiere un

conocimiento que abarca muchos temas, incluidas las ciencias de la computación, la inteligencia artificial, las ciencias del aprendizaje y las interacciones humano-computadora. Las limitaciones están ligadas al acceso a este tipo de tecnologías, que, si bien se ha incrementado en años recientes, aún se encuentra restringido a algunos cuantos. De igual manera, cualquier tecnología está ligada a estas restricciones.

A pesar de estos desafíos, el uso de Tecnologías Inmersivas en la formación ofrece numerosas oportunidades de innovación. Por ejemplo, podrían proporcionar una plataforma para el aprendizaje remoto, lo que puede ser especialmente beneficioso en lugares con acceso limitado a equipos o instalaciones especializadas. El sector educativo podría crear atractivas experiencias de aprendizaje inmersivas que soporten el contenido presentado en aulas.

### **El empleo aumentado**

Desde la perspectiva industrial, el uso de las tecnologías ya mencionadas podría habilitar la transformación digital como una forma estructurada de mejorar constantemente la experiencia del usuario y aumentar las competencias digitales. Estos cambios requieren que los líderes identifiquen de forma continua las brechas y preparen a los empleados para realizar nuevas funciones laborales, lo que representa un gran desafío para las empresas, y para los mismos empleados. Esta confluencia de tecnologías propicia la idea del *empleo aumentado*, en quien el aumento de las capacidades humanas se genera proporcionando la información correcta, en el momento correcto, y transmitida al usuario correcto a través de comunicación multimodal (Peña-Rios *et al.* 2018). Esto puede ayudar como soporte hacia el pensamiento multidimensional y la resolución de problemas (Warwick 2014).

La Figura 6 muestra cuatro áreas donde las Tecnologías Inmersivas y la computación espacial y cognitiva podrían ser utilizadas en este contexto:



Figura 6. El empleado aumentado. Elaboración propia.

- a) Soporte experto en el trabajo. Las Tecnologías Inmersivas han demostrado capacidad para mitigar la creciente complejidad de las tareas de mantenimiento y reparación industrial; con la consecuente mejora del entorno de trabajo al proporcionar soporte en cualquier momento (Wang *et al.* 2017). A través de la RA se permite el acceso a la información que se superpone en el campo de visión del usuario. Además, al utilizar videollamadas mejoradas con datos aumentados, es posible proporcionar diagnósticos remotos, colaboración remota y capacitación en el trabajo.
- b) Capacitación y entrenamiento. La capacitación inmersiva es el concepto de capacitar a las personas centrándose en la participación del alumno basado en enfoques de aprendizaje constructivistas (por ejemplo, aprendizaje experimental, aprendizaje exploratorio y aprendizaje basado en problemas). La participación o *engagement* es un constructo psicológico que hace referencia a cuán activamente se encuentra involucrado el individuo en una determinada actividad (Kahn 1990). Las características de la RV ofrecen muchos beneficios sobre la

- formación tradicional en el aula, en particular para los roles operativos que se centran más en actividades prácticas que en conocimientos teóricos. Algunos de estos beneficios incluyen acceso a simulaciones que permiten recrear eventos que en el mundo físico serían imposibles (por ejemplo, situaciones de riesgo físico para los aprendices, como incendios o momentos históricos).
- c) Gestión del conocimiento. Como herramienta educativa, el uso de la RA podría resultar en menos tiempo en capacitación y menos tiempo para dominar una tarea (Wild 2016). Sin embargo, para lograr esto es necesario identificar los escenarios correctos y diseñar experiencias de aprendizaje apropiadas (Wang *et al.* 2017). Por ejemplo, la RA puede proporcionar disparadores para la información clave en los procedimientos de mantenimiento que no sólo permite a los trabajadores menos calificados completar tareas críticas de manera competente, sino que también acelera su transición de novato a experto.
  - d) Herramientas de apoyo a la toma de decisiones. El uso de Gemelos Digitales, monitoreo y control remoto de equipos. Por ejemplo, la capacidad de la RA de proporcionar contenido virtual tridimensional (3D) en un espacio de trabajo del mundo real tiene el potencial de 1) reducir la carga de trabajo mental y 2) ser una herramienta cognitiva efectiva para aprender a mejorar el recuerdo y la comprensión, acelerando la curva de aprendizaje, y reduciendo la carga cognitiva (Wang *et al.* 2017).

El gran desafío actual para las empresas es aumentar la experiencia de los empleados y mantener/capturar el conocimiento experto y transferirlo a otros colegas. Convertir conocimiento tácito (experiencia) en explícito (tangible o documentado) no es una tarea fácil. Las Tecnologías Inmersivas y la computación espacial y cognitiva podrían proporcionar ayuda con estos retos.

## Casos de estudio

El campo de las operaciones de servicio es complejo y requiere conocimientos especializados, por ejemplo, en el mantenimiento y reparación de equipos y procedimientos especializados. El mantenimiento manual de los equipos es costoso y requiere tiempo, además de que existe un alto riesgo de mal funcionamiento del equipo, lo que reduce la productividad. En este sentido, el concepto de *Servicio 4.0* tiene como objetivo apoyar y promover la innovación en las operaciones de servicio al utilizar tecnología emergente (Peña-Rios *et al.* 2018). La gestión de las operaciones de servicio incluye la detección de una necesidad de servicio de campo (a través de monitoreo remoto u otros medios, inspección o un cliente que detecta una falla); programación y optimización de tareas; asignación de técnicos de campo; despacho, entrega de información y/o piezas al campo, y soporte de procesos de interacciones de técnicos de campo (Gartner s.f.). Además, las innovaciones tecnológicas actuales presentan una oportunidad significativa para brindar soporte en el sitio y en tiempo real a los profesionales de servicio de campo en muchas áreas.

A continuación, se presentan casos de uso en el que las Tecnologías Inmersivas se pueden usar en el proceso de entrenamiento de un ingeniero o técnico de campo en una empresa de servicios (Figura 7).



Figura 7. Casos de uso. Elaboración propia.

*Reclutamiento: simulador video inmersivo 360° de realidad virtual*

Atraer talento a una empresa es una tarea compleja. Los reclutadores enfrentan diferentes desafíos al tratar de encontrar al candidato adecuado para el trabajo adecuado. Estos desafíos van desde la gestión de las expectativas de los candidatos hasta el cumplimiento de perfiles de trabajo especializados. La RV puede ser utilizada para mostrar y promover lo que implica trabajar como parte de la empresa, al presentar simulaciones para probar el conocimiento de un candidato o para desplegar roles específicos al usar representaciones de las áreas de trabajo. Cuando se hace correctamente, el impacto del uso de la RV en el reclutamiento puede ser significativo. Por ejemplo, el caso del ejército británico que utilizó RV para exponer el entrenamiento militar resultó ser un éxito, con un aumento de 65% en el número de solicitudes en febrero de 2018 (Weekes 2018).

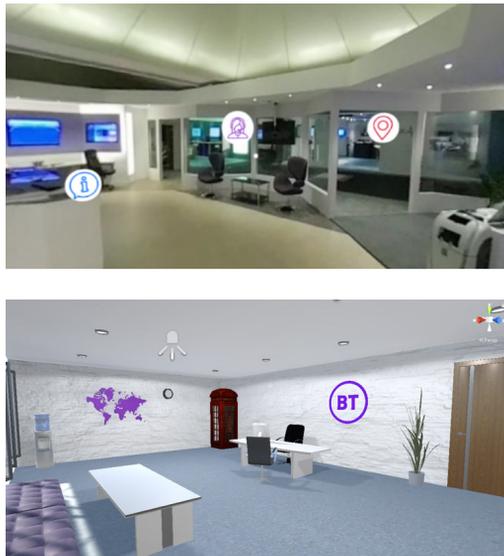


Figura 8. Simulador video inmersivo 360° (Peña-Rios *et al.* 2020a).

Elaboración propia.

La Figura 8 muestra una aplicación para dispositivo móvil que muestra un simulador de video inmersivo 360°, es decir, una combinación de realidad virtual y escenas multimedia de 360°, que permite recorridos virtuales y la actualización y personalización del contenido según el perfil de los roles que se promocionarán (Peña-Rios *et al.* 2020a).

Se completó un ensayo con 14 participantes en total, 7 hombres y 7 mujeres, dentro de un rango de edad entre 18 y 40 años, usando un teléfono móvil con unos lentes de RV (Peña-Rios *et al.* 2020a). El contenido utilizado en las pruebas se centró en describir los roles técnicos de I + D utilizando cinco escenas de 360° que presentaban contenido relacionado con roles específicos para los que la empresa desea atraer talento. Los participantes proporcionaron retroalimentación positiva con respecto al uso de esta herramienta para ilustrar de una manera más entendible y realista lo que significa trabajar en un rol en específico, lo cual puede ayudar a mejorar las expectativas acerca de lo que el rol involucra.

### ***Entrenamiento en aula: capacitación inmersiva usando realidad virtual***

Después de concluir la búsqueda de talento, es necesario asegurarse de que un nuevo empleado comprenda sus responsabilidades laborales. Una formación inadecuada puede conducir a una eventual rotación de empleados, lo que significa repetir todo el proceso. La capacitación inmersiva proporciona muchos beneficios en comparación con la capacitación tradicional en el aula, particularmente con roles operativos que se enfocan más en actividades prácticas que en conocimientos teóricos. Algunos de estos beneficios incluyen (Peña-Rios *et al.* 2020b):

- *Entrenamiento activo a través de equipos físicos simulados*, lo que permite el acceso a equipos clave en un entorno seguro, y reduce riesgos para el personal y el equipo. Además, los alumnos pueden completar las tareas a su propio ritmo, y

crear un entorno de apoyo para las personas con menos experiencia en el tema, así como para las personas con discapacidades de aprendizaje.

- *Soporte logístico optimizado para planeación de actividades de capacitación*, lo que reduce: la necesidad de viajar a una ubicación específica, los costos de viaje, las emisiones de CO<sub>2</sub> y el tiempo de inactividad de los ingenieros.
- *Recreación de eventos poco comunes*, que son difíciles de replicar con fines de capacitación debido a factores como acceso restringido a ciertas áreas, regulaciones de seguridad o baja disponibilidad de conocimiento (por ejemplo, habilidades especializadas que están en conocimiento de unos pocos empleados), lo que dificulta que la formación en el puesto de trabajo sea suficiente, ya que los alumnos pueden terminar su periodo de formación sin experimentar muchos escenarios posibles.
- *Captura de métricas para cumplimiento y auditoría*, las cuales permiten asegurar que el trabajador esté preparado por completo para realizar la actividad en el ambiente de trabajo. Además, estos análisis se pueden utilizar posteriormente para mejorar la experiencia de los usuarios.

Por ejemplo, Walmart adoptó el aprendizaje inmersivo basado en la realidad virtual como una nueva herramienta en la capacitación de servicio al cliente en 200 academias de Walmart. La simulación de RV coloca al asociado en un piso de ventas para practicar sus habilidades, esto incluye la gestión de situaciones estresantes como las ventas de *Black Friday* en un entorno frenético simulado. Las pruebas informaron que los empleados que tomaron la capacitación en RV presentaron 30% más de satisfacción, obtuvieron 70% más en las pruebas y mostraron entre 10 a 15% más de retención de conocimientos (Sisson 2018).

La Figura 9 presenta un simulador para entrenamiento inmersivo que tiene como objetivo mejorar y facilitar el proceso de formación de los ingenieros de servicio de campo. Se diseñó

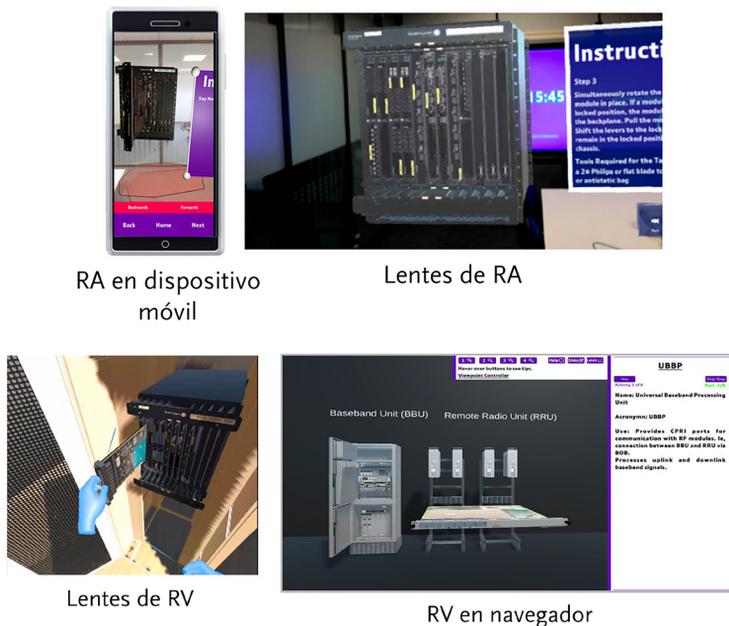


Figura 9. Simulador para capacitación inmersiva en diferentes plataformas (Peña-Rios *et al.* 2022). Elaboración propia.

como una aplicación para varias plataformas: RV inmersiva, RA, RV basada en navegador, y RA para dispositivo móvil. La aplicación muestra simulaciones en tercera dimensión (3D) de equipo clave para la red de telecomunicaciones, lo que permite la exploración del equipo y sus componentes, así como actividades de configuración. Todas las plataformas muestran el mismo contenido, ya que éste se encuentra de manera centralizada en una base de datos. Sin embargo, el contenido es adaptado a las diferentes características del hardware mediante el cual el empleado interactúa. En este sentido, se realizaron diferentes pruebas con usuarios para identificar las fortalezas y los puntos a mejorar en cada una de las plataformas (Peña-Rios *et al.* 2020b, Bahceci *et al.* 2021). Por ejemplo, usar lentes de RV o de RA aumentada

permite una experiencia más inmersiva, por lo tanto, el usuario se puede concentrar mejor en las actividades. Acceder a ellas mediante un dispositivo móvil limita la cantidad de interacciones e información que se presenta debido al tamaño de la pantalla. En el caso de la experiencia de RV en navegador, moverse por un entorno 3D en una pantalla 2D puede confundir a los participantes porque les resulta difícil orientarse y percibir la profundidad (Peña-Rios *et al.* 2020b). Sin embargo, el acceso a contenido de entrenamiento inmersivo mediante dispositivos móviles y computadoras personales con navegadores de Internet ayuda a democratizar el contenido poniéndolo a disposición de una audiencia más vasta.

### ***Experto remoto para capacitación en campo***

La capacidad de conectar a los trabajadores de primera línea con información relevante es clave para la transformación digital de una organización. Este tipo de trabajadores hacen su mejor trabajo cuando pueden obtener la información correcta en el momento adecuado. A veces esa información sólo puede provenir de un experto o mentor externo. El llamado *experto remoto* es la capacidad de conectarse con especialistas remotos al compartir video en vivo, audio, telestración y contenido aumentado en tiempo real. La telestración se define como una técnica que permite el dibujo de marcas a mano alzada (anotaciones) sobre una imagen o video (Budrionis *et al.* 2013). Los expertos remotos pueden ayudar a los trabajadores de primera línea desde cualquier parte del mundo, ya sea que se encuentren en la oficina central, en casa o incluso en otra zona horaria. Esta colaboración en vivo puede capturar datos como grabaciones e imágenes que pueden ser usadas en auditorías o en capacitaciones futuras.

La implementación de un *experto remoto* se puede hacer mediante lentes de RA o dispositivos móviles. Ambos están equipados con una cámara y dispositivos de salida y captura de audio que permiten a los trabajadores en el campo colaborar con

expertos remotos. Al conectarse fácilmente y trabajar con un experto remoto, los trabajadores de primera línea pueden resolver desafíos difíciles o únicos de manera más fácil y rápida. Algunos de los beneficios de los expertos remotos incluyen:

- *Mejorar las tasas de reparación a la primera y reducir el tiempo de inactividad.* No todas las soluciones se pueden encontrar en un documento técnico. Un técnico experimentado a menudo puede identificar y diagnosticar problemas desafiantes de un vistazo y ayudar a guiar el mantenimiento o la reparación de forma rápida y sencilla. Las resoluciones rápidas son cruciales para minimizar el tiempo de inactividad del equipo y garantizar la máxima productividad.
- *Reducir las necesidades de viaje y los costos operativos.* Enviar expertos en todo el mundo para supervisar tareas a corto plazo es costoso, requiere mucho tiempo e interrumpe el equilibrio entre la vida laboral y personal de los empleados. A través de la colaboración remota, los expertos pueden recorrer las instalaciones, abordar problemas específicos o interactuar con los trabajadores de forma virtual. Y como no necesitan viajar, los expertos pueden “visitar” más sitios sin los gastos de transporte, alojamiento y comidas asociados.
- *Acelerar el proceso de transferencia de habilidades.* Los mentores son fundamentales para desarrollar nuevos talentos. Pero cuando la asistencia presencial es difícil de escalar en una empresa, un experto remoto puede ayudar a muchos trabajadores a la vez en varios sitios.

Por ejemplo, Volkswagen anunció en Reino Unido el lanzamiento a nivel nacional de lentes de RA para técnicos en toda su red de centros de reparación autorizados de vehículos comerciales. Esto para ayudar a acelerar los trabajos de reparación complejos. Una prueba exitosa les ahorró a los clientes un año de tiempo de inactividad con una eficiencia de reparación de 93%. No tener que

desplazarse físicamente también supuso una reducción de más de 2.5 toneladas de CO2 para el TSC (RealWear 2020).

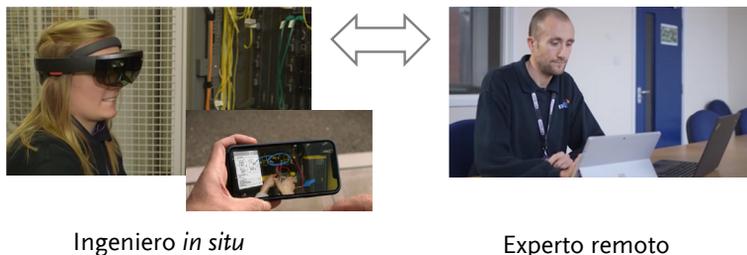


Figura 10. Experto remoto. Elaboración propia.

La Figura 10 muestra una sesión entre un ingeniero *in situ* y un experto remoto. Para evaluar este caso de uso se completaron una serie de sesiones entre ingenieros de telecomunicaciones en Reino Unido entre diciembre de 2016 y febrero de 2017, quienes utilizaron dispositivos portátiles inteligentes (Peña-Rios *et al.* 2018). La muestra estuvo formada por 5 hombres y 2 mujeres con edades comprendidas entre los 20 y los 56 años (media = 32). Todos los participantes estaban familiarizados con las computadoras. En esta evaluación, se utilizó un dispositivo Microsoft HoloLens versión 1, que es un dispositivo de realidad mixta sin ataduras que usa lentes holográficos transparentes, combinados con mapeo espacial, voz, habla y reconocimiento de gestos. El objetivo era evaluar la eficacia percibida y la usabilidad del soporte uno-a-uno en tiempo real. Los resultados de la evaluación mostraron que los participantes pensaban que la tecnología usada era útil para completar las tareas asignadas. Sin embargo, los participantes reportaron desventajas del hardware relacionadas con la reducida amplitud del campo de visión, el peso y el costo del dispositivo (Peña-Rios *et al.* 2018). Comentarios adicionales reflejaron preocupación por el hecho de que los usuarios percibían que el equipo era muy sofisticado y, por lo tanto, no

todos los ingenieros estarían dispuestos a probarlo, especialmente aquellos que no tienen un sentimiento afín con la tecnología. Otra preocupación se refirió a la logística y disponibilidad de usar el equipo, ya que el poder del equipo reside en la posibilidad de usarlo en eventos no programados cuando es difícil encontrar con rapidez a la persona adecuada para el trabajo.

### ***Asignación de tareas: asistentes virtuales***

Los asistentes virtuales son agentes de software que facilitan la comunicación hombre-máquina al realizar tareas para un usuario en función de comandos específicos mediante el uso de computación cognitiva, incluidos el Procesamiento del Lenguaje Natural (PLN), el Aprendizaje Automático, el reconocimiento visual de objetos, el análisis acústico y el procesamiento de señales. Algunos asistentes virtuales pueden responder a través de voces sintetizadas, y algunos solamente pueden comunicarse mediante el chat en línea (*bots* conversacionales). En el contexto del ambiente de trabajo, los asistentes virtuales pueden ofrecer a los empleados información personalizada a las actividades que realizan y su nivel de conocimiento.

En combinación con Tecnologías Inmersivas y dispositivos prendables (por ejemplo, lentes de RA), pueden proporcionar instrucciones paso a paso para que los empleados realicen procedimientos complejos en el trabajo. En algunos dispositivos es posible que los operadores controlen la interfaz con la mirada o mediante instrucciones de voz, lo que les deja las manos libres para hacer el trabajo. Al usar RA es posible presentar información personalizada al usuario, como tarjetas de instrucciones, diagramas y esquemas, videos e inclusive hologramas que señalen las herramientas y piezas que necesitan y les muestran exactamente cómo y dónde utilizarlas. De esta manera, el asistente virtual puede guiar a los usuarios a través de las tareas asignadas, impulsando el siguiente paso en el flujo de trabajo, haciendo que sea más rápido y fácil completar las tareas, al tiempo que

garantiza una captura de datos eficiente a la que se puede acceder de inmediato a través de registros digitales. El beneficio de este tipo de experiencias incluye:

- *Reducción de errores en captura de datos*, ya que la documentación manual de las tareas y los datos generados (por ejemplo, bitácoras o formularios) puede verse obstaculizada por la velocidad, la precisión y la interpretación del usuario. Mediante la ayuda de un asistente virtual sería posible capturar información de manera verbal y convertirla en registros de datos.
- *Acelerar la finalización de las tareas a la primera*, ya que los operadores ven exactamente qué se debe hacer y dónde, para que puedan hacer el trabajo más rápido, con menos errores.
- *Mayor retención de habilidades*, ya que se considera que el aprendizaje es un “proceso mediante el cual el conocimiento se crea mediante la transformación de la experiencia” (Kolb 1984). De esta manera, el asistente virtual al guiar a usuarios novatos para completar las tareas contribuye a la capacitación de los operadores.

Por ejemplo, la empresa Dotmatics está impulsando la automatización de los flujos de trabajo de datos de laboratorio para proporcionar a los científicos de I + D la capacidad de registrar, acceder y rastrear datos dentro de un cuaderno de laboratorio electrónico (*electronic laboratory notebook*, o ELN por sus siglas en inglés). A través de esta herramienta los científicos podrían operar en un entorno de laboratorio de manos libres, al utilizar su voz para solicitar el estado de los instrumentos, clasificar muestras, capturar medidas y ajustar experimentos; todo en tiempo real, mejorando la integridad de los datos y el cumplimiento del usuario (News Medical 2020).

Para realizar una valoración inicial de este caso de uso, se construyó una aplicación para lentes de realidad mixta (Microsoft HoloLens), con el objetivo de proporcionar instrucciones y objetos

holográficos a los ingenieros de campo a aprender durante el flujo de trabajo, además de proporcionarles información relevante cuando y donde se necesiten. Estas tarjetas de instrucciones están unidas visualmente al lugar donde se realiza el trabajo y pueden incluir imágenes, videos y modelos holográficos en 3D, y siguen al operador mientras se desplaza por el equipo

Aunque la retroalimentación por parte de los usuarios fue positiva, uno de los retos principales para la implementación de este tipo de herramientas es contar con flujos de trabajo bien establecidos, así como procedimientos específicos bien documentados. Sin embargo, aunque esta información se encuentre disponible, es necesario adaptar la experiencia y el conocimiento tácito de los usuarios expertos para alimentar este tipo de sistemas cuando un usuario presenta un problema no considerado en el flujo de las tareas, y generar asistentes virtuales que se adapten a situaciones inesperadas dependiendo del contexto y al permitir al usuario una mayor personalización de las herramientas.

### ***Ejecución de tareas: Gemelos Digitales***

Como se mencionó antes, un Gemelo Digital es un modelo virtual altamente complejo que es la contraparte exacta (o gemelo) de un objeto físico o de un proceso. Este objeto podría ser un automóvil, equipo especializado en fábricas u hospitales, un puente e incluso ciudades enteras. Esta conexión entre el modelo virtual y el activo físico se produce mediante sensores conectados en el activo físico y que recopilan datos que se pueden asignar al modelo virtual, lo que permite monitorear información crucial del activo en tiempo real.

Los Gemelos Digitales pueden representar la realidad física en diferentes niveles de precisión y fidelidad, dependiendo de su propósito, casos de uso y tres elementos principales (Bolton *et al.* 2018):

- Datos. La calidad y disponibilidad de los datos en los que se basa el gemelo.

- **Modelo.** La confiabilidad de los algoritmos y la implementación tecnológica de la representación digital alineados a la validez de los supuestos.
- **Visualización.** La fidelidad de presentación de la salida.

Sus beneficios incluyen capacidades de modelado y simulación para probar conceptos y servicios, planificar, tomar decisiones, investigar tecnologías y generar colaboración comunitaria. A su vez, los Gemelos Digitales pueden crear nuevas posibilidades para mejorar la creatividad y la innovación al vincular modelos digitales y simulaciones con datos del mundo real, además de utilizar un diseño centrado en el ser humano. Muchos podrían argumentar que la toma de decisiones a través de Gemelos Digitales se puede hacer simplemente utilizando análisis de datos (2D), pero no debemos olvidar que los humanos están en el centro de las operaciones; así, los Gemelos Digitales 3D presentados a través de Tecnologías Inmersivas como la realidad mixta, la realidad aumentada o la realidad virtual pueden proporcionar una confirmación visual y un contexto espacial, lo que hace que sea más fácil de entender, no sólo visualizando los datos, sino experimentándolos en primera persona y en tres dimensiones (Peña-Rios 2022).

Por ejemplo, Singapur Virtual es un proyecto de colaboración entre la Fundación Nacional de Investigación de Singapur y Dassault Systèmes (National Research Foundation 2018). Desde 2015 el proyecto ha operado un modelo digital 3D dinámico de la ciudad de Singapur que conecta a todas las partes interesadas en un entorno seguro y controlado. Con imágenes y datos recopilados de varias agencias públicas (incluyendo geoespaciales y de topología), así como datos históricos y en tiempo real (demografía, movimiento o clima), los usuarios pueden crear modelos visuales ricos y simulaciones realistas a gran escala de Singapur y explorar digitalmente el impacto de la urbanización en la ciudad-estado y desarrollar soluciones que optimicen la logística, la gobernanza y las operaciones relacionadas con la gestión

ambiental y de desastres, la infraestructura, la seguridad nacional o los servicios comunitarios.

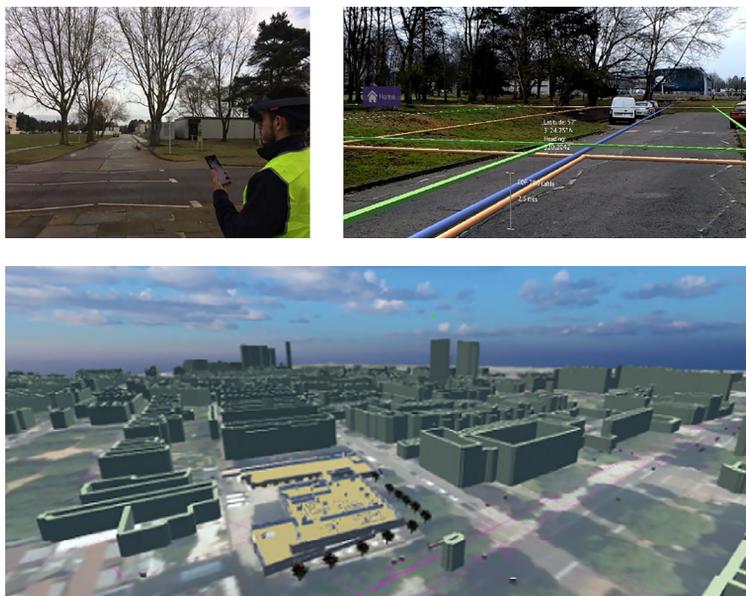


Figura 11. Visualización de infraestructura de la red de telecomunicaciones (Peña-Rios *et al.* 2018). Elaboración propia.

Los Gemelos Digitales nos brindan la capacidad de visualizar, ejecutar simulaciones y escenarios “qué pasaría si...”, y proporcionar instantáneas digitales para trabajos futuros de sistemas verdaderamente complejos. Estos pueden ser usados para monitorear infraestructura mediante interfaces de RV y RA (Figura 11). Los Gemelos Digitales podrían respaldar la transformación digital en diferentes industrias de servicio, al permitir un mejor uso, operación, mantenimiento, planificación y entrega de los activos, sistemas y servicios. Algunos ejemplos son la gestión de la infraestructura energética, que comprende el uso de energía para optimizar los recursos; el modelado de activos, que proporciona

una visión clara de los activos físicos al usar visualizaciones en 3D basadas en información geoespacial precisa para crear modelos de simulación optimizados, y el diagnóstico remoto, que permite la supervisión y el diagnóstico remotos al utilizar datos de telemetría. Esto, combinado con el Aprendizaje Automático y el análisis de datos, puede respaldar las actividades proactivas (en lugar de reactivas) de mantenimiento y crear una infraestructura inteligente.

Estos Gemelos Digitales nacientes pueden brindar conocimientos y cambios que no anticipamos al mejorar la eficiencia, optimizar Sistemas Complejos, reducir errores y proporcionar comprensión a nivel humano de problemas complejos.

### **Reflexiones e implicaciones éticas**

Las Tecnologías Inmersivas están emergiendo como uno de los impulsores clave de la economía tecnológica. Éstas tienen el potencial de generar un impulso de 1.5 billones de dólares a la economía mundial para 2030 (PWC 2019). Se espera que los avances en lentes inteligentes, dispositivos prendables y tecnologías habilitadoras críticas (como la Inteligencia Artificial y las redes 5G) respalden las Tecnologías Inmersivas para proporcionar experiencias de vanguardia, allanando el camino para varios casos de uso. Según un informe de Immerse UK, una organización de Tecnologías Inmersivas impulsada por la agencia de investigación del gobierno de Reino Unido, 45% de las industrias de alto impacto, donde la vida humana o la calidad está en juego, reconoce las Tecnologías Inmersivas como parte de su crecimiento futuro. El informe también indica que la formación basada en realidad virtual reduce el tiempo de formación en 40% y mejora el rendimiento de los empleados en 70% (Immerse UK and Digital Catapult 2019).

A medida que el poder de la computación cognitiva aumente, se prevé contenido de alta calidad que altere la realidad aumentada. Es posible que ya no necesitemos equipo de efectos visuales extremadamente costoso para crear contenido de realidad

aumentada, lo que propiciará que las líneas entre la realidad física y la virtual se desdibujen aún más. Ejemplo de esto será el uso de técnicas de manipulación de video para alterar las expresiones faciales, los movimientos de los labios y el habla, con la desventaja potencial de permitir aún más la difusión de noticias falsas y propaganda. Por ello, es crucial pensar en las implicaciones que estas tecnologías tienen en la sociedad. Es necesario que estas experiencias tengan un diseño bien pensado, y presten atención a las lecciones del pasado.

Un sello distintivo del lugar de trabajo del futuro serán los empleados que trabajen codo con codo con la Inteligencia Artificial. Encontrar el éxito en este nuevo modelo requiere un enfoque holístico de la habilitación digital. El uso de las Tecnologías Inmersivas en este contexto podría mejorar potencialmente los entornos de trabajo y llevar a una transferencia de conocimiento mucho más rápida que elimine las restricciones de tiempo y ubicación. Para la industria, el uso de Tecnologías Inmersivas podría ayudar a reducir los costos operativos y así mantener su crecimiento e innovación.

Es bien sabido que el activo más importante con que cuenta una empresa es su capital humano, por lo que para las empresas es vital contar con herramientas para potenciar su desarrollo. Y aunque la tecnología no es el único elemento que define el lugar de trabajo del futuro, es un elemento crítico que permite a una organización atraer el talento adecuado, integrar datos en el proceso de toma de decisiones y servir como puente para unir espacios de trabajo físicos y digitales.

En particular, la transferencia de conocimiento representa un continuo desafío para las empresas y los propios empleados. Aunado a esto, la creciente evolución de los procesos, y la complejidad y el número de variantes en las tecnologías subyacentes complica el proceso de garantizar que la fuerza laboral esté permanentemente actualizada. El riesgo de perder conocimiento crucial que no se puede reemplazar con facilidad cuando los

empleados abandonan una organización (por ejemplo, con la jubilación de la fuerza laboral) es otro factor que debe considerarse. El conocimiento debe estar disponible de forma continua y distribuirse entre todos los actores de la fuerza laboral, así como en los soportes documentales.

Sin embargo, la adopción de estas tecnologías presenta diversos desafíos, entre estos se encuentran la creciente complejidad de los componentes tecnológicos automatizados y la aceptación de esta nueva tecnología por parte de los empleados. Por ejemplo, el personal de servicio en campo presenta mayor movilidad que otro tipo de trabajadores, ya que generalmente necesita viajar a diferentes lugares para tareas de mantenimiento e instalación. Esto representa un desafío considerable ya que aumenta los niveles de incertidumbre debido a diversos factores (por ejemplo, el medio ambiente o el comportamiento humano) que la tecnología no puede controlar. Otros factores como los cambios en los conocimientos y habilidades necesarios, los cambios demográficos que enfrentan las empresas cuando sus empleados expertos se retiran, incorporación de nuevo talento, factores externos como la movilidad laboral y migración o eventos globales (por ejemplo, la emergencia sanitaria por coronavirus) también están cambiando la naturaleza y la calidad del trabajo en el panorama global.

El trabajo colaborativo a distancia es otro de los retos en que las Tecnologías Inmersivas y la computación espacial podrían ser de ayuda. La mayoría de las organizaciones aceptan que la colaboración eficaz es esencial para un alto rendimiento. Sin embargo, situaciones como la emergencia sanitaria por COVID-19 forzó a millones de personas a trabajar desde casa, e hizo que el trabajo en equipo y colaborativo fuera más desafiante, ya que el trabajo remoto oscurece el lenguaje corporal y distorsiona las señales verbales que pueden ser cruciales para comprender la intención de los mensajes. Las videollamadas formales y programadas, o los mensajes instantáneos o mensajes de texto más frecuentes, no sustituyen los intercambios de información rápidos

y espontáneos. Un entorno compartido en 3D donde los empleados puedan ver representaciones de ellos mismos y de sus colegas (avatares) y realizar reuniones informales pueden ser útiles para sesiones interactivas de co-creación o ideación (Salako, Gardner y Callaghan 2020).

Si bien la tecnología puede tener impactos positivos, se deben tener en cuenta las amenazas que puede plantear. Más allá del mareo y las náuseas causadas por las limitaciones tecnológicas actuales, los efectos a largo plazo deben evaluarse a medida que las experiencias se vuelven cada vez más realistas, ya que podrían conducir a una adicción severa y, al mismo tiempo, hacer que las relaciones del mundo real sean menos atractivas. Por lo tanto, el constante avance en las tecnologías y su incorporación en todos los aspectos de nuestra vida ya sea laboral o personal “más que nunca requiere del aprendizaje de nuevas habilidades relacionadas a la cultura digital, para tener una vida online responsable y segura, y para convivir de manera armoniosa en la era digital” (Peña-Rios 2021). Por esto las sociedades deben promover el desarrollo de las “habilidades del siglo XXI” (Jenkins 2009), que son habilidades digitales necesarias para adaptarse a la vida actual, e incluyen el saber construir una identidad en línea, el uso balanceado de los medios digitales, poseer cierto conocimiento sobre seguridad informática, tener la capacidad de comunicarse a través de medios digitales, estar familiarizado con términos computacionales básicos, tener conocimiento sobre los derechos de los usuarios y la privacidad de datos, y poseer la habilidad para mantener un equilibrio emocional al interactuar y relacionarse con otros por medios digitales (Peña-Rios 2021). De igual manera, se debe promover un uso responsable de las tecnologías y del uso de los datos por parte de las empresas, así como respetar la privacidad de sus empleados y asegurar su bienestar físico y mental.

La transformación digital aún se encuentra en sus primeras etapas y la investigación interdisciplinaria es clave para identificar

beneficios y riesgos. En general, la promesa de utilizar nuevas tecnologías reside en la aplicación adecuada de la tecnología, no en la tecnología en sí. A medida que las Tecnologías Inmersivas se vuelven más populares debido a los avances en la tecnología y la reducción del costo de los dispositivos, la adopción de la computación espacial podría proporcionar un entorno seguro para mejorar las habilidades, así como ayudar a mantener y transformar el conocimiento. Este cambio combinado en los modelos de percepción e interacción podría conducir a la creación del *humano aumentado*, por tanto, es necesario abordar los desafíos tecnológicos, sociales y éticos con una visión interdisciplinaria y éticamente responsable.

### Agradecimientos

Agradezco a mis colegas de BT Research Labs y estudiantes visitantes por sus contribuciones a la prueba de conceptos aquí presentadas.

### Referencias

- BAHCECI, Ozkan Cem, Anasol Peña-Rios, Vikas Gupta, Anthony Conway y Gilbert Owusu (2021). *Work-in Progress-Using Immersive Virtual Reality in Field Service Telecom Engineers Training*. 7th International Conference of the Immersive Learning Research Network (iLRN 2021). Immersive Learning Research Network.
- BANAKOU, Domna, Alejandro Beacco, Solène Neyret, Marta Blasco-Oliver, Sofia Seinfeld y Mel Slater (2020). "Virtual body ownership and its consequences for implicit racial bias are dependent on social context". *Royal Society* (open Science) 7, n.º 12.
- BOLTON, Alexandra, Lorraine Butler, Ian Dabson, Mark Enzer, Matthew Evans, Tim Fenemore, Fergus Harradence, Emily Keaney, Anne Kemp, Alexandra Luck, Neill Pawsey, Steve Saville, Jennifer Schooling, Miranda Sharp, Tom Smith, Jeni Tension, Jennifer Whyte, Alan Wilson y Chara Makri (2018).

- “The Gemini Principles: Guiding values for the national digital twin and information management framework”. <https://doi.org/10.17863/CAM.32260>.
- BUDRIONIS, Andrius, Knut Magen Augestad, Hiten Patel y Johan-Gustav Bellika (2013). “An Evaluation Framework for Defining the Contributions of Telestration in Surgical Telementoring”. *Interactive Journal of Medical Research* 2, n.º 2: e14. <https://doi.org/10.2196/ijmr.2611>.
- DEDE, Chris (1995). “The evolution of constructivist learning environments: Immersion in distributed, virtual worlds”. *Educational Technology* 35, n.º 5: 46-52.
- EL SADDIK, Abdulmotaleb (2018). “Digital Twins: The Convergence of Multimedia Technologies”. *IEEE Multimedia* 25, n.º 2: 87-92.
- ELOR, Aviv y Sri Kurniawan (2020). “The ultimate display for physical rehabilitation: A bridging review on immersive virtual reality”. *Frontiers in Virtual Reality* 1, 585993.
- ERTMER, Peggy A. y Timothy J. Newby (2013). “Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective”, *Performance Improv.t Qtr* 26, n.º 2: 43-71, DOI:10.1002/PIQ.21143.
- GARTNER (s.f.). “Field Service Management”. Acceso 16 de marzo de 2020. <https://www.gartner.com/en/information-technology/glossary/field-service-management>.
- GREENWOLD, Simon (2003). “Spatial computing”. Tesis de maestría. Massachusetts Institute of Technology. <https://doi.org/10.1145/2756547>.
- IJSSELSTEIJN, Wijnand A. (2005). “History of Telepresence”. En *3D Videocommunication: Algorithms, Concepts and Real-Time Systems in Human Centred Communication*, editado por Oliver Schreer, Peter Kauff y Thomas Sikora, 5-21. John Wiley & Sons. <https://doi.org/10.1002/0470022736.ch1>.
- IMMERSE UK AND DIGITAL CATAPULT (2019). “The immersive economy in the UK 2019”. <https://www.immerseuk.org/wp-content/>

uploads/2019/11/The-Immersive-Economy-in-the-UK-Report-2019.pdf.

- IOANNOU, Androniki, Evridiki Papastavrou, Marios N. Avraamides y Andreas Charalambous (2020). "Virtual reality and symptoms management of anxiety, depression, fatigue, and pain: a systematic review". *SAGE open nursing* 6. <https://doi.org/10.1177/23779608209361>.
- JAMES, Laura K., Chien-Yu Lin, Anthony Steed, David Swapp y Mel Slater (2003). "Social anxiety in virtual environments: results of a pilot study". *Cyberpsychol Behav* 6, n.º 3: 237-243.
- JONASSEN, David H. (1991). "Evaluating constructivistic learning". *Ed. Tech.* 31, n.º 9: 28-33.
- JENKINS, Henry (2009). *Confronting the Challenges of Participatory Culture*. The MIT Press. <https://doi.org/10.7551/mitpress/8435.001.0001>.
- KAHN, William A. (1990). "Psychological conditions of personal engagement and disengagement at work". *Academy of Management Journal* 33, n.º 4: 692-724. <https://doi.org/10.5465/256287>.
- KISHORE, Sameer, Bernhard Spanlang, Guillermo Iruretagoyena, Shivashankar Halan, Dalila Szostak y Mel Slater (2022). "A Virtual Reality Embodiment Technique to Enhance Helping Behavior of Police Towards a Victim of Police Racial Aggression". *Presence - Virtual and Augmented Reality* 28, n.º 1: 5-27.
- KOLB, David (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood: Prentice-Hall.
- LATHAN, Corinna E. y Geoffrey Ling (2020). "Spatial Computing Could Be the Next Big Thing". *Scientific American*. <https://www.scientificamerican.com/article/spatial-computing-could-be-the-next-big-thing/>.
- MATAMALA-GOMEZ, Marta, Mel Slater y Maria V. Sanchez-Vives (2022). "Impact of virtual embodiment and exercises on functional ability and range of motion in orthopedic rehabilitation". *Scientific reports* 12, n.º 1.

- MILGRAM, Paul y Fumio Kishino (1994). “A taxonomy of mixed reality visual displays”. *IEICE Transactions on Information and Systems* 77, n.º 12: 1321-1329.
- NATIONAL RESEARCH FOUNDATION (NRF) (2018). “Virtual Singapore”. <https://www.nrf.gov.sg/programmes/virtual-singapore>.
- NEWS MEDICAL (2020). “Dotmatics collaborates to enable voice-assisted workflows”. <https://www.news-medical.net/news/20201112/Dotmatics-collaborates-with-LabVoice-to-enable-voice-assisted-laboratory-workflows.aspx>.
- PARROTT, Aaron y Lane Warshaw (2017). “Industry 4.0 and the Digital Twin”. *Deloitte Insights*. <https://dupress.deloitte.com/dup-us-en/focus/industry-4-0/digital-twin-technology-smart-factory.html>.
- PEÑA-RIOS, A. Alejandra (2021). *Las habilidades para la era digital (Ensayo)*. Posgrado en Artes y Diseño. Universidad Nacional Autónoma de México (UNAM).
- PEÑA-RIOS, Anasol, Hani Hagrass, Gilbert Owusu y Michael Gardner (2018). “Furthering Service 4.0: Harnessing Intelligent Immersive Environments and Systems”. *IEEE Systems, Man, and Cybernetics Magazine* 4, n.º 1: 20-31. <https://doi.org/10.1109/MSMC.2017.2769199>.
- PEÑA-RIOS, Anasol (2019). “The Augmented Employee: Immersive Technologies in Field Service Operations”. En *Workshop, Long and Short Paper, and Poster Proceedings from the Fifth Immersive Immersive Learning Research Network Conference*, editado por D. Beck, A. Peña-Rios, T. Ogle, D. Economou, M. Mentzelopoulos, L. Morgado, C. Eckhardt, J. Pirker, R. Koitz-Hristov, J. Richter, C. Gütl y M. Gardner, 3-4. Technischen Universität Graz.
- PEÑA-RIOS, Anasol, Tomas Oplatek, Hani Hagrass, Anthony Conway y Gilbert Owusu (2020a). “Work-in-Progress — Measuring Engagement in Virtual Reality for Talent Attraction Purposes”. En *6th International Conference of the Immersive Learning Research Network (iLRN 2020)*, editado por Daphne Economou, Alexander Klippel, Heather Dodds, Anasol Peña-Rios,

- Mark J. W. Lee, Dennis Beck, Johanna Pirker, Andreas Dengel, Tiago M. Peres y Jonathon Richter, 324-327. Immersive Learning Research Network.
- PEÑA-RIOS, Anasol, Ozkan Cem Bahceci, Vikas Gupta, Anthony Conway y Gilbert Owusu (2020b). “Work-in-Progress — A Web-based Virtual Reality Training Simulator for Field Service Telecommunications Engineers”. En *6th International Conference of the Immersive Learning Research Network (iLRN 2020)*, editado por Daphne Economou, Alexander Klippel, Heather Dodds, Anasol Peña-Rios, Mark J. W. Lee, Dennis Beck, Johanna Pirker, Andreas Dengel, Tiago M. Peres y Jonathon Richter, 316-319. Immersive Learning Research Network.
- PEÑA-RIOS, Anasol, Carla Di Cairano-Gilfedder, Louise Krug, Tim Glover, Anthony Conway y Gilbert Owusu (2022). “Uncovering the Power of Digital Twins for Telcos.” *The ITP Journal* 16, n.º 1: 38-44.
- PEÑA-RIOS, Anasol (2022). “Digital Twins, Immersive Technologies and the Workplace of the Future”. En *Adjunt Proceedings of the 1st Intl. Workshop on Analytics, Learning & Collaboration in eXtended Reality (XRWALC 2022)*, editado por Rene Kaiser, Anasol Peña-Rios, Heather Dodds, Johanna Pirker, Teresa Chambel, Christian Gütl y Krzysztof Pietroszek, 197-204. ACM.
- PWC (2019). “Seeing is believing”. <https://www.pwc.com/gx/en/technology/publications/assets/how-virtual-reality-and-augmented-reality.pdf>.
- REALWEAR (2020). “Volkswagen Rolls Out RealWear Augmented Reality Headsets in UK”. <https://realwear.com/blog/volkswagen-commercial-vehicles-rolls-out-realwear-augmented-reality-headsets-across-uk/>.
- RUTHENBECK, Greg y Karen J. Reynolds (2015). “Virtual reality for medical training: the state-of-the-art”. *Journal of Simulation* 9: 16-26.
- SALAKO, Oluwatimilehin, Michael Gardner y Vic Callaghan (2020). “Generating Virtual Worlds for Collaborative Innovation

- Activities: A Responsive Templating Approach”. En *EAI International Conference on Technology, Innovation, Entrepreneurship and Education*, editado por Ping Zheng, Victor Callaghan, David Crawford, Tiina Kymäläinen y Angelica Reyes-Munoz, 37-53. Springer International Publishing.
- SISSON, Patrick (2018). “Walmart adopts virtual reality to train its workforce for Black Friday”. *Vox*. <https://www.vox.com/the-goods/2018/11/15/18092456/walmart-virtual-reality-black-friday-vr>.
- SLATER, Mel y Sylvia Wilbur (1997). “A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments”. *Presence: Teleoperators and Virtual Environments* 6, n.º 6: 1-20. <http://discovery.ucl.ac.uk/79956/>.
- WANG, Minjuan, Vic Callaghan, Jodi Bernhardt, Kevin White y Anasol Peña-Rios (2017). “Augmented reality in education and training: pedagogical approaches and illustrative case studies”. *Journal of Ambient Intelligence and Humanized Computing*: 1-12. <https://doi.org/10.1007/s12652-017-0547-8>.
- WARWICK, Kevin (2014). “Human Enhancement—The Way Ahead: The way ahead: The technological singularity (Ubiquity symposium). *Ubiquity*, n.º October: 1-7.
- WEEKES, Sue (2018). “Time to don the headset?”. <https://www.recruiter.co.uk/opinion/2018/02/time-don-headset>.
- WILD, Fridolin (2016). “The future of learning at the workplace is augmented reality”. *Computer* 49, n.º 10: 96-98.

## **Sobre la autora**

### **Dra. Anasol Peña-Rios**

Doctora en Ciencias de la Computación con más de 20 años de experiencia en investigación y desarrollo en la academia y en la industria. Actualmente trabaja como gerente de investigación (*Research Manager*) en los laboratorios de investigación de British Telecom (BT) en Reino Unido, donde ha sido reconocida como Ingeniera Distinguida (*Distinguished Engineer*). Sus áreas de

especialización incluyen Gemelos Digitales, XR (realidad mixta, realidad aumentada, realidad virtual), Inteligencia Artificial, internet de las cosas, inteligencia ambiental, interfaces hombre-computadora y sistemas ciber físicos, sobre los cuales ha publicado extensamente y producido múltiples patentes. También está interesada en el trabajo colaborativo remoto, el aprendizaje mejorado con tecnología, la innovación y el proceso de generación de ideas.

Es académica visitante (Honorary Senior Lecturer) en la Escuela de Ciencias de la Computación e Ingeniería Electrónica dentro de la Universidad de Essex (Reino Unido). Es Miembro Senior de la IEEE, miembro del Consejo Directivo del Comité Técnico de Entornos de Aprendizaje Inmersivo de la IEEE Education Society y Editor Asociado de *IEEE Transactions on Learning Technologies* (TLT). Forma parte de la junta ejecutiva de dos organizaciones sin fines de lucro: Immersive Learning Research Network (iLRN) y Creative Science Foundation, donde contribuye a convertir sus objetivos estratégicos en recomendaciones prácticas. También forma parte del Grupo Asesor de la iniciativa Digital Twin Hub, apoyada por Innovate UK (la agencia de innovación del gobierno de Reino Unido). Es ganadora del premio “WeAreTechWomen’s TechWomen100” en noviembre de 2020 y finalista en los premios “Computing Women in IT Excellence Awards 2018”.

# Tecnología, capital y persona: entre Prometeo y Polifemo. Notas críticas desde el marxismo

ENRIQUE G. GALLEGOS

## Resumen

En la sociedad capitalista contemporánea ha podido promoverse la idea de que las tecnologías son benéficas, y que contribuyen a hacer la vida laboral, empresarial y cotidiana más eficaz, productiva, fácil, duradera y placentera. Sin embargo, lo que queda oculto en ese relato entusiasta, ilustrado y optimista son los efectos perniciosos que generan las tecnologías, particularmente porque están inscritas en la lógica de acumulación del capital. A este fenómeno el marxismo lo denomina como *fetichismo*. Sin desconocer su dimensión positiva, en este capítulo se hace un análisis, desde lo que Marx denominaba como la *historia crítica de la tecnología*, para ponderar los efectos irracionales de las tecnologías, desmitificando y alertando sobre las consecuencias y usos negativos y destructivos que le son inherentes por estar determinadas por la sociedad capitalista.

**Palabras clave:** filosofía de la tecnología, tecnología en el capitalismo, marxismo, Walter Benjamín, Hannah Arendt.

## Preliminar

“Según diversas tradiciones de la mitología griega, Prometeo robó el fuego y se lo entregó a los mortales; de la misma manera hurtó las artes de Hefesto para dárselas a éstos. Por ese motivo, Zeus le impuso un castigo terrible: estar encadenado a una montaña donde un águila le devoraba el hígado; pero como era inmortal, en la noche le volvía a crecer y en la mañana el águila se lo comía de nuevo” (Esquilo 2000). Como Prometeo, la humanidad se desenvuelve en una tecnología capitalista que de día le devora el corazón, mientras que de noche le vuelve a crecer con la esperanza de un mundo libre de explotación y miseria. De eso trata este capítulo: de la dialéctica de la tecnología, que construye y destruye.

### 1

La tecnología puede ser analizada desde diferentes perspectivas: a partir de sus propios desarrollos e innovaciones internas, desde la historia de sus procesos y productos o también revisada como una historia social en la que se trazan las relaciones entre sociedad y tecnología. *Grosso modo* podemos señalar que estas perspectivas parten del supuesto de que la tecnología contribuye a generar mejores condiciones en la vida, facilita las relaciones sociales y, por lo tanto, desde este punto de vista, resultaría claramente beneficiosa para la humanidad.

Esto aparentemente es corroborado por el sentido común; basta pasar la mirada a nuestro alrededor. Tenemos celulares que facilitan la comunicación; computadoras portátiles que ayudan en las tareas; internet para acceder a revistas y resultados de investigación; aplicaciones para aprender idiomas y comprar comidas, y drones para realizar entregas de productos, vigilar las fronteras y filmar eventos deportivos, etcétera. Pero no sólo el sentido común parece corroborar esa benignidad. Las diversas ciencias también la corroborarían, por ejemplo, gracias a las aplicaciones de la inteligencia artificial a la medicina se logran diagnósticos más efectivos y precisos que salvan la vida de las personas.

Entonces, a primera vista pareciera que la tecnología tiene, en efecto, el objetivo de hacer la vida más fácil, placentera, duradera y llevadera. El fondo de esta visión optimista y positiva de la tecnología descansa en la vieja idea de la Ilustración que sostenía que la razón y sus productos —las ciencias, las tecnologías, el Estado de Derecho, las instituciones, los procedimientos para normar la vida social, el mercado, la pedagogía—, si son adecuadamente usados y aplicados, finalmente llevarían a la modernización y liberación de la humanidad. Así, la emancipación de los hombres y mujeres no sería producto de las revoluciones y luchas de trabajadores y la tradición de los oprimidos (indígenas, mujeres, pobres, colonizados, obreros, etcétera), para usar una expresión de Walter Benjamin —revoluciones que pasaban por crueles, sangrientas, irracionales y costosas—, sino de la capacidad de la razón para ordenar y organizar la sociedad y aprovechar los resultados de la ciencia y la tecnología (Horkheimer y Adorno 2006). El hambre, los desastres, la explotación, la miseria, las opresiones, las enfermedades, las incomodidades, las desigualdades y todos los males que aquejan a la humanidad serían finalmente superados. La filósofa Hannah Arendt, aunque no dejaba también de alertar sobre los peligros por los usos bélicos de la tecnología, también era de esta opinión (Arendt 2006a, 27 y 29; 2006b, 151).

De entrada, planteadas las cosas de esta manera, debemos señalar que esa descripción optimista produce cierta extrañeza. Mediante una adecuada pócima de razón, ciencia y tecnología, cuando menos desde la Ilustración, se prometía el cielo, la dicha y prodigalidad. Por supuesto, no se niega que la computadora, en efecto, nos haya facilitado hacer las tareas escolares; que el internet haya simplificado las comunicaciones y la transmisión de bienes e información; que la Inteligencia Artificial medicinal otorgue rápidos y eficaces diagnósticos de enfermedades. Seríamos sumamente irracionales si lo negáramos, y se malinterpretaría lo que estoy tratando de plantear.

Pero preguntemos, ¿acaso no existen cientos de niños muriéndose de hambre en África?, ¿no tenemos miles de mujeres, niños y hombres sometidos a hacinamiento y maltrato en la frontera con Guatemala?, ¿no existen cientos de desempleados que padecen subalimentación y enfermedades de todo tipo?, ¿no se están gestando nuevas enfermedades asociadas a las tecnologías digitales?, ¿no se derriten los glaciares a causa de los efectos del uso de las tecnologías en las grandes industrias? Sin dejar de reconocer las aportaciones de las tecnologías, y si dejamos a un lado las visiones entusiastas e hiperpositivizadas de éstas, ¿acaso la tecnología no tiene también su lado oscuro y calamitoso?, ¿acaso no produce sus propios monstruos?, ¿acaso no padecemos la condena de Prometeo?

Estas preguntas son las que vuelven sospechosa una mirada a la tecnología desde su propia historia, desarrollos e innovaciones y la parcialidad de sus consecuencias. Como observaron Adorno y Horkheimer, la tecnología admite, en su misma racionalidad, usos y efectos irracionales (Horkheimer y Adorno 2006). No basta con describir las maravillas del producto tecnológico, sino que es necesario afinar la mirada para situarnos en lo que Marx denominaba como la “historia crítica de la tecnología”, es decir, un análisis que dé cuenta de los procesos de producción capitalista, de las relaciones sociales, de las relaciones con la naturaleza y de las representaciones que acontecen en esas relaciones (Marx 2001, 453).

De otra manera, si sólo la consideramos en su momento racional, la constituimos en otra modalidad del fetichismo y ocultamos de esa manera su momento irracional o destructivo. Debe entenderse que no estamos rechazando la tecnología *per se*, sino situándonos en un plano de *alerta crítica*. Por ello, para responder esas preguntas debemos dar algunos rodeos para emplazarnos en una tradición crítica que, si bien no se desatiende del potencial positivo de las tecnologías, también señala sus *efectos* y usos negativos y destructivos.

## 2

Desde diferentes tradiciones filosóficas se ha sostenido que la nuestra es la época de la técnica. Así lo plantearon dos pensadores fundamentales del siglo XX: Martín Heidegger y Ortega y Gasset. El primero sostenía que la técnica era uno de los fenómenos de acceso al ente “como decisión sobre la esencia de la verdad” (Heidegger 2006, 598); el segundo, por su parte, afirmaba que era una especie de “sobrenaturaleza” (Ortega y Gasset 2006, 598). Si bien ambos filósofos levantaron acta constitutiva de lo fundamental de la técnica, pasaban de alto la praxis desfetichizadora y las críticas a ésta que habían sido ganadas con el marxismo. Analicemos esto.

Formalmente se puede afirmar que dos productos tecnológicos, la escritura realizada por los monjes antes de la invención de la imprenta y la ejecutada a través del ordenador de finales del siglo XX, responden a idénticos principios: establecer la comunicación, facilitar su procesamiento y transmisión, poco importa que uno sea en pieles de cordero y el otro en formato digital. Para mi argumentación, dejó de lado su manifiesta diferencia tecnológica y su sentido. Si situamos esos productos tecnológicos en sus respectivas sociedades, se constata que su *estatuto óntico* es radicalmente heterogéneo. El estatuto de un producto difiere, sea que haya sido producido en una sociedad “tradicional” (precapitalista) o en una sociedad capitalista. Esta distinción, aparentemente general y abstracta, es la que nos permite problematizar algunas de las aporías y contradicciones de la tecnología que mencionamos más arriba.

En efecto, las sociedades tradicionales —anteriores al modo de producción capitalista— se caracterizaban, en cuanto a los productos tecnológicos, por enfatizar su valor de uso. Lo que destacaba de los objetos elaborados por la tecnología —cama, martillo, jícara, guitarra, mueble, etcétera— es que eran producidas para el *consumo*. No es que no existiera el intercambio, no es que no existiera una suerte de comercio, sino que en la medida en que

aún no se habían constituido las fuerzas sociales y materiales que darán vida a las sociedades modernas capitalistas —el trabajo asalariado y capital que se valoriza— se mantenían bajo el régimen del *valor de uso*. Sólo cuando en el transcurso de los siglos XV, XVI y XVII son “liberadas” las grandes masas de campesinos para constituirseles como trabajadores asalariados y se comienza a formar el capital, tenemos las condiciones para la aparición de lo que propiamente es una sociedad productora de mercancías (Marx 1978).

Es en ese periodo cuando los productos manufactureros y tecnológicos van a invertir su función: de ser *valores de uso*, se constituyen en *valores de cambio*. La posibilidad, pues, de que dos productos sean intercambiados en el mercado es que sean valores de cambio (Marx 2001, 45).<sup>1</sup> Por supuesto, no desaparece el valor de uso de los productos en los que se cristalizan los desarrollos e innovaciones tecnológicas: el celular es producido para usarse: comunicarse y conectarse a las redes sociales; pero para las sociedades capitalistas eso pasa a segundo término, lo que importa es que el ciclo económico se mantenga fluido (capital, producción, intercambio y consumo) para la obtención del plusvalor.

Esta modificación en el estatuto de los productos —que pasan de valores de uso a valor de cambio o mercancías— es lo que va a determinar la función de la tecnología en las sociedades capitalistas. Para decirlo de otra manera, la historia de la tecnología no puede dissociarse de la propia transformación de la sociedad capitalista. No es casualidad, como menciona el historiador social de la tecnología, Headrick, en su obra *Los instrumentos del imperio. Tecnología e imperialismo en el siglo XIX* (1998), que durante el imperialismo del siglo XIX y principios del XX, cuando se dieron los procesos brutales de acumulación del capital asociado con las colonias, también fueron momentos de apogeo de la tecnología. Por supuesto, no siempre existe esa relación *inmediata* entre

---

<sup>1</sup> Dejamos de lado en la argumentación la problemática de que ese valor de cambio descansa a su vez en el trabajo abstracto.

el desarrollo de la tecnología y la acumulación del capital, pues en no pocos ámbitos de investigación, innovación y aplicación científica y tecnológica —como en las universidades públicas, por ejemplo, que no persiguen la generación de plusvalor, aún cuando forman la fuerza de trabajo— parece que no responden de forma inmediata a las directrices del capital y, por lo tanto, se deben establecer mayores mediaciones para precisar en qué puntos acontece la subsunción real de la tecnología al capital en esos espacios “alejados” de la producción del plusvalor.

### 3

Si he sostenido que existe una suerte de aporía o contradicción entre las promesas liberadoras de la tecnología (y la ciencia) y el mundo real en el que vivimos, y que parte de las razones las encontramos en la función social y económica de aquellas en el modo de producción capitalista, ¿cuál es entonces el estatuto de la tecnología en las sociedades capitalistas? Para responder a esta pregunta debemos hacer algunas precisiones, pues, como indiqué al inicio, aquí no interesa analizar los desarrollos e innovaciones tecnológicas por sí mismas ni tampoco hacer la historia de sus transformaciones y productos, sino partir de una mirada crítica que tome en cuenta sus posibles usos y efectos sociales, culturales y políticos en una sociedad determinada: el capitalismo.

Si aceptamos que nuestro mundo moderno se caracteriza, entre otras cosas, por la existencia de mercados, flujo de bienes, productos y servicios, lo que se conoce como una sociedad capitalista —por más que el capitalismo haya mudado en algunos aspectos del siglo XIX al XXI—, entonces debemos de aceptar que se caracteriza por el conflicto entre capital y trabajo. Para decirlo de otra manera: el capitalismo es el modo social de producción cuyo último objetivo es la *valorización del capital en detrimento del trabajo*, incluso desvalorizándolo. Esto no quiere decir que no existan acciones altruistas, generosos filántropos y ministros que cumplan piadosamente con el Estado de Derecho, que no se encuentren

ámbitos más o menos al margen del mercado o que no sea posible pensar y practicar espacios no regidos totalmente por el valor de cambio. Lo que se quiere significar es que en la medida en que el objetivo o la causa final de las sociedades capitalistas es el plusvalor, se dan amplios y englobantes procesos de subsunción real al capital (Veraza 2008). Y de estas lógicas de subsunción no pueden estar al margen los productos y desarrollos de la tecnología.

Recordemos que parte de los resortes de la revolución industrial estuvieron en la necesidad de incrementar los márgenes de ganancia del capital. En algún momento del siglo XVIII la *directa* dependencia del capital de la mano de obra se volvió una limitante (por sus costos, intensidad y constante rebeldía del obrero) y entonces de esa contradicción entre capital y trabajo surgirá el periodo de la Gran Industria (Marx 2009, 467-468). En ese periodo se van a invertir los papeles: ya no será la máquina la que dependa de las habilidades del obrero, sino que éste se convertirá en un apéndice de las máquinas y la tecnología. A través de la producción en serie y la producción en línea, el fordismo intensificará el sometimiento de los trabajadores a las máquinas y tecnología para ampliar los márgenes de ganancia (Lara 2015). Entonces, en el modo de producción capitalista, la tecnología en gran medida está determinada por la capacidad que tiene para generar plusvalor en los diferentes sectores económicos —una suerte de ADN de las sociedades capitalistas—, independientemente de que nuestras sociedades neoliberales estén fuertemente financiarizadas y concentradas en economías de servicios e informacionales.

Esto no impide, insistamos, que no puedan existir centros de investigación tecnológica cuyos objetivos *inmediatos* no son el plusvalor (como las universidades públicas), pero la penetración social de los productos e invenciones tecnológicas está en gran medida determinada por el impacto que generan en las tasas de ganancias. Por ejemplo, sólo cuando se dio una relación aceptable para el capital entre costos/ganancia en la producción y venta

de celulares y fueron tecnológicamente accesibles a las personas, se masificó su consumo. Por supuesto, también debe existir una cultura material que posibilite su uso y extensión cotidiana. Digamos que si dejamos de lado la idea de que la tecnología facilita la vida y la vuelve más sencilla y placentera —y espero haya quedado claro que no estoy negando esto— encontraremos que el fondo de la implementación de la tecnología obedece a la lógica del mercado y las posibilidades de obtener mayores beneficios, de tal manera que los anuncios publicitarios de una vida moderna y cómoda apoyada en las tecnologías suelen encubrir que el resorte oculto es la cruda racionalidad mercantil de la ganancia.

Esta misma lógica del “frío interés” del capitalismo que subsume la tecnología en sus procesos de valorización del capital la encontramos en múltiples ámbitos y con consecuencias de largo alcance. Sobre esto, analizaría uno de ellos, a modo de breve y marginal ilustración, en el que se muestra la potencialidad destructiva de la tecnología (digo potencialidad porque, sin ser una necesidad, es algo que aguarda su momento para desplegar sus efectos). Se podría contraargumentar que estos efectos se pueden controlar mediante la vigilancia, códigos éticos y límites impuestos al desarrollo y uso de las tecnologías a través de acuerdos, comisiones o regulaciones del Estado o asociaciones nacionales e internacionales. También se podría insistir en cómo la Inteligencia Artificial médica a no pocas personas les ha salvado la vida y cómo grandes empresas usan tecnologías “amigables” con el ambiente y las personas.

No dudo que estas puntualizaciones puedan trazar parcialmente algún horizonte; sin embargo, tengo mis dudas sobre sus *efectos globales y generales*, pues si nos atenemos a lo que he llamado como el ADN del capitalismo —la valorización del capital—, es decir, a la forma en que la tecnología termina finalmente subsumida en el capital, no puede no estar condicionada a la generación de la ganancia y es éste el umbral que determina el uso o no de cierta tecnología y no tanto la ideas de “buena vida”,

“modernización”, “progreso”, “protección del medio ambiente”, “economía verde”, etcétera, más allá de las buenas intenciones y el voluntarismo inocente. Para decirlo de otra manera, si no se pone en perspectiva cómo opera la valorización del capital en nuestra sociedad, hay que desconfiar de una tecnología promovida bajo las ideas de “modernización”, “comodidad”, “calidad de vida”, “sustentabilidad”, etcétera. Si no se develan estos substratos del capital, la tecnología se vuelve una manifestación fetichizada de las sociedades capitalistas.

Este ámbito en el que ejemplifico mi análisis se relaciona con la tecnología aplicada a los procesos de automatización, robotización y aplicación de la inteligencia artificial en los diferentes sectores de la economía. Empresas, comercios, librerías, bancos, entidades de gobierno están sometiendo sus procesos, servicios y productos a las automatizaciones y robotizaciones. Por ejemplo, una entidad bancaria mexicana —pero que forma parte de las multinacionales— comenzó a implementar una aplicación (*app*) para hacer operaciones bancarias mediante el celular y en 2019 anunció un programa de modernización de sus procedimientos, cuyo eje es la digitalización de sus servicios (Fariza 2019); todo lo cual fue anunciado —en un lenguaje edulcorado, optimista y modernizante— como medidas para hacer que los clientes se sientan más cómodos y seguros en sus operaciones, con lo que también refirió, como nota menor y casi a pie de página, que tendrán que despedir a 1 500 empleados.

Usualmente estas medidas de digitalización, virtualización, puesta en línea, automatización, robotización o implementación de Inteligencia Artificial son presentadas de esa manera: vanguardismo, modernidad, seguridad, comodidad y toda una suerte de adjetivos que pretenden instalarnos en el futuro o el paraíso terrenal. Pero lo que se oculta —y de aquí el carácter fetichizado de la tecnología— son dos cosas: que el fondo de esas medidas suele descansar en la cruda, calculada y fría pretensión de ampliar el margen de ganancias y que es una medida insensible y

puede terminar siendo cruel e inhumana: el despido de trabajadores y empleados, que son lanzados al desempleo, la incertidumbre, pobreza o muerte (porque el capitalismo es un régimen de muerte, como veremos más adelante).

Estos usos y efectos sociales de la automatización, robotización y aplicación de la Inteligencia Artificial en los diferentes sectores de la economía deben ser encuadrados en un horizonte más general del desarrollo de la sociedad capitalista, por lo que señalo a continuación algunos datos adicionales para luego proseguir con mi argumentación.

En 2019, la Organización Internacional del Trabajo (OIT), a través de la Comisión Mundial sobre el Futuro del Trabajo, rindió su informe, *Trabajar para un futuro más prometedor*, en el que recupera diversos estudios que resaltan, entre otras cosas, que “47 por ciento de los trabajadores de los Estados Unidos corren el riesgo de verse sustituidos en sus puestos de trabajo por la automatización”; mientras que “56 por ciento de los puestos de trabajo corren el riesgo de automatizarse en los próximos veinte años”. En ese mismo informe se refiere que el Banco Mundial estima que “dos tercios de los puestos de trabajo de los países en desarrollo podrían ser automatizados”. Asimismo, se menciona que para el Foro Económico Mundial “casi el 50 por ciento de las empresas esperan que la automatización lleve a una reducción de su fuerza de trabajo a tiempo completo en 2022” (OIT 2019, 19).<sup>2</sup>

Frente a este potencial escenario de destrucción del trabajo y del que dependen miles de personas, la población mundial no ha dejado de incrementarse. Según cifras del Banco Mundial, en 1980 había 4 431 mil millones de personas; en 1990, 6 115 mil millones y para 2015, 7 341 mil millones, es decir, en 35 años la población ha aumentado casi 70%. A esto hay que sumar la

---

<sup>2</sup> Téngase en cuenta que este informe es previo a la pandemia que aquejó al mundo entre 2019 y 2022. Y uno de los efectos de la pandemia fue precisamente intensificar y ampliar la robotización, automatización y digitalización del mundo social y económico, incluido el cotidiano y familiar.

desigual distribución de la riqueza; de acuerdo a Oxfam, “el 1% más rico de la población mundial posee más riqueza que el 99% restante de las personas del planeta”. Según este estudio, “la riqueza en manos de las 62 personas más ricas del mundo se ha incrementado en un 44% en apenas cinco años, algo más de medio billón de dólares (542.000 millones) desde 2010, hasta alcanzar 1,76 billones de dólares”, mientras que “la riqueza en manos de la mitad más pobre de la población se redujo en más de un billón de dólares en el mismo periodo, un desplome del 41%” (Oxfam 2016).

Por el mismo periodo, diferentes periódicos daban cuenta de que se habían ahogado decenas de migrantes en su intento de cruzar el Mediterráneo. Según el Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUR 2016a, 2016b), en 2015 “más de 2.000 personas al día arriesgan sus vidas y las de sus hijos cruzando el Mediterráneo” y en 2016 se habían ahogado 5 000 personas (casi 14 personas cada día) en su intento de llegar a las costas de Europa.

La guerra, el desempleo, la crisis climática, la pobreza y el hambre hacen que las personas migren a los llamados países desarrollados para buscar mejores condiciones de vida. Noticias que dan cuenta de hechos similares no han parado, incluidas aquellas que relatan las migraciones de miles de personas en su intento por cruzar México y llegar a Estados Unidos. Pero a veces no hace falta tratar de llegar por mar a un país para encontrar la muerte o el asesinato. Según cifras conservadoras, en México existían más de 250 mil asesinatos en el marco de la llamada guerra al narcotráfico y más de 40 mil desaparecidos desde 2006. También el país se ha constituido en un lugar donde el feminicidio está a la orden del día. Según algunas fuentes, se cometen entre 7 y 10 por día.<sup>3</sup>

Automatizaciones, robotización, aplicación de Inteligencia Artificial, aumento de la población, migraciones, muerte, asesinato,

---

<sup>3</sup> Estas cifras corresponden al momento de escribir este artículo.

feminicidio, desigual distribución de la riqueza, inseguridad, a lo que habría que agregar la radical crisis medioambiental, ¿qué relación guardan estos datos aparentemente inconexos? Si las automatizaciones y robotizaciones en los sectores de la economía se han incrementado y muestran una tendencia a asumir mayores espacios sociales y económicos, y si la población ha aumentado, si se prolonga la injusta distribución de la riqueza social y continúa la inseguridad y muerte, entonces estamos frente a un problema estructural del sistema capitalista al que no le preocupan las personas y que funciona con la figura de la “personas descartable”, justamente porque si el ADN del capitalismo es la valorización del capital desvalorizando al trabajador, implica registros inimaginables de desprecio a las personas, a la vida, al medioambiente y a las mínimas condiciones de reproducción de una vida digna.

Para comprender lo anterior es necesario insistir en que las sociedades capitalistas se definen por el conflicto entre capital y trabajo. En el modo social de reproducción capitalista no hay forma de que las mayorías no terminen por ser trabajadores, empleados u oficinistas, es decir, asalariados que deben vender su fuerza de trabajo para reproducir su vida, incluidos aquellos que se consideran freelancers o aprendices del emprendurismo. Ciertamente existe toda una superestructura que sostiene y defiende la existencia de los derechos, los procedimientos jurídicos y la idea de que las personas tienen derechos humanos “inherentes a todos los seres humanos” y, por lo tanto, que deben ser objeto de respeto y protección. Al seguir esos argumentos también se exige que las implementaciones tecnológicas respeten esos derechos, a las personas y al medio ambiente, pero todas estas ideas “humanistas” operan en determinadas condiciones sociales y económicas de vida que —en nuestro mundo, cuando menos desde el siglo XVI hasta este año 2019—, están encuadradas por el conflicto entre capital y trabajo, es decir, por una sociedad productora de mercancías, cuyo fin es la acumulación del capital.

Se debe comprender que con lo anterior no se están desechando esas ideas humanistas ni desconociendo sus contribuciones, sino situando sus limitaciones y determinaciones. Para decirlo peregrinamente: en el capitalismo si no trabajas no comes y si trabajas comes mal o te subalimentas con comida chatarra.

Lo que importa destacar para mi argumentación son tres cosas: 1) en las sociedades productoras de mercancías, las personas se constituyen en fuerza de trabajo y es en esa condición que son mediadas en sus relaciones económicas y laborales, por lo tanto, como cualquier insumo productivo o comercial, se constituyen en descartables; 2) el uso o no de las tecnologías en los sectores de la economía está en relación directa con la generación de plusvalor, y 3) si en esa ecuación se determina que existen mayores márgenes de ganancia para aplicar automatizaciones, robotización o Inteligencia Artificial, entonces no importará destruir el trabajo como fuente de ingreso de las personas, sencillamente porque la fuerza de trabajo necesaria para generar los productos, bienes o servicios, la puede realizar una máquina o un aplicación tecnológica, y en la medida en que las personas son reducidas a “fuerza de trabajo” se vuelven intercambiables y, en última instancia, en desechables.

Por la reducción de las personas a fuerza de trabajo (y finalmente su constitución en descartables), es por lo que el neoliberalismo, como fase actual capitalista de suma agresividad, se transforma de biopolítico en necropolítico, en el que ya no se gestiona la vida como en la biopolítica, sino la muerte (Foucault 2002).<sup>4</sup> ¿Qué va a ser de este mundo si sigue incrementándose la población, si se cumplen las estimaciones de las automatizaciones y robotizaciones y si el medio ambiente continúa su curso de crisis ecológica con las implicaciones que de ello se siguen (menos bosques, agua, recursos naturales, etcétera)?

---

<sup>4</sup> El uso académico del término “necropolítica” se suele adjudicar a Mbembe, pero como idea ya está en Marx en sus análisis del capitalismo, pues el capitalismo no puede sino producir muerte, o más bien, reproducirse produciendo muerte.

En alguna parte de *El capital*, Marx realiza una analogía religiosa: recuerda que Adán fue expulsado del paraíso por morder la manzana del conocimiento del bien y del mal. Con ello sugiere que el paraíso eran las sociedades tradicionales previas al capitalismo en las que prevalecía el valor de uso, mientras que su expulsión y caída era el paso de éste al valor de cambio, es decir, para Marx lo que verdaderamente era el infierno, es el capitalismo.<sup>5</sup>

Por supuesto, las tecnologías también tienen otros efectos y usos, igualmente nocivos, que dejé de lado. Desde la tecnología militar para exterminar a las personas o pueblos, hasta la vigilancia y control mediante la biometría, drones, satélites y web. Sobre este último punto, uno de los casos más paradigmático para nuestra época está en las grandes empresas globales de información como Google, que “controla el 69% del mercado mundial de motores de búsqueda de Internet” (Oxfam 2016), con lo cual puede definir qué buscar y encontrar, qué información priorizar y en qué orden; por no mencionar las formas en que se pueden usar las redes sociales (Facebook, Twitter, Instagram, etcétera) para espiar, controlar y manipular la opinión pública.

#### 4

En 1937 un filósofo alemán, Walter Benjamin, consideraba que las tecnologías eran *medios* y, en tanto que tales, podían servir a fines conservadores o a fines revolucionarios (Benjamin 2008). Tanto el fascismo como los movimientos revolucionarios, por ejemplo, aprovecharon las tecnologías aplicadas al cine para apuntalar sus campos de lucha, pero el filósofo pasó por alto que el umbral de las tecnologías está determinado por las condiciones materiales de producción de la vida, o sea, en las sociedades capitalistas, por la valorización del capital y la desvalorización de las personas.

---

<sup>5</sup> Por supuesto, esta comparación debe entenderse no en el sentido de que Marx esté haciendo apología de las sociedades tradicionales (que también albergaban sus propias opresiones y desastres) o de que creyera que fueran mejores, sino como una comparación para resaltar el cambio del valor de uso al valor de cambio y sus implicaciones.

Y en tanto no comprendamos lo radical de este modo de producción, que modifica las mismas raíces de la existencia humana sin importar cualquier otra consideración (vida, medioambiente, personas, etcétera), en tanto no tracemos programas para salir del capitalismo, el ADN del capital informará —subsumirá— el uso de las tecnologías a los fines del plusvalor y, con ello, potencialmente instalará la posibilidad de la destrucción de la vida.

Esta es la irracionalidad implícita en la racionalidad de la tecnología. Por supuesto, de cuando en cuando, podremos enterarnos de cómo la Inteligencia Artificial médica salvó la vida de las personas que tuvieron los medios económicos para pagarla, pero esa información oculta el otro rostro: las potencias destructivas de la tecnología al ponerse al servicio del monstruo del capital que, como Polifemo —aquel que devoraba a los compañeros de Ulises, el mítico héroe griego—, no tiene llenadera y está sediento de más sangre humana.

## Referencias

- ACNUR (2016a). “Una media de más de 2.000 personas cruzan el Mediterráneo cada día”. Acceso el 12 de octubre del 2019. <https://eacnur.org/es/actualidad/noticias/emergencias/una-media-de-mas-de-2000-personas-cruzan-el-mediterraneo-cada-dia>.
- \_\_\_\_ (2016b). “2016, récord de muertes en el mar Mediterráneo”. Acceso el 12 de octubre del 2019. <https://eacnur.org/es/actualidad/noticias/emergencias/2016-record-de-muertes-en-el-mar-mediterraneo>.
- ARENDETT, Hannah (2006a). *Sobre la violencia*. Madrid: Alianza.
- \_\_\_\_ (2006b). *Sobre la revolución*. Madrid: Alianza.
- BANCO MUNDIAL (2019). “Población, total”. Acceso el 12 de octubre del 2019. <https://datos.bancomundial.org/indicador/SP.POP.TOTL>.
- BENJAMIN, Walter (2008). “La obra de arte en la época de su reproducibilidad técnica”. *En Obras libro I, vol. 2*, 51-85. Madrid: Abada.
- ESQUILO (2000). *Tragedias*. Barcelona: Biblioteca Básica Gredos.

- FARIZA, Ignacio (2019). “BBVA despide a 1.500 trabajadores en México en plena transición digital”. *El País*, 12 de septiembre. Acceso el 12 de octubre 2019. [https://elpais.com/economia/2018/09/12/actualidad/1536768059\\_026347.html](https://elpais.com/economia/2018/09/12/actualidad/1536768059_026347.html).
- FOUCAULT, Michel (2002). *Defender la sociedad: curso en el College de France, 1975-1976*. México: Fondo de Cultura Económica.
- HEADRICK, Daniel (1998). *Los instrumentos del imperio: tecnologías e imperialismo europeo en el siglo XIX*. Barcelona: Altaya.
- HEIDEGGER, Martin (2006). *Ser y tiempo*. Barcelona: Trotta.
- HORKHEIMER, Max y Theodor Adorno (2006). *Dialéctica de la Ilustración. Fragmentos filosóficos*. Madrid: Trotta.
- LARA SANCHEZ, Miguel Ángel (2015). *Del fordismo a la automatización del trabajo mental*. México: UNAM-FES Aragón.
- MARX, Karl (1978). *Elementos fundamentales para la crítica de la economía política (Grundrisse). 1857-1858*. México: Siglo XXI.
- (2001). *El Capital. Crítica de la economía política. Tomo I/ Vol. 1. Libro primero. El proceso de producción del capital*. México: Siglo XXI.
- (2009). *El Capital. Tomo I/Vol. 2. Crítica de la economía política. Libro primero. El proceso de producción del capital*. México: Siglo XXI.
- ORGANIZACIÓN INTERNACIONAL DEL TRABAJO, COMISIÓN MUNDIAL SOBRE EL FUTURO DEL TRABAJO (2019). *Trabajar para un futuro más prometedor*. Bruselas: OIT.
- ORTEGA Y GASSET, José (2006). *Obra completas, Tomo V (1932-1940)*. Madrid: Fundación Ortega y Gasset/Taurus.
- OXFAM (2016). “Informe: Una economía al servicio del 1%. Acabar con los privilegios y la concentración de poder para frenar la desigualdad extrema”. [https://www-cdn.oxfam.org/s3fs-public/file\\_attachments/bp210-economy-one-percent-tax-havens-180116-es\\_0.pdf](https://www-cdn.oxfam.org/s3fs-public/file_attachments/bp210-economy-one-percent-tax-havens-180116-es_0.pdf).
- SCHNEIDER, Uri (2019). “Paraíso o robocalipsis”. Documental de DW. Acceso el 9 de octubre del 2019. <https://www.youtube.com/watch?v=sHVwwriaT6k>.

VERAZA, Jorge (2008). *Subsunción real del consumo al capital. Dominación fisiológica y psicológica en la sociedad contemporánea*. México: Ítaca.

## **Sobre el autor**

### **Enrique G. Gallegos**

Es poeta, filósofo, crítico literario y militante sindical. Trabaja como profesor-investigador adscrito al Departamento de Ciencias Sociales de la División de Ciencias Sociales y Humanidades de la Universidad Autónoma Metropolitana, Unidad Cuajimalpa. Es miembro del Sistema Nacional de Investigadores. Sus líneas de investigación son temas y autores de la filosofía política, teorías críticas, estética (particularmente poética). Es autor de varios libros, capítulos y artículos de investigación; su último libro es *Sartre y la filosofía de la subjetivación*. Ha coordinado diversos libros con otros investigadores, entre los que destacan *Ráfagas de dirección múltiple*, *Abordajes de Walter Benjamin*, y *Das Kapital. Marx, actualidad y crítica*. Recientemente se publicó su antología de poemas y aforismos para celebrar sus 25 años de producción poética: *Obra mínima. Poemas y aforismos* (2021).

# La evaluación tecnológica como un vínculo entre la política y la tecnología

CAMILO BECERRIL RAMÍREZ

## Resumen

Como objetivo principal, intentamos mostrar que la evaluación tecnológica puede funcionar como un vínculo que conecte la actividad política o democrática con el fenómeno de la tecnología. Para llevar acabo esto, exponemos el concepto de tecnología que concibe a ésta como un sistema de acciones intencionales. Por otro lado, hacemos referencia a algunas posibles relaciones entre la tecnología y la actividad política. Por último, a grandes rasgos, exponemos un enfoque de la noción de evaluación externa de la tecnología.

**Palabras clave:** evaluación tecnológica, política, filosofía de la tecnología, León Olivé, Miguel A. Quintanilla.

## Introducción

Una encuesta realizada en 2019, en Europa, por The Center for the Governance of Change (2019), revela que “25% de los europeos está parcial o totalmente a favor de permitir que una Inteligencia

Artificial (IA) tome decisiones importantes sobre el funcionamiento de su país”, es decir, decisiones sobre asuntos de interés público o político. Más allá de tomar los datos de esta encuesta como un síntoma general en los regímenes democráticos, y más allá de que sea o no posible, factualmente, que un sistema de IA tome decisiones importantes en los asuntos públicos, ¿qué reflexiones podríamos realizar, a partir de dicho ejercicio, acerca de la relación entre los usuarios de tecnologías y la actividad política en la actualidad?

Ya es bien sabido que la tecnología actual ejerce un poderoso influjo no sólo en la manera como pretendemos obtener conocimiento de las diversas dimensiones de nuestra realidad, sino también en la forma como definimos esta misma realidad y hasta en la manera en que actuamos políticamente, como es evidente. Por lo tanto, está claro que si la tecnología tiene tanto peso en nuestras condiciones vitales presentes sería deseable evitar eso que podríamos nombrar como la *aceptación irreflexiva* de tecnologías por parte de los usuarios comunes.

Si un grupo de personas o ciudadanos, en un contexto específico, estuviera de acuerdo en permitir que una IA realizara decisiones importantes en el ámbito político, sería deseable que esas personas participaran en el proceso de evaluación de dicha decisión tecnológica. Es decir, que exigieran a las comunidades científicas y tecnológicas información clara y asequible acerca de la naturaleza o cualidades del funcionamiento de dicho desarrollo tecnológico.

El objetivo principal de este texto es mostrar que la evaluación tecnológica externa puede ser considerada como un medio que vincula la actividad política o democrática con el fenómeno de la tecnología. En este sentido, primero expondremos un concepto filosófico de la tecnología, que la concibe como un sistema de acciones intencionales y no como algo que sea susceptible de ser valorado neutralmente. Después, haremos referencia a algunas posibles relaciones entre la tecnología y la actividad política. Por

último, a grandes rasgos, presentaremos un enfoque de la noción de evaluación externa de la tecnología.

### **El concepto de tecnología**

El concepto de tecnología que aquí utilizaremos es el que propone y desarrolla Miguel Ángel Quintanilla en su libro *Tecnología: Un enfoque filosófico y otros ensayos de la filosofía de la tecnología* (2005). El autor formula un concepto genérico de técnica. En el marco de esta caracterización afirma que una técnica es una entidad de naturaleza abstracta, y su “aplicación o realización concreta” es un sistema técnico concreto, es decir, “la técnica en cuanto tal sería una entidad abstracta que podríamos definir como el conjunto de todas las realizaciones técnicas concretas posibles” (2005, 46).

Una realización o sistema técnico concreto es un conjunto de acciones orientadas de manera intencional a la transformación de objetos concretos con la finalidad de obtener resultados valiosos de una manera eficiente (Quintanilla 2005, 64). De tal modo que este concepto genérico de técnica implica no sólo la noción de un sistema de acciones intencionales, sino también el conjunto de resultados de esas acciones —que son los artefactos— y el “sistema de reglas operacionales o de conocimientos teóricos y prácticos que son necesarios para llevar a cabo las acciones” (Quintanilla 2005, 51-52).

Esta noción de “acción intencional”, como lo señala León Olivé (1998), nos permite tener una visión amplia de la tecnología, que no reduce el concepto de tecnología a “un conjunto de aparatos, los cuales son neutrales respecto a los fines que se persiguen” (1998, 149). Por lo tanto, esta concepción permite refutar la “neutralidad valorativa” de la tecnología en determinadas situaciones (1998, 148).

En este sentido, la tecnología no puede ser pensada únicamente como una herramienta, un medio o un conjunto de artefactos, sino también como las acciones de individuos que son capaces de

formular creencias o ideas y representaciones acerca de la realidad que desean modificar o intervenir; individuos que son capaces también de realizar valorizaciones acerca de determinados “estados, acontecimientos o cosas”, y de desear que se lleven a cabo los que ellos creen que son más valiosos (Quintanilla 2005, 78). Por lo tanto, las personas, sus fines, creencias y conocimientos son constitutivos de las acciones intencionales que, a su vez, conforman los sistemas tecnológicos concretos.

Dado que la configuración o aplicación de un sistema tecnológico requiere que las personas que intervienen en ellos posean un sistema de “conocimientos representacionales”, aparte de los operacionales (Quintanilla 2005, 57), se podría decir que los individuos encargados de alguna innovación o configuración tecnológica tienen un conocimiento representacional sobre las características de los objetos que desean transformar, es decir, en cierta medida saben que “tal y cual cosa tiene, en tal y cual momento, tal propiedad” (Quintanilla 2005, 54). También tienen estipulados los posibles resultados que quieren obtener o que, de hecho, se podrían obtener.

Así pues, los individuos encargados de la configuración, operación o decisión de tal o cual tecnología, en algunas ocasiones saben o suponen que hay indicios razonables de que la operación de determinadas tecnologías podría provocar consecuencias catastróficas o violentas en el ambiente o en las personas, y sin embargo realizan su propagación en la realidad social. Miguel Á. Quintanilla (2005) nos informa también que, gracias al desarrollo de los sistemas tecnológicos y del conocimiento científico, podemos darnos cuenta de que las consecuencias de la operación tecnológica son múltiples, es decir, no son siempre calculables. Por lo tanto, son “sistemas en desarrollo”: “nunca están completos, o dicho de otra manera, nunca es posible controlar completamente todas las variables que intervienen en el sistema” (2005, 29). En este sentido, también hay tecnologías cuya operación provoca efectos violentos “como consecuencias no intencionales de la

operación de sistemas técnicos“ (Olivé 1998, 153), es decir, hay tecnologías cuya operación genera efectos destructivos o aniquiladores sin la intención de haberlos provocado, y esto ocurre debido a la imposibilidad de controlar o prever totalmente las variables o consecuencias que pudieran emerger a causa del desarrollo u operación de un determinado sistema tecnológico. De tal modo que es necesario disponer de “criterios y métodos efectivos” (Quintanilla 2005, 140) para efectuar la evaluación tecnológica.

### **Algunos enfoques sobre la relación tecnología-política**

Miguel Ángel Quintanilla identifica tres diversos modos en los que puede estar relacionada la tecnología con la política. Uno de ellos tiene que ver con lo que él llama “la tecnología como marco condicionante de la acción política” (Quintanilla 2002, 638). En esta dimensión están incluidas todas aquellas reflexiones que establecen que las nuevas tecnologías de la información y las comunicaciones (TIC) podrían contribuir al desarrollo de una democracia electrónica, de un ágora virtual y, por lo tanto, a la existencia de una “ciudadanía digital” (Robles 2009, 11). Todo esto se encuentra vinculado, pues, con el aumento de la participación ciudadana en los asuntos de interés público por medio del espacio virtual, el cual podría permitir tener un tipo de relación comunicativa entre los representantes políticos y los ciudadanos.

En este sentido, en Nueva Zelanda crearon un nuevo tipo de tecnología que podría fungir como marco condicionante de la acción política: se trata de Sam, el llamado “primer político virtual” desarrollado con IA. Ésta es una primera experimentación que quizá pueda ser el brote que anuncie un inédito modo de intentar influir en las decisiones de los asuntos públicos.

Sam, el político artificial, es uno de los llamados bots de conversación o chatbots. La definición clásica de un chatbot establece que es “un programa de computadora que procesa la entrada en lenguaje natural de un usuario y genera respuestas inteligentes y relativas que luego se envían de vuelta al usuario”. En la

actualidad, los chatbots funcionan con “motores de IA que interactúan principalmente con los usuarios a través de una interfaz basada en texto” (Khan y Das 2018, 1).

Uno de los objetivos de Sam es “aprender y representar las opiniones políticas de los neozelandeses” (Ministerio de Gobierno Digital de Nueva Zelanda 2020). Este conocimiento se irá incrementando, según sus creadores, a medida que vaya interactuando cada vez más con las personas. Incluso pretenden que participe, de algún modo, en las elecciones generales de 2020 en Nueva Zelanda.<sup>1</sup>

Sam puede actuar y hablar como una persona. También es capaz, entre otras cosas, de responder a preguntas sobre política y razonar sobre políticas de manera inteligente. “Sam utiliza técnicas de inferencia del lenguaje natural para comprender realmente el sentimiento de la opinión escrita o hablada de una política” (Ministerio de Gobierno Digital de Nueva Zelanda 2020). Además tiene acceso a diversas bases de datos para construir su capital cultural sobre los asuntos políticos, y el listado de sus características se podría extender aún más.

La existencia de este fenómeno tecnológico podría sugerir algunas cuestiones: ¿podríamos llamar a este tipo de relación entre el chatbot y los ciudadanos una genuina interacción política?, ¿puede una interfaz conversacional con IA ser un genuino representante político? Este fenómeno, evidentemente, puede generar mucha controversia en el ámbito de la teoría política.

Un segundo tipo de relación entre tecnología y política es lo que Quintanilla designa como “la tecnología como instrumento”. En esta relación se concibe a la tecnología como un medio importante sin el cual determinados fines de políticas públicas en un sistema democrático no se podrían llevar cabo (Quintanilla 2002, 640). La tercera variante de la relación tecnología política es lo que nombra como “la tecnología como objetivo de la acción

---

<sup>1</sup> Este artículo describe información actualizada a octubre de 2019.

política”; esta relación tiene que ver con el hecho de que es necesario reivindicar lo que el autor llama “la democracia tecnológica plena”, es decir, que es necesario establecer mecanismos adecuados y efectivos en donde los ciudadanos puedan participar plenamente en la toma de decisiones sobre el desarrollo tecnológico (Quintanilla 2002, 644). En este sentido, el poder político puede alentar el desarrollo tecnológico para obtener beneficios económicos y sociales que convengan a los ciudadanos (Quintanilla 2002, 644).

### **La evaluación externa de la tecnología**

Como ya hemos mencionado, en el diseño de una tecnología intervienen diversos fines establecidos por un grupo social específico, y los efectos de su aplicación pueden ser múltiples y nunca previsible totalmente, puesto que toda tecnología es un “sistema en desarrollo”. Por lo tanto, estos fenómenos hacen necesario el papel de la evaluación de una tecnología.

Un proyecto tecnológico se puede evaluar de dos maneras: interna y externamente, y ambas contribuyen al desarrollo tecnológico. La evaluación interna tiene que ver con la consideración de factores de eficiencia, factibilidad, eficacia y fiabilidad (Olivé 2000, 97). Este tipo de evaluación está vinculado con el trabajo especializado de ingenieros o tecnólogos, por ejemplo.

Por otro lado, la evaluación externa de una tecnología está relacionada con la utilidad o el valor que tal o cual tecnología tienen para la sociedad en su totalidad (Quintanilla 2005, 126). Dado que una tecnología se desarrolla en un contexto específico, este último es alterado por los efectos que genera la aplicación de la misma. Este tipo de evaluación se hace necesaria en un orden democrático, ya que si la aplicación de una tecnología puede afectar a cierto grupo de ciudadanos, éste tiene derecho a participar en los procesos de evaluación y a decidir sobre la posible implementación de tal o cual tecnología: “El desarrollo tecnológico puede favorecer la realización de intereses generales auténticos de

las comunidades humanas, por ejemplo, obtener energía limpia para una población, o puede ir en contra de ellos y promover intereses particulares de sólo ciertos grupos” (Olivé 2000, 99).

Así pues, en un ámbito democrático es necesaria la evaluación externa de la tecnología en donde participen activamente la mayoría de los ciudadanos porque las tecnologías actuales afectan a gran parte de la sociedad, y “sobre todo a las posibilidades futuras de desarrollo económico, social y cultural de la humanidad” (Quintanilla 2005, 140). También porque el cambio tecnológico es vertiginoso y es de vital importancia prever todos los posibles efectos que pudiera causar sobre las sociedades futuras, y porque el desarrollo tecnológico está subordinado a decisiones humanas que pueden orientar el desarrollo tecnológico “en múltiples direcciones, de acuerdo con nuestros intereses, o en contra de ellos” (Quintanilla 2005, 140).

Dicha evaluación externa, dice Quintanilla, posee dos variantes: una tiene que ver con los criterios de idoneidad y la otra con las consecuencias que su aplicación pueda generar. La evaluación externa de las consecuencias es la que nos importa exponer. En este tipo de evaluación se pueden distinguir tres criterios: los riesgos, el impacto ambiental y el impacto social (Quintanilla 2005, 148). El riesgo como criterio de evaluación externa tiene que ver con la previsión de las posibles consecuencias funestas que la aplicación de una tecnología pueda provocar en “la vida humana, la salud o el bienestar de la población potencialmente afectada” (Quintanilla 2005, 148).

El criterio relacionado con el impacto ambiental tiene que ver con las posibles consecuencias que una aplicación tecnológica pueda provocar en el entorno físico en que se desarrolla, “desde variables físicas (geológicas, químicas, biológicas, atmosféricas) hasta variables estéticas (impacto sobre el paisaje)” (Quintanilla 2005, 149). Por último, la evaluación externa de las consecuencias sociales guarda relación con la cada vez más importante determinación que la tecnología ejerce sobre la estructura de la sociedad:

“En efecto, cualquier tecnología de cierta importancia terminará alterando en mayor o menor medida la estructura social, las costumbres, la vida cotidiana, etc.” (Quintanilla 2005, 150).

Con todo esto, resulta evidente que, como dice Quintanilla, la evaluación externa obtiene cada vez más una importancia política, y que la participación de un sector amplio de la ciudadanía es imprescindible en los procesos de evaluación y decisión tecnológica.

### **Reflexiones finales**

Hemos podido notar que, en un régimen democrático, es importante que un gran número de personas o ciudadanos tengan acceso al conocimiento científico-tecnológico y al conocimiento sobre las ventajas o desventajas que puedan generar determinados planes de desarrollo tecnológico.

Según el enfoque de la filosofía de la técnica (Quintanilla 2005), una de las formas más valiosas de modificar e intervenir la realidad es mediante las acciones técnicas, acciones que se adaptan a los deseos y necesidades de los humanos. En este sentido, es importante conocer de qué diversas maneras la tecnología está modificando violentamente la realidad social y humana. Una vez identificados estos aspectos podríamos estar mejor capacitados para ver qué tipos de efectos negativos que emergen de la operación tecnológica afectan nuestra realidad y, más importante aún, qué tipo de efectos violentos podrían afectar a las sociedades futuras.

En México, recientemente se originó un programa de Consulta Ciudadana de Ciencia, Tecnología e Innovación a nivel nacional,<sup>2</sup> en donde se busca que los ciudadanos participen activamente en la toma de decisiones sobre estos asuntos relevantes. Sin embargo, más allá de este tipo de programas, considero que es deber de cada uno de nosotros, los usuarios, evitar la aceptación y uso irreflexivos de las tecnologías.

---

<sup>2</sup> <http://agendaiberoamericana.org/mexico/index.php>.

También me gustaría dejar una pregunta abierta: ¿qué consecuencias podría tener el hecho de que los ciudadanos permitieran, de manera automática y sin reflexión, que una IA tomara decisiones políticas importantes en un contexto democrático?

## Referencias

- BOBBIO, Norberto (1986). *El futuro de la democracia*. México: Fondo de Cultura Económica.
- KHAN, Rashid y Anik Das (2018). *Build better chatbots. A complete guide to getting started with chatbots*. Berkeley: Apress.
- MINISTERIO DE GOBIERNO DIGITAL DE NUEVA ZELANDA (2020). SAM: Meet your politician of the future. 10 de agosto. <https://www.digital.govt.nz/showcase/sam-meet-your-politician-of-the-future/>
- OLIVÉ, León (1998). “Tecnología y violencia”. En *El mundo de la violencia*, editado por Adolfo Sánchez Vázquez, 147-163. México: Fondo de Cultura Económica/UNAM.
- \_\_\_\_\_. (2000). *El bien, el mal y la razón: facetas de las ciencias y la tecnología*. México: Paidós/UNAM.
- QUINTANILLA, Miguel Ángel (2002). “La democracia tecnológica”. *Arbor* CLXXIII: 683-684.
- \_\_\_\_\_. (2005). *Tecnología: Un enfoque filosófico y otros ensayos de la filosofía de la tecnología*. México: Fondo de Cultura Económica.
- ROBLES, José Manuel (2009). *Ciudadanía digital. Una introducción a un nuevo concepto de ciudadano*. Barcelona: Editorial UOC.
- THE CENTER FOR THE GOVERNANCE OF CHANGE (2019). European Tech Insights 2019. <https://docs.ie.edu/cgc/European-Tech-Insights-2019.pdf>.

## Sobre el autor

### Mtro. Camilo Becerril Ramírez

Licenciado en humanidades por la División de Ciencias Sociales y Humanidades de la Universidad Autónoma Metropolitana, Unidad Cuajimalpa, y maestro en el posgrado en Ciencias

Sociales y Humanidades de la misma dependencia, con la idónea comunicación de resultados “Las implicaciones políticas de la técnica en el pensamiento de Hannah Arendt”.

# Aplicaciones de IA: perspectivas de análisis para la discusión

KARINA GUTIÉRREZ FRAGOSO, ELÍAS RUIZ HERNÁNDEZ  
Y GERMAN CUAYA SIMBRO

## Resumen

En este capítulo se presenta un análisis de las aplicaciones de técnicas de Inteligencia Artificial en diferentes ámbitos. En la primera sección se revisa a la Ciencia de datos vista como una rama de la Inteligencia Artificial. Se destacan tres niveles de ésta y de cómo la tarea de ubicarse en el nivel apropiado debe analizarse con cuidado a partir de los datos disponibles. Se explora desde una visión estadística que los datos se pueden separar en información *a priori* y en información mayormente obtenida por la evidencia. La segunda sección describe las perspectivas para mantener la competitividad global, la cual estará vinculada de manera directa a la industrialización de la Inteligencia Artificial, dado que se prevé que ésta, y en específico el Aprendizaje Automático, remodelarán la fabricación, la gestión de la energía, el transporte urbano, la producción agrícola, los mercados laborales y la gestión financiera. Por último, en la sección final se comentan ejemplos de aplicaciones en diferentes áreas de la

medicina, pero también se exponen las preocupaciones del personal médico y los pacientes, así como los retos de las implicaciones sociales y éticas de los modelos de Inteligencia Artificial en el ámbito médico.

**Palabras clave:** Ciencia de datos, industria, Inteligencia Artificial, medicina.

## Introducción

En años recientes las tecnologías de la información y la comunicación (TIC) han impactado en mayor medida nuestras vidas. El mundo genera millones de datos por segundo por lo que el análisis de éstos ya no resulta fiable en manos y mentes humanas. En este sentido, las estrategias de automatización de procesos por máquinas están creciendo de forma muy fuerte en todo el orbe. Hoy es evidente que necesitamos máquinas para analizar estos datos y para tomar mejores decisiones en diversos campos de la industria y de la ciencia. Los aspectos sociales también pueden verse beneficiados. A nivel internacional los Objetivos de Desarrollo Sostenibles (ODS) (Gamez 2015) y a nivel nacional los Programas Nacionales Estratégicos (Pronaces) requieren implícitamente procesar datos que deriven en una mejor toma de decisiones ambientales, energéticas, de vivienda, etcétera. Los campos de la Inteligencia Artificial (IA) son clave para poder diseñar algoritmos que permitan simplificar y aportar en estas tareas masivas y que sean desarrolladas por máquinas en lugar de por personas. En este capítulo, se exploran nociones y conceptos de la emergente área denominada Ciencia de datos que busca repensar el tratamiento de éstos desde el siglo XX hasta su manejo eficiente y automatizado en el siglo XXI. De igual manera, se muestra cómo la IA forma parte de la vida cotidiana de los individuos en las sociedades modernas, y resuelve problemas cognitivos complejos asociados con la inteligencia humana o, en un contexto más común, es capaz de reconocer problemas y crear soluciones para el beneficio de las personas y la sociedad, como

los ODS y los Pronaces antes referidos. Por último, se analizan los avances y retos que las soluciones basadas en IA enfocadas en diversas áreas de la medicina han presentado hasta ahora. Por una parte, se mencionan ejemplos recientes de aplicaciones en radiología, oftalmología, cardiología y patología, así como algunos de los productos que se han liberado al mercado; pero también se exponen las preocupaciones del personal médico, los pacientes y la necesidad de un esfuerzo disciplinar conjunto.

### **Los datos están en todos lados**

¿Qué necesito para tomar una decisión en mi vida? Todos los seres vivos, aún los animales, toman decisiones de acuerdo con dos elementos de información que, según la teoría bayesiana, podríamos separar: la información *a priori*, es decir, nuestras creencias, nuestros pensamientos internos nuestros gustos intuitivos, y a partir de la evidencia disponible, es decir, del contexto en el que nos encontramos. Seguramente un vaso de agua en medio del desierto valdrá más que estando en nuestras casas al lado del refrigerador. Como quiera sea, nuestras decisiones siempre están basadas en datos, de muchos tipos. Un *dato* es una representación simbólica de una variable cualitativa o cuantitativa. Esta definición cabe muy bien dentro de un enfoque filosófico que nos dice que nuestra mente es una estructura análoga a una computadora, que procesa símbolos de entrada y obtiene como salidas otros símbolos (decisiones, acciones) (Johannes 2010, Innis 2009). Si bien existe un debate amplio sobre si nuestra mente es algo más simbólica o algo más conexionista (Haugeland 1997), lo que es inevitable es que siempre usa datos para descubrir conocimiento, para generar acciones o para tomar decisiones en nuestras vidas. Bajo este concepto, vale la pena preguntarse si parte de este universo de datos es cuantificable y, ¿por qué no?, trasladable a una computadora para que a escala limitada las computadoras decidan algunas cosas o puedan realizar algunas acciones en su contexto.

Esta situación hoy ya existe. Entonces, ¿cómo hacen las máquinas para tomar decisiones (con las restricciones que pueda implicar)? De principio, diremos que es posible clasificar los datos que son procesables en computadora en tres tipos de manera genérica:

- Datos textuales
- Datos auditivos (señales de audio)
- Datos visuales (imágenes y videos)

Como nota, aunque el gesto es un tipo de dato, se puede ubicar en la categoría de datos visuales, puesto que las computadoras reconocen gestos a partir de fotografías o videos de los rostros de las personas. Así que, en el afán de ser un tanto genérico, éstos son los tres tipos básicos de información que puede ser procesables por una computadora para generar conclusiones extraídas de algoritmos computacionales. Ésta es la información de entrada de los modelos que procesan datos. Estos últimos tienen varios nombres en los que, al momento, no queda claro un consenso de cuáles palabras usar. Algunos de estos conceptos del análisis de datos para que una computadora tome una decisión o apoye una decisión reciben alguno de estos nombres:

- Ciencia de datos (*Data science*)
- Inteligencia de negocio (*business intelligence*)
- Minería de Datos (*Data Mining*)
- Visualización de datos (*data viz*)
- Aprendizaje máquina (*machine learning*)
- Aprendizaje profundo (*deep learning*)
- Reconocimiento de patrones (*pattern recognition*)
- Aprendizaje supervisado y no supervisado (*supervised & unsupervised learning*)
- Aprendizaje predictivo y descriptivo (*predictive and descriptive learning*)
- Selección de atributos (*feature selection*)

- Aprendizaje cognitivo (*cognitive learning*)
- Aprendizaje estadístico (*statistical learning*)
- Cómputo suave (*soft computing*)

Cada uno de estos conceptos define un área más particular de la Inteligencia Artificial. Si tomamos como nombre de esta área *Ciencia de datos* es posible mirarla desde cierta perspectiva, pues sugiere un punto intermedio en el que las bases de datos y los métodos de IA confluyen y permiten extraer conocimiento de datos estructurados o no estructurados.

### Pirámide de los Datos

Con lo anterior es posible construir una Pirámide de los Datos que va desde tareas más simples hasta tareas más complejas en la minería:

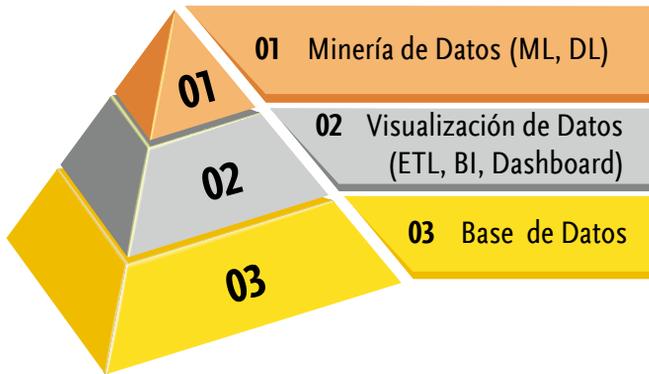


Figura 1. La Pirámide de los Datos, desde el guardado y consulta de los datos crudos, pasando por la visualización, hasta los modelos estadísticos que permiten estimar la probabilidad de ocurrencia de alguna variable a partir de la información recolectada. Elaboración propia.

En esta pirámide siempre empezamos con un sistema gestor de bases de datos (SGBD) (Nevado 2010). Los SGBD almacenan la

información, establecen relaciones entre los datos y permiten un acceso estructurado y coherente de la información. Sin embargo, este nivel aún no es tan fuerte para hablar de tomar decisiones, sino que solamente colecta y almacena información.

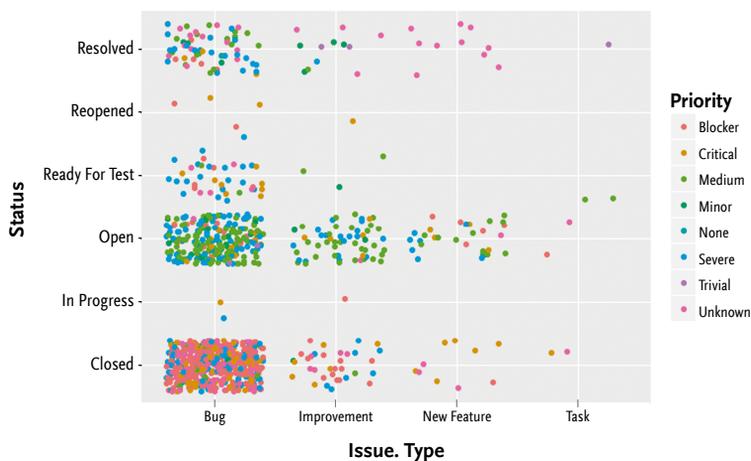


Figura 2. Diagrama de dispersión de más de dos variables (acontecimiento [issue], estado [status] y prioridad [priority]). En este análisis, se busca encontrar algunos patrones al agrupar los datos en tres variables categóricas. En la imagen se muestran muchos bugs de desarrollo de software abiertos y cerrados. Los datos provienen de una tabla de casos de probadores (testers) en la industria de software. En principio, si se observan muchos bugs críticos, la alarma de atención y de mover recursos y personal a atender estos problemas se activará. Con ello, la visualización de datos apoya a la toma de decisiones de las organizaciones. Elaboración propia.

El segundo nivel es el de representación, visualización y exploración de los datos. Su visualización, que en algunos contextos es conocida como *inteligencia de negocio*, se referiera a tener una infraestructura para mostrar los datos, así como a la transformación

de éstos para tener una representación temporal y su correspondiente aplicación en gráficas (se ilustra un ejemplo sencillo en la Figura 2). En la visualización hay varias tareas, tal como la limpieza de datos, la extracción transformación y carga (ETL) de éstos y la presentación de cuadros de mando (*dashboards*), mediante diversas aplicaciones de la industria tal como RapidMiner (Hofmann y Klinkenberg 2016), Tableau (Murray 2013), PowerBI, Python y R, entre otros. Este segmento de la pirámide de datos es particularmente útil para los indicadores clave (KPI) (Peral, Maté y Marco 2017) de muchos negocios, ya que, una vez descubiertos estos KPI, la ruta para obtenerlos es mediante analítica de datos o inteligencia de negocios. En la actualidad este sector está en crecimiento puesto que se ha encontrado que todo negocio necesita de datos para poder subsistir, al mismo tiempo que de los mismos datos estos negocios podrían mejorar sus procesos y decisiones. Estos indicadores clave o KPI, no sólo tienen sentido e impacto en el contexto industrial o comercial, también lo tienen en el sector social (como indicadores para los niveles de contaminación, la calidad del agua, la educación con inclusión, la revisión del cambio climático, la calidad en las viviendas, la presencia y medición de agentes tóxicos y procesos contaminantes, entre otros).

Finalmente, en la cúspide se encuentra la Minería de Datos o la Ciencia de datos. En este último nivel se encuentra, la estadística inferencial, el análisis predictivo y descriptivo de los datos y las tareas más recientes del conocido aprendizaje profundo (LeCun, Bengio y Hinton 2015). El análisis más simple se puede reducir a uno de regresión entre dos variables, una independiente y una dependiente (Figura 3). Sin embargo, no siempre es posible realizar este tipo de técnicas estadísticas para predecir variables a partir de datos. En muchas ocasiones, puede llegar a suceder el fenómeno GIGO (Robbert y Senne 2003), es decir, *garbage in, garbage out*, el cual indica que los datos no están ni siquiera correlacionados (Cohen *et al.* 2013) o que, al menos, existe

dependencia condicional (Dendukuri y Joseph 2001). En estos casos no podremos predecir un evento o variable de nuestro interés por mejor que se encuentre el algoritmo de Inteligencia Artificial, aprendizaje máquina o aprendizaje profundo. Este es un error en el que se puede incurrir al querer analizar los datos. Por eso, como se dijo al principio del párrafo, no siempre será posible estimar con precisión algunas variables si los datos de entrada no son apropiados o, en el peor caso, ni siquiera existen. Esto a primera vista podría reducir el universo de aplicaciones donde la Minería de Datos puede aportar (por ejemplo, en la estimación de enfermedades a partir de síntomas explícitos, por ejemplo, algo como en Wu *et al.* [2019]). Es importante destacar que recientemente se está usando este tipo de métodos para problemas más subjetivos, donde a pesar de ello, prevalecen algunas distribuciones de probabilidad que pueden ayudar a segmentar y predecir (Santoso *et al.* 2020, Almerexhi *et al.* 2020).

Como quiera, se puede decir que si toda organización genera datos (que así sucede), al menos puede ubicarse en alguno de los tres niveles de la pirámide. Lo anterior lleva a la pregunta: ¿cómo escalo en la pirámide de datos para mi organización? Si bien en principio podríamos decir que al escalar se esperaría un manejo de datos más productivos, esto no siempre es así. La recomendación clave es subir al nivel ideal a fin de incrementar la productividad del negocio o de la organización. Una vez que se sube en el escalón, se debe explorar si hay más áreas en dicho escalón por mejorar. Por contraste, cuando hablamos de fenómenos de impacto social, como Pronaces, la recomendación es subir al nivel en donde los datos o el análisis de los mismos arroja una interpretación de un riesgo a corto, mediano o largo plazo para la sociedad (por ejemplo, un agente contaminante muestra presencia a la alza en una ciudad o región determinada al medirlo en una sucesión de años).

Es importante señalar que los problemas de la cúspide pueden ser un tanto difíciles (tal como predecir los riesgos político-sociales

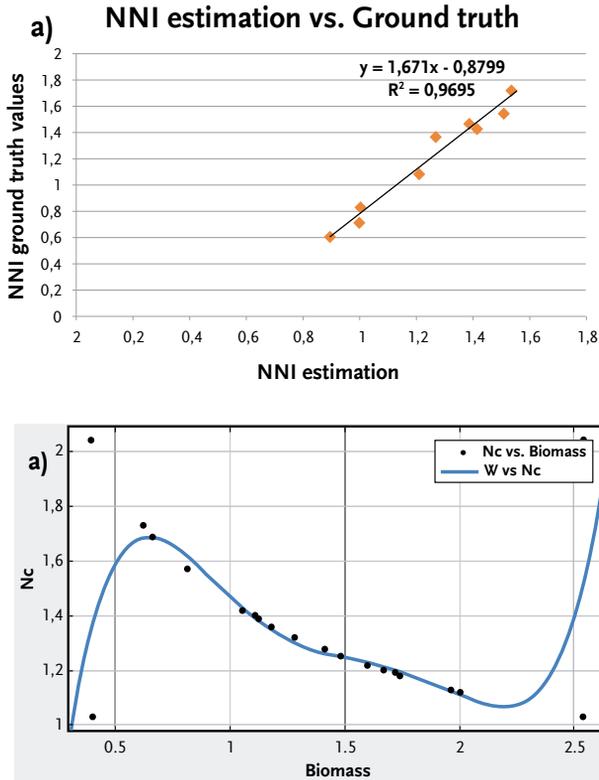


Figura 3. Análisis de regresión para dos variables. En *a)* se muestra una regresión lineal de dos variables, mientras en *b)* se muestra una regresión no lineal usando un polinomio de ajuste a los datos (la curva es de grado polinómico cinco). En ambos casos se deben cumplir aspectos de homocedasticidad: la varianza no se mueve a lo largo de sucesivas observaciones. Si la varianza cambia un modelo de este tipo no podría predecir apropiadamente la variable dependiente. Posiblemente un ajuste sea requerido. El ejemplo *a)* trata de estimar valores de nitrógeno contrastado a los reales. El modelo *b)* trata de ajustar valores de biomasa de una planta (por ejemplo, hojas y tallos) a la cantidad de nitrógeno crítico (Nc). Se interpreta en el segundo ejemplo una necesidad de nitrógeno a la baja conforme la planta crece. Elaboración propia.

en los próximos años para hacer valuaciones en el mercado de futuros y en el mercado de seguros), además de que requieren una gran cantidad de datos o herramientas para el manejo de macrodatos (*big data*) (Arfat *et al.* 2020) (tal como el cálculo del impacto de un nuevo producto con cierto grado de innovación a partir del estado actual del mercado). En los escenarios de macrodatos, si bien se ha avanzado a nivel hardware para acelerar los cálculos, una estrategia que aún debe tomarse en cuenta es la de selección de atributos / selección de características (*feature selection*), la cual consiste en evaluar los grados de dependencia condicional estadística de cada variable que se tenga en el modelo. Si no se observa que ésta tenga un impacto en la predicción podemos descartarla del modelo (ver ejemplo en Figura 4). El resultado de aplicar estas técnicas puede derivar en una considerable reducción de los tiempos de entrenamiento (Al-Thanoon, Algamal y Qasim 2021).

id1	d1	d2	d3	clase
1	0.87	0.0	-1.1	Yes
2	0.5	0.0	-0.2	Yes
8	0.4	0.01	0.35	No
6	0.2	0.0	0.43	No



id1	d1	d3	clase
1	0.87	-1.1	Yes
2	0.5	-0.2	Yes
8	0.4	0.35	No
6	0.2	0.43	No

Figura 4. Se muestra el caso de selección de atributos para un modelo de datos tabular. Esta técnica permite reducir el número de dimensiones (también llamados columnas o atributos) de la tabla. Se eliminan aquellas columnas que estadísticamente se consideren no relevantes para simplificar el modelo en términos de tamaño o para evitar intervención de variables espurias. Elaboración propia.

Una situación un tanto riesgosa en términos económicos para una organización que desee realizar Minería de Datos es si debe ocupar muchos recursos para recolectar buenos datos y hacer buenos análisis. Si lo hace y aun así no funciona la estrategia

(porque los datos no son tan buenos como se esperaban o aún faltan algunos casi imposibles de recolectar y a un altísimo costo), la empresa habrá perdido tiempo y recursos económicos. Estrategias más recientes sugieren que, una vez recolectados los datos (por malos o buenos que sean), se deben explorar los caminos del *crowdsourcing* (Filice *et al.* 2020) para tantear las posibilidades que tiene la Minería de Datos en el sector. Si se observan resultados prometedores en términos de que retribuye el análisis para la organización, entonces hay que avanzar hacia el siguiente estado y buscar conformar un laboratorio de análisis de datos con personal especializado. Debe tenerse en cuenta que, por el alto perfil del personal requerido (experiencias en estadística inferencial, algoritmia, programación, manejo de macrodatos, análisis de variables, comprensión del negocio y conocimiento multidisciplinario) la búsqueda para encontrar al personal con el perfil requerido para la organización podría ser lenta.

### **Inferencia de conocimiento como punto clave en la Ciencia de datos**

Como último punto se hablará de la inferencia, en particular de la bayesiana, ya que muchos problemas de la Ciencia de datos pueden ser descritos como variables aleatorias que interactúan a partir de reglas de dependencia condicional. Existen otras estrategias, aunque la esencia radica en interpretar el dato a predecir o estimar como una variable aleatoria.

Para ello podemos considerar un caso de análisis estadístico de escala social. Tenemos una base de datos tabular con los síntomas de pacientes con diabetes. Esta patología representa una cuestión de riesgo para la sociedad en términos de salud y se vuelve un área estratégica de investigación. Acotemos esto a una situación específica. Consideremos que dicha base de datos alimenta a un modelo computacional de clasificación supervisada que aprende de los datos para poder estimar un diagnóstico clínico (tal como lo haría un médico a partir de mirar los síntomas

del paciente). Esta estrategia puede ayudar a un médico a tomar una decisión mejor informada, ya que las computadoras trabajan mejor con un conjunto amplio de datos. Esta estimación computacional se puede escribir en forma bayesiana usando la probabilidad *a posteriori* con un enfoque *naïve* o suponiendo que las variables no dependen unas de otras y sólo tienen dependencia con la variable que se desea predecir (Pearl 1988):

$$P(\text{Síntomas}) = \frac{P(\text{Diagnóstico}_i) P(\text{Síntomas} \vee \text{Diagnóstico}_i)}{P(\text{Síntomas})} \quad (4)$$

donde el denominador puede ser estimado mediante una normalización de los datos:

$$P(\text{Diagnóstico}) = \sum_{x=1}^n P(\text{Diagnóstico}_i) P(\text{Diagnóstico}_x) \quad (5)$$

Así, si suponemos independencia condicional, y que los datos de cada paciente son un vector de la forma Síntomas=[d\_1,d\_2,-d\_3,...d\_m] el numerador se puede reescribir de forma más sencilla como

$$P(\text{Diagnóstico}_i) = \prod_{x=1}^m P(\text{Diagnóstico}_i) \quad (6)$$

Así que encontrar la probabilidad posterior de la ecuación 4 se reduce a encontrar una razón de un producto de probabilidades con los datos del vector de Síntomas como entrada [d\_1,-d\_2,d\_3,...d\_m]. Si bien esta interpretación puede parecer muy simple (estimar un diagnóstico a partir de síntomas), consideremos que puede tener un impacto modesto en facilitar la tarea de diagnóstico “previo o preliminar” en casos donde no se cuente con infraestructura de atención hospitalaria primaria.

En otro tenor, la Tabla 1 muestra un ejemplo para estimar si podría (o no) comprar una cerveza un cliente en un supermercado a partir de los productos que compra. En este

escenario, nuevamente se podría ocupar el esquema bayesiano anterior.

ID cliente	Leche (d1)	Piña (d2)	Carne (d3)	Café (d4)	Pan (d5)	Snacks (d5)	¿Comprará cerveza? (clase)
5	0	1	0	0	1	1	0
4	0	1	0	1	0	1	1
8	1	0	0	1	0	0	0
27	1	1	1	0	1	1	1
35	0	0	0	1	0	0	?

*Tabla 1.* En esta tabla, los datos son compras de un cliente y el objetivo es predecir qué tan probable es que el cliente también compre cerveza. Igual se puede modelar como la probabilidad de comprar cerveza dado lo que ya ha comprado al momento. Una variante de este problema con n productos se denomina reglas de asociación (Agrawal *et al.* 1996)

### Puntos clave de la Ciencia de datos

Por último, el enfoque de la Ciencia de datos puede realizarse con todo dato que se tenga disponible. Sin embargo, no debe perderse de vista que si éstos presentan pobre dependencia condicional, y no están involucrados con la variable que se desea predecir, de poco servirá un modelo computacional de aprendizaje más complejo (tal como redes bayesianas, bosques de árboles aleatorios [*random forest*], máquinas vectoriales supervisadas [*svm*], o aprendizaje profundo), pues el resultado será igualmente malo. Hay casos donde a partir de la correlación se pueden “pseudo-predecir” variables no asociadas a los fenómenos mostrados en los atributos. Esto último se trata de manipulación de datos para producir salidas a modo. Si bien es un buen ejercicio como diversión, es algo que no debe realizarse en la práctica. Si quiere ver algún ejemplo de correlaciones extrañas, busque en internet algo como: *bizarre correlations*.

Los datos siguen creciendo y las aplicaciones basadas en datos cada vez son más notorias para todos. En las siguientes secciones se hablará de algunos casos de países que están incorporando la IA de manera institucional para emprender proyectos más ambiciosos en materia de Ciencia de datos bajo un enfoque de Industria 4.0, además de que buscan incrementar el impacto de estas técnicas en problemas nacionales del área de salud, tal como el análisis de imágenes médicas y medicina en general.

## **Inteligencia Artificial, industria y sustentabilidad**

### ***La inteligencia Artificial y la industria***

La IA ha entrado en una etapa de comercialización, por lo que ejerce diferentes impactos en los actores de las industrias tradicionales e impulsa cambios en los ecosistemas de éstas. De acuerdo con el reporte realizado por Deloitte (2019), de manera general se pueden observar tales cambios en tres niveles:

- *Cambio empresarial:* la IA está involucrada en los procesos de administración y producción de la empresa, con una tendencia a comercializarse cada vez más, y algunas empresas han realizado aplicaciones inteligentes relativamente maduras. Estas aplicaciones han recopilado y utilizado la información obtenida del usuario a través de diversos medios tecnológicos para proporcionar a los consumidores productos y servicios pertinentes, al mismo tiempo que satisfacen sus necesidades potenciales al comprender sus preferencias y comportamientos, y optimizar los datos.
- *Cambio de la industria:* la transformación traída por la IA impulsa cambios fundamentales en la relación de los sectores ascendentes y descendentes en la cadena industrial tradicional, es decir, permite expandir los tipos de proveedores de productos ascendentes, y los usuarios también pueden cambiar de consumidores individuales a consumidores

empresariales, o ambos, debido al cambio de las características del producto. Esto es conseguido a través del uso de Minería de Datos.

- *Cambio laboral*: la aplicación de nuevas tecnologías, como la IA, está mejorando la eficiencia del uso de la información y reduciendo el número de empleados para labores repetitivas. Esto último, puede alcanzarse con el uso más amplio de los robots, lo cual implicaría aumentar el porcentaje de personal tecnológico y de gestión, lo que traería cambios en las estructuras laborales de las empresas.

Por otro lado, y de manera más específica, el principal aporte de la IA en la industria se centra en la automatización y búsqueda de la eficiencia en los procesos de producción, lo cual se conoce como la cuarta revolución industrial o industria 4.0, desarrollo de fábricas inteligentes.

Existen empresas pioneras en el desarrollo de soluciones de IA industrial basadas en procesos. Éstas emplean la IA para permitir a los fabricantes predecir y prevenir ineficiencias inesperadas de algún proceso que continuamente daña el rendimiento y la calidad de la producción.

A continuación, se muestra un resumen que cubre varios aspectos o aplicaciones de la Inteligencia Artificial dentro del paradigma de la Industria 4.0, y cómo los fabricantes ya utilizan esta poderosa tecnología para impulsar la eficiencia, mejorar la calidad y gestionar mejor las cadenas de suministro (SEEBO 2019):

- *Mantenimiento/OEE (overall equipment effectiveness/efectividad total de los equipos)*. El mantenimiento predictivo se ha convertido en un caso de uso muy solicitado para los fabricantes que buscan avanzar a la Industria 4.0. En lugar de realizar el mantenimiento de acuerdo con un cronograma predeterminado, el predictivo utiliza algoritmos para determinar la próxima falla de un componente / máquina / sistema y alerta al personal

para que realice procedimientos para evitar la falla, pero no demasiado pronto para desperdiciar el tiempo de inactividad innecesariamente.

- *Calidad 4.0.* A los fabricantes les resulta más difícil que nunca mantener consistentemente altos niveles de calidad. Esto se debe, en parte, a una creciente complejidad en los productos y objetivos de tiempo de comercialización muy cortos. Los problemas de calidad cuestan mucho dinero a las empresas, pero con el uso de algoritmos de IA desarrollados a través del Aprendizaje Automático, los fabricantes pueden recibir alertas de problemas iniciales menores que causan caídas de calidad, de forma similar a la que se crean las alertas para el mantenimiento predictivo. Calidad 4.0 permite a los fabricantes mejorar continuamente la calidad de su producción, al tiempo que recopila datos de uso y rendimiento de productos y maquinaria en el campo. Estos datos se convierten en una fuente vital de información que forma la base para el desarrollo de productos y decisiones comerciales cruciales.
- *Colaboración humano-robot.* En este aspecto, el enfoque general es que a medida que los robots asuman los trabajos, los trabajadores recibirán capacitación para puestos de alto nivel en programación, diseño y mantenimiento. Mientras tanto, la eficiencia del trabajo colaborativo humano-robot se está mejorando a medida que los robots de fabricación están aprobados para trabajar junto con los humanos. Conforme aumente la adopción de la robótica en la fabricación, la IA desempeñará un papel importante para garantizar la seguridad del personal humano y otorgará a los robots más responsabilidad para tomar decisiones que puedan optimizar aún más los procesos en función de los datos en tiempo real recopilados en la planta de producción.
- *Diseño generativo.* Los fabricantes también pueden hacer uso de la IA en la fase de diseño. Con un resumen de diseño

claramente definido como entrada, los diseñadores e ingenieros pueden utilizar un algoritmo de IA, generalmente conocido como software de diseño generativo, para explorar todas las configuraciones posibles de una solución. El resumen puede incluir restricciones y definiciones para tipos de materiales, métodos de producción, limitaciones de tiempo y presupuestarias. El conjunto de soluciones generadas por el algoritmo se puede probar utilizando Aprendizaje Automático. La fase de prueba proporciona información adicional sobre qué ideas/decisiones de diseño funcionaron y cuáles no. De esta manera, se pueden realizar mejoras adicionales hasta que se encuentre una solución óptima.

- *Adaptación del mercado/cadena de suministro.* El uso de algoritmos de IA para optimizar la cadena de suministro de las operaciones de fabricación puede ayudar a responder mejor y anticipar los cambios en el mercado. Para construir estimaciones de la demanda del mercado, un algoritmo puede tener en cuenta los patrones de demanda clasificados por fecha, ubicación, atributos socioeconómicos, comportamiento macroeconómico, estado político, patrones climáticos y más. Esto es innovador para los fabricantes que pueden usar esta información para optimizar el control de inventario, la dotación de personal, el consumo de energía, las materias primas y tomar mejores decisiones financieras con respecto a la estrategia de la empresa.

### **La Inteligencia Artificial y la sustentabilidad**

La última aplicación mencionada de la IA en la industria puede ser extrapolada a otros dominios en los cuales se requiera la optimización de recursos y la reducción de residuos, por ejemplo, en el desarrollo sostenible. Muchos de los retos a los que se enfrenta el ser humano (tales como el cambio climático, la escasez de agua, las desigualdades o el hambre) sólo se pueden resolver desde una perspectiva global y al promover el desarrollo sostenible.

La IA puede hacer grandes aportes a tales retos, a través de análisis predictivos y de causa raíz, con lo cual podría prevenir el desperdicio de recursos naturales (agua, energía, tierra, etcétera) al identificar áreas de pérdida y proponer acciones enfocadas a reducir dichos desperdicios.

Lo anterior es una preocupación a nivel internacional y nacional. En el primer caso, los objetivos de desarrollo sostenible (ODS), también conocidos como los *objetivos globales*, son una llamada de Naciones Unidas a todos los países para afrontar los grandes desafíos a los que se enfrenta la humanidad y garantizar que todas las personas tengan las mismas oportunidades y puedan llevar una vida mejor sin comprometer nuestro planeta. En el segundo caso, a nivel nacional, los Pronaces (o Programas Nacionales Estratégicos, del Consejo Nacional de Humanidades, Ciencias y Tecnologías [Conahcyt]) buscan promover la investigación centrada en el desarrollo de proyectos que resuelvan o aborden temas de sustentabilidad, salud, seguridad y economía.

Tanto los 17 ODS de la Organización de Naciones Unidas (Gamez 2015) como los Pronaces (2024) están definiendo la agenda de desarrollo para diferentes países en el mundo. Con las discusiones cada vez más urgentes sobre el cambio climático y el calentamiento global, existe una creciente conciencia de la importancia de las actividades sostenibles. Pero tratar el medio ambiente cuidadosamente no es sólo una cuestión moral: la degradación ambiental también es un grave peligro para la economía global. Por ello consideramos que la IA puede ser una herramienta importante en el esfuerzo por adoptar prácticas comerciales responsables, es decir, si incorporamos de forma adecuada esta tecnología podemos alcanzar cambios significativos con respecto a la sostenibilidad.

La IA se ha incorporado de diversas formas en los ODS, primero a través de la experimentación y luego en programas de gestión y liderazgo sostenibles. A continuación, se presentan tres

proyectos relevantes que utilizan IA para aportar en alguno de estos objetivos (Goralski y Keong 2020).

### ***La IA y la crisis del agua***

Las elecciones que hacen las comunidades, ciudades y naciones con respecto al manejo de los recursos hídricos tienen grandes implicaciones en nuestro bienestar futuro. El agua debe tratarse y transportarse adecuadamente de acuerdo con las normas de higiene y salud para garantizar que su calidad y propiedades cumplan con las especificaciones exigidas por los usuarios finales. En los últimos años, existe un movimiento para aprovechar las tecnologías emergentes para ofrecer soluciones sostenibles para el tratamiento, el transporte y el reciclaje y la reutilización del agua.

La humanidad debe aprovechar los avances tecnológicos y la innovación en IA para poder satisfacer la demanda económica a corto plazo y al mismo tiempo salvaguardar la sostenibilidad ambiental a largo plazo de los ecosistemas naturales y las bioregiones. El progreso en la nueva tecnología de IA podría hacer que el monitoreo en las plantas de tratamiento de agua sea más barato y fácil, y ayudar a salvaguardar la salud pública.

Los datos de entrada en sistema inteligente se correlacionan con las salidas conocidas para permitir que los algoritmos aprendan con el tiempo. Luego, en la “fase operativa”, el programa comienza a dar sentido a los patrones a medida que se introducen nuevos datos. Debido a la capacidad de IA de adaptar y procesar constantemente grandes cantidades de datos en tiempo real, es una herramienta ideal para administrar los recursos hídricos en un entorno en constante cambio y permite a los administradores de servicios de agua maximizar los ingresos actuales y planear para los años venideros.

### ***La IA y la agricultura***

Las enfermedades de las plantas y las plagas locales pueden reducir en 40 por ciento o más el cultivo de diferentes productos

como el maíz, la cebada o el trigo, por mencionar algunos. Por lo anterior el interés del desarrollo de tecnología que permita apoyar en la determinación de la salud de los cultivos.

PlantVillage (Digging Deeper: Plant Village 2020), es un proyecto fundado por el entomólogo David Hughes de la Universidad de Penn State en conjunto con el epidemiólogo digital Marcel Salathé del Instituto Federal Suizo de Tecnología (EPFL) en Lausana, Suiza. El proyecto intenta diagnosticar enfermedades de las plantas mediante el uso de una aplicación en un teléfono inteligente de bajo costo común por un agricultor en el campo. Al mismo tiempo, el equipo está creando una base de datos que puede reconocer e identificar enfermedades de las plantas mediante el uso de una imagen de teléfono móvil. Dado que la IA es una tecnología de predicción y descubrimiento de bajo costo, se pueden utilizar grandes cantidades de datos para identificar patrones y hacer predicciones. Ésta es la premisa sobre la cual se fundó PlantVillage.

### ***IA, saneamiento y salud***

Obtener agua limpia en un mundo donde miles de millones de personas se ven afectadas por enfermedades transmitidas por este líquido es un problema crítico. Es un desafío perenne que es difícil de controlar, ya que necesita respaldo financiero y tiempo para aprender cómo utilizar mejor un nuevo sistema de agua limpia.

Lant (2020) muestra cómo utilizar tecnología para mejorar la vida humana. El autor comenzó a trabajar en el proyecto Clean Water AI en diciembre de 2017 para el concurso World Virtual GovHack. Uno de los desafíos de este concurso fue la escasez de agua limpia y alimentos, por lo que una directriz de su investigación fue concentrarse principalmente en cómo las tecnologías pueden mejorar la vida de las personas, en lugar de sólo lo que las tecnologías pueden hacer. El sistema desarrollado de AI de agua limpia puede realizar un análisis en tiempo real e identificar contaminantes sin un enlace a Internet. La prueba

es autocontrolada y consta de componentes simples y económicos que se pueden comprar en el mercado: un microscopio digital, una computadora económica que ejecuta el sistema operativo Ubuntu, y un Intel Movidius Neural Compute Stick que activa el Aprendizaje Automático e IA en tiempo real. La prueba en su totalidad se puede construir por menos de US\$500, lo que la pone al alcance de muchas organizaciones que no pueden permitirse un sistema tradicional costoso.

### **Inteligencia Artificial en medicina**

Esta sección se articula con el ODS 3 Salud y Bienestar y el Pronaces de salud. En este sentido, se puede advertir que la aplicación de las técnicas de IA en un campo como la medicina es fértil, por lo que es previsible que resulte crucial en los desafíos que la población mundial está por enfrentar en este aspecto. Por una parte, esto tiene estrecha relación con el principio de la medicina de preservar la condición de salud y la vida humana. Por otra parte, aborda un objeto de estudio sumamente complejo tanto a nivel individual como poblacional: el ser humano. Estos componentes detonaron el interés por aplicar IA en este terreno desde sus inicios en la década de 1960, cuando se desarrollaron sistemas que intentaban representar las decisiones de los médicos a través de la formulación de reglas.

Los sistemas desarrollados hasta ese momento no tuvieron el éxito esperado y eso constituyó la primera lección sobre el dominio de la medicina para la IA: es un campo de conocimiento muy complejo. Las preguntas subyacentes no se hicieron esperar. ¿Cómo representar ese tipo de conocimiento? ¿Qué observaba el personal médico que no se lograba modelar con estos métodos? Aparentemente, la solución llegaría con los métodos que aprenden a partir de los ejemplos de datos del fenómeno en cuestión.

En una interesante aproximación, Chin-Yee y Upshur (2019) abordan tres problemas relacionados con el análisis de grandes volúmenes de datos (macrodatos o *big data*) y la aplicación de la

IA en medicina desde una perspectiva filosófica. Entre los planteamientos de los autores, resulta relevante la crítica a los algoritmos de IA en términos de que la experiencia del personal médico no se reduce a datos cuantitativos. Esto revela la necesidad de comprender que la investigación en IA ha buscado desarrollar técnicas para contender con datos tanto cuantitativos como cualitativos, justamente por reconocer la complejidad y riqueza de la experiencia humana.

Sin embargo, es cierto que la historia de la IA se ha construido por etapas de grandes expectativas y episodios de grises inviernos, pero al menos en los últimos años se puede percibir un ambiente más bien primaveral con una expansión creciente de aplicaciones en los más diversos ámbitos, no sin la natural oposición de algunos. En este sentido, la convergencia disciplinar por lo general tiende un puente que no está exento de confusiones conceptuales y diferentes perspectivas. De manera que no debería sorprender que en el binomio IA-medicina existan ciertas apreciaciones erróneas sobre la comprensión de esta relación. Chang (2019) señala con tino estos errores comunes, pero además los califica de forma indulgente como humanos y comprensibles: *a)* el personal médico será reemplazado por algoritmos de IA, *b)* las técnicas de IA se pueden aplicar en todos los procesos y niveles del cuidado de la salud, *c)* las redes neuronales profundas son las herramientas por excelencia, *d)* abordar un problema biomédico desde la IA siempre requiere grandes volúmenes de datos, y *e)* el área bajo la curva ROC es un buen indicador del desempeño de un algoritmo para el dominio médico.

En síntesis, las actividades del personal médico constituyen una integración compleja de información y la intención no es prescindir de los especialistas. La aplicación de técnicas de IA no se traduce necesariamente en beneficios o generación de valor. Esto porque el abanico de técnicas de IA es muy amplio, y existen fenómenos de morbilidad en los que no siempre se pueden obtener grandes volúmenes de datos. Además, la conclusión sobre la evaluación de

un método en el dominio médico debe integrar aspectos cruciales como la prevalencia de la enfermedad (Chang 2019).

Aunque todos estos aspectos quizás no se han considerado de manera consciente por ambas disciplinas, hasta ahora se ha reportado un copioso volumen de aplicaciones de IA en medicina y algunos trabajos más ambiciosos aún se encuentran en proceso de desarrollo. En Amisha *et al.* (2019) se describen algunas de estas aplicaciones en áreas como radiología, entrenamiento médico, identificación de casos negativos, detección de infecciones adquiridas de manera intrahospitalaria, atención médica en línea, terapia psicológica, cirugía, monitoreo de tratamientos, monitoreo de la salud, procesos de higiene de manos en médicos y enfermeras, entre otros.

En el área farmacéutica, un ejemplo relevante fue el análisis de opciones de moléculas activas contra el virus del Ébola en 2014 a través de modelos bayesianos (Anantpadma *et al.* 2019). Recientemente, ante la pandemia de COVID-19 que enfrentamos, también se han realizado importantes esfuerzos que incluyen la aplicación de técnicas de IA para la detección de brotes, modelos de predicción de diseminación, prevención y desarrollo de vacunas, detección y seguimiento tempranos de casos, predicción del pronóstico de los pacientes afectados y desarrollo de fármacos (Bansal *et al.* 2020).

En cuanto a los aspectos más promisorios, se ha señalado que es en el nivel de atención primaria donde se puede tener mayor impacto de la aplicación de IA, debido a que el personal médico por lo general invierte más tiempo en los registros hospitalarios y trabajo de escritorio que en la atención personal a los pacientes (Amisha *et al.* 2019). En este sentido, el uso de IA se justifica más allá de la automatización, con el fin de que el personal médico pueda enfocarse primordialmente en observar al paciente, generar una comunicación efectiva y establecer el vínculo profesional que contribuya a un mayor compromiso de adherencia a los tratamientos y el seguimiento del paciente (Reddy 2018).

Sin embargo, es fundamental señalar la importancia del control de calidad en los procesos para coleccionar, almacenar y procesar los datos provistos a los modelos. Lo anterior porque si los métodos de aprendizaje computacional reciben datos que se introducen “ciegamente”, la interpretación de los resultados debe considerarse con las reservas necesarias (Bansal *et al.* 2020). En Becker (2019) se menciona el ejemplo de un estudio que sesgaba la sobreestimación o sub-estimación de riesgo de enfermedad cardiovascular en una población no-blanca. Además, se explica que el sesgo racial posiblemente podría haber sido introducido accidental o intencionalmente por vía humana con implicaciones adversas en los protocolos de tratamiento. De Camp y Lindvall (2020) abordan el concepto de sesgos latentes en la implementación de soluciones basadas en IA para la medicina y proponen maneras de gestionarlos. Las propuestas incluyen la caracterización de los sesgos, marcos regulatorios para monitorearlos en el desempeño de los modelos y la integración de los involucrados en el desarrollo de soluciones de IA en medicina. En particular, el primer aspecto mencionado constituye una veta de investigación que requiere explorarse para diseñar los mecanismos de identificación proactiva y evaluar su efecto en el desempeño de los modelos.

Pero no todo es un reto pendiente, el tratamiento de imágenes médicas es una de las áreas más prolíficas de aplicación de técnicas de IA, lo que incluso ha generado inquietud entre algunos expertos del área al punto de sugerir que eventualmente serán reemplazados. Esa narrativa se basa en que el desempeño de los algoritmos ha alcanzado niveles equiparables a los de expertos humanos en algunas tareas específicas. Sin embargo, un sector más optimista contempla que la IA formará parte de la rutina de la vida diaria de los radiólogos (y otros especialistas), y hará su trabajo más eficiente, preciso y valioso.

En Ahuja (2019) se detallan ejemplos de aplicaciones en radiología, patología, oftalmología y cardiología, los cuales dan cuenta de que la aplicación de IA en medicina se encuentra en franca

expansión desde algoritmos para diagnóstico hasta robots para cirugía. También se han reportado aplicaciones en medicina pulmonar, endocrinología, nefrología, gastroenterología, neurología e histopatología (Briganti y Le Moine 2020). En diez años, se espera que los algoritmos de IA realicen un análisis previo sobre los conjuntos de imágenes antes de que sean revisadas por los especialistas incluyendo la aplicación de técnicas para segmentación, conteo y reconocimiento de patrones (Recht y Bryan 2017).

El argumento hacia los más escépticos es que liberar a los especialistas de estas actividades les permitirá integrar información clínica de los pacientes en la interpretación, participar en el desarrollo de nuevas tecnologías y preservar el uso ético de las mismas al considerar su relevancia clínica en favor del cuidado de los pacientes (Kruskal *et al.* 2017). Derivado de esta tendencia de la medicina moderna, algunas escuelas ya ofrecen cursos sobre infraestructura tecnológica, aprendizaje computacional y gestión de datos (Ahuja 2019).

Desafortunadamente, esta visión transdisciplinaria aún no es generalizada, pero es necesaria para favorecer un terreno más neutro que permita una efectiva integración de expertos de ambas disciplinas. La constante comparación del desempeño entre los métodos computacionales y el personal médico han generado la impresión de que son contrapartes en competencia, pero la clave es comprender la potencia de la sinergia ante una evolución inevitable de la práctica clínica moderna.

En este sentido, conviene tratar de entender el señalamiento que el área médica manifiesta en términos de que las implementaciones de soluciones basadas en IA en medicina tienen un enfoque marcadamente tecnológico y que no consideran aspectos de la relación profesional médico-paciente, las cuales forman parte sustancial de la práctica clínica. En este sentido, nos referimos a la aplicación del pensamiento crítico, las habilidades interpersonales y de comunicación, la inteligencia emocional y la creatividad. Al respecto, Acampora *et al.* (2013) conceptualizan el

entorno clínico inteligente como un espacio sensible, adaptativo y receptivo que rodea al personal médico y al paciente. En concordancia, Chang (2019) explica que es justamente la aplicación más amplia de la IA en el ámbito médico la que permitirá que en el espacio en el que se encuentran el médico y su paciente se privilegie la interacción humana.

Con este enfoque han surgido trabajos como el de Cáceres, Rosário y Amaya (2019), quienes presentaron una propuesta para la optimización de la operación de un hospital basado en las características de lo que se denomina como Salud 4.0. En particular, se utilizó como caso de estudio el Servicio de Urgencias. Según los autores, dada la naturaleza crítica de la operación de este tipo de áreas, resultan un escenario retador pero propicio para validar propuestas en e-Salud.

En Becker (2019) se presenta una revisión amplia de ejemplos de aplicaciones de IA en medicina publicadas en PubMed a partir de 2015. Los resultados se agruparon en cuatro categorías con base en la manera en que se está aplicando las técnicas de IA en el ámbito médico: 1) en la evaluación del riesgo de aparición de la enfermedad y en la estimación del éxito del tratamiento antes del inicio; 2) en un intento de manejar o aliviar las complicaciones; 3) para ayudar con el cuidado del paciente durante el tratamiento activo o la fase de procedimiento, y 4) en investigaciones destinadas a dilucidar la patología o el mecanismo y/o el tratamiento ideal de una enfermedad.

En general, en el trabajo de Becker se observa una muy variada gama de aplicaciones de las técnicas de IA, se explican algunas de las preocupaciones desde la perspectiva de los pacientes y el personal médico, también se incluyen ejemplos de estudios en los cuales los sistemas de diagnóstico asistido presentaron clasificaciones incorrectas de los casos y los productos para el cuidado de la salud que se ofrecen hoy en el mercado.

Entre las preocupaciones de los pacientes se destacan la reducción del contacto humano, el incremento en los sentimientos

de objetivación y pérdida de control, así como la pérdida de privacidad y libertad personal. En otros contextos, la adopción de tecnologías basadas en IA resulta muy atractiva, sin embargo, cuando se trata de una condición relativa a la salud, la sensación de vulnerabilidad cambia por completo el panorama.

Es posible que con el paso del tiempo la adopción y confianza en tecnologías basadas en IA alcancen un mayor nivel de aceptación, pero es necesario continuar investigando para identificar aquellas tareas que representan mayores dificultades; consolidar las áreas en las que se han logrado avances sustanciales mediante una integración disciplinar efectiva, y mantener la discusión sobre las implicaciones sociales, legales y éticas encaminadas a una regulación consensuada, pero también para encaminar esfuerzos hacia un acceso igualitario.

En relación con este aspecto, Ferryman (2020) advierte la necesidad de considerar las disparidades en el acceso a los servicios de salud como parte de las buenas prácticas en los modelos de aprendizaje computacional y los marcos regulatorios tanto a nivel de software como de dispositivos médicos. De esta manera, se pueden anticipar y mitigar los efectos en los grupos poblacionales con vulnerabilidad sanitaria. Esta noción considera indicadores a nivel individual tales como condición de pobreza, pertenencia a minorías étnicas o raciales, presencia de enfermedades mentales crónicas, condición de discapacidad y falta de cobertura médica. Asimismo, incluye indicadores a nivel sistémico relacionados con el acceso a servicios de salud, pruebas diagnósticas y tratamiento. El riesgo inherente es que las soluciones basadas en IA perpetúen las condiciones de desigualdad en el acceso a las prestaciones y servicios de salud.

Aunque los marcos regulatorios aún están en desarrollo, una oferta reducida pero creciente de productos basados en técnicas de IA dirigidos al ámbito médico ya han sido liberados al mercado. Estas soluciones tecnológicas han recibido recientemente la aprobación de la Administración de Alimentos y Medicamentos

de Estados Unidos (FDA, por sus siglas en inglés) por lo que podrían implementarse. Algunos ejemplos de estos productos se enfocan en el análisis de imágenes para detectar síntomas de pérdida de visión asociada con diabetes; asistencia en el diagnóstico de enfermedades retinales; alertamiento de accidentes cerebrovasculares; análisis de imágenes para buscar signos de fracturas radio distales; mejoramiento de la calidad de imagen durante la tomografía de emisión de positrones y por resonancia magnética mediante adquisición automática; identificación de hemorragia intracraneal en tomografía computarizada, y lectura concurrente de escaneos automáticos de ultrasonido de mama, entre otros (Becker 2019). No obstante, actualmente sólo algunos entornos específicos se benefician de la aplicación de IA en la práctica clínica (Briganti y Le Moine 2020).

Para incidir en la reducción de esta brecha es fundamental promover una visión integradora de los perfiles disciplinares, desde la formación educativa hasta el diseño e implementación de soluciones tecnológicas basadas en IA para el campo de la medicina. Las aportaciones a la luz de este enfoque fortalecen la competencia profesional humana, facilitan la transferencia de conocimiento, y promueven la participación necesaria de expertos para validación y monitoreo de las soluciones tecnológicas, propician una discusión abierta y diversa para la generación de marcos normativos que atiendan cuestiones y dilemas éticos. La idea clave es que la combinación de la IA y la inteligencia humana supere el desempeño de ambas por separado (Kelly *et al.* 2019).

## Conclusiones

De acuerdo con lo visto, la IA y la Ciencia de datos poco a poco seguirán insertándose en campos diversos del conocimiento. En algunos casos lo harán a escala baja (almacenamiento y visualización); en otros lo harán a una escala más alta (predicción; generación de ejemplos; análisis de información; toma de decisiones; agrupamiento de datos semejantes; simplificación de grandes

volúmenes de datos; apoyo en la solución de problemas sociales de escala nacional e internacional, y asistencia en medicina, industria, agricultura, servicios, etcétera). Solamente es cuestión de tiempo para que, por un lado, tecnologías de procesamiento y almacenamiento sigan creciendo en cantidad y disponibilidad para automatizar más procesos y, por otro, se dé un incremento en la generación de datos especializados que permitan trabajar adecuadamente el análisis de dicha información (bases de datos médicas, BD de monitoreo industrial, entre otras).

Por otro lado, cada año que pasa se irá incrementando el personal con conocimiento en tecnologías que trabajen con datos. Personas capaces de comprender, analizar y diseñar algoritmos para distintos dominios serán requeridos. Una máxima que debe suponerse es que, para poder trabajar con todo ello, necesitamos un enfoque multidisciplinario para conformar equipos de trabajo especializados en su ramo y al mismo tiempo capaces de comprender terminología y procesos de otras áreas del saber.

También, en este capítulo, se han presentado diferentes aplicaciones de IA en el ámbito industrial las cuales, en su mayoría, están orientadas a la mejora de los procesos productivos, al procurar la optimización de recursos y disminuir los residuos generados en los procesos de producción. También, se ha mostrado la aplicación de IA en un ámbito diferente a la industria, en el desarrollo sostenible, en el que se comentaron diferentes proyectos en los cuales se está haciendo uso de la IA para aportar en la solución de algún ODS (en el contexto internacional) o Pronaces (en el contexto nacional). Asimismo, se ha hecho énfasis en las áreas de oportunidad y retos para el uso y mejoramiento de herramientas basadas en IA.

En el ámbito médico se ponen de relieve varios aspectos tales como la importancia de la calidad de los datos suministrados a los modelos de IA, la crucial participación de los especialistas para la validación de los resultados, el rol de la caracterización de sesgos y la evaluación de sus efectos en el desempeño

de las soluciones. Aunado a ello, está la necesidad de continuar el esfuerzo para modificar las condiciones de desigualdad en el acceso a servicios de salud; la demanda legítima de considerar aspectos asociados a la vulnerabilidad sanitaria para mejorar el diseño e implementación de productos y servicios médicos basados en técnicas de IA, e integrar en las métricas de evaluación del desempeño de los modelos de IA indicadores (como la prevalencia de la enfermedad, entre otros), relacionados al dominio de aplicación, la tendencia de formación transdisciplinaria en el personal médico con la consecuente evolución de la práctica clínica y una combinación efectiva de la IA y humana. Lo anterior, sin afán comparativo que se traduzca en competencia infructuosa, sino más bien una interacción que encamine los esfuerzo bajo la premisa común: preservación de la salud y mejor calidad de vida humana.

## Referencias

- ACAMPORA, Giovanni, Diane Cook, Parisa Rashidi y Athanasios Vasilakos (2013). "A survey on ambient intelligence in health care". *Proc IEEE Inst Elect Electron Eng*. 101: 2470-2494. DOI: 10.1109/JPROC.2013.2262913.
- AGRAWAL, Rakesh, Heikki Mannila, Ramakrishnan Srikant, Hannu Toivonen y A. Inkeri Verkamo (1996). "Fast discovery of association rules". *Advances in knowledge discovery and Data Mining* 12, n.º 1: 307-328.
- AHUJA, Abhimanyu (2019). "The impact of artificial intelligence in medicine on the future role of the physician". *PeerJ* 7: e7702. <http://doi.org/10.7717/peerj.7702>.
- ALMEREKHI, Hind, Haewoon Kwak, Joni Salminen y Bernard J. Jansen (2020). "Are These Comments Triggering? Predicting Triggers of Toxicity in Online Discussions". En *Proceedings of The Web Conference 2020*, 3033-3040. <https://doi.org/10.1145/3366423.3380074>.

- AL-THANOON, Niam, Zakariya Yahya Algamal y Omar Saber Qasim Algamal (2021). "Feature selection based on a crow search algorithm for big data classification". *Chemometrics and Intelligent Laboratory Systems* 212: 104288.
- AMISHA, Paras Malik, Monika Pathania y Vyas Kumar Rathaur (2019). "Overview of artificial intelligence in medicine". *Journal of family medicine and primary care* 8, n.º 7: 2328-2331. [https://doi.org/10.4103/jfmprc.jfmprc\\_440\\_19](https://doi.org/10.4103/jfmprc.jfmprc_440_19).
- ANANTPADMA, Manu, Thomas Lane, Kimberley M. Zorn, Mary A. Lingerfelt, Alex M. Clark, Joel S. Freundlich, Robert A. Davey, Peter B. Madrid y Sean Ekins (2019). "Ebola Virus Bayesian Machine Learning Models Enable New in Vitro Leads". *ACS Omega* 4, n.º 1:2353-2361.
- ARFAT, Yasir, Sardar Usman, Rashid Mehmood e Iyad Katib (2020). "Big data tools, technologies, and applications: A survey". En *Smart Infrastructure and Applications*, 453-490. Springer.
- BANSAL, Agam, Rana Prathap Padappayil, Chandan Garg, Anjali Singal, Mohak Gupta y Allan Klein (2020). "Utility of Artificial Intelligence Amidst the COVID 19 Pandemic: A Review". *Journal of medical systems* 44, n.º 9: 156. <https://doi.org/10.1007/s10916-020-01617-3>.
- BECKER, Aliza (2019). "Artificial intelligence in medicine: What is it doing for us today?". *Health Policy and Technology* 8, n.º 2:198-205. <https://doi.org/10.1016/j.hlpt.2019.03.004>.
- BRIGANTI, Giovanni y Oliver Le Moine (2020). "Artificial Intelligence in Medicine: Today and Tomorrow". *Front. Med.* 7: 27. DOI: 10.3389/fmed.2020.00027.
- CÁCERES, Camilo, Joao Mauricio Rosário y Dario Amaya (2019). "Towards Health 4.0: e-Hospital Proposal based industry 4.0 and Artificial Intelligence Concepts". En *Artificial Intelligence in Medicine. 17th Conference on Artificial Intelligence in Medicine, AIME 2019 Poznan, Poland, June 26-29, 2019 Proceedings*, editado por David Riaño, Szymon Wilk y Annette ten Teije, 84-89. Springer.

- CHANG, Anthony (2019). “Common Misconceptions and Future Directions for AI in Medicine: A Physician-Data Scientist Perspective”. *En Artificial Intelligence in Medicine. 17th Conference on Artificial Intelligence in Medicine, AIME 2019 Poznan, Poland, June 26-29, 2019 Proceedings*, editado por David Riaño, Szymon Wilk y Annette ten Teije, 3-6. Springer.
- CHIN-YEE, Benjamin y Ross Upshur (2019). “Three Problems with Big Data and Artificial Intelligence in Medicine”. *Perspectives in Biology and Medicine* 62, n.º 2: 237-256. DOI:10.1353/pbm.2019.0012.
- COHEN, Jacob, Patricia Cohen, Leona S. Aiken y Stephen G. West (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.
- DECAMP, Matthew y Charlotta Lindvall (2020). “Latent bias and the implementation of artificial intelligence in medicine”. *J Am Med Inform Assoc.* 27, n.º 12: 2020-2023. DOI: 10.1093/jamia/ocaa094.
- DELOITTE (2019). “Global artificial intelligence industry White paper”. <https://www2.deloitte.com/cn/en/pages/technology-media-and-telecommunications/articles/global-ai-development-white-paper.html>.
- DENDUKURI, Nandini y Lawrence Joseph (2001). “Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests”. *Biometrics* 57, n.º 1: 158-167.
- DIGGING DEEPER: PLANT VILLAGE (2020). “WPSU Penn state video”. <https://news.psu.edu/video/426943/2016/09/20/wpsu-penn-states-digging-deeper-plantvillage>.
- FILICE, Ross, Anouk Stein, Carol C. Wu, Veronica A. Arteaga, Stephen Borstelmann, Ramya Gaddikeri, Maya Galperin-Aizenberg, Ritu R. Gill, Myrna C. Godoy, Stephen B. Hobbs, Jean Jeudy, Paras C. Lakhani, Archana Laroia, Sundeep M. Nayak, Maansi R. Parekh, Prasanth Prasanna, Palmi Shah, Dharshan Vummidi, Kavitha Yaddanapudi y George Shih (2020). “Crowdsourcing pneumothorax annotations using machine learning

- annotations on the NIH chest X-ray dataset”. *Journal of digital imaging* 33, n.º 2: 490-496.
- FERRYMAN, Kadija (2020). “Addressing health disparities in the Food and Drug Administration’s artificial intelligence and machine learning regulatory framework”. *Journal of the American Medical Informatics Association* 27, n.º 12: 2016-2019.
- GAMEZ, María José (2015). “Objetivos y metas de desarrollo sostenible, Desarrollo Sostenible”. Acceso el 10 de enero de 2024. <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>.
- GORALSKI, Margaret y Tay Keong (2020). “Artificial intelligence and sustainable development”. *The International Journal of Management Education*, 18, n.º 1. <https://doi.org/10.1016/j.ijme.2019.100330>.
- HAUGELAND, John (ed.) (1997). *Mind design II: philosophy, psychology, artificial intelligence*. The MIT Press.
- HOFMANN, Markus y Ralf Klinkenberg (eds.) (2016). *RapidMiner: Data Mining use cases and business analytics applications*. Boca Raton: CRC Press.
- INNIS, Robert E. (2009). *Susanne Langer in focus: The symbolic mind*. Bloomington: Indiana University Press.
- JOHANNES, Bronkhorst (2010). *Ritual, holophrastic utterances, and the symbolic mind*. Wiesbaden: Harrassowitz, 159-202.
- KELLY, Christopher, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado y Dominic King (2019). “Challenges for delivering clinical impact with artificial intelligence”. *BMC Med.* 17: 195. doi:10.1186/s12916-019-1426-2.
- KRUSKAL, Jonathan, Seth Berkowitz, J. Raymond Geis, Woojin Kim, Paul Nagy y Keith Dreyer (2017). “Big data and machine learning-strategies for driving this bus: a summary of the 2016 intersociety summer conference”. *Journal of American College of Radiologists* 14, n.º 6: 811-817 DOI:10.1016/j.jacr.2017.02.019.

- LANT, Karla (2020). "Clean Water AI Puts Prevention of Waterborne Disease In The Palm Of Your Hand". *Environmental Monitor*. <https://www.fondriest.com/news/clean-water-ai-puts-prevention-waterborne-disease-palm-hand.htm>.
- LECUN, Yann, Yoshua Bengio y Geoffrey Hinton (2015). "Deep learning". *Nature* 521, n.º 7553: 436-444.
- MURRAY, Dan (2013). *Tableau your data!: fast and easy visual analysis with tableau software*. John Wiley & Sons.
- NEVADO CABELLO, María Victoria (2010). *Introducción a las bases de datos relacionales*. Madrid: Vision Libros.
- PEARL, Judea (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.
- PERAL, Jesús, Alejandro Maté y Manuel Marco (2017). "Application of Data Mining techniques to identify relevant key performance indicators". *Computer Standards & Interfaces* 54: 76-85.
- PRONACES (2024). *Programas Nacionales Estratégicos*. Consejo Nacional de Humanidades, Ciencias y Tecnologías. Gobierno de México. Acceso el 9 de enero de 2024. <https://conahcyt.mx/pronaces/>.
- RECHT, Michael y R. Nick Bryan (2017). "Artificial intelligence: threat or boon to radiologists?". *Journal of American College of Radiologists* 14, n.º 11: 1476-1480. DOI:10.1016/j.jacr.2017.07.007.
- REDDY, Sandeep (2018). "Use of Artificial Intelligence in Healthcare Delivery". En *eHealth. Making Healthcare Smarter*, editado por Thomas Heston, 81-97. Reino Unido: IntechOpen.
- ROBBERT, Mary-Ann y Linda Senne (2003). "Teaching GIGO: Data Quality in the Curriculum". *AMCIS 2003 Proceedings*, 296. <http://aisel.aisnet.org/amcis2003/296>.
- SANTOSO, Heru, Eko Hari Rachmawanto, Adhitya Nugraha, Akbar Aji Nugroho, De Rosal Ignatius Moses Setiadi y Ruri Suko Basuki (2020). "Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization". *Telkomnika* 18, n.º 2: 799-806.

- SEEBO (2019). "Artificial Intelligence - The Driving Force of Industry 4.0 [White paper]". <https://www.augury.com/blog/industry-insights/artificial-intelligence-the-driving-force-of-industry-4-0/>.
- WU, Chieh-Chen, Wen-Chun Yeh, Wen-Ding Hsu, Md. Mohaimenul Islam, Phung Anh Nguyen, Tahmina Nasrin Poly, Yao-Chin Wang, Hsuan-Chia Yang y Yu-Chuan Li (2019). "Prediction of fatty liver disease using machine learning algorithms". *Computer methods and programs in biomedicine* 170: 23-29.

## **Sobre los autores**

### **Dra. Karina Gutiérrez Fragoso**

Cuenta con una licenciatura en Informática por el Instituto Tecnológico de Pachuca. A su vez, tiene una Maestría en Inteligencia Artificial por la Universidad Veracruzana. Pertenece al Comité para la elaboración y validación de reactivos del Examen General para el Egreso de la Licenciatura en Ingeniería de Software EGEL-ISOFT desde 2015. Su área de investigación es el procesamiento digital de imágenes y aprendizaje computacional. Tiene diferentes proyectos en proceso y uno concluido llamado Identificación de microcalcificaciones en imágenes de mamografía para detección de cáncer de mama mediante la aplicación de técnicas de procesamiento digital de imagen. Ha impartido distintos cursos de Fundamentos de investigación, Arquitectura de computadoras, Inteligencia Artificial, Ingeniería de Software y Simulación de sistemas programables, entre otros.

### **Dr. Elías Ruiz Hernández**

Es egresado del Instituto Tecnológico de Minatitlán en la Licenciatura en Ingeniería en Sistemas Computacionales. Tiene una maestría y un doctorado en Ciencias Computacionales por el Instituto Nacional de Astrofísica, Óptica y Electrónica. Es miembro de la Red de Inteligencia Computacional Aplicada. Sus áreas de investigación son el Aprendizaje Automático y aplicaciones, visión computacional, modelos gráficos probabilistas, cómputo inteligente y representación

del conocimiento, Ciencia de datos, Minería de Datos, gramáticas visuales. Ha impartido cursos de Integración de tecnologías de la información, calidad de software, y bases de datos y bases de conocimientos.

**Dr. German Cuaya Simbro**

Tiene una maestría y un doctorado en Ciencias Computacionales por el Instituto Nacional de Astrofísica, Óptica y Electrónica. También tiene una licenciatura en Ciencias Computacionales por la Benemérita Universidad Autónoma de Puebla. Sus áreas de investigación son Aprendizaje Automático, aprendizaje profundo, computación cognitiva y procesamiento de bioseñales y computación médica. Tiene una larga trayectoria en la que ha impartido distintos cursos, participado en congresos y seminarios, cuenta con diversas certificaciones IBM, así como proyectos y publicaciones. Es colaborador con el grupo de investigación de análisis de movimiento del INR y colaborador con grupo de investigación. Es socio en empresas de IT.

# Respuestas emocionales en la interacción con dispositivos tecnológicos

GLORIA ADRIANA MENDOZA FRANCO

## Resumen

Los dispositivos interactivos están cada vez más presentes en la vida cotidiana y su diversidad es cada vez mayor. Los robots son los más recientes de ellos y de su interacción con personas se encarga la disciplina denominada Interacción Humano Robot. Derivado de ella, surge el problema de las emociones involucradas en esa relación entre dispositivos robóticos y seres humanos. Aquí se introduce el tema en general y se presenta un par de casos donde se abordan las emociones desde perspectivas distintas: la evaluación y la experiencia artística. Finalmente, se reflexiona sobre el estado actual de avance en el conocimiento en torno a las emociones como una cuestión interdisciplinaria, y el futuro desarrollo de robots que ayuden a resolver problemáticas sociales.

## Introducción

En nuestro mundo tecnologizado, los dispositivos interactivos comienzan a inundar nuestra vida cotidiana. La variedad de

dispositivos interactivos que nos rodea aumenta constantemente, iniciando por los teléfonos inteligentes, consolas de videojuego, quioscos de compras electrónicas, asistentes de comando por voz personales, de casa o automóvil, dispositivos de internet de las cosas (IoT, por sus siglas en inglés) y, de manera más reciente, robots. Disciplinas como el diseño de interacciones y el diseño de experiencias se han enfocado al desarrollo y evaluación de las formas en que nos comunicamos con estos dispositivos para buscar que los usuarios logren sus objetivos de manera eficiente y agradable al hacer uso de ellos. Es claro que las emociones juegan un papel muy importante en el proceso, no sólo porque se busca desarrollar experiencias que sean agradables y despierten estados de ánimo positivos, sino porque afectan nuestra memoria, atención y personalidad (Lim *et al.* 2008). Estos impactos pueden ser más relevantes en tecnologías de reciente introducción y en pleno desarrollo como son los robots sociales, en los que centraré este capítulo.

El objetivo de este texto es analizar de una forma más profunda el rol que juegan las emociones en nuestra vida cotidiana y en nuestros procesos cognitivos con el fin de bosquejar algunos caminos alternos para la inclusión del componente emocional en el desarrollo de robótica social que permita desarrollar aplicaciones que beneficien de múltiples formas a los usuarios y el contexto social. Sin embargo, este análisis es completamente extensible al desarrollo de cualquier dispositivo interactivo.

La primera parte de este texto presenta generalidades de la disciplina Interacción Humano Robot (HRI, por sus siglas en inglés) y la importancia que el estudio de las emociones tiene dentro de ella. Después se desglosan algunos de los cuestionamientos más importantes dentro de las ciencias afectivas con el objetivo de mostrar el estado del arte en los estudios emocionales para, posteriormente, con ejemplos derivados de las ciencias afectivas, proponer nuevos enfoques en el desarrollo de robótica social con una perspectiva emocional. En la tercera y cuarta parte

del texto se presentan dos proyectos dentro de la disciplina de HRI que abordan el problema de las emociones desde dos perspectivas distintas: la evaluación y la experiencia artística. Estos ejemplos ayudan a mostrar la necesidad de enfoques metodológicos inter y transdisciplinarios para abordar las emociones dentro de la disciplina. Por último, se abordan algunos de los problemas éticos que trae consigo el desarrollo de tecnología emocional. Derivado de los cuestionamientos éticos surge también la responsabilidad ética que tiene el campo para el desarrollo de tecnología para el futuro. El desarrollo de tecnología enfocada en los procesos emocionales humanos tiene un enorme potencial para proponer soluciones a problemas de salud, de aprendizaje, de desempeño, incluso de discriminación. Así, es necesario acortar la brecha entre la teoría y la práctica, y promover la colaboración entre disciplinas a fin de que los efectos de la inclusión de componentes emocionales en la tecnología dejen de ser indicadores de satisfacción y se conviertan en indicadores de bienestar.

### **De una disciplina llamada Interacción Humano-Robot**

Uno de los campos de aplicación de la Inteligencia Artificial (IA) es la robótica, o, mejor dicho, parte de la IA se desarrolla para ser implementada en sistemas robóticos. Las aplicaciones de estos sistemas son muy variadas; de entre ellas las más conocidas son en sistemas de fabricación y ensamblaje industrial, robots de búsqueda y rescate, robots exploradores, robots de limpieza y, últimamente, los llamados robots sociales.

Es esta última categoría de la que estaremos hablando a lo largo de este capítulo. La disciplina HRI es el campo dedicado a los robots denominados sociales, ya que se trata de aquellos robots que fueron diseñados para interactuar con humanos de forma cotidiana. Una definición más exacta de la HRI sería el campo dedicado al entendimiento, diseño y evaluación de sistemas robóticos para su uso por y con humanos (Goodrich y Schultz 2007). Se trata de un campo interdisciplinario donde se requiere

de aportaciones de expertos en el área de la ingeniería, el diseño, la antropología y la psicología, entre otras, y que se encuentra en constante crecimiento desde su consolidación a finales de la década de 1990 (Goodrich y Schultz 2007).

De forma general, el principal problema que la HRI intenta resolver es desarrollar robots que sean eficientes, pero también aceptables para la gente; que cubran las necesidades sociales y emocionales de sus usuarios, además de respetar los valores humanos (Dautenhahn 2013). Algunas de las aplicaciones potenciales de los robots sociales en el futuro son robots de servicio en establecimientos comerciales, como restaurantes, hoteles o tiendas. Pero también algunas otras aplicaciones que requieren no sólo de interacción social, sino de un profundo entendimiento de la conducta humana, como robots profesores o niñeras que estarán en contacto con niños y tendrán una influencia en la crianza y desarrollo de habilidades sociales (Deshmukh *et al.* 2019, Han *et al.* 2005, Tanaka *et al.* 2006, Westlund *et al.* 2016). También se plantea el desarrollo de robots con fines terapéuticos para personas que sufren deterioros cognitivos o discapacidades mentales (Balasuriya *et al.* 2019; Freire, Valente y Filipe 2020, Shamsuddin *et al.* 2012); así como robots virtuales con fines de intervención terapéutica y estudio de fenómenos sociales como el prejuicio (Tassinari, Auallbach y Jasinskaja-Lahti 2022), estrés post-traumático (Han *et al.* 2021, Tielman *et al.* 2017) o salud mental en general (Ahn y Fox 2017; Yokotani, Takagi y Wakashima 2018).

De acuerdo con esta definición del problema, es evidente que para diseñar robots que satisfagan necesidades sociales y emocionales es indispensable poder definir con claridad dichas necesidades. De hecho, una gran parte del trabajo desarrollado dentro de la disciplina está enfocado en encontrar dichas necesidades a través de la evaluación de los sistemas en ambientes cercanos al contexto real de uso.

Kerstin Dautenhahn recopiló una serie de distintas definiciones de lo que es un robot social y las características que distinguen

a este tipo de robots, planteadas por diversos investigadores, y que en su conjunto muestran diferentes formas en que un robot puede ser considerado social (Dautenhahn 2007):

*Socialmente evocativo*: robots que provocan emociones debido a su aspecto antropomorfizado; los humanos tienden a cuidar o criar a su “creación” (Breazeal, 2004).

*Socialmente situado*: robots dentro un entorno social que pueden percibir y al que reaccionan. Son capaces de distinguir a otros agentes sociales y objetos varios en su entorno (Fong, Nourbakhsh y Dautenhahn 2003).

*Sociable*: robots que propician relacionarse con las personas para satisfacer necesidades sociales internas (por ejemplo, emociones). Estos robots requieren de modelos avanzados de cognición social (Breazeal, 2003).

*Socialmente inteligentes*: Robots que muestran rasgos de inteligencia social humana; cuentan con modelos de cognición humana y competencia social avanzados (Dautenhahn, 1998).

*Robots socialmente interactivos*: para estos robots la interacción social es básica para la HRI de igual a igual, es decir, comparten el espacio con los usuarios (Fong, Nourbakhsh y Dautenhahn 2003).

En cada una de las definiciones anteriores se enfatiza la interacción social como parte de la tarea principal u objetivo de un robot social. Una interacción humano-robot exitosa es lograda si se atienden de forma global las necesidades humanas relacionadas con nuestra capacidad de interacción y de creación y mantenimiento de lazos sociales entre humanos. Las emociones juegan un papel fundamental en este contexto, ya que es a través de ellas que logramos regular las interacciones sociales, por lo tanto, también estar en contacto con robots sociales.

De entre los robots que se han desarrollado con fines emocionales encontramos a PARO, un robot foca bebé pensado como herramienta terapéutica para adultos mayores con deterioro cognitivo. PARO ha demostrado disminuir la depresión y el sentimiento de soledad, además de mejorar la socialización en casas de retiro, e

incluso mejorar el estado de salud de los usuarios (Chang, Šabanovic y Huber 2014; Chang y Šabanović 2015). El robot se ha probado también con adultos mayores sanos, y ha mostrado efectos positivos posteriores a la interacción (McGlynn *et al.* 2017).

Otro muy claro ejemplo del desarrollo de robots emocionales es el caso de interacciones robóticas con niños dentro del espectro autista. Se ha encontrado, a través de las investigaciones en el campo, que los niños autistas muestran entusiasmo hacia los robots (Begum, Serna y Yanco 2016; Hamid *et al.* 2013; Kim *et al.* 2012; Sammonds, Arora y Arora 2022), razón por la cual se propone su uso para enseñar y evaluar habilidades comunicativas y de interacción social. Las herramientas robóticas propuestas van desde robots muy sofisticados hasta herramientas robóticas más simples, como peluches con movimientos motorizados, e incluso agentes digitales o robots virtuales. La gran variedad de posibles soluciones robóticas demuestra que a pesar de ser una solución de alta tecnología, podría ser accesible a muchos sectores de la población. Los resultados de las intervenciones robóticas en niños dentro del espectro autista son muy prometedoras y parece que sólo es cuestión de tiempo para el desarrollo de protocolos estándar que permitan la implementación de este tipo de robots de forma cotidiana.

Finalmente, el tercer ejemplo que quiero discutir es el uso de agentes virtuales para estudiar o intervenir fenómenos sociales complejos. Este tipo de agentes se han utilizado para hacer investigación en temas como trastornos alimenticios, ansiedad social o síndrome de estrés post-traumático. El uso de robots virtuales en estos temas permitirá llevar a cabo experimentación de forma más precisa y objetiva (Ahn y Fox, 2017, Tielman *et al.* 2017), además de que se pueden implementar intervenciones terapéuticas dentro de entornos virtuales. Por ejemplo, se pueden diseñar robots virtuales que actúen como reclutadores de recursos humanos de una empresa y lleven a cabo una simulación de entrevista con una persona que está tratando de superar su fobia social

(Emmelkamp, Meyerbröker y Morina 2020). Estas aplicaciones permiten simular situaciones humanas de forma sistemática y repetitiva haciendo posible intervenciones que de otra forma son muy costosas y requieren de logística muy demandante.

Todos los casos anteriormente descritos requieren de un claro entendimiento de las emociones humanas como fenómeno. Así que el estudio de las emociones es un área profundamente explorada en la HRI; pues es necesario que un robot social comprenda, reaccione y genere emociones en su interlocutor humano.

Derivado de este entendimiento de la importancia que tienen las emociones en las relaciones sociales, diversos estudios y proyectos se han enfocado en el diseño y desarrollo de robots capaces de identificar estados emocionales, expresar o simular emociones, provocar respuestas emocionales e incluso modificar su comportamiento en función de la respuesta emocional de los usuarios. Sin embargo, al recorrer estos proyectos y la diversidad de aplicaciones planteadas, parecen surgir muchas preguntas respecto a la relación robots- emociones, algunas incluso de carácter ético. Ejemplos de ellas podrían ser: ¿los robots necesitan expresar emociones? Si es así, ¿qué tipo de emociones?, ¿existen otras formas de estimular ciertas respuestas emocionales?, ¿existen otras necesidades emocionales que van más allá de la interacción social? O incluso, como Kate Darling lo describió en una de sus conferencias en 2016, un dispositivo que puede atraerte podría también manipularte: “another word for an engaging technology is a manipulative technology” (Darling 2016a), ¿cuáles podrían ser entonces las implicaciones éticas de lograr establecer un lazo emocional con un robot? Para tratar de dar respuesta a estas preguntas habremos de ahondar primero en lo que son las emociones y sus funciones.

### **De la dimensión emocional humana**

Durante mucho tiempo disciplinas como la filosofía, la psicología o la sociología se han enfrentado al reto de estudiar las emociones,

definirlas, caracterizarlas y entender sus muchas funciones e impactos en otros procesos cognitivos. De forma más reciente, interdisciplinas llamadas *ciencias afectivas* han comenzado a colaborar con el objetivo de comprender, modelar y predecir respuestas emocionales; disciplinas como la psicología, la neurociencia, la filosofía, la economía, la historia, la sociología y las ciencias computacionales son partícipes de este proceso (Coppin y Sander 2016). Sin embargo, lejos de comprender a profundidad el fenómeno, ha sido cada vez más evidente lo poco que se sabe sobre los procesos emocionales y su rol en la vida cotidiana. A diferencia de muchos otros procesos cognitivos (como la toma de decisiones, atención o el lenguaje), el estudio de las emociones es incipiente, no porque no fuera de interés científico antes, sino porque el nivel de complejidad que representa no ha permitido un avance más rápido en su estudio. Es hasta la aparición de la neurociencia y del uso de las técnicas de imagen para estudiar el cerebro que se ha podido progresar de forma importante en el campo (Feldman, 2017). Muestra de ese rezago es la falta de un consenso respecto a la definición del concepto base *emoción*.

Sin intención de profundizar mucho en el tema, ya que es muy amplio, pretendo señalar los aspectos más importantes relacionados con el problema de definir lo que es una emoción, ya que nos permitirá más adelante bosquejar posibilidades de desarrollo en el campo de la HRI y otras tecnologías interactivas.

Si intentamos buscar una definición de la palabra “emoción”, seguramente encontraremos una gran variedad de propuestas que van desde la clásica definición de diccionario: “Alteración del ánimo intensa y pasajera, agradable o penosa, que va acompañada de cierta conmoción somática” (RAE 2023), hasta libros enteros que muestran las muchas posibles definiciones desde diversas disciplinas y teorías. No poder definir un fenómeno pone en gran riesgo el desarrollo científico (Feldman 2016) ligado a dicho fenómeno al no permitir la colaboración entre expertos de diferentes disciplinas (Scherer 2005) y, en consecuencia, tampoco el

desarrollo tecnológico. La falta de acuerdos entre teóricos y académicos en la definición de la palabra “emoción” provoca que las teorías de la emoción, lejos de complementarse, debatan incansablemente por intentar definir el concepto (algunos ejemplos se pueden revisar en Lindquist *et al.* [2013], Izard [2007], Panksepp y Watt [2011] o en Barret [2019]).

Estos debates provocan falta de claridad en la investigación aplicada que busca rastrear y caracterizar respuestas cerebrales, fisiológicas, psicológicas y conductuales ligadas a los procesos emocionales, ya que deben apegarse a algún modelo teórico, lo que fragmenta sus posibles aportes. Esta fragmentación existente en el campo permea hasta la investigación tecnológica, donde se busca utilizar el conocimiento científico para generar aplicaciones que permitan resolver problemas. Y es que, ¿cómo podemos planear el desarrollo de tecnología “emocional” si no hay un consenso en lo que la palabra “emocional” pudiera significar o implicar? Uno de los objetivos de este texto es presentar un marco de referencia que permita identificar las muchas posibilidades de desarrollo de tecnología que dialogue con los procesos emocionales.

Dentro de los acuerdos que sí existen en torno a la definición de la *emoción*, se entiende que se trata de un evento con una duración determinada y que consiste en dos procesos: 1) un mecanismo que provoca una emoción al enfrentarse a eventos relevantes y 2) la configuración de una respuesta emocional con diversos componentes (tendencia a la acción, reacción fisiológica, expresión y sentimiento) (Coppin y Sander 2016). Otra forma de definir el concepto es a través de sus componentes, Scherer propone cinco (en Mulligan y Scherer 2012):

- Cognitivo o evaluación: cada emoción surge a partir de una evaluación del contexto. Cuando una persona se hace consciente de su respuesta emocional a un estímulo, puede calificar dicho estímulo como bueno/malo, peligroso, agradable o sorprendente.

- Neurofisiológico: se refiere a la respuesta fisiológica ligada a cada estado emocional. Cambios en la frecuencia cardiaca, frecuencia respiratoria, actividad eléctrica de la piel, contracción muscular, etcétera, son ejemplos de este componente.
- Motivacional: cada emoción conlleva cierta motivación a la acción; hay algunas emociones que despiertan una gran necesidad de acción (como la euforia), mientras otras provocan más bien la pasividad (como la melancolía).
- De expresión: éste es probablemente el componente más estudiado de las emociones y se refiere a las expresiones faciales, posturales y vocales asociadas a cada emoción.
- De sentimientos subjetivos: se refiere propiamente a la experiencia emocional, a la forma como cada uno siente y se hace consciente de su propio estado emocional.

Esta definición por componentes ha permitido el avance de las ciencias afectivas en el desarrollo de técnicas para medir y caracterizar las emociones. Además de sus componentes, entender cuáles son sus funciones también es crucial, ya que se trata de un fenómeno que permea una gran parte de nuestra vida. Algunos científicos proponen que las emociones interfieren en el razonamiento, la toma de decisiones, las acciones, las relaciones sociales, el bienestar y la salud (Izard 2010). Tooby y Cosmides hacen un análisis bastante completo de todos los procesos cognitivos que pudieran ser dirigidos por las emociones, de hecho, ellos proponen que son precisamente las emociones las que dirigen y coordinan todas nuestras respuestas y comportamientos frente a los retos que encontramos en el entorno. Algunos de los procesos cognitivos que probablemente son comandados por las emociones son (Tooby y Cosmides 2008):

- El establecimiento de objetivos
- Las motivaciones que permiten asignar prioridades
- La motivación por explorar

- Los marcos conceptuales fundamentales
- Los mecanismos de la percepción
- La memoria
- La atención
- Las respuestas fisiológicas
- La comunicación y las expresiones emocionales
- El comportamiento
- La capacidad de hacer inferencias
- Los reflejos
- El aprendizaje
- Los estados de ánimo, niveles de energía, administración de los esfuerzos
- La depresión

De manera general los teóricos de la emoción están divididos en tres grandes grupos (Coppin y Sander 2016). Uno de ellos defiende que las emociones pueden ser descritas en grandes categorías “básicas” que incluso pudieran ser universales y que son innatas, es decir, están inscritas en algún lugar en nuestros genes o cerebro (ejemplo: felicidad, tristeza, enojo, etcétera). El segundo grupo prefiere identificar a las emociones como resultado de un proceso de evaluación del entorno; en este caso también se consideran como mecanismos con los que nacemos (ejemplo: estrés, sorpresa, miedo). Por último, el tercer grupo se autodenomina constructivista y considera que las emociones son mecanismos que desarrollamos como fruto de las experiencias que afrontamos en nuestro día a día, es decir, están fuertemente influenciadas por el entorno, así que, aunque dos personas califiquen su estado emocional como “enojado”, en realidad, no están experimentando el mismo estado emocional, y no lo expresarán de la misma forma. En este último grupo, las emociones se representan más bien en modelos dimensionales y no en categorías absolutas (ejemplo de posibles dimensiones: activación, valencia, intensidad).

Es importante mencionar que, en realidad, lo más probable es que todas las teorías sean ciertas. Como ya se mencionó antes, las emociones son un fenómeno altamente complejo, que tal vez se encuentra en todas las dimensiones de la consciencia humana, hace uso de muchos sistemas corporales y tiene múltiples componentes. Así que cada teoría podría estar abarcando sólo una parte del fenómeno, y entre todas estarían abarcando de forma más completa el estudio de las emociones. Así que todos aquellos que no estamos enfocados en construir los modelos teóricos de la emoción, sino en la evaluación de las respuestas emocionales y en el desarrollo de aplicaciones enfocadas en las experiencias emocionales, utilizaremos uno u otro modelo teórico, en función de los elementos que nos interesa estudiar (Coppin y Sander 2016).

Posiblemente en un futuro tendremos un mapeo global de las emociones, similar al mapeo del genoma humano, sin embargo, mientras llega ese día, tenemos que adaptarnos y utilizar la información disponible.

Para finalizar este apartado dedicado a explicar el estado del arte en el estudio de las emociones y sobre todo la importancia que las mismas tienen en la vida de cualquier persona, quiero mostrar tres ejemplos provenientes de diferentes disciplinas que explican con claridad las relaciones que hay entre las emociones y otros procesos cognitivos y sociales.

El primer ejemplo es fruto de las investigaciones de Jaak Panksepp y sus colaboradores. Panksepp es particularmente reconocido por el descubrimiento de siete circuitos emocionales en el cerebro de los mamíferos. Lo que muestra en primer lugar que, en efecto, nacemos con al menos estos siete mecanismos emocionales que, por cierto, son fundamentales para nuestro desarrollo, y en segundo lugar que todos los mamíferos compartimos dichos circuitos, es decir, los seres humanos no somos los únicos animales con emociones. Los sistemas emocionales de Panksepp son: SEEK (o el impulso por buscar, explorar, conocer el mundo), PLAY (el juego social), CARE (el amor materno),

PANIC (la desolación, el temor a estar solos), FEAR (miedo de lo que nos pueda dañar), RAGE (la ira relacionada a la frustración) y LUST (el deseo sexual).

Voy a profundizar sólo en el sistema emocional PLAY, que, aunque se mantiene a lo largo de toda la vida, está presente con mayor fuerza durante la niñez. Una imagen clásica del PLAY sería un par de niños jugando: el juego rápidamente se vuelve rudo, de repente los niños se están revolcando en el piso y, de un momento a otro, uno de ellos sale llorando por un golpe que fue más fuerte de lo que esperaba. Este tipo de “juego rudo” es el que Panksepp califica como un juego social (Panksepp, Siviy y Normansell 1984), ya que se trata de un juego que es necesario para desarrollar la capacidad de establecer relaciones sociales y mediar la forma en que nos relacionamos con los demás; afectando también nuestras percepciones, pensamientos y memorias (Davis y Panksepp 2011). El juego social permite que los niños aprendan las dinámicas sociales, a partir del CARE por los otros y motivados por el SEEK para conocer el mundo (Panksepp 2007). Una vez entendido este proceso, resulta evidente no sólo la importancia del juego, sino que las condiciones en las que éste se realiza, el contexto cultural o la frecuencia de juego son factores que pueden tener una enorme repercusión en el desarrollo de habilidades sociales e incluso de nuestra personalidad (Panksepp 2008). Existe la teoría de que el contexto cultural del juego moldea las relaciones sociales en ese contexto; incluso podría ser una de las causas de las diferencias en habilidades sociales presentadas entre géneros (Fry 2005).

Desde un punto de vista clínico hay evidencia que sugiere que la falta de juego social, y en consecuencia de estimulación del PLAY en la niñez, puede derivar en déficit de atención o incluso en desórdenes de personalidad (Panksepp 2007). Existe particularmente un interés creciente en el estudio del juego social y su rol en el tratamiento del déficit de atención, como alternativa a la medicación, así como también en el desarrollo de terapias para

la adquisición de habilidades sociales en niños dentro del espectro autista (Panksepp 2007).

Este ejemplo nos muestra cómo los sistemas emocionales son fundamentales para el desarrollo humano, para el establecimiento de nuestras relaciones sociales, e incluso para el tratamiento de algunas condiciones neuro-diversas. Dentro de la HRI existe una tendencia a desarrollar aplicaciones en robots sociales que favorezcan, a modo de terapia, el desarrollo de habilidades sociales en niños dentro del espectro autista (Cabibihan *et al.* 2013). Dichos estudios se llevan a cabo siguiendo un proceso de prueba y error para tratar de encontrar aquellos escenarios (tipos y formas de robots, formas de interacción, número de intervenciones) que son más favorecedores. Lo anterior al utilizar una aproximación basada en el desarrollo tecnológico de los robots y no en las necesidades clínicas de los pacientes (Kim *et al.* 2012).

Este enfoque no permite la generalización de los resultados, ya que cada proyecto individual utiliza el robot al que tiene acceso, diseña sus propias interacciones, hace pruebas con participantes con características distintas y evalúa los resultados también de forma distinta (Begum, Serna y Yanco 2016). Un mejor entendimiento del rol que tiene el sistema emocional PLAY en este contexto podría ayudar a sistematizar la investigación, así como favorecer a la extensión de este tipo de investigaciones a otras poblaciones con necesidades similares como niños con parálisis cerebral severa o niños en ambientes riesgosos y aislados.

El segundo ejemplo está relacionado con el papel que las emociones tienen en el desarrollo de algunos procesos cognitivos como la toma de decisiones y resolución de problemas. Este campo de estudio ha encontrado evidencia que señala un impacto de las emociones “no intensas” cotidianas en los procesos de pensamiento (Isen 2008). Es decir, no es necesario experimentar una emoción negativa intensa (como el miedo o la tristeza) para observar un impacto en diversos procesos de pensamiento. La presencia de emociones positivas ayuda a memorizar ciertos eventos con

facilidad (Teasdale y Fogarty 1979), pero también ayuda a tener un pensamiento más flexible, lo que beneficia el pensamiento creativo y la resolución de problemas (Fernández-Abascal y Martín 2013). Hay estudios que demuestran que las emociones positivas ayudan a las personas a ser pensadores más flexibles y mejores tomadores de decisiones, ya que incrementa las habilidades para atender a diversas situaciones, lo que propicia respuestas o soluciones más eficientes ante circunstancias complejas y cambiantes (Isen 1999).

Estos estudios demuestran la influencia que las emociones cotidianas pueden llegar a tener en los procesos de innovación, así como en el desempeño en tareas que requieren de pensamiento flexible o de respuestas creativas. Además, también sugiere que las emociones negativas podrían afectar de forma negativa procesos como la atención y en consecuencia el desempeño en tareas que requieren de ésta. Asimismo, estos hallazgos demuestran que dentro del diseño de experiencias las emociones tienen un papel más relevante que simplemente ser indicadores de satisfacción. Ahora bien, dentro de la disciplina de la HRI no se ha realizado mucho trabajo serio en torno al rol que las emociones dentro de una experiencia de interacción con un robot puedan llegar a tener en cuanto al desempeño de una tarea realizada por el humano. Existe una idea de que los robots sociales agregan un componente atrayente a ciertos contextos como el aprendizaje (Westlund *et al.* 2016) o rehabilitación (Begum, Serna y Yanco, 2016), logrando así mejores resultados de desempeño, sobre todo con poblaciones infantiles. Sin embargo, poco se ha ahondado en el tipo de emociones que requieren ser experimentadas para tener los mejores resultados, los escenarios de interacción que llevan a dichas emociones o en cómo evitar tener efectos adversos en interacciones a largo plazo (Seibt 2016).

El último ejemplo corresponde de nuevo al rol que tienen las emociones en las interacciones sociales (en esta ocasión, me refiero no a las emociones individuales sino a las grupales). Se

considera a una emoción como grupal cuando surge dentro de un grupo y es compartida por sus miembros. Los integrantes de éstos comparten rasgos identitarios como tener el mismo origen étnico, las mismas creencias religiosas o incluso el mismo género. Las emociones grupales permiten no sólo mantener el sentido de identidad dentro del grupo, sino regular las interacciones con otros grupos (Reynolds y Klik 2016). Este mecanismo permitía a los primeros humanos mantenerse alerta contra amenazas provenientes de otros grupos humanos que pudieran poner en riesgo sus recursos o supervivencia. El contexto social actual es muy distinto al de entonces, sin embargo, el mecanismo emocional para regular las relaciones intergrupales es el mismo.

Dentro de los posibles efectos de este mecanismo de regulación está el surgimiento de los prejuicios y discriminación entre grupos (Talaska, Fiske y Chaiken 2008). Hay evidencia que muestra que tanto las acciones positivas para mantener a un grupo unido (empatía) como las negativas hacia otros grupos que puedan parecer amenazantes (antipatía) están reguladas por respuestas emocionales compartidas entre los miembros del mismo grupo (Mendes *et al.* 2002). Si bien, no se han encontrado con claridad las emociones que despiertan, tanto de forma consciente como inconsciente, los prejuicios entre diferentes grupos, es claro que el estudio de éstas permitirá en un futuro generar estrategias específicas para minimizar la discriminación y promover una sociedad más incluyente (Han 2018).

La investigación en HRI en torno a la empatía está dirigida principalmente a desarrollar robots que simulen empatía hacia los humanos, o robots que despierten respuestas empáticas (Paiva *et al.* 2017). Sin embargo, poco se ha explorado en el desarrollo de aplicaciones que intenten promover la empatía entre seres humanos, tal vez, regulada o mediada por la tecnología, un ejemplo sería el presentado por Zuckerman (Zuckerman y Hoffman 2015).

El entendimiento de las respuestas emocionales grupales como reguladores de las relaciones intergrupales permitirá el diseño de experiencias emocionales que ayuden al desarrollo de respuestas empáticas hacia los otros.

Estos tres ejemplos nos muestran que las respuestas emocionales son fenómenos transversales, es decir, están relacionadas con todo lo que hacemos, pensamos, creemos y somos. Es por ello que deben ser consideradas no sólo como indicadores de la experiencia, sino como objeto de estudio y objetivo de las mismas experiencias. Por supuesto que ello implica un nivel más alto de trabajo interdisciplinario. Sin embargo, la HRI es una disciplina nacida dentro de este tipo de contexto, así que, a diferencia de otras, debiera poder dialogar con mayor facilidad con otras aproximaciones teóricas y metodológicas.

### **El problema de la evaluación**

Uno de los obstáculos más importantes en el avance de las investigaciones que intentan agregar componentes emocionales en la interacción con dispositivos tecnológicos, y en particular en la disciplina de la HRI, es la falta de métodos confiables para la evaluación de las respuestas emocionales (Mendoza 2015). La gran mayoría de los proyectos utiliza herramientas etnográficas como observaciones, entrevistas y cuestionarios para evaluar las respuestas emocionales. Si bien estos instrumentos son valiosos para recuperar información relevante para interpretar las respuestas emocionales, no permiten la replicación de los resultados (Dautenhahn 2007). En consecuencia, los hallazgos de dichos experimentos no pueden ser generalizados en principios o teorías.

La búsqueda de métodos de evaluación que permitan tener certeza sobre las respuestas emocionales de las personas ha llevado a la creación de numerosos métodos que van desde cuestionarios validados hasta métodos de imagen (como tomografías o resonancias magnéticas). A continuación, presento una breve

clasificación en función de la tecnología o tipo de recursos que utiliza cada método:

- Métodos de auto-reporte: piden directamente a los participantes que reporten su estado emocional. Dentro de ellos encontramos a los cuestionarios, entrevistas, test y métodos gráficos que evalúan ya sea los estados emocionales directamente, o indicadores o componentes de dichos estados emocionales. La principal desventaja de estos es que no permiten medir los estados emocionales inconscientes y sólo se pueden aplicar *a posteriori*, anulando la posibilidad de evaluar transiciones entre estados emocionales.
- Métodos observacionales: analizan el comportamiento, gestos y lenguaje corporal de los participantes para identificar sus estados emocionales. Su principal desventaja es que se limita a expresiones muy precisas y medibles, dejando fuera muchas sutilezas, especificidades de la personalidad y expresiones que no son fácilmente clasificables.
- Métodos psicofisiológicos: miden la respuesta fisiológica ligada a un estado emocional por medio de la evaluación de parámetros como la frecuencia cardíaca o actividad eléctrica de la piel. Son considerados bastante confiables porque son cuantitativos y permiten medir emociones inconscientes. Sin embargo, no hay suficiente evidencia para la identificación de estados emocionales con precisión, más bien permiten medir la intensidad de algunos componentes emocionales como la valencia o la activación, lo que dificulta la interpretación de los resultados en contextos de evaluación poco controlados.
- Métodos neuronales: finalmente, existen algunos métodos que intentan identificar los estados emocionales a partir de la actividad neuronal que los produce, por ejemplo, las resonancias magnéticas funcionales y tomografías. Si bien prometen ser los más precisos en la identificación de las respuestas emocionales, el nivel de conocimiento del funcionamiento del

cerebro no es suficiente para la correcta interpretación de los resultados. Otra gran desventaja es que para obtener una imagen cerebral los participantes deben estar inmóviles dentro de una máquina por tiempo prolongado, por lo que el contexto para las mediciones está lejos de ser similar a la vida cotidiana.

Dentro de la disciplina de la HRI la evaluación de las respuestas emocionales es particularmente difícil, ya que, aunado a los obstáculos ya mostrados en los diferentes métodos, las interacciones con robots se hacen por lo general en ambientes controlados, una única vez y por tiempos muy limitados (alrededor de 15 minutos) (Dautenhahn 2013). Este contexto particularmente difícil para la evaluación de respuestas emocionales debe verse no como un obstáculo, sino como un reto a superar por las ciencias afectivas. Uno de los resultados más interesantes de este tipo de experimentaciones es que revelan necesidades que deben ser cubiertas por otras disciplinas, para ayudar al desarrollo de las ciencias afectivas, en este caso.

Uno de los objetivos principales de la colaboración interdisciplinaria es que las disciplinas participantes puedan retroalimentarse, y empujarse mutuamente a seguir avanzando a través de la localización de estos huecos o carencias (Baxter *et al.* 2016). Esto porque muchos temas de investigación en la HRI han provocado otros en disciplinas como la psicología, la antropología o la filosofía, al proponer nuevas preguntas que no habían sido abordadas con anterioridad.

En el tema particular de la HRI y las evaluaciones emocionales hay que estar conscientes de que algunos de los métodos de evaluación no pueden utilizarse porque el contexto no lo permite. Por ejemplo, no es posible tomar una tomografía o resonancia magnética al mismo tiempo que se está interactuando con un robot; o, en el caso de los métodos psicofisiológicos, se puede presentar un contexto “hipertecnologizado”, ya que los participantes no sólo estarán interactuando, tal vez por primera vez, con un

robot, sino que además estarán conectados a electrodos y equipos de cómputo, lo que puede afectar de forma negativa las respuestas emocionales. Sin embargo, existen muchos otros métodos que pueden utilizarse para sistematizar la evaluación de respuestas emocionales dentro de la HRI.

En mi proyecto de investigación de maestría ese fue el tema principal. Al abordar el problema del diseño de interacciones robóticas y buscar respuestas emocionales específicas, me di cuenta de que más importante que el diseño de las expresiones del robot o de la experiencia misma era abordar el problema de la falta de sistematización en la evaluación de las respuestas emocionales en los usuarios. Como propuesta decidí explorar los métodos más utilizados desde el diseño, la psicología y otras disciplinas para la identificación de estados emocionales, y decidí probar un método gráfico diseñado por Desmet, las Emocards (Desmet 2005), con base en el modelo afectivo de Russell (Russell 1980). Con este método logré diseñar expresiones emocionales en un robot Nao que representaban dos estados emocionales con diferentes grados de activación en el mismo nivel de valencia. Utilizando el mismo modelo llevé a cabo un proceso de evaluación de la respuesta emocional de los participantes después de jugar lotería con el robot en cada uno de los dos escenarios. Utilicé el mismo método para que el participante tratara de identificar el estado emocional del robot.

Los resultados mostraron que el método es efectivo para la evaluación de la respuesta emocional del usuario, pero también para la identificación de estados emocionales en los robots sociales (Mendoza 2015). Lo interesante de este experimento es que se utilizó un modelo emocional constructivista, es decir, no se utilizaron categorías emocionales para describir las emociones (por ejemplo: feliz, triste, enojado), además de que el robot Nao cuenta con muy pocos elementos de expresión facial (ya que no tiene motores faciales, sólo se pueden diseñar algunos gestos basados en parpadeos y cambios de color en los ojos). Por lo tanto,

el tipo de emociones abordadas en el experimento no correspondía a estados absolutos, sino a estados emocionales más sutiles, más cotidianos y, por lo tanto, más difíciles de identificar, ya que no hay una sonrisa o una mueca de por medio, sólo “una actitud” mediada por el lenguaje corporal.

Estos resultados muestran el potencial de los métodos ya existentes para la evaluación de respuestas emocionales dentro del contexto de la HRI. La generalización en el uso de métodos validados en otras disciplinas puede ayudar a la sistematización del proceso de evaluación emocional dentro de la HRI y otras disciplinas tecnológicas. Por otro lado, este experimento mostró también la importancia que tiene abordar las emociones desde las teorías de la emoción y las ciencias afectivas, ya que me permitieron construir una interacción emocional sutil, más cercana a las interacciones sociales cotidianas. Además, las expresiones emocionales desarrolladas no requieren de complejos mecanismos para la expresión facial en los robots y, por lo tanto, son más viables a ser implementadas en sistemas robóticos comerciales.

### **Los robots, las artes y las emociones**

Dentro de los muchos posibles contextos para crear interacciones sociales entre humanos y robots se encuentran las artes escénicas. Diversos proyectos que intentan incluir robots sociales en obras de teatro o fílmicas se han propuesto con el objetivo de explorar las posibilidades creativas en el uso de HRI en las artes. Asimismo, es un contexto que permite experimentar comportamientos no cotidianos en los robots, además de que es una estrategia que permite a niños y adultos el acercamiento a este tipo de tecnologías sin formar parte de un protocolo experimental. Desde un punto de vista científico, las aproximaciones artísticas utilizando dispositivos robóticos permite reflexionar en torno a las diversas formas que podemos relacionarnos con las máquinas desde un punto de vista filosófico y sociológico (Vorn 2016).

Dentro de un proceso transdisciplinario con artistas visuales, coreógrafos, dramaturgos, performers e ingenieros se desarrolló la obra escénica SAGA, en la que un robot humanoide actúa en escenario junto con un intérprete humano. Uno de los objetivos académicos de dicho proyecto fue identificar a qué grado y bajo qué mecanismos un robot es capaz de mantener la convención teatral. Para que dicha convención exista, es necesario que suceda una conexión emocional entre el actor y el público y que dicha conexión pueda mantenerse a lo largo del fenómeno escénico.

Para lograr este cometido, se hizo un planteamiento metodológico derivado del diseño de experiencias que derivó en una serie de nuevas propuestas que abarcan desde una estructura dramática específica para robots, hasta un cuestionamiento del lenguaje escénico tal cual lo conocemos hoy (Mendoza y Sánchez 2016).

El tema que se desarrolla en SAGA es el aislamiento social en niños y adolescentes presionados por las exigencias sociales del siglo XXI. Una de las muestras más extremas de este fenómeno es el de los llamados “hikikomoris”, personas que se aíslan dentro de sus casas por más de seis meses sin tener contacto con la sociedad (en algunas ocasiones estas personas se mantienen activos en el mundo digital [Kato, Kanba y Teo 2018]). SAGA muestra un escenario paradójico, en el que un robot que vive originalmente en un mundo virtual cobra vida en el mundo real para mostrarle a un niño en aislamiento lo importante que son las relaciones sociales y mantenerse conectado con aquellos a los que ama. Una de las intenciones de la obra es mostrar que los dispositivos tecnológicos pueden ser utilizados para mejorar nuestros estados de ánimo, así como para optimizar nuestras relaciones sociales, y así promover una sociedad mejor comunicada. Después de múltiples presentaciones en diferentes escenarios con niños y adolescentes de diferentes edades, quedó evidenciada la enorme capacidad de establecer relaciones emocionales que tienen los robots sociales.

Por supuesto que no se trata sólo de colocar un robot en un escenario, es necesario diseñar la experiencia completa: guión, coreografía, escenografía, audio, iluminación y, por supuesto, las expresiones y actitudes del robot para lograrlo. También es importante resaltar que la presencia del actor humano es indispensable para que la convención teatral suceda y el robot logre establecer una relación con el público. Y es que los robots carecen de intencionalidad: es el compañero humano junto con todos los demás elementos los que dotan al robot de dicha intencionalidad (Mendoza 2019). Por lo tanto, lo que sucede a lo largo de la obra es una convención teatral entre el robot y el público mediada por los agentes humanos. Esta propuesta resulta interesante ya que muestra nuevas facetas de la relación humano-robot; por un lado, se muestra un escenario utópico en el que las relaciones humano-humano pueden ser mediadas por un agente tecnológico y, por otro lado, SAGA es un ejemplo de una relación humano-robot mediada por un agente humano.

El trabajo conjunto entre la HRI y las artes siempre va a tener de por medio motivaciones emocionales, ya que el arte busca despertar emociones en la audiencia. Por lo que se trata de un contexto ideal para la exploración de diferentes dinámicas emocionales entre humanos y robots (Velonaki y Rye 2016).

Desafortunadamente el formato de las artes escénicas no favorece procesos de evaluación, por lo que no es posible obtener mediciones de respuestas emocionales válidas. Aun así, este tipo de colaboraciones, además de favorecer la exploración de situaciones o paradigmas fuera de los contextos cotidianos (Sandry 2016), ayuda a acercar la tecnología a los futuros usuarios de la misma y, por lo tanto, es un paso importante en el proceso de generación de convenciones comunicativas entre robots y humanos.

### **Emociones y tecnologías interactivas en el futuro**

La HRI junto con otras disciplinas enfocadas en el diseño de tecnologías interactivas, como la interacción humano computadora

o la realidad virtual, utilizan aproximaciones metodológicas similares para la resolución de problemas relacionados con las experiencias en los usuarios (Dautenhahn 2013). Todas consideran las emociones componentes indispensables en el diseño de interacciones (Lim *et al.* 2008). Sin embargo, la mayor parte de los proyectos conceptualiza las respuestas emocionales de los usuarios como un indicador de satisfacción (Preece, Sharp y Rogers 2015), facilidad de uso (Walker y Prytherch 2008) o simplemente de una experiencia positiva (Hassenzahl 2013).

Pocos estudios abordan los componentes emocionales como medio para lograr otro tipo de efectos en los usuarios que impacten de forma positiva en la salud y bienestar de manera profunda y a largo plazo. Algunas posibles aplicaciones emocionales de los robots sociales son aquellas dirigidas a la investigación e implementación con fines terapéuticos en situaciones de salud mental como la ansiedad social o los trastornos alimenticios. También se propone su uso con niños dentro del espectro autista y con personas de la tercera edad en casas de retiro. Algunas aplicaciones más simples, pero que también requieren de interacciones emocionales fructíferas son los robots de compañía dirigidos a personas solitarias, familias divididas o simplemente aquellos que se sienten solos en ocasiones. Los robots sociales deben ser vistos como herramientas que podrán ayudarnos a establecer relaciones emocionales más sanas, a aprender más sobre nosotros, a alcanzar el bienestar emocional y ayudar a aquellos que tienen algunas limitaciones.

Si bien los robots sociales pueden ser utilizados en beneficio de muchas personas, también son cotidianamente vistos como herramientas mercadológicas que, gracias a lo llamativos que son y sus habilidades para establecer lazos emocionales con humanos, son utilizados para aumentar la venta de productos. De hecho, algunos autores afirman que el componente emocional de las interacciones tiene como objetivo establecer una relación agradable entre el usuario y el dispositivo al asegurar el consumo

del producto y el uso constante del mismo (Jordan 2003). En ese sentido, la utilización del estado emocional de los usuarios con fines mercadológicos propone una serie de conflictos éticos que aún no se han resuelto.

Afortunadamente, al mismo tiempo que avanza la HRI y otras disciplinas en la intersección entre la tecnología y las personas, también se hacen progresos en materia ética. Existe cierta claridad en la importancia que tiene la protección del consumidor en los contextos donde se establecen relaciones emocionales con dispositivos tecnológicos (Darling 2016a). Por ello, es indispensable desarrollar marcos normativos que permitan establecer límites y dirigir el tipo de relaciones emocionales que se proponen, así como los mecanismos para alimentar y mantener dichas relaciones. Efectos como manipulaciones para llevar a cabo compras, como el *upgrade* de un sistema operativo para un robot, o desarrollo de dependencias emocionales con un dispositivo, como podría ser el caso de los robots terapéuticos, son ejemplo de los escenarios distópicos que pretenden evitarse.

Incluso en este contexto se vuelve evidente lo importante que es el estudio profundo de los mecanismos emocionales, ya que el mejor entendimiento de estos permitirá anticipar estos efectos no deseados a tiempo. Si las disciplinas que estudian las interacciones tienen un enfoque centrado en el ser humano, debieran entonces también adoptar una postura filosófica centrada en el ser humano, y buscar el bienestar como objetivo último de toda interacción (Buchanan 2001). Y por bienestar no me refiero a una experiencia agradable de corto plazo, sino a un efecto profundo que ayude a resolver problemas de salud, sociales, económicos, ambientales o de cualquier otra naturaleza. El estado actual de avance en el conocimiento en torno a las emociones no parece muy prometedor. Por un lado, las disciplinas tecnológicas desean diseñar dispositivos útiles para resolver problemas profundos, como es el caso de la investigación en HRI y niños en el espectro autista; sin embargo, las propuestas no logran llegar al

contexto de uso porque los resultados obtenidos no son replicables y no existe sistematización en el campo de estudio (Kim *et al.* 2012). Por otro lado, ciencias afectivas, como la neurociencia, buscan que los modelos teóricos que se construyen dentro de la disciplina puedan aterrizar en aplicaciones y productos finales; desafortunadamente la forma de trabajo y de financiamiento de la ciencia no permite en la mayoría de los casos ir más allá de la publicación de resultados experimentales (Grimshaw *et al.* 2012).

Es claro que lo que hace falta es más investigación interdisciplinaria, de otra forma es probable que este estado de poca comunicación entre el contexto científico básico y el aplicado permanezca obstaculizando el avance de las disciplinas y el desarrollo de tecnología con impacto social. Así que deseo terminar este texto haciendo una invitación a todos los interesados en el tema de las emociones y las interacciones con sistemas robóticos o simplemente tecnológicos a ir más allá de sus disciplinas y preguntarse de qué formas podemos diseñar experiencias emocionales que tengan impacto a largo plazo, tanto a nivel individual como social. O cuáles de nuestros problemas sociales, de personalidad, de comportamiento, de cognición, podrían estar asociados con procesos emocionales y de qué forma la Inteligencia Artificial podría ayudar a superarlos.

## Referencias

- AHN, Sun Joo y Jesse Fox (2017). "Immersive virtual environments, avatars, and agents for health". En *Oxford Research Encyclopedia of Communication*. <https://oxfordre.com/communication/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-325>.
- BALASURIYA, Saminda Sundeepa, Laurianne Sitbon, Margot Brereton y Stewart Koplick (2019). "How can social robots spark collaboration and engagement among people with intellectual disability?". *Proceedings of the 31st Australian Conference on*

- Human-Computer-Interaction*, 209-220. <https://doi.org/10.1145/3369457.3370915>.
- BARRET, Lisa Feldman (2019). "In search of emotions". *Current Biology* 29, núm.º 5: R140-R142.
- BAXTER, Paul, James Kennedy, Emmanuel Senft Séverin Lemaignan y Tony Belpaeme (2016). "From characterising three years of HRI to methodology and reporting recommendations". En *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 391-398. IEEE. DOI: 10.1109/HRI.2016.7451777.
- BEGUM, Momotaz, Ruchard Serna y Holly Yanco (2016). "Are robots ready to deliver autism interventions? A comprehensive review". *International Journal of Social Robotics* 8, n.º 2: 157-181.
- BREAZEAL, Cynthia (2003). "Toward sociable robots". *Robotics and autonomous systems* 42, n.º 3-4: 167-175.
- (2004). *Designing sociable robots*. Cambridge: The MIT Press.
- BUCHANAN, Richard (2001). "Human dignity and human rights: Thoughts on the principles of human-centered design". *Design issues* 17, n.º 3: 35-39.
- CABIBIHAN, John John, Hifza Javed, Marcelo Ang Jr. y Sharifah Mariam Aljunied (2013). "Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism". *International journal of social robotics* 5, n.º 4: 593-618.
- CHANG, Wan-Ling y Selma Šabanović (2015). "Interaction expands function: Social shaping of the therapeutic robot PARO in a nursing home". *HRI '15: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*: 343-350. <https://doi.org/10.1145/2696454.2696472>.
- CHANG, Wan-Ling, Selma Šabanovic y Lesa Huber (2014). "Observational study of naturalistic interactions with the socially assistive robot PARO in a nursing home". En *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 294-299. Edinburgo: IEEE.

- COPPIN, Géraldine y David Sander (2016). "Theoretical approaches to emotion and its measurement". En *Emotion measurement*, editado por Herbert L. Meiselman, 3-30. Elsevier.
- DARLING, Kate (2016a). "Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects". En *Robot law*, editado por Ryan Calo, A. Michael Froomkin e Ian Kerr, 213-232. Edward Elgar Publishing.
- \_\_\_\_ (2016b). "What Are the Rules of Human-Robot Interaction?". <https://www.youtube.com/watch?v=k7NNK-nQquo>.
- DAUTENHAHN, Kerstin (1998). "The art of designing socially intelligent agents: science, fiction, and the human in the loop". *Applied Artificial Intelligence* 12, n.º 7-8: 573-617. <https://doi.org/10.1080/088395198117550>.
- \_\_\_\_ (2007). "Socially intelligent robots: dimensions of human-robot interaction". *Philosophical Transactions of the Royal Society B: Biological Sciences* 362, n.º 1480: 679-704. <https://doi.org/10.1098/rstb.2006.2004>.
- \_\_\_\_ (2013). "Human Robot Interaction". En *The Encyclopedia of Human-Computer Interaction*. <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-robot-interaction>.
- DAVIS, Kenneth y Jaak Panksepp (2011). "The brain's emotional foundations of human personality and the Affective Neuroscience Personality Scales". *Neuroscience & Biobehavioral Reviews* 35, n.º 9: 1946-1958.
- DESMET, Pieter (2005). "Measuring emotion: Development and application of an instrument to measure emotional responses to products". En *Funology*, editado por Mark Blythe y Andrew Monk, 111-123. Springer. [http://link.springer.com/content/pdf/10.1007/1-4020-2967-5\\_12.pdf](http://link.springer.com/content/pdf/10.1007/1-4020-2967-5_12.pdf).
- DESHMUKH, Amol, Sooraj Babu, R. Unnikrishnan, Shanker Ramesh, Parameswari Anitha y Rao R. Bhavani (2019). "Influencing hand-washing behaviour with a social robot: HRI study

- with school children in rural India”. *28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1-6. IEEE. DOI: 10.1109/RO-MAN46459.2019.8956367.
- EMMELKAMP, Paul M. G., Katharina Meyerbröker y Nexhmedin Morina (2020). “Virtual Reality Therapy in Social Anxiety Disorder”. *Current Psychiatry Reports* 22, n.º 7: 32. <https://doi.org/10.1007/s11920-020-01156-1>.
- FELDMAN, Lisa (2016). “Navigating the Science of Emotion”. En *Emotion Measurement*, editado por Herbert L. Meiselman, 31-63. Woodhead Publishing. <https://doi.org/10.1016/B978-0-08-100508-8.00002-3>.
- (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.
- FERNÁNDEZ-ABASCAL, Enrique y María Martín (2013). “Affective induction and creative thinking”. *Creativity Research Journal* 25, n.º 2: 213-221.
- FONG, Terrence, Illah Nourbakhsh y Kerstin Dautenhahn (2003). “A survey of socially interactive robots”. *Robotics and Autonomous Systems* 42, n.º 3-4: 143-166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X).
- FREIRE, Ana, António Valente y Vítor Filipe (2020). “Can NAO Robot Influence the Eye Gaze and Joint Attention of Mentally Impaired Young Adults?”. En *Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion*, 111-115. <https://doi.org/10.1145/3439231.3440619>.
- FRY, Douglas (2005). “Rough-and-tumble social play in humans”. En *The nature of play: Great apes and humans*, 54-85.
- GOODRICH, Michael y Alan Schultz (2007). “Human-Robot Interaction: A Survey”. *Foundations and Trends® in Human-Computer Interaction* 1, n.º 3: 203-275. <https://doi.org/10.1561/1100000005>.

- GREENE, Terry y Helga Noice (1988). "Influence of positive affect upon creative thinking and problem solving in children". *Psychological reports* 63, n.º 3: 895-898.
- GRIMSHAW, Jeremy, Martin Eccles, John Lavis, Sophie Hill y Janet Squires (2012). "Knowledge translation of research findings". *Implementation science* 7, n.º 1: 50.
- HAMID, Aminullah Che, Mohd Azfar Miskam, Hanafiah Yussof, Syamimi Shamsuddin, Hafizan Hashim y Luthffi Ismail (2013). Human-Robot Interaction (HRI) for Children with Autism to Augment Communication Skills. *Applied Mechanics and Materials* 393: 598-603. <https://doi.org/10.4028/www.scientific.net/AMM.393.598>.
- HAN, Shihui (2018). "Neurocognitive basis of racial ingroup bias in empathy". *Trends in Cognitive Sciences* 22, n.º 5: 400-421.
- HAN, Hee Jeong, Sanjana Mendu, Beth K. Jaworski, Jason E. Owen y Saeed Abdullah (2021). "PTSDialogue: Designing a conversational agent to support individuals with post-traumatic stress disorder". *Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 198-203. <https://doi.org/10.1145/3460418.3479332>.
- HAN, Jeonghye, Miheon Jo, Sungju Park y Sungho Kim (2005). "The educational use of home robots for children". *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication*, 2005, 378-383. DOI: 10.1109/ROMAN.2005.1513808.
- HASSENZAHL, Marc (2013). "User experience and experience design". *The encyclopedia of human-computer interaction* 2. [https://www.researchgate.net/profile/Marc-Hassenzahl/publication/259823352\\_User\\_Experience\\_and\\_Experience\\_Design/links/56a7352d08ae997e22bbc807/User-Experience-and-Experience-Design.pdf](https://www.researchgate.net/profile/Marc-Hassenzahl/publication/259823352_User_Experience_and_Experience_Design/links/56a7352d08ae997e22bbc807/User-Experience-and-Experience-Design.pdf).

- ISEN, Alice (1999). "On the relationship between affect and creative problem solving". En *Affect, creative experience, and psychological adjustment*, editado por Sandra W. Russ, 3-17. Routledge.
- (2008). "Some ways in which positive affect influences decision making and problem solving". *Handbook of emotions* 3: 548-573.
- IZARD, Carroll (2007). "Basic emotions, natural kinds, emotion schemas, and a new paradigm". *Perspectives on psychological science* 2, n.º 3: 260-280.
- (2010). "The many meanings/aspects of emotion: Definitions, functions, activation, and regulation". *Emotion Review* 2, n.º 4: 363-370.
- JORDAN, Patrick (2003). *How to make brilliant stuff that people love... And make big money out of it*. Hoboken: John Wiley & Sons.
- KATO, Takahiro, Shigenobu Kanba y Alan Teo (2018). "Hikikomori: experience in Japan and international relevance". *World Psychiatry* 17, n.º 1: 105.
- KIM, Elizabeth, Rhea Paul, Frederick Shic y Brian Scassellati (2012). "Bridging the research gap: Making HRI useful to individuals with autism". *Journal of Human-Robot Interaction* 1, n.º 1: 26-54.
- LIM, Youn-kyung, Justin Donaldson, Heekyoung Jung, Breanne Kunz, David Royer, Shruti Ramalingam, Sindhia Thirumaran y Erik Stolterman (2008). "Emotional Experience and Interaction Design". En *Affect and Emotion in Human-Computer Interaction. Lecture Notes in Computer Science*, editado por Christian Peter y Russell Beale, 116-129. Berlín: Springer. [https://doi.org/10.1007/978-3-540-85099-1\\_10](https://doi.org/10.1007/978-3-540-85099-1_10).
- LINDQUIST, Kristen, Erika Siegel, Karen Quigley y Lisa Feldman (2013). "The hundred-year emotion war: Are emotions natural kinds or psychological constructions? Comment on Lench, Flores, and Bench (2011)". *Psychological Bulletin* 139, n.º 1: 255-263. <https://doi.org/10.1037/a0029038>.
- McGLYNN, Sean A., Shawn Kemple, Tracy L. Mitzner, Chih-Hung Aaron King y Wendy A. Rogers (2017). "Understanding the

- potential of PARO for healthy older adults”. *International Journal of Human-Computer Studies* 100: 33-47.
- MENDES, Wendy, Jim Blascovich, Brian Lickel y Sarah Hunter (2002). “Challenge and threat during social interactions with White and Black men”. *Personality and Social Psychology Bulletin* 28, n.º 7: 939-952.
- MENDOZA, Gloria (2015). “Evaluation of the Emotional Answer in HRI on a Game Situation”. En *Proceedings of the Latin American Conference on Human Computer Interaction*, 1-7. <https://doi.org/10.1145/2824893.2824897>.
- (2019). “Del autómata al robot actor. Reflexiones en torno a SAGA”. *¡Cuerpo, Máquina, Acción!, Estudios sobre cuerpo, performance y tecnologías emergentes*, 31-38.
- MENDOZA, Gloria y Alicia Sánchez (2016). “Lenguajes híbridos y mediaciones en las poéticas del cuerpo”. *Interdanza* 36: 31-38.
- MULLIGAN, Kevin y Klaus Scherer (2012). “Toward a working definition of emotion”. *Emotion Review* 4, n.º 4: 345-357.
- PAIVA, Ana, Iolanda Leite, Hana Boukricha e Ipke Wachsmuth (2017). “Empathy in virtual agents and robots: a survey”. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, n.º 3: 1-40.
- PANKSEPP, Jaak (2007). “Can PLAY Diminish ADHD and Facilitate the Construction of the Social Brain?”. *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 16, n.º 2: 57-66.
- (2008). “Play, ADHD, and the Construction of the Social Brain: Should the First Class Each Day Be Recess?”. *American Journal of Play* 1, n.º 1: 55-79.
- PANKSEPP, Jaak, Steve Siviy y Larry Normansell (1984). “The psychology of play: theoretical and methodological perspectives”. *Neuroscience & Biobehavioral Reviews* 8, n.º 4: 465-492.
- PANKSEPP, Jaak y Douglas Watt (2011). “What is basic about basic emotions? Lasting lessons from affective neuroscience”. *Emotion review* 3, n.º 4: 387-396.

- PREECE, Jenny, Helen Sharp e Yvonne Rogers (2015). *Interaction design: beyond human-computer interaction*. New Jersey: John Wiley & Sons.
- REAL ACADEMIA ESPAÑOLA (RAE) (2023). “Emoción | Diccionario de la lengua española”. Acceso el 22 de julio de 2020. <https://dle.rae.es/emoción>.
- REYNOLDS, Katherine y Kathleen Klik (2016). “New developments in prejudice research: from its neural basis and impact on well-being to prejudice reduction”. *Current Opinion in Psychology* 11: 115-119.
- RUSSELL, James (1980). “A circumplex model of affect”. *Journal of personality and social psychology* 39, n.º 6: 1161-1178.
- SAMMONDS, Andrew, Anshu Saxena Arora y Amit Arora (2022). “Robotic Anthropomorphism and Intentionality Through Human-Robot Interaction (HRI): Autism and the Human Experience”. En *Managing Social Robotics and Socio-cultural Business Norms: Parallel Worlds of Emerging AI and Human Virtues*, editado por Anshu Saxena Arora, Sabine Jentjens, Amit Arora, John R. McIntyre, Mohamad Sepehri, 55-72. Springer International Publishing. [https://doi.org/10.1007/978-3-031-04867-8\\_5](https://doi.org/10.1007/978-3-031-04867-8_5).
- SANDRY, Eleanor (2016). “The Potential of Otherness in Robotic Art”. En *Robots and Art. Exploring an Unlikely Symbiosis*, editado por Damith Herath y Christian Kroos, 177-189. Springer.
- SCHERER, Klaus (2005). “What are emotions? And how can they be measured?”. *Social science information* 44, n.º 4: 695-729.
- SEIBT, J. (2016). “An interdisciplinary approach to improving cognitive human-robot interaction—a novel emotion-based model”. *What Social Robots Can and Should Do: Proceedings of Robophilosophy 2016/TRANSOR 2016* 290: 195.
- SHAMSUDDIN, Syamimi, Hanafiah Yussof, Luthffi Idzhar Ismail, Salina Mohamed, Fazah Akhtar Hanapiah y Nur Ismarrubie Zahari (2012). “Humanoid robot NAO interacting with autistic children of moderately impaired intelligence to augment communication skills”. *Procedia Engineering* 41: 1533-1538.

- TALASKA, Cara, Susan Fiske y Shelly Chaiken (2008). "Legitimizing racial discrimination: Emotions, not beliefs, best predict discrimination in a meta-analysis". *Social justice research* 21, n.º 3: 263-296.
- TANAKA, Fumihide, Javier Movellan, Bret Fortenberry y Kazuki Aisaka (2006). "Daily HRI evaluation at a classroom environment: Reports from dance interaction experiments". En *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 3-9. <https://doi.org/10.1145/1121241.1121245>.
- TAPUS, Adriana y Maja Mataric (2008). "Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance". En *AAAI spring symposium: emotion, personality, and social behavior*, 133-140.
- TASSINARI, Matilde, Matthias Burkard Aulbach e Inga Jasinska-ja-Lahti (2022). "The use of virtual reality in studying prejudice and its reduction: A systematic review". *PloS One* 17, n.º 7: e0270748.
- TEASDALE, John y Sarah Fogarty (1979). "Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory". *Journal of abnormal psychology* 88, n.º 3: 248.
- TIELMAN, Myrthe L., Mark A. Neerincx, Rafael Bidarra, Ben Kybartas y Willem-Paul Brinkman (2017). "A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories". *Journal of Medical Systems* 41: 1-10.
- TOOBY, John y Leda Cosmides (2008). "The evolutionary psychology of the emotions and their relationship to internal regulatory variables". En *Handbook of emotions*, editado por Michael Lewis, Jeannette Haviland-Jones y Lisa Feldman Barrett, 114-137. The Guilford Press.
- VELONAKI, Mari y David Rye (2016). "Designing robots creatively". En *Robots and Art. Exploring an Unlikely Symbiosis*, editado por Damith Herath y Christian Kroos, 379-401. Springer.

- VORN, Bill (2016). "I Want to Believe—Empathy and Catharsis in Robotic Art". En *Robots and Art. Exploring an Unlikely Symbiosis*, editado por Damith Herath y Christian Kroos, 365-377. Springer.
- WALKER, Shane y David Prytherch (2008). "How is it for you? (A case for recognising user motivation in the design process)". En *Affect and Emotion in Human-Computer Interaction*, editado por Christian Peter y Russell Beale, 130-141. Springer.
- WESTLUND, Jaqueline, Goren Gordon, Samuel Spaulding, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das y Cynthia Breazeal (2016). "Lessons from teachers on performing HRI studies with young children in schools". En *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 383-390. IEEE. DOI: 10.1109/HRI.2016.7451776.
- YOKOTANI, Kenji, Gen Takagi y Kobun Wakashima (2018). "Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design". *Computers in Human Behavior* 85: 135-145. <https://doi.org/10.1016/j.chb.2018.03.04>.
- ZUCKERMAN, Oren y Guy Hoffman (2015). "Empathy objects: Robotic devices as conversation companions". En *Proceedings of the ninth international conference on tangible, embedded, and embodied interaction*, 593-598. <https://doi.org/10.1145/2677199.2688805>.

## **Sobre la autora**

### **MDI Gloria Adriana Mendoza Franco**

Ingeniera en biónica egresada de la Unidad Profesional Interdisciplinaria en Ingenierías y Tecnologías Avanzadas del Instituto Politécnico Nacional. Maestra en Diseño Industrial por la Universidad Nacional Autónoma de México. Se interesa en la ergonomía, en el área de Interacción Humano Robot. Es experta en biomecánica y en el estudio del movimiento humano, así como en el procesamiento de señales biológicas. Profesora de carrera en el Posgrado en Diseño Industrial de la UNAM, donde

coordina el Laboratorio de ergonomía, con las líneas de investigación en: diseño y discapacidad, ergonomía física, ergonomía cognitiva, ergonomía para el diseño, robótica y diseño, y diseño e interacción. Actualmente es doctorante en el Departamento de Neurociencia e Ingeniería Biomédica (Neuroscience and Biomedical Engineering) en la Universidad de Alto.

# Interfaces cerebro computadora con perspectivas a su aplicación en robots de servicio doméstico

MONTSERRAT ALVARADO GONZÁLEZ

## Resumen

Una interfaz cerebro computadora (BCI) le permite a sus usuarios interactuar con su entorno únicamente mediante la información que puede ser extraída de su cerebro con el fin de controlar un dispositivo electrónico. Esta interacción le ofrecería a personas con ciertas condiciones médicas (como síndrome de enclaustramiento o esclerosis lateral amiotrófica) independencia de sus cuidadores y, con ello, la posibilidad de mejorar su calidad de vida. En este capítulo resumimos los avances que nuestro grupo de investigación ha alcanzado hasta el momento en el desarrollo de una BCI que le permitiría a un usuario controlar un robot de servicio doméstico. En particular, nos hemos enfocado en generar e implementar algoritmos computacionales basados en inspiración biológica (por ejemplo, redes neuronales artificiales y computación evolutiva) para: *i*) extraer y clasificar distintos tipos de señales fisiológicas y cerebrales para controlar al robot de servicio doméstico, y *ii*) proveer de autonomía al robot de servicio

para que lleve a cabo tareas de reconocimiento visual, navegación (por ejemplo, planeación de rutas y generación de mapas) y manipulación de brazos mecánicos.

## Introducción

Una interfaz cerebro computadora (BCI, por las siglas en inglés de *brain-computer interface*) es un conjunto de hardware y software que le permite a sus usuarios comunicarse e interactuar con su entorno únicamente mediante la información que puede ser extraída de su cerebro (Wolpaw *et al.* 2000). Las BCI permiten auxiliar a personas con alteraciones en las vías nerviosas periféricas, en las placas neuromusculares o en los músculos mismos, como consecuencia de daños sufridos por accidentes o enfermedades, y que, a pesar de ello, conservan funcionales sus capacidades cognitivas (Hosler *et al.* 2000, Fager *et al.* 2006). El proceso general que se sigue para que una BCI convierta señales cerebrales en instrucciones para controlar un dispositivo es el siguiente: durante un periodo fijo de tiempo se adquiere un conjunto de señales cerebrales de un usuario mediante electrodos (en nuestro caso, superficiales). Dichas señales se digitalizan con un convertidor que contiene un amplificador analógico digital. Luego, las señales digitalizadas se filtran para minimizar el ruido que viene de las líneas eléctricas, así como de movimientos musculares (incluyendo latidos del corazón y movimientos oculares). Posteriormente, se emplean algoritmos de extracción de características y de clasificación para identificar características particulares de la señal, las cuales serán traducidas en instrucciones para controlar algún dispositivo. El usuario puede supervisar el estado del dispositivo a través de retroalimentaciones que le permitan determinar el resultado de sus esfuerzos por controlarlo.

Las investigaciones acerca de estas señales son muy recientes y la precisión en la clasificación aún es baja (hasta 68% como máximo en algunos usuarios [Bobrov *et al.* 2011]). Se han desarrollado distintas aplicaciones para permitir que una persona

en las condiciones mencionadas anteriormente pueda comunicarse y tener más independencia de sus cuidadores. Por ejemplo, aquellas que permiten deletrear texto (Donchin, Spender y Wijesinghe 2000; Klobassa *et al.* 2009; Nijboer *et al.* 2008; Sellers y Donchin 2006; Zhang, Guan y Wang 2005), hacer búsquedas en Internet (Sirvent *et al.* 2010), dar comandos de navegación a una silla de ruedas tanto virtual como real (Acevedo *et al.* 2009, Atum *et al.* 2009, Atum *et al.* 2010, García y Gentiletti 2008, Gareis *et al.* 2011, Gentiletti *et al.* 2009, Long *et al.* 2012), controlar un brazo mecánico (Pathirage, Alqasemi y Dubey 2012) o controlar distintos tipos de robots.

En este proyecto se propone desarrollar una BCI para controlar un robot de servicio doméstico. Se busca que éste eventualmente cuente con autonomía en la navegación y en la manipulación de objetos, seguimiento y búsqueda de personas, y orientación hacia la fuente del sonido. El objetivo de que este dispositivo tenga las características de un robot de servicio es que se requiera de una intervención mínima del usuario para controlarlo.

### **Trabajos previos**

Hasta el momento, las BCI que se han desarrollado para controlar robots son las siguientes. En Millán *et al.* (2004) presentan una BCI basada en tareas mentales que reconoce tres estados, elegidos por el usuario, para mover a un robot en un ambiente controlado. En Zhao, Li y Cui (2007) proponen una BCI basada en las ondas alfa del EEG (que tienen una frecuencia en el rango de 8 a 13 Hz) para controlar a un robot de servicio; sin embargo, no mencionan nada relacionado con cómo explotar características de este tipo de robots. La BCI sólo le da instrucciones de movimiento básico al robot (arriba, abajo, izquierda y derecha) y el usuario es estimulado mediante un tablero con 5 LEDs. En Valbuena *et al.* (2007) implementan una BCI basada en potenciales visuales de estado estacionario (SSVEP) para controlar un robot semiautónomo; esta BCI muestra un menú mediante el cual

el usuario puede darle instrucciones de alto nivel al robot, como pedirle que sirva una bebida en un vaso. En Bell *et al.* (2008) presentan una BCI basada en la detección del P300 para controlar un robot humanoide parcialmente autónomo que lleva a cabo tareas de navegación que le permiten tanto dirigirse a una ubicación como tomar un objeto indicado por el usuario; la matriz de estimulación se genera de manera dinámica, ya que ésta se conforma de las imágenes de los objetos que son captados por las cámaras del robot. En Bryan *et al.* (2011) proponen una BCI basada en SSVEP para controlar un robot humanoide. También proponen un paradigma de estimulación que le permite a un usuario controlar el brazo de un robot. Dicho paradigma permite enseñarle al robot nuevos movimientos y a navegar. En Carlson *et al.* (2013) reportan una BCI híbrida basada tanto en ERD/ERS como en la actividad muscular residual para controlar un robot que cuenta con navegación automática y evasión de obstáculos. La BCI híbrida permite indicarle al robot que inicie el movimiento y que se detenga. En Chen, Zheng y Wang (2015) implementan una BCI para controlar un robot mediante cuatro opciones de imaginación motora.

En Burget *et al.* (2017) presentaron una BCI basada en cinco tareas mentales (movimiento de un dedo de la mano derecha, movimiento de los dedos gordos de ambos pies, rotación de objetos, generación de palabras y descanso) para controlar un planeador de tareas de alto nivel. El robot navega de forma autónoma de una ubicación a otra, sujeta objetos, suelta objetos (en una repisa o mesa), vierte líquido de una botella en una taza y provee al usuario con una bebida. Las tareas se actualizan constantemente con la información que se adquiere con los sistemas de percepción del robot.

### **Motivación y objetivos**

Este proyecto se enfoca en potenciar la calidad de vida de personas que experimentan dificultades de comunicación debido a alteraciones en las vías nerviosas periféricas, en las placas

neuromusculares o en los músculos mismos, como consecuencia de daños sufridos por accidentes o por condiciones médicas como el síndrome de enclaustramiento o la esclerosis lateral amiotrófica. Para ello se propone el desarrollo de interfaces cerebro-computadora que amplíen sus capacidades comunicativas. Esta iniciativa está alineada con los objetivos estratégicos para el mejoramiento de la salud a nivel nacional, según lo establecido en el Plan Nacional de Desarrollo y los Programas Nacionales Estratégicos del Conahcyt (Pronaces 2024). Como se explicó anteriormente, existen muy pocas BCI dirigidas a controlar robots de servicio y las que existen no explotan el potencial de interacción que puede darse entre el usuario y el robot. Por esta razón, los objetivos de este proyecto son:

1. Desarrollar matrices de estimulación basadas en los paradigmas de P300 y tareas mentales que permitan la interacción entre el usuario de la BCI y el robot de servicio.
2. Generar e implementar algoritmos computacionales basados en inspiración biológica (por ejemplo, redes neuronales artificiales y computación evolutiva) para *i*) extraer y clasificar distintos tipos de señales fisiológicas y cerebrales para controlar un robot de servicio doméstico, y *ii*) proveer de autonomía al robot de servicio para que lleve a cabo tareas de reconocimiento visual, navegación (por ejemplo, planeación de rutas) y manipulación de brazos robóticos.
3. Construir un robot de servicio de bajo costo al que se le pueda dar mantenimiento en la propia UAM.

### **Metodología**

A continuación, se describen los avances actuales de cada una de las etapas del proyecto. Es importante señalar que es un proyecto que cubre tanto investigación como docencia a nivel licenciatura, por lo que algunos de los algoritmos implementados son muy básicos y no necesariamente del estado del arte.

### **Adquisición de las señales cerebrales**

En este proyecto se trabaja con señales fisiológicas que se adquieren con un sistema de bajo costo llamado OpenBCI (Frey 2016), el cual cuenta con las tarjetas Cyton y Daisy que amplifican y transforman la señal analógica a digital. Estas tarjetas permiten la adquisición de 16 canales, tienen un procesador de 32 bits y los datos se muestrean a 250Hz en cada uno de los canales.

Desarrollamos un programa para adquirir señales directamente de las tarjetas y almacenarlas en disco duro. Adicionalmente, el programa segmenta la serie de tiempos de acuerdo con el protocolo de estimulación y la etiqueta, e indica el tipo de estímulo recibido en cada momento.

### **Estimulación**

Para estimular visualmente a los usuarios, en este proyecto se desarrollaron dos tipos de matrices: la matriz de selección general y la matriz de selección específica. El objetivo de ambas es reproducir el paradigma *oddball*, el cual busca generar una respuesta en el usuario al presentar un elemento raro o poco frecuente (Donchin *et al.* 2000, Klobassa *et al.* 2009). Dicha respuesta es un pico positivo que se puede detectar, en general, 300 ms posteriores a la presentación del estímulo. En García y Gentiletti (2008) se sugiere que el P300 está compuesto por la onda P3a y P3b. La onda P3a se origina en los mecanismos de atención frontal dirigidos por el estímulo durante el procesamiento de tareas. La onda P3b se origina en la actividad parietal-temporal asociada con la atención y parece estar relacionada con el consiguiente procesamiento en la memoria.

La matriz de selección general le permite al usuario darle instrucciones generales al robot, como que se dirija a un lugar, que busque un objeto y que siga o busque a una persona. Está compuesta de una matriz de 3×4 con iconos de colores que representan de forma abstracta instrucciones que el usuario puede darle al robot. En este deletreador el usuario fija su atención en la celda

que contiene el icono a ser comunicado mientras se intensifican aleatoriamente las filas y las columnas de la matriz. A este paradigma se le conoce como paradigma de filas o columnas, o paradigma RC (Donchin *et al.* 2000). La fila y la columna que contienen la celda atendida son elementos raros (o poco frecuentes), por lo que sólo estos eventos provocan la aparición del P300 (Donchin *et al.* 2000).

Por otro lado, la matriz de selección específica le permite al usuario darle instrucciones detalladas al robot, como que seleccione un objeto específico de una escena, por ejemplo, que elija una bebida entre varias que se encuentren en la cocina. Esta matriz se genera dinámicamente, es decir, los objetos de una imagen tomada por el robot en un momento específico se convierten en los iconos de la matriz (Bell *et al.* 2008) por lo que no sólo los iconos, sino el tamaño de la matriz, son variables.

Para identificar los objetos en la imagen adquirida por las cámaras del robot, se utiliza el algoritmo SIFT (Lowe 1999) que extrae características invariantes a factores de escala, traslación, rotación y parcialmente invariantes a cambios de iluminación y afinidades.

De la misma forma, se extraen características en las imágenes de distintos objetos aislados y previamente almacenados en una base de datos. Después, se emparejan los descriptores de características de ambos conjuntos y se calcula la distancia entre ellos. El objeto que esté en la escena será aquél con la menor distancia.

Una vez que se identifican los objetos en la escena, se dibuja un contorno que permitirá implementar el llamado paradigma de una sola columna (Guan, Thulasidas y Wu 2004) en el cual se intensifican aleatoriamente cada celda en lugar de cada fila y columna.

### ***Algoritmo de clasificación***

Debido a que las señales de ERP tienen una gran cantidad de ruido, es necesario que el usuario repita varias veces el proceso de estimulación para aumentar la relación señal/ruido [30]. Este proceso puede volverse inaceptablemente lento y agotador para el

usuario, por ello, gran parte del esfuerzo en el desarrollo de las BCI es estimular al usuario el menor número de veces posible, de preferencias sólo una vez (es decir, en una época única). Por otro lado, sería deseable que una BCI pudiera ser utilizada por cualquier usuario sin que fuera necesaria una etapa de calibración inicial. Esto podría lograrse si pudiera generalizarse la información adquirida con anterioridad de otros usuarios. La detección de P300 basada en la información recuperada durante la etapa de calibración por un solo usuario se conoce como clasificación de época única del usuario, mientras que la detección basada en la información recuperada por otros usuarios se conoce como clasificación entre usuarios.

Con el objetivo de conseguir una clasificación de época única tanto para un usuario como entre usuarios, en este proyecto desarrollamos una arquitectura basada en red neuronal convolucional (CNN) simple, llamada SepConv1D, que consiste en un bloque convolucional 1D separable en profundidad seguido por un bloque de clasificación Sigmoide (Alvarado-González, Fuentes-Pineda y Cervantes-Ojeda 2021).

Encontramos que la arquitectura propuesta no sólo puede resolver estos problemas tan bien como los métodos basados en CNN de última generación, sino que puede hacerlo con el costo computacional más bajo. Además, proporcionamos elementos para demostrar que hacer que la red neuronal sea más profunda y altamente compleja no conduce a resultados más precisos en todos los casos. Para probar nuestra hipótesis, llevamos a cabo experimentos con cuatro conjuntos de datos de referencia procesados con diez métodos de última generación basados en CNN, una red neuronal completamente conectada de capa oculta con dos neuronas, y el SepConv1D. Comparamos los valores promedio de AUC resultantes, junto con sus desviaciones estándar. Además, analizamos las diferencias entre el número de parámetros, la cantidad de operaciones de coma flotante por segundo (FLOPS) y el tiempo de inferencia requerido por los modelos.

Los resultados mostraron que, en todos los casos, SepConv1D requiere el menor número de parámetros: en el mejor de los casos, solo usó 225 mientras que EEGNet, el modelo con el número más cercano de parámetros usó 1 474, incluso cuando no había diferencia estadísticamente significativa entre los valores promedio de AUC. Además, la diferencia entre el número de FLOPS y el tiempo de inferencia requerido por su entrenamiento y clasificación fue notoriamente diferente: EEGNet requirió 2 801 FLOPS y  $2.74 \times 10^{-4}$  ms de tiempo de inferencia, mientras que SepConv1D requirió solo 443 FLOPS y menos de la mitad del tiempo de inferencia. En dos conjuntos de datos, la diferencia entre los valores promedio de AUC de EEGNet y SepConv1D fueron estadísticamente significativos para calcular la clasificación de época única de un usuario. La ganancia de EEGNet fue de 0.04 pero requirió, en el mejor de los casos, 1814 FLOPS adicionales y  $4.74 \times 10^{-4}$  milisegundos adicionales. Por otro lado, observamos que no hubo diferencias estadísticamente significativas entre ninguno de los métodos para la clasificación entre usuarios. Estos hallazgos son importantes porque se pueden construir dispositivos más simples, más baratos, más rápidos y, por lo tanto, más portátiles.

### ***Construcción del robot de servicio***

El robot de servicio doméstico que se utiliza en este proyecto está compuesto de dos brazos robóticos antropomorfos, una base robótica y un soporte que los une. Ninguna de las partes físicas ha sido diseñada por nosotros, sin embargo, han sido construidas desde sus componentes más elementales. Por esta razón, no sólo logramos reducir los costos, sino que podemos darle mantenimiento dentro de la propia UAM.

El diseño de los brazos es de Ryan Gross (2018). Cada brazo está compuesto por un hombro, un codo, un antebrazo, y una mano con cinco dedos. El hombro tiene tres ejes para obtener un movimiento humano completo. El codo utiliza un servomotor grande conectado a él por una correa ranurada (el mismo tipo de

correa utilizada en la mayoría de las impresoras 3D FDM); la correa permite que el rango completo de 180 grados del servomotor se canalice a los grados que se mueve el codo. La mano está compuesta de seis servomotores capaces de obtener un movimiento de muñeca de 180 grados; tiene un pulgar oponible y un movimiento de dedo individual para todos los dedos, excepto el índice y el pulgar que se mueven como una unidad.

La base robótica es un Arlo Robot de la empresa Parallax. Es un robot móvil de interior que utiliza el microcontrolador Propeller Parallax de ocho núcleos. La base robótica está compuesta de dos ruedas de aluminio y una placa de control que requiere controladores de motor externos. Adicionalmente, tiene tres sensores ultrasónicos al frente y uno en la parte trasera.

### **Navegación**

La navegación autónoma de robots en un ambiente dinámico incluye la creación de un plan de acción basado en el conocimiento sobre el medio ambiente y un sistema reactivo que permita una respuesta rápida a los cambios en el entorno (Dudek y Jenkin 2010). Básicamente, el sistema reactivo evita colisiones con obstáculos al reconocer el entorno en función de la información obtenida por los sensores del robot y reacciona modificando los actuadores del robot. El sistema reactivo se puede construir con base en el diseño de una solución específica. En la actualidad se ha desarrollado un algoritmo básico que le permite al robot evitar obstáculos visibles por cuatro sensores ultrasónicos. Sin embargo, se tiene la desventaja de tener que pensar en todos los factores que deben considerarse, y si falta uno de ellos la solución no sería tan eficiente como podría. Por el contrario, en el enfoque cognitivo como el que proponemos, las soluciones proporcionadas tienen en cuenta toda la información detectada por los robots.

En este proyecto desarrollamos una metodología que permite la evolución simultánea de neurocontroladores para múltiples robots (López-Jaimes *et al.* 2018). Debido a que contamos con

un simulador para robots similares a automóviles, la solución la enfocamos a este tipo de problemas. Sin embargo, puede ser fácilmente aplicado al robot de servicio doméstico. La metodología consta de tres módulos: el módulo algoritmo evolutivo basado en RankGA (Cervantes y Stephens 2009), el módulo de controladores de robots basado en una red neuronal *feedforward*, y el módulo de robots, donde se utilizó el simulador TORCS para los experimentos. La ventaja de evolucionar de forma simultánea los neurocontroladores resultantes es que las simulaciones son más rápidas. Con respecto a los resultados obtenidos, observamos que los neurocontroladores pudieron manejar de manera adecuada el volante para avanzar y tomar curvas. También aprendieron a cambiar la marcha y el pedal del acelerador para reducir o aumentar la velocidad. Sin embargo, se necesita más trabajo para producir neurocontroladores de desempeño satisfactorio. Por ejemplo, los desarrollados con nuestra metodología no aprendieron a evitar obstáculos con eficiencia. Parece ser muy difícil obtener los pesos adecuados para que la red neuronal aprenda a conducir en la pista lo más rápido posible mientras intenta evitar un choque con otros automóviles. Creemos que incrementar las neuronas intermedias en la red neuronal podría ayudar. Además, ejecutar el algoritmo evolutivo durante más generaciones debería producir mejores resultados.

Por otro lado, se desarrolló un control remoto utilizando un *leap motion* para controlar el robot de forma natural al utilizar las manos.

### **Brazos robóticos**

Por último, se desarrolló un modelo cinemático rotacional para manipular un brazo robótico antropomórfico con un *leap motion*. Esto permitirá que un usuario realice de manera natural movimientos de agarre de distintos objetos. Los datos generados por el agarre conforman tramas que guardan información sobre el tiempo, velocidad y movimiento angular de la mano. Las

configuraciones de agarre determinan el ángulo que debe girar cada uno de los servomotores para que el brazo robótico pueda sujetar un determinado objeto. Con esta información se podrán generar casos de uso para entrenar un algoritmo de clasificación. De esta manera, eventualmente la mano robótica podrá funcionar de manera autónoma con movimientos muy parecidos a aquellos de los humanos.

## **Conclusión**

En este capítulo se presentó el avance del proyecto interfaz cerebro computadora para controlar un robot de servicio doméstico. Este proyecto podría incrementar las capacidades comunicativas, así como proveer de cierta independencia y calidad de vida a personas con alteraciones en las vías nerviosas periféricas, en las placas neuromusculares o en los músculos mismos, como consecuencia de daños sufridos por accidentes o por condiciones médicas como el síndrome de enclaustramiento o la esclerosis lateral amiotrófica. Esta iniciativa está alineada con los objetivos estratégicos para el mejoramiento de la salud a nivel nacional, según lo establecido en el Plan Nacional de Desarrollo y los Programas Nacionales Estratégicos del Conahcyt. Las interfaces cerebro computadora son herramientas que le permiten a un usuario comunicarse con distintos dispositivos mediante sus señales cerebrales. En este proyecto se han desarrollado interfaces de estimulación para generar potenciales evocados relacionados con eventos que le permiten a un usuario dar instrucciones generales y específicas a un robot de servicio doméstico. Las señales cerebrales son adquiridas con un equipo de bajo costo. Posteriormente son procesadas y clasificadas con algoritmos basados en redes neuronales convolucionales que han demostrado tener un desempeño semejante a los algoritmos del estado del arte y con un uso de recursos computacionales significativamente menor. Adicionalmente, se construyó un robot de servicio doméstico de bajo costo y al que se le puede dar mantenimiento dentro de la

propia UAM. Dicho robot cuenta con una base robótica con tres sensores ultrasónicos que puede navegar de forma autónoma y evitar obstáculos de forma muy básica, la cual podrá controlarse con un control remoto, ya funcional, utilizando un *leap motion*. Para mejorar la navegación autónoma, se ha desarrollado un sistema reactivo basado en un neurocontrolador que evita colisiones con obstáculos al reconocer el entorno en función de la información obtenida por los sensores del robot y modificando sus actuadores cuando es necesario. El robot cuenta con dos brazos antropomorfos. Estos brazos pronto serán controlados con modelo cinemático rotacional ya desarrollado que permitirá manipularlos con un *leap motion*. Como trabajo futuro, se agregará a la BCI el paradigma de tareas mentales y se generará un clasificador para identificar las instrucciones de control. Como mecanismo de retroalimentación, se incorporará el potencial relacionado al error y las señales fisiológicas que reflejan estrés en el usuario al momento de generarse un error en la clasificación del P300.

Con respecto a la construcción del robot, se le agregará una cara que permita una interacción más natural con el usuario. También se integrarán los siguientes sensores: un LIDAR, una cámara omnidireccional y una cámara de seguimiento y profundidad RealSense. Se modificará el diseño para ayudarlo a mejorar su apariencia. Se buscarán materiales más resistentes y se agregará tacto mediante la incorporación del diseño textil.

Por otro lado, se incorporará un neurocontrolador basado en optimización multiobjetivo (que se ya se encuentra en desarrollo) para mejorar la navegación reactiva.

Adicionalmente, se desarrollará un algoritmo para controlar las trayectorias de dos brazos robóticos con el objetivo de que sean capaces de colaborar para sujetar distintos tipos de objetos.

## **Agradecimientos**

Quisiera agradecer a los miembros del grupo Cerebral (Computational nEuRoscience, Evolutionary roBotics and InteRfAces

Laboratory): al co-responsable Dr. Antonio López, a los alumnos Liliana López, Eric Rovelo, Roberto Silva, Ana Paula Trujillo, María de Jesús Sánchez, Guillermo Lizalde, Francisco Álvarez, Antonio Guerrero y Diego González.

Así como a los colaboradores del proyecto: Dr. Luis Franco (UAM), Dr. Gibrán Fuentes (IIMAS-UNAM) y al M. en C. Hernando Ortega (IIMAS-UNAM). Entre todos hemos desarrollado las distintas partes del proyecto.

## Referencias

- ACEVEDO, Rubén, Gerardo Gentiletti, Verónica Medina y Leonardo Rufiner (2009). “Extracción de características en ICC mediante métodos basados en diccionarios óptimos: Resultados preliminares”. En *Anales de las II Jornadas Argentinas sobre Interfaces Cerebro Computadora JAICC 2009*, 49-52.
- ALVARADO-GONZÁLEZ, Montserrat, Gibrán Fuentes-Pineda y Jorge Cervantes-Ojeda (2021). “A few filters are enough: Convolutional Neural Network for P300 Detection”. *Neurocomputing* 425: 37-52.
- ATUM, Yanina, Gerardo Gentiletti, Rubén Acevedo y Leonardo Rufiner (2009). “Detección de P300 en Interfaz Cerebro Computadora mediante Algoritmos Genéticos y Maquinas de Soporte Vectorial”. *Memorias del XVII Congreso Argentino de Bioingeniería (SABI 2009)*, n.º 146: 51-55.
- ATUM, Yanina, Iván Gareis, Gerardo Gentiletti, Rubén Acevedo y Leonardo Rufiner (2010). “Genetic feature selection to optimally detect P300 in brain computer interfaces”. En *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. DOI: 10.1109/IEMBS.2010.5627254.
- BELL, Christian J., Pradeep Shenoy, Rawichote Chalodhorn y Rajesh P. Rao (2008). “Control of a humanoid robot by a noninvasive brain-computer interface in humans”. *Journal of Neural Engineering* 5, n.º 2: 214.

- BOBROV, Pavel, Alexander Frolov, Charles Cantor, Irina Fedulova, Mikhail Bakhnyan y Alexander Zhavoronkov (2011). "Brain-computer interface based on generation of visual images". *PLoS One* 6, n.º 6. <https://doi.org/10.1371/journal.pone.0020674>.
- BRYAN, Matthew, Joshua Green, Mike Chung, Lillian Chang, Reinhold Scherert, Joshua Smith y Rajesh P. Rao (2011). "An adaptive brain-computer interface for humanoid robot control". En *11th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 199-204. DOI: 10.1109/Humanoids.2011.6100901.
- BURGET, Felix, Lukas Dominique Josef Fiederer, Daniel Kuhner, Martin Vólker, Johannes Aldinger, Robin Tibor Schirrmester, Chau Do, Joschka Boedecker, Bernhard Nebel y Tonio Ball (2017). "Acting thoughts: Towards a mobile robotic service assistant for users with limited communication skills". En *2017 European Conference on Mobile Robots (ECMR)*, 1-6. IEEE. DOI: 10.1109/ECMR.2017.8098658.
- CARLSON, Tom, Luca Tonin, Serafeim Perdakis, Robert Leeb y José del R. Millan (2013). "A hybrid BCI for enhanced control of a telepresence robot". En *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2013*, 3097-3100. DOI: 10.1109/EMBC.2013.6610196.
- CERVANTES, Jorge y Chirs Stephens (2009). "Limitations of existing mutation rate heuristics and how a rank GA overcomes them". *IEEE Transactions on Evolutionary Computation* 13, n.º 2: 369-397.
- CHEN, Yan, Shuhua Zheng y Xiangzhou Wang (2015). "A Brain-Robot Interface by BCI based on Repeated Binary CSP". En *2015 Chinese Automation Congress*, 826-830. IEEE. DOI: 10.1109/CAC.2015.7382612.
- DONCHIN, Emanuel, Kevin M. Spencer y Ranjith Wijesinghe (2000). "The mental prosthesis: Assessing the speed of a P300-based brain-computer interface". *IEEE Transactions on Rehabilitation Engineering* 8: 174-179.

- DUDEK, Gregory y Michael Jenkin (2010). *Computational Principles of Mobile Robotics*. New York: Cambridge University Press.
- FAGER, Susan, Dave Beukelman, Renee Karantounis y Tom Jakobs (2006). "Use of safe-laser access technology to increase head movement in persons with severe motor impairment: a series of case reports". *Augmentative and Alternative Communication* 22, n.º 3: 222-229.
- FREY, Jérémy (2016). "Comparison of a consumer grade EEG amplifier with medical grade equipment in BCI applications". En *International BCI meeting*. Asilomar.
- GARCÍA, Eliana y Gerardo Gabriel Gentiletti (2008). "Interfaz Cerebro Computadora (ICC) basada en el potencial relacionado con eventos P300: análisis del efecto de la dimensión de la matriz de estimulación sobre su desempeño". *Revista Ingeniería Biomédica* 2, n.º 4: 26-33.
- GAREIS, Iván, Gerardo Gentiletti, Rubén Acevedo y Leonardo Rufiner (2011). "Feature Extraction on Brain Computer Interfaces using Discrete Dyadic Wavelet Transform: Preliminary Results". *Journal of Physics: Conference Series* 313, n.º 1. DOI 10.1088/1742-6596/313/1/012011.
- GENTILETTI, Gerardo, José Guillermo Gebhart, Rubén Acevedo, Óscar Yáñez y Verónica Medina-Bañuelos (2009). "Command of a simulated wheelchair on a virtual environment using a Brain-Computer Interface". *Ingénierie et Recherche Biomédicale* 30, n.º 5-6: 218-225.
- GROSS, Ryan (2018). "Humanoid robotic torso proto1". <https://www.myminifactory.com/object/3d-print-humanoid-robotic-torso-proto1-48754>.
- GUAN, Cuntai, Manoj Thulasidas y Jiankang Wu (2004). "High performance P300 speller for brain-computer interface". En *IEEE International Workshop on Biomedical Circuits and Systems*, 13-16. IEEE. DOI: 10.1109/BIOCAS.2004.1454155.
- HOSLER, Betsy A., Teepu Siddique, Peter C. Sapp, Wen Sailor, Michael C. Huang, Anwar Hossain, Jasper R. Daube, Martha Nance,

- Chaohong Fan, Jocelyn Kaplan, Wu-Yen Hung, Diane McKenna-Yasek, Jonathan L. Haines, Margaret A. Pericak-Vance, H. Robert Horvitz y Robert H. Brown, Jr. (2000). "Linkage of familial amyotrophic lateral sclerosis with frontotemporal dementia to chromosome 9q21-q22". *Journal of the American Medical Association* 284, n.º 13: 1664-1669.
- KLOBASSA, Daniela Sabine, Theresa Vaughan, Peter Brunner, N. E. Schwartz, Jonathan Wolpaw, Christa Neuper y Eric Sellers (2009). "Toward a highthroughput auditory P300-based Brain-Computer Interface". *Clinical Neurophysiology* 120, n.º 7: 1252-1261.
- LONG, Jinyi, Yuanqing Li, Hongtao Wang, Tianyou Yu, Jiahui Pan y Feng Li (2012). "A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair". *IEEE Transactions on Neural Systems & Rehabilitation Engineering* 20, n.º 5, p. 720.
- LÓPEZ, Antonio, Jorge Cervantes, Maria Gómez-Fuentes y Montserrat Alvarado (2018). "Simultaneous evolution of neuro-controllers for multiple car-like robots". *Research in Computing Science* 147, n.º 10: 29-44.
- LOWE, David G. (1999). "Object recognition from local scale-invariant features". En *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1150- 1157. IEEE. DOI: 10.1109/ICCV.1999.790410.
- MILLÁN, José del R., Frédéric Renkens, Josep Mouriño y Wulfram Gerstner (2004). "Noninvasive brain-actuated control of a mobile robot by human EEG". *IEEE Transactions on Biomedical Engineering* 51: 1026-1033.
- NIEDERMEYER, Ernst y Fernando Lopes da Silva (2005). *Electroencephalography: Basic Principles Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins.
- NIJBOER, Femke, Eric W. Sellers, Jürgen Mellinger, Mary Ann Jordan, Tamara Matuz, Adrian Furdea, Sebastian Halder, Ursula Mochty, D. J. Krusienski, T. M. Vaughan, J. R. Wolpaw, N. Birbaumer y

- A. Kübler (2008). “A P300-based Brain-Computer Interface for people with Amyotrophic Lateral Sclerosis”. *Clinical Neurophysiology* 119, n.º 8: 1909-1916.
- PATHIRAGE, Indika Upashantha, Redwan Alqasemi y Rajiv Dubey (2012). “P300 Brain Computer Interface Based Task Oriented Control of a WheelChair Mounted Robotic Arm”. *IEEE EMB/CAS/SMC Workshop on Brain-Machine-Body Interfaces*.
- PRONACES (2024). “Programas Nacionales Estratégicos”. Consejo Nacional de Humanidades, Ciencias y Tecnologías. Gobierno de México. Acceso el 9 de enero de 2024. <https://conahcyt.mx/pronaces/>.
- SELLERS, Eric W. y Emanuel Donchin (2006). “A P300-based Brain-Computer Interface: Initial tests by ALS patients”. *Clinical Neurophysiology* 117, n.º 3: 538-548.
- SIRVENT, José L., José M. Azorín, Eduardo Iáñez, Andrés Úbeda y Eduardo Fernández (2010). “P300-Based Brain-Computer Interface for Internet Browsing”. En *Trends Practical Applications of Agents and Multi-agent Systems*, editado por Yves Demazeau, Frank Dignum, Juan Manuel Corchado, Javier Bajo, Rafael Corchuelo, Emilio Corchado, Florentino Fernández-Riverola, Vicente Julián, Pawel Pawlewski y Andrew Campbell, 615-622. Springer Berlin Heidelberg.
- VALBUENA, Diana, Marco Cyriacks, Ola Friman, Ivan Volosyak, Axel Graser (2007). “Brain-computer interface for high-level control of rehabilitation robotic systems”. En *IEEE 10th International Conference on Rehabilitation Robotics*, 619-625. IEEE. DOI: 10.1109/ICORR.2007.4428489.
- WOLPAW, Jonathan R., Niels Birbaumer, William J. Heetderks, Dennis McFarland, P. Hunter Peckham, Gerwin Schalk, Emanuel Donchin, Louis A. Quatrano, Charles J. Robinson y Theresa M. Vaughan (2000). “Brain-computer interface technology: a review of the first international meeting”. *IEEE Transactions on Rehabilitation Engineering* 8, n.º 2: 164-173.

- ZHANG, Haihing, Cuntai Guan y Chuanchu Wang (2005). "A statistical model of brain signals with application to brain-computer interface". *Proceedings of the IEEE Engineering Medicine and Biology Society* 5: 5388-5391.
- ZHAO, Li, Chuo Li y Shigang Cui (2007). "Service Robot System Based on Brain-computer Interface Technology". En *Third International Conference on Natural Computation (ICNC 2007)*, 349-353. IEEE. DOI: 10.1109/ICNC.2007.657.

### **Sobre la autora**

#### **Dra. Montserrat Alvarado**

La Dra. Montserrat Alvarado González obtuvo una maestría y doctorado en Ciencias de la Computación del Programa de Posgrado en Ingeniería y Ciencias de la Computación de la Universidad Nacional Autónoma de México (UNAM). Actualmente, es profesora-investigadora del Departamento de Matemática Aplicada y Sistemas de la Universidad Autónoma Metropolitana (UAM). Es codirectora del Laboratorio de neurociencia computacional, robótica evolutiva e interfaces (Cerebral), donde se desarrollan hardware y algoritmos para una interfaz cerebro-computadora que controla un robot de servicio. Adicionalmente, se desarrollan biosensores y dispositivos para adquirir energía de la interacción plantas-microbios-tierra. Durante su carrera académica ha trabajado en las áreas de visión computacional, reconocimiento de patrones, robótica e interfaces cerebro-computadora.

# Sistemas Multi-Agente y Modelación Basada en Agentes: una herramienta interdisciplinaria

WULFRANO ARTURO LUNA RAMÍREZ

## Resumen

La Inteligencia Artificial desde sus inicios se ha caracterizado como una ciencia de naturaleza interdisciplinaria. Bajo esta perspectiva, los Sistemas Multi-Agente (SMA) y la Modelación Basada en Agentes (MBA) representan tanto una propuesta unificadora del área, como una posibilidad de vinculación con otras disciplinas, de forma muy prometedora con las Ciencias Sociales y las Humanidades. De manera muy clara, estas interacciones evidencian el potencial de las convergencias interdisciplinarias en torno a objetos de estudio comunes: los sistemas sociales como Sistemas Complejos.

Por un lado, los Sistemas Multi-Agente y la Modelación Basada en Agentes se benefician de la interacción con los conocimientos y objetos de estudio provenientes de las humanidades y las ciencias sociales al obtener bases sólidas para realizar sus indagaciones y desarrollos, configurar abstracciones y, aún más, crear marcos de referencia para elaborar sus propias teorías.

En forma recíproca, dichos enfoques proporcionan al análisis humanístico social una herramienta de representación y simulación que permite determinar distintas escalas de observación, tanto estática como dinámica; representar leyes o hipótesis, y condiciones socioambientales, y plantear distintas interacciones entre los elementos que la constituyen que pueden denotar tanto individuos como grupos o poblaciones. Además, posibilitan la diferenciación entre individuos considerados de la misma naturaleza, además de favorecer la observación y medición de fenómenos como la autoorganización y la emergencia (estructural o de comportamientos).

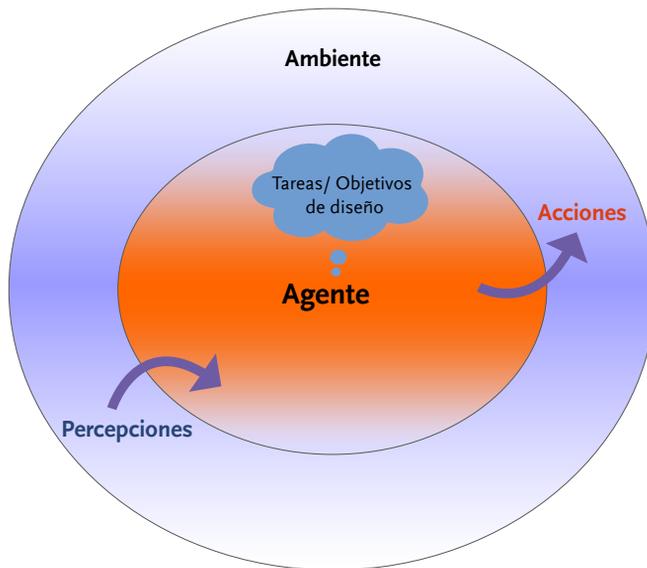
Tal intercambio ha dado lugar a nuevos enfoques y disciplinas híbridas con distintos énfasis. Aquí se presenta una breve descripción de los SMA y la MBA haciendo hincapié en las posibilidades de interacción interdisciplinaria.

### **Inteligencia Artificial**

La Inteligencia Artificial (IA) puede entenderse como aquella disciplina orientada al estudio y síntesis (es decir, construcción) de sistemas que aspiran a (por lo menos) exhibir un comportamiento inteligente. Para tal propósito, los investigadores en el área se han planteado distintos enfoques, como los inspirados en modelos naturales a los que se les atribuye inteligencia, o bien aquellos referidos a una noción abstracta de inteligencia (Russel y Norvig 2004). Su aportación, liderada en buena medida por Sistemas Multi-Agente (SMA), incluye tanto el diseño y desarrollo de simulaciones para representar sistemas físicos o biológicos, como sociales, por lo que esta disciplina también ha contribuido al estudio de los Sistemas Complejos (Wooldridge y Ciancarini 2000, Jennings 1999, Railsback y Grimm 2019).

El componente principal de tales sistemas es el agente. Un agente puede concebirse como una entidad (artificial o biológica) con un propósito de diseño específico que sostiene interacciones con su medio ambiente a través de percepciones (lectura de datos) y

acciones (modificación del entorno o emisión de datos) (Russel y Norvig 2004, Gilbert 2019) como se muestra esquemáticamente en la Figura 1. Entre sus características relevantes se encuentran la persistencia (búsqueda continua del logro de sus objetivos), su racionalidad (hacer aquello que lo lleve a la consecución de tales objetivos), su autonomía (no requieren la operación directa constante de un usuario), su capacidad de aprendizaje (la mejora de su operación en pro del logro de objetivos, a través de la experiencia) y su habilidad social (comunicación con pares y con otros elementos del entorno) (Russel y Norvig 2004; Singh, Padgham y Logan 2016; Rao 1996). Esta última característica es la que faculta a tales sistemas para constituirse en ensambles compuestos de varios agentes (siendo posible que ostenten una arquitectura y configuración homogénea o heterogénea, incluso con diseños y objetivos distintos o contrapuestos), es decir, SMA.



*Figura 1.* Esquema de un agente artificial. Elaboración propia.

A través del uso de agentes, se puede obtener un alto grado de realismo al representar conglomerados sociales. Además, los SMA son un enfoque de desarrollo de software que integra herramientas de análisis y diseño con un nivel de abstracción y expresividad superiores a las de otros paradigmas —como el Orientado a Objetos—, lo que facilita el desarrollo de implementos computacionales aplicables a distintos dominios (Jennings 1999).

Como se puede observar, hay una intersección multidisciplinaria en las concepciones y aplicaciones de los SMA. En la Figura 2 se muestra de manera gráfica, y se hace énfasis en las áreas mencionadas y las Humanidades y las Ciencias Sociales.

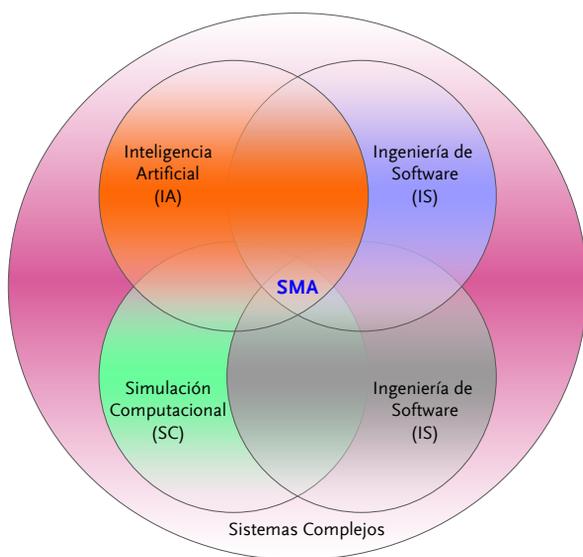


Figura 2. Esquema de SMA-MBA. Elaboración propia.

### Simulación Computacional de Sistemas Complejos

Una subárea de los SMA es la Modelación Basada en Agentes (MBA), que configura una técnica computacional orientada a la representación y explicación generativa de Sistemas Complejos

(Gilbert 2019). Ésta tiene profundamente imbricada la idea del surgimiento emergente de comportamientos durante el curso de la simulación. En otras palabras, es factible observar comportamientos globales a partir de acciones simples de sus componentes individuales (los agentes) que superan la mera composición de partes en un todo, pues se presentan en ausencia de un control centralizado y en un ambiente en constante cambio.

La MBA permite representar directamente los elementos del sistema, así como las interacciones que se presentan durante su operación, ya sea entre sus mismos componentes, o con el ambiente donde están situados los agentes. Además, se puede considerar la dimensión temporal, de ahí que se pueda observar la aparición de fenómenos emergentes.

Por lo anterior, esta técnica es una candidata idónea para la representación de sistemas sociales, lo que la sitúa por encima de los modelos puramente estadísticos o de aquellos basados en la Teoría de Juegos (Gilbert 2019): por un lado, las descripciones verbales son flexibles, pero imprecisas, en tanto que las descripciones matemáticas son rigurosas, pero es difícil incluir heterogeneidad en ellas, por no hablar de la no-linealidad de los fenómenos y su eventual intratabilidad. Por otro lado, los modelos computacionales son rigurosos, precisos y facilitan la representación heterogénea y las interacciones de componentes, además de su reproducibilidad ad hoc con las necesidades del experimento.

Esta capacidad de representación es consecuencia del elemento que la hace posible, es decir, el formalismo: la informática. Un sistema de símbolos (los programas de computadora) faculta el diseño de componentes y mecanismos particulares, que en su ejecución permitirán observar el espacio de acciones posibles, inaccesible *a priori* ya sea por su extensión, o porque su aparición requiere del concurso de las relaciones entre individuos, objetos y su medio ambiente. Este formalismo implica, es cierto, además del conjunto de habilidades matemáticas y del dominio de aplicación, algún grado de conocimiento de informática, pues

crear un modelo implica conocer los supuestos y puntos de partida de la teoría del fenómeno, y generar reglas y algoritmos que representen sus interacciones, así como dotar de información del dominio con el que operar (Gilbert 2019, Railsback y Grimm 2019). Todo ello reafirma el carácter interdisciplinario de este tipo de esfuerzos.

### **Microscopio y telescopio para las Ciencias Sociales y las Humanidades**

Por las ventajas antes mencionadas, para las Ciencias Sociales y las Humanidades, la MBA representa una herramienta de investigación empírica interdisciplinaria cuyo potencial principal se encuentra en la posibilidad de representar —con distintos niveles de detalle— individuos, interrelaciones y ambientes, y es aplicable, por ejemplo, para analizar teorías sociales, o bien las consecuencias de la aplicación de leyes o reglamentos en la vida pública de una determinada población. Aunado a esto, los modelos pueden ser alimentados con datos reales obtenidos de las propias investigaciones en el área, ya sean de carácter histórico-demográfico, sociológico y económico, entre muchos otros. Así se han abordado modelos para analizar fenómenos como segregación racial o étnica, migración y crecimiento poblacional, dinámica de opinión y preferencias electorales, entre otros.

Si se acepta la metáfora aquí propuesta (Luna-Ramírez 2016), la MBA es análoga al microscopio, en cuanto que permite analizar los detalles más representativos de entidades sociales (grupos o individuos) y sus relaciones latentes. Asimismo es análoga al telescopio, en cuanto que permite observar en perspectiva las características de los componentes, dentro del conglomerado completo donde se ubican, incluso dentro de una cierta temporalidad. Además de poder repetir el fenómeno la cantidad de veces que se requiera. Por último, tal representación y posibilidad de repetición se realizan a un costo asequible, por no mencionar que, en ocasiones, en cuestiones sociales, es imposible reproducir el

fenómeno más allá de la especulación argumental o numérica (pongamos por caso una elección, un conflicto bélico o la aplicación de medidas económicas o sanitarias). En la sección de “dominios de aplicación” se presentan otros casos en que las MBA han sido utilizadas.

### **Desarrollo de una MBA y experimentación**

La experimentación vía MBA comienza con el desarrollo de ésta, en términos computacionales (Gilbert 2019, Railsback y Grimm 2019). Para su desarrollo es menester conocer sus componentes: los agentes (considerados unidades de comportamiento que toman decisiones individuales o colectivas); el ambiente (ubicación donde los agentes están situados, espacios geográficos o físicos, pero también conceptuales o virtuales inclusive); el comportamiento (realizado tanto por agentes como por ambientes y por otros elementos de la simulación, como objetos de un ecosistema o fenómeno); las interacciones (determinan la cadena detonante de acciones que generan nuevas acciones, las influencias entre los elementos de la simulación, es decir, del fenómeno simulado), y el tiempo (impone la dinamicidad del sistema, se estratifica comúnmente en unidades discretas, y a través de él se realizan las decisiones y se norma y observa el comportamiento).

En primer término, desarrollar una MBA es un ciclo que parte de su *especificación*, es decir, la creación del modelo: definición de los agentes, sus tipos y comportamientos, el ambiente y sus interacciones y temporalidad, etcétera. Continúa con su *formalización*, esto es su traducción a una representación lógico-matemática y algorítmica. En ambas fases se embebe la teoría social que se pretende analizar en las características de los agentes o grupos de ellos, así como las relaciones que éstos sostendrán entre sí en el curso de la simulación. Este proceso prosigue con la *implementación*, en la que la especificación se traduce a una representación computacional. Así, se está en condiciones de iniciar la *experimentación* (corrida de la simulación y su repetición a voluntad). Por último,

el proceso culmina con las dos últimas fases: *verificación*, en la que se prueba qué tanto se acerca la implementación a la especificación (de corte más informático), y *validación*, en la cual se observan y analizan los resultados (concerniente más con el dominio de la simulación), lo que permite contrastar y, en su caso, reiniciar el ciclo.

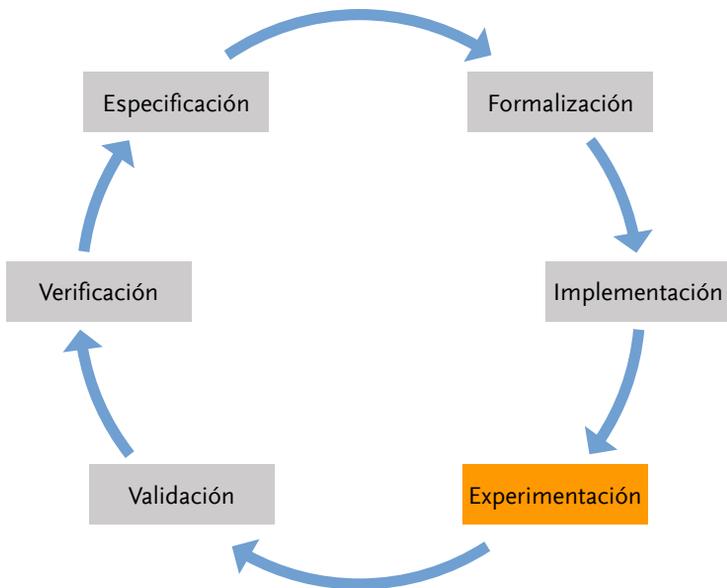


Figura 3. Etapas del desarrollo de una MBA. Elaboración propia.

Las dos últimas etapas conciernen a la puesta a punto de la MBA, lo que implica adecuar la implementación que se obtuvo (es decir, el programa computacional) a la especificación definida en un inicio. Como se comentó, esta adecuación se realiza tanto en el sentido de la Ingeniería de Software (lo que se conoce como verificación y depuración), como en el sentido de las Ciencias Sociales y las Humanidades (esto conlleva determinar la validez que la simulación tiene dentro del dominio y para el fenómeno estudiado, so pena de convertir el tele/microscopio en un caleidoscopio).

### **Análisis estadístico**

Hay que indicar que en una MBA no hay punto de predicción, cuando mucho estimación gruesa o relativa (Gilbert 2019). La emergencia se observa sólo durante la operación, y no puede ser explicada a partir de la descripción de sus componentes. Esto es una consecuencia de la dinámica de los experimentos (en particular los de carácter social): es la imposibilidad práctica de su anticipación (entre otras razones por los posibles comportamientos emergentes). Por ello no es realista esperar que se obtenga un resultado en plena correspondencia con los datos empleados en la calibración de la simulación. De forma alternativa, es común emplear análisis estadísticos para determinar la validez de ésta y, en su caso, ajustar los modelos utilizados dotándolos de un mayor número de variables (o modificar las existentes), o bien, variar el grado de detalle de la representación. Además, debe considerarse que a mayor nivel de detalle menos poder explicativo de la simulación. En otras palabras, como abstracción de un fenómeno (como modelo de él), la MBA necesariamente deja fuera detalles que no se consideran suficientemente relevantes para el experimento, quedándose con aquellos que son de mayor interés conforme las observaciones previas o la teoría bajo la cual se investiga el fenómeno. Me permito apoyarme en otra metáfora en un intento de ilustrar lo anterior. Para ello, apelo a “Del rigor en la ciencia”, de Jorge Luis Borges (2023), donde se cuestiona la conveniencia de tener un mapa del tamaño mismo del territorio representado: al lograr el máximo nivel de detalle se obtiene desafortunadamente la mínima utilidad, pues para observar el mapa ¡hay que recorrer la misma distancia del territorio de interés! En términos de Guillermo de Ockham, diríase que se prefiere la explicación más sencilla (es decir, la que menos elementos irrelevantes integra) (Russell y Norvig 2004).

### **Dominios de aplicación**

Dentro de los dominios donde se han aplicado las MBA se cuenta una amplia gama que va de las Ciencias Naturales y Exactas a

las Humanidades y Ciencias Sociales (Railsback y Grimm 2019, Gilbert 2019, Luna-Ramírez y Fasli 2017, Adam y Gaudou 2016). Algunos de los dominios típicos son: *modelos urbanos*. En este caso, es célebre el experimento de segregación social de Schelling (Wilensky 1999), que ha sido implementado en diversos sistemas de modelado: plantea dos poblaciones, representadas por individuos de distintos tipos, representados comúnmente por colores diferentes, con varios parámetros de convivencia (por ejemplo, tolerancia o aversión al diferente) que, como resultado de la operación, configura una territorialidad distinta. Otros modelos se enfocan en el análisis de poblaciones, migración, y movilidad poblacional (aunque en el último caso también se pueden aplicar a poblaciones de personas o animales en entornos no urbanos).

En cuanto a la participación política y la dinámica, y variación de la opinión pública sobre algún punto de interés (como las inclinaciones electorales, entre otros), en estos modelos la opinión puede mapearse como un número entre un rango, definido por un límite inferior y uno superior, en tanto que la variación se establece conforme reglas estocásticas o de incertidumbre, y se utiliza la vecindad espacial o afinidad de características. Aunque parece simplista, las observaciones sobre estos modelos revelan, por ejemplo, que la influencia de ciertas opiniones extremas (con valores cercanos a las cotas inferiores o superiores) afectan a la mayoría, con lo cual se han aplicado al análisis de ideologías extremistas, religiosas y políticas, principalmente.

Un caso similar al anterior son los análisis de comportamiento de consumidores, las redes industriales y las cadenas de suministro. Aquí las relaciones e interacciones entre los agentes cobran gran importancia. De tal suerte que de las redes sociales que se forman entre ellos depende la preferencia en el consumo de ciertos productos en el modelo. De manera análoga, las redes determinan aquellos productos que tendrán mayor éxito, así como la preferencia o aceptación de éstos, lo que a su vez influencia en su fabricación, que puede variar en la inclusión de ciertos componentes

(más económicos, con ciertas características o estándares de calidad o cualquier otro criterio analizado). En las cadenas de suministro, los agentes pueden representar a los proveedores de determinados productos o servicios, y así llevar un seguimiento puntual de existencias, y flujos de éstos a través de la red de relaciones que se establezcan entre ellos (Jaimez-González y Luna-Ramírez 2013).

En cuanto a los modelos para analizar la aplicación y seguimiento de políticas públicas o programas sociales, de infraestructura o servicios públicos (como el suministro eléctrico o hídrico, que ha llevado a la modificación espacial del alojamiento de recursos, su planificación acorde con las poblaciones, el establecimiento de leyes o su modificación, entre otras), los agentes representan grupos de individuos cuyas características están en función de los elementos de análisis. Así, por ejemplo, la aplicación de una ley de matrimonio entre personas implicará la determinación de género/sexo, edad, costumbres sociales, población disponible, recursos necesarios para las uniones, etcétera, mientras que las implicaciones de la ley serán variables o restricciones (representadas posiblemente como reglas) para los agentes y sus interacciones o relaciones mutuas (Walker y Davis 2013).

Otros experimentos tienen que ver con el análisis de dispersión de enfermedades, de rumores o la eficacia en la distribución de vacunas veterinarias. Un dominio bastante publicitado es el de la simulación de tráfico en carreteras y zonas urbanas.

Como se podrá ver, las posibilidades se amplían tanto como los problemas mismos. Sin embargo, no en todos los casos se recomienda desarrollar una MBA. He aquí una lista de razones para emplear una MBA (Gilbert 2019, Railsback y Grimm 2019):

1. Interés en modelar interacciones y realimentación entre actores, y entre los actores y su medio.
2. Si la heterogeneidad de los actores es importante en el sistema.
3. Si nos interesa la dinámica espacial del sistema.

4. Si la historia (acciones pasadas) son importantes para el desarrollo de los acontecimientos presentes/futuros del sistema.
5. Si los actores tienen comportamientos cambiantes o se adaptan a lo largo del tiempo.

En las siguientes secciones se aborda la importancia de las MBA respecto de las problemáticas nacionales, que impactan o se ven afectadas también por las circunstancias y esfuerzos internacionales para la atención de problemas sociales.

### **Problemáticas prioritarias nacionales**

Como ya se indicó, los SMA y las MBA tiene un ámbito de aplicación para analizar y plantear soluciones en torno a problemas sociales diversos, lo que no exime a los planteados por los Programas Nacionales Estratégicos del Consejo Nacional de Humanidades, Ciencias y Tecnologías (Conahcyt) (Pronaces 2024), expedidos a mediados de 2023 por el gobierno mexicano, cuyo interés es organizar los esfuerzos investigativos sobre algunas problemáticas nacionales diagnosticadas por él como de atención urgente y que en sus lineamientos textualmente indica que se debe diagnosticar, prospectar y proponer acciones y medidas para su atención y solución. Lo anterior, además, supone la generación de nuevos conocimientos y desarrollo científico-tecnológico (Secretaría de Gobernación 2023). Si se atiende a lo expuesto en la Figura 3 y a la lista de razones antedicha, se podrán observar —sin sobreinterpretar la idoneidad de los SMA y las MBA respecto de lo planteado por los Pronaces tanto en las temáticas (problemas sociales) como en las formas de operación y sus finalidades científico-tecnológicas— medios y métodos para analizar y superar dichos problemas. Dentro de las temáticas enunciadas por los Programas se cuentan: salud, agua, educación, cultura, vivienda, soberanía alimentaria, agentes tóxicos y procesos contaminantes, seguridad humana, sistemas socio-ecológicos y energía y cambio climático (incluida la transición energética) (Secretaría de Gobernación 2023).

A continuación se exponen un dominio de aplicación que se alinea con los Pronaces, de manera directa con seguridad humana, y de manera indirecta con educación y cultura. Además de tener incidencia también en los Objetivos de Desarrollo Sostenible definidos por la Organización de Naciones Unidas (Gamez 2015), a saber: 3. Salud y bienestar, particularmente en su meta 3.d: reforzar la capacidad de todos los países, en particular los países en desarrollo, en materia de alerta temprana, reducción de riesgos y gestión de los riesgos para la salud nacional y mundial.

### ***Simulación de desastres***

Mediante la MBA es posible estudiar *in silico* fenómenos sociales presentes y pasados en diversos ámbitos (Gilbert 2019). Un dominio de aplicación que ha despertado gran interés en los sectores académico, gubernamental y empresarial es la simulación de contingencias y desastres naturales (Luna-Ramírez y Fasli 2017).

Al simular las condiciones de un entorno, ya sea un edificio, o una demarcación citadina o rural de tamaño diverso, se pueden poner a prueba los planes de rescate y administración de la contingencia, o bien ayudar a entender el comportamiento de grupos de personas durante ella. De tal manera, se pueden simular los flujos de personas durante una evacuación en edificios, escuelas o complejos habitacionales. Además de la simulación del tráfico de automóviles y personas (incluyendo a las mismas autoridades en misiones de salvamento o mitigación), también se pueden representar las situaciones de peligro en un determinado inmueble, así como el comportamiento de los individuos durante y después de los eventos, con lo que se pueden representar casos de trauma o ayudar a planificar los requerimientos de auxilio médico y rescate, además del seguimiento post-evento, ya sea médico o psicológico.

Como ya se dijo, mediante los resultados observados en la simulación, pueden anticiparse políticas de acción para generar una respuesta apropiada, incluso gestionar la creación de

infraestructura. Pero se puede ir más allá: las simulaciones proporcionan una herramienta digital para la enseñanza y entrenamiento de personal de rescate. Los grupos de rescate pueden tener acceso a una representación de ciertas situaciones donde su intervención es requerida. Dado que dichas simulaciones pueden integrar diversas tecnologías como la realidad mixta (combinación de realidad virtual o aumentada), dotaría a los rescatistas de una forma de tener estas experiencias previas, con una inversión de recursos significativamente menor, y con ello prepararse para enfrentar los casos en campo de una manera más realista (pues las MBA no impide que se hagan trabajos simulados en campo, sino que ofrece una herramienta complementaria).

Desde luego, en México, país que padece el azote de diversos fenómenos naturales como sismos, huracanes e inundaciones, el uso adecuado de esta clase de simulaciones tendría verdadera utilidad. No sólo para que las autoridades competentes planifiquen y realicen acciones preventivas y de respuesta, sino incluso para ayudar a la concientización de la población en general.

### ***Herramientas de MBA para modelar desastres***

Como se dijo anteriormente, la MBA se puede emprender desde cero, es decir, con un desarrollo del sistema informático que utilice lenguajes de programación de propósito general, y ex profeso para una simulación en particular. No obstante, la alternativa más socorrida es utilizar una plataforma diseñada para tal fin. En el caso de los desastres naturales, a nivel internacional, existen diversas iniciativas para el desarrollo de MBA y herramientas computacionales para este propósito. Entre ellas se encuentra la *RoboCup Rescue League* (s.f.), una versión de la famosa competencia de robots futbolistas orientada a la simulación de situaciones de desastre. Ésta permite representar ciudades a partir de mapas reales (con una variante del formato *OpenStreetMaps* [2015]) y herramientas de simulación de incendios, temblores e inundaciones, entre otros. Los participantes deben desarrollar un SMA

capaz de resolver las tareas de ubicación de víctimas y salvamento bajo restricciones de tiempo y un ambiente cambiante.

Otra plataforma usada frecuentemente en MBA, aunque incipientemente en simulación de desastres, es NetLogo (Wilensky 1999). Ésta integra tanto un ambiente de desarrollo como un lenguaje de programación para la confección de la simulación con un alto nivel de detalle. En las Figuras 4 y 5 pueden apreciarse estos entornos de simulación en mapas reales de la Ciudad de México (Luna-Ramírez 2016).

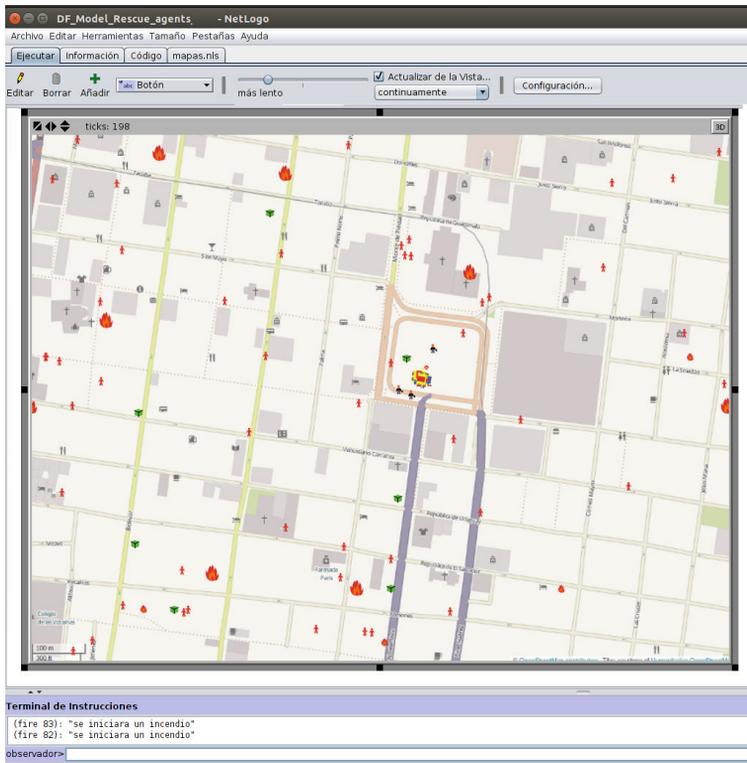


Figura 4. Representación del Zócalo de la Ciudad de México en un entorno NetLogo. Elaboración propia.

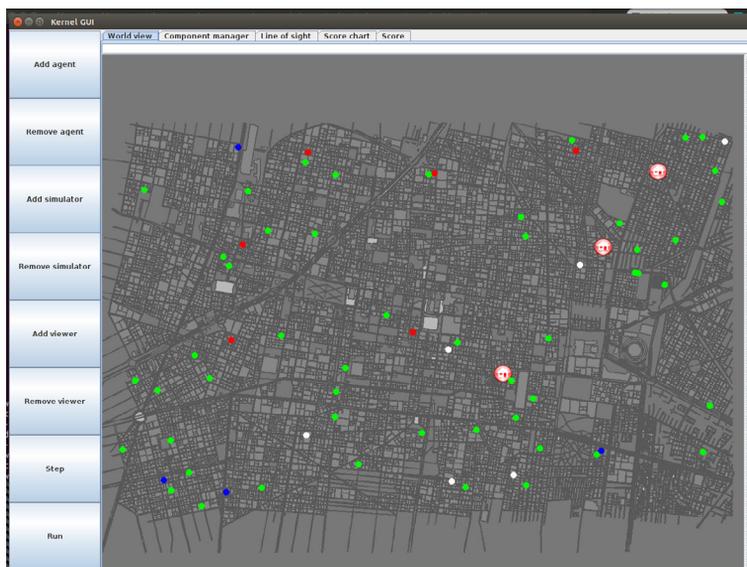


Figura 5. Representación del Zócalo de la Ciudad de México en un entorno RoboCup Rescue. Elaboración propia.

En nuestro país el uso de ABM, y en particular de simulaciones de desastres, es aún muy incipiente. Sería un esfuerzo digno de ser impulsado por sectores como el académico y gubernamental, en aras de desarrollar estas simulaciones en beneficio de la sociedad toda.

## Conclusión

La IA, de suyo interdisciplinaria, integra como parte de sus subdivisiones a los Sistemas Multi-Agente, y estos a su vez a la Modelación Basada en Agentes. Tal modelación abreva de los conocimientos provenientes de muy diversas disciplinas, y adopta sus objetos de estudio con distintas escalas de observación y temporalidad. Tal es el caso de las Humanidades y las Ciencias Sociales que, al trabajar con Sistemas Complejos, pueden beneficiarse de esta técnica, cuyas ventajas incorporan las de la modelación matemática convencional y el formalismo de la informática para

representar teorías y conocimientos sólidos en los distintos dominios de aplicación con el fin de realizar investigación social. Son una herramienta de representación y simulación que incorpora abstracciones y teorías tanto para aplicarlas a su objeto de estudio, como para modificarse.

Dentro de los diversos dominios de aplicación están los desastres naturales, que pueden concebirse desde puntos de vista interdisciplinarios y son, además, instancias de Sistemas Complejos en toda regla. En nuestro país deben impulsarse estos desarrollos por su importancia tanto académica como por su relevancia social.

La Modelación Basada en Agentes representa una herramienta de ayuda metodológica y es en sí una forma de integración interdisciplinaria para la informática y las disciplinas científicas y humanísticas.

## Referencias

- ADAM, Carole y Benoit Gaudou (2016). "BDI agents in social simulations: a survey". *The Knowledge Engineering Review* 31, n.º 3: 207-238.
- BALKE, Tina y Nigel Gilbert (2014). "How do agents make decisions? a survey". *Journal of Artificial Societies and Social Simulation* 17, n.º 4: 13
- BORGES, Jorge Luis (2023). "Del rigor en la ciencia". *Ciudad Seva*. <https://ciudadseva.com/texto/del-rigor-en-la-ciencia/>.
- EPSTEIN, Joshua y Robert Axtell (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Cambridge: The MIT Press.
- GILBERT, Nigel (2019). *Agent-based models*. Sage Publications.
- GAMEZ, M. J. (2015). "Objetivos y metas de desarrollo sostenible, Desarrollo Sostenible". Acceso el 10 de enero de 2024. <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>.
- JAIMEZ-GONZÁLEZ, Carlos R. y Wulfrano A. Luna-Ramírez (2013). "An Agent-Based Architecture for Supply Chain Management".

- Proceedings of the 11th IEEE International Symposium on Autonomous Decentralized Systems (ISADS 2013)*, 109-114. IEEE. DOI: 10.1109/ISADS.2013.6513405.
- JENNINGS, Nicholas R. (1999). "Agent-oriented software engineering". En *European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, 1-7. Berlín: Springer.
- LUNA-RAMÍREZ, Wulfrano Arturo (2016). "Modelación Basada en Agentes. Telescopio y microscopio para las ciencias sociales". *Ciencia y Luz* 1. <https://www.uv.mx/cienciauv/blog/modelacionbasadaenagentestelescopioymicroscopio/>.
- LUNA-RAMIREZ, Wulfrano A. y Maria Fasli (2017). "Integrating NetLogo and Jason: a disaster-rescue simulation". En *2017 9th Computer Science and Electronic Engineering (CEECE)*, 213-218. IEEE. DOI:10.1109/CEECE.2017.8101627.
- OPENSTREETMAP contributors (2015). "OpenStreerMap". <https://planet.openstreetmap.org>.
- PRONACES (2024). "Programas Nacionales Estratégicos. Consejo Nacional de Humanidades, Ciencias y Tecnologías". Gobierno de México. Acceso: el 9 de enero de 2024. <https://conahcyt.mx/pronaces/>.
- RAILSBACK, Steven F. y Volker Grimm, V. (2019). *Agent-based and individual-based modeling: a practical introduction*. Princeton /Oxford: Princeton University Press.
- RAO, Anand S. (1996). "AgentSpeak (L): BDI agents speak out in a logical computable language". En *European workshop on modelling autonomous agents in a multi-agent world*, 42-55. Berlín: Springer.
- ROBOCUP RESCUE LEAGUE (s. f.). "RoboCupRescue". <https://www.robo-cup.org/domains/2>.
- RUSSELL, Stuart y Peter Norvig (2004). *Inteligencia Artificial, un enfoque moderno*. Madrid: Pearson Educación.
- SECRETARÍA DE GOBERNACIÓN (2023). "Lineamientos para la operación de los Programas Nacionales Estratégicos del Consejo Nacional de Humanidades, Ciencias y Tecnologías". Diario Oficial de la

- Federación. Acceso el 9 de enero de 2024. [https://www.dof.gob.mx/nota\\_detalle.php?codigo=5693775&fecha=29/06/2023#gsc.tab=0](https://www.dof.gob.mx/nota_detalle.php?codigo=5693775&fecha=29/06/2023#gsc.tab=0).
- SINGH, Dharendra, Lin Padgham y Brian Logan (2016). "Integrating BDI agents with agent-based simulation platforms". *Autonomous Agents and Multi-Agent Systems* 30, n.º 6: 1050- 1071.
- WALKER, Lyndon y Peter Davis (2013). "Modelling 'Marriage Markets': A Population-Scale Implementation and Parameter Test". *Journal of Artificial Societies and Social Simulation* 16, n.º 1: 6. <http://jasss.soc.surrey.ac.uk/16/1/6.html>. DOI:10.18564/jasss.2106.
- WILENSKY, Uri (1999). "NetLogo". <http://ccl.northwestern.edu/netlogo/>.
- WOOLDRIDGEY, Michael y Paolo Ciancarini (2000). "Agent-oriented software engineering: The state of the art". En *International workshop on agent-oriented software engineering*, 1-28. Berlín: Springer.

## **Sobre el autor**

### **Dr. Wulfrano Arturo Luna-Ramírez**

Cuenta con un doctorado en Ciencias de la Computación en la Escuela de Ciencias de la Computación e Ingeniería Electrónica de la Universidad de Essex en el Reino Unido. Ha fungido como desarrollador de sistemas en los sectores públicos y privados. Actualmente se encuentra adscrito al Departamento de Tecnologías de la Información de la UAM-Cuajimalpa, al cual pertenece desde el año 2009. Además, imparte diversos cursos de la Licenciatura en Tecnologías y Sistemas de Información (de la que actualmente es coordinador), y dirige diversos proyectos terminales y proyectos de servicio social. Es responsable del proyecto de investigación Sistema de Gestión de Riesgo Personal Basado en Agentes Autónomos (DCCD.TI.PI-64).

Entre sus intereses se encuentran los Sistemas Multi-Agente, el Aprendizaje Automático, los sistemas inteligentes aplicados a la educación y el procesamiento de lenguaje natural, la ética y filosofía de la Inteligencia Artificial y la computación.

***Inteligencia Artificial y Problemas Sociales***

Versión electrónica.

Agosto 2024.

En su formación se utilizó la tipografía Scala Offic Pro  
y su variante Scala Sans Offic.

El presente volumen muestra un panorama interdisciplinario de la discusión actual sobre el impacto de las ciencias y tecnologías relacionadas con la Inteligencia Artificial en la vida cotidiana de nuestra sociedad. Reúne puntos de vista de diversos especialistas tanto de esta disciplina como de las Ciencias Sociales y Humanidades. Parte de una exposición introductoria de corte histórico que muestra los vericuetos a los que se ha enfrentado la Inteligencia Artificial y sus implicaciones ético-morales, para posteriormente analizar las dificultades del propio desarrollo del área, por ejemplo, para incluir de manera efectiva al usuario final (individual o colectivo) en los procesos de investigación, desarrollo y evaluación de las soluciones propuestas.

A lo largo del capitulado se tocan diversos problemas y se evidencia la necesidad de un enfoque integrador en torno de objetos de estudio comunes que permiten y exigen ejercer de facto la indagación interdisciplinaria. Se exponen distintas áreas como el Aprendizaje Automático y el Análisis de Datos, las Tecnologías Inmersivas y la Modelación Basada en Agentes, así como aplicaciones a sectores tan demandantes e importantes como la medicina o la industria.

En esta obra también se pone de manifiesto que la resolución de problemas locales, nacionales o regionales no implica privarse de la participación en las discusiones internacionales, sino que, por el contrario, esto forma parte fundamental de ellas, además de contribuir al cumplimiento de la misión de nuestras universidades y, en particular, de la Universidad Autónoma Metropolitana, Unidad Cuajimalpa, siendo la División de Ciencias de la Comunicación y Diseño un espacio natural para su consecución.

Tanto el lector que quiera una introducción ágil e informada a la Inteligencia Artificial, como quien desde un ojo especialista revise estas páginas podrán beneficiarse de la información que aquí se ofrece para construirse un juicio y contrastar sus propios puntos de vista.

