Seminario Inteligencia Artificial y Generación de Conocimiento

Santiago Yarce Prince, Santiago Andrés Bolaños Cruz, Javier Andrés Peña Vargas, Freddy Santiago Galán, Julián Colmenares.

Trabajo de Grado para Optar por el Título de Ingeniero de Sistemas

Directora

Sonia Cristina Gamboa Sarmiento

Doctora en Educación

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingeniería de Sistemas e Informática

Bucaramanga

2026

Dedicatorias

A Dios,

Por darme siempre la destreza necesaria y las fortalezas para poder sobrellevar cada etapa de mi camino académico que ahora me llevan a convertirme en un profesional,

A mis padres,

Quienes me dieron su apoyo incondicional en todo momento y ofrecieron sus fuerzas y sus recursos para poder cumplir un sueño que no es solo mío, sino nuestro y que ahora vemos cumplido gracias a la responsabilidad y la paciencia que desde pequeño me inculcaron.

A mi hermana,

Quien ha sido mi apoyo emocional en cada etapa de mi vida, especialmente en esta que culmina y quien con su amor su cariño y su carisma ha logrado ayudarme a soportar los momentos más difíciles y complicados.

A mis profesores,

Quienes han aportado a mi formación integral y profesional, y han compartido de manera crítica y reflexiva sus conocimientos para estimular mi pensamiento, cuestionar mis ideas y fortalecer mi proceso de aprendizaje.

Freddy Santiago Galán

A Dios,

por ser mi guía constante, mi refugio en los momentos de duda y la fortaleza que me sostuvo cuando el camino parecía difícil. Gracias por cada oportunidad, por cada lección y por permitirme llegar hasta aquí.

A mis padres,

quienes, con amor incondicional, sacrificio incansable y una fe inquebrantable me sacaron adelante aun en medio de las adversidades. Cada logro mío es reflejo de su esfuerzo, de sus desvelos, de sus renuncias silenciosas y de su lucha diaria por darme un mejor futuro. Gracias por creer en mí cuando todo parecía incierto, por enseñarme el valor del trabajo honesto, la perseverancia y la humildad. Este logro es tan suyo como mío.

A mi familia,

por el apoyo constante, las palabras de aliento y la confianza depositada en mí a lo largo de este proceso. Su compañía y respaldo fueron fundamentales para no rendirme y seguir adelante hasta alcanzar esta meta.

Este trabajo es el resultado del amor, la fe y el esfuerzo de todos ustedes.

Javier Peña

A Dios, por darme la fortaleza en momentos de grande dificultad, cuando más difícil y bifurcado se estaba colocando el camino.

A mis padres, por estar ahí con su apoyo y no sucumbir ante los diferentes retos que se presentaron, por ser tan grato ejemplo de responsabilidad y resiliencia.

A mis compañeros, que me apoyaron durante mi formación académica y personal que fue de gran ayuda, a mi familia que estuvo acompañando el proceso también de forma personal.

Por ultimo y no menos importante a la profesora Sonia Gamboa, por su tiempo, compromiso y creer en mi durante el final de mi formación académica.

Julián Colmenares

Agradezco a Dios por darme la fortaleza y la sabiduría durante todo este proceso de aprendizaje, por mantenerme fuerte en los momentos donde quería rendirme y por guiarme cuando más lo necesitaba.

Agradezco a mi familia, pero sobre todo a mis padres, gracias a ellos he podido llegar a donde estoy, su cariño, apoyo incondicional, consejos y acompañamiento en todo este proceso formativo han sido fundamental para mí, son mi motor para seguir adelante con mis sueños.

Agradezco a mi pareja por ser mi lugar seguro en los momentos difíciles de mi vida, por demostrarme a través de su amor un apoyo incondicional, por caminar a mi lado en cada paso que doy, por escucharme con paciencia y por darme la fuerza y motivación necesarias cuando más lo he necesitado.

Agradezco a mi gatita Chloe por ser mi fiel compañera durante tantas noches en vela, su compañía siempre fue reconfortante y me brindó la calma que necesitaba para el desarrollo de este proyecto, eso me ayudó a seguir enfocado.

Por último, agradezco a la docente Sonia Gamboa por su apoyo a lo largo de este proyecto, su disposición, actitud y enseñanzas fueron claves para plasmar nuestras ideas en este libro y sin su orientación no hubiera sido posible.

Santiago Bolaños

Primeramente, agradezco a Dios por permitirme llegar a este momento tan importante de mi vida, en el que se ve reflejado el esfuerzo conjunto de mis padres y el mío. A mis padres, por su apoyo incondicional, sus consejos y su guía constante, y por enseñarme que todo se logra con disciplina, perseverancia y esfuerzo.

A mi abuela, quien ha sido como una segunda madre para mí y ha estado presente desde mi infancia, brindándome siempre amor y apoyo. A mi hermana, a quien considero un ejemplo a seguir tanto en lo personal como en lo académico.

De manera especial, agradezco a Natalia Álvarez, quien ha estado conmigo desde el primer día de la universidad y ha comprendido verdaderamente lo exigente y desafiante que ha sido este proceso académico.

Finalmente, expreso mi agradecimiento a la profesora Sonia Gamboa, por su tiempo, apoyo y motivación, los cuales fueron fundamentales y aportaron significativamente a mi formación durante este seminario.

Santiago Yarce

Agradecimientos

Crear algo desde cero y poder aportar de manera tan significativa a través de un proyecto como lo fue este seminario solo fue posible gracias al apoyo indispensable de muchas personas, a quienes es importante mencionar y agradecer de forma sincera y especial.

Agradecemos a nuestros compañeros de niveles superiores y al grupo de investigación, quienes, con sus conocimientos, brindaron aportes valiosos para la culminación de este seminario.

Agradecemos, de manera reiterada, a nuestra directora de proyecto, PhD. Sonia Cristina Gamboa Sarmiento, cuya labor como guía y cuyo compromiso se mantuvieron como un pilar fundamental para el desarrollo de este proyecto.

Finalmente, agradecemos a la Universidad Industrial de Santander por brindarnos el espacio y las herramientas necesarias para culminar nuestra formación profesional.

Este proyecto va dedicado a ellos y a muchos otros con quienes compartimos este logro, gracias a su compañía, sus aportes y su indiscutible confianza para llegar hasta este punto.

Santiago Andrés Bolaños Cruz

Santiago Yarce Prince

Javier Andrés Peña Vargas

Julián David Colmenares Rodríguez

Freddy Santiago Galán Figueroa

Tabla de Contenido

Introducción 13
1. Generalidades del seminario de investigación 14
1.1.¿Qué es el seminario de investigación? 14
1.2. Objetivo del Seminario de Investigación 14
1.3. Características del Seminario de Investigación 14
1.3.1. Descripción de los roles 15
2. Planteamiento del problema 16
3. Justificación 16
4. Objetivos del seminario 18
4.1.Objetivo general 18
4.2. Objetivos específicos 18
5. Metodología 19
5.1. Contenido: temas y subtemas 20
5.2. Organización de las sesiones 22
6. Fundamentos de IA y generación de conocimiento 25
6.1. Introducción 25
6.2. Fundamentos de la inteligencia artificial y la generación de conocimiento 27
6.3. Creatividad y generación de conocimiento en la inteligencia artificial 30
6.4. Conocimiento, agencia epistémica y cognición extendida en la inteligencia artificial 34
6.5. Conclusión 39
7. IA como generadora de conocimiento 40
7.1. Contexto actual: Machine science y la sobrecarga de información 40
7.2. La IA como generadora de Hipótesis: mecanismos y ejemplos 42
7.2.1. Machine Science: de la asistencia al descubrimiento automatizado 43
7.2.2. El modelo ABC de Sawson: Un antecedente fundacional 46
7.2.3. Aplicaciones a gran escala y expansión del conocimiento 48
7.3. Marco conceptual: Los niveles de abstracción (LoA) de Floridi 49
7.4. Riesgos y Limitaciones 50
7.4.1. Riesgos y consideraciones 51
7.5. Dependencia tecnológica 52

7.6. Discusión critica con autores 54
7.7. Perspectivas de futuro 55
8. Impacto de la IA en la cognición humana. 56
8.1. Fundamentos de la cognición humana. Algunos principios de filosofía de la mente sobre cómo aprenden y procesan información los humanos. 56
8.2. Análisis de cómo la IA está cambiando la manera en que los humanos aprenden y procesan información. 61
8.3. Análisis del riesgo de la IA en estos procesos cognitivos a partir de la dependencia tecnológica y del desplazamiento que la IA está causando en los empleos humanos. 65
9. Conciencia y metacognición en IA. 73
9.1. Introducción 73
9.2. La conciencia como modelo de atención 74
9.3. Conciencia como resultado del aprendizaje 76
9.4.¿Esquema de atención o aprendizaje autorreferencial? 78
9.5. De la conciencia humana a la cognición artificial: aportes de dehaene y flavell 79
9.6. Posibilidades de conciencia y metacognición IA 82
9.7. Estado actual de la IA 86
10. Colaboración humana-IA en la generación de conocimiento 89
10.1. Introducción 89
10.2. La IA como motor del conocimiento y la innovación 90
10.3. La IA como sistema de extracción y control 92
10.4. Entre el progreso y la dominación: el doble filo del conocimiento artificial 95
10.5. Hacia una inteligencia artificial humana 98
10.6. Conclusión 102
11. Conclusiones 103
12. Recomendaciones 104
Referencias Bibliográficas 107

Lista de Tablas

Tabla 1 Cronograma de Sesiones 22

Resumen

Título: Seminario Inteligencia artificial y generación de conocimiento. Posibilidades de construcción de nuevo conocimiento con y por la IA*

Autores: Santiago Yarce Prince, Javier Peña Vargas, Santiago Bolaños Cruz, Freddy Santiago Galán, Julián David Colmenares**

Palabras Clave: Inteligencia artificial, epistemología, inteligencia artificial generativa, teoría computacional de la mente, psicología computacional, conexionismo.

Descripción: En la actualidad, la carrera tecnológica para desarrollar algoritmos y sistemas de inteligencia artificial (IA) cada vez más precisos y complejos busca no solo imitar, sino superar, la eficiencia de los procesos cognitivos y físicos humanos. Si bien el alcance de esta tecnología ha permeado todos los ámbitos de la vida, su acelerado avance coloca a la humanidad ante profundos dilemas y riesgos que podrían desafiar el dominio y la existencia misma de nuestra especie. En este contexto, surge una pregunta fundamental para la ciencia contemporánea: ¿es la IA capaz de actuar como creadora de conocimiento nuevo, válido y consistente, más allá de las intenciones de su diseño original?

Es nuestra responsabilidad como ingenieros asumir un papel crítico respecto a los posibles impactos, tanto positivos como negativos, que esta tecnología ejerce sobre nuestro futuro. Se requiere un análisis profundo de los fenómenos que la IA genera en la construcción del conocimiento y en el desarrollo de las capacidades cognitivas humanas. Esta tesis de pregrado, desarrollada bajo la modalidad de seminario de investigación, se propone como un espacio de reflexión sistemática y discusión razonada para abordar la IA desde una perspectiva epistemológica y social. El estudio gira en torno a cuatro ejes fundamentales: la generación autónoma de conocimiento por la IA, su impacto en la cognición, las posibilidades de autoconciencia e intencionalidad en las máquinas, y el potencial de una colaboración humano-IA en la producción científica.

---

* Trabajo de Grado

** Facultad de Fisicomecánicas. Escuela de Ingeniería de Sistemas. Director: Sonia Cristina Gamboa Sarmiento. Doctora en Educación.

Abstract

Title: Seminar on Artificial Intelligence and Knowledge Generation. Possibilities for the Construction of New Knowledge With and By AI*

Authors: Santiago Yarce Prince, Javier Peña Vargas, Santiago Bolaños Cruz, Freddy Santiago Galán, Julián Colmenares**

Key words: artificial intelligence, epistemology, generative artificial intelligence, computational theory of mind, computational psychology, connectionism

Description: At present, the technological race to develop increasingly precise and complex artificial intelligence (AI) algorithms and systems seeks not only to imitate but to surpass the efficiency of human cognitive and physical processes. Although the scope of this technology has permeated all areas of life, its accelerated advancement places humanity before profound dilemmas and risks that could challenge the very dominance and existence of our species. In this context, a fundamental question arises for contemporary science: is AI capable of acting as a creator of new, valid, and consistent knowledge beyond the intentions of its original design?

As systems engineering students in training, it is our responsibility to assume a critical role regarding the potential impacts—both positive and negative—that this technology exerts on the future of humanity. It is not enough to understand its algorithmic functioning; a deep examination is required of the phenomena that AI generates in the construction of knowledge and in the development of human cognitive abilities. This undergraduate thesis, developed under the research seminar modality, is proposed as a space for systematic reflection and reasoned discussion to address AI from an epistemological and social perspective. The study is structured around four fundamental axes: the autonomous generation of knowledge by AI, its impact on cognition, the possibilities of self-awareness and intentionality in machines, and the potential for human–AI collaboration in scientific production.

---

* Degree Work

** Faculty of Physics and Mechanics. School of Systems Engineering. Director: Sonia Cristina Gamboa Sarmiento. PhD in Education.

Introducción

La inteligencia artificial (IA) ha dejado de ser un área marginal y se ha convertido en un elemento esencial de la vida actual. Facilita el acceso a la información, automatiza tareas cognitivas y transforma prácticas académicas y profesionales. En este contexto, es importante examinar si la IA puede generar conocimiento y no solo procesar información. Esto implica que puede ofrecer contribuciones válidas y rigurosas a la ciencia. También plantea la cuestión de qué significa que esta tarea, que ha sido históricamente humana, sea compartida o delegada a sistemas algorítmicos.

Desde esta perspectiva, el proyecto se sitúa en una tensión actual. Por un lado, existe una carrera tecnológica por desarrollar algoritmos más precisos. Por otro, está la necesidad de entender sus riesgos, repercusiones y efectos sobre los procesos cognitivos y sobre la investigación científica misma, que es fundamental para garantizar la validez y el rigor.

El objetivo general del seminario es analizar, desde diferentes enfoques, las posibilidades de que la inteligencia artificial genere conocimiento y los impactos que esto tiene en la humanidad, tanto a nivel individual como social. Para ello, se estudian fundamentos de epistemología y epistemología de la tecnología para definir qué se entiende por conocimiento científico y práctico. Se construye un marco de referencia sobre las capacidades actuales de la IA. Se analiza su impacto positivo y negativo en los procesos cognitivos humanos desde la filosofía de la tecnología, la filosofía de la mente y la psicología. También se resumen las posturas críticas informadas que surgen de las discusiones del grupo. A través de este enfoque, el seminario busca ofrecer pautas para diferenciar entre información generada algorítmicamente y conocimiento significativo.

1. Generalidades del seminario de investigación

1.1. ¿Qué es el seminario de investigación?

El Reglamento estudiantil de pregrado1 –REP– propone, entre otras, la modalidad de trabajo de grado denominada seminario de investigación 2, la cual se define como "un proceso reflexivo, sistemático y crítico que tiene como propósito fortalecer en el estudiante las habilidades requeridas en el manejo de la información y la comunicación para desarrollar investigación científica" (REP; p. 61). Estas habilidades hacen referencia a la búsqueda y selección de fuentes bibliográficas, a la lectura y escritura crítica de textos y a la discusión ordenada y argumentada, sobre algún tema específico que sea de interés para el campo de conocimiento en el cual están optando por su título profesional.

1.2. Objetivo del Seminario de Investigación

Si bien, la investigación científica implica tanto el desarrollo de otras habilidades adicionales a las ya mencionadas (como identificación y formulación de problemas, diseño metodológico, recolección y análisis de datos, entre otras), como la formación gradual de cierta disciplina, el seminario alemán, como se conoce por su origen, busca transformar los espacios de docencia tradicionales en espacios en los que sea posible, al mismo tiempo, aprender sobre determinado tema, explorado al mismo nivel jerárquico entre profesores y estudiantes, y desarrollar nuevo conocimiento, a la luz tanto de teorías como del estado del arte en el tema, de manera que se logra un cierto nivel de formación científica en los estudiantes, al tiempo que se aborda "el estudio de nuevos objetos de investigación de interés para la Escuela" (REP; Id.).

1.3.Características del Seminario de Investigación

El REP establece que esta modalidad la podrán realizar entre 3 y 5 estudiantes para el mismo seminario, quienes conjuntamente, y bajo la dirección de un profesor, elaboran un plan del Seminario, sobre un tema específico de interés para la Escuela, el cual se organiza por subtemas a estudiar "alrededor del problema seleccionado, la bibliografía a consultar, la programación de sesiones, la asignación de responsabilidades en cada sesión y los relatores respectivos" (REP; p. 68), y "mediante una dinámica que comprende actividades de relatoría, correlatoría, discusión y elaboración de un documento síntesis" (REP; p. 61)

1.3.1. Descripción de los roles

El relator es quien tiene la responsabilidad principal durante una sesión; su función es estudiar a profundidad el tema asignado, preparar una exposición clara y fundamentada, y presentar el contenido al grupo. Este rol exige una preparación rigurosa, ya que el relator debe dominar la bibliografía y los aspectos teóricos y metodológicos del tema para guiar el análisis colectivo. El correlator complementa la labor del relator, aportando observaciones, críticas constructivas y ampliaciones al tema expuesto. Su función es enriquecer la discusión con aportes adicionales y ayudar a clarificar puntos que requieran mayor profundidad, facilitando así un debate más completo y crítico. La discusión es el espacio donde todos los participantes intervienen activamente, aportando sus ideas, cuestionamientos y reflexiones. Este momento es clave para fomentar el pensamiento crítico, la argumentación y la confrontación respetuosa de diferentes puntos de vista, lo que enriquece el aprendizaje colectivo. Finalmente, el protocolante o responsable del protocolo tiene la tarea de documentar detalladamente cada sesión. Esto incluye registrar el tema tratado, los asistentes, las funciones cumplidas, el desarrollo de la discusión, los acuerdos alcanzados y los

interrogantes surgidos. Este documento es fundamental para el seguimiento del seminario y para la elaboración del producto final.

2. Planteamiento del problema

A partir de las concepciones teóricas y del estado del arte de la IA, se busca identificar su potencial para generar autónomamente nuevo conocimiento, riguroso y válido, que pueda ser considerado aporte a la ciencia, así como las implicaciones que la colaboración en o el reemplazo de esta ocupación, hoy exclusivamente humana, pudiera tener tanto en procesos cognitivos humanos como nuevos hallazgos científicos.

3. Justificación

En cuanto la inteligencia artificial ha logrado permear todos los ámbitos de la vida, se contraponen, como intereses, por una parte, la carrera tecnológica por lograr algoritmos y automatismos cada vez más precisos y más complejos que lleguen a ejecutar de mejor forma y con mucha más eficiencia procesos de fenómenos naturales, incluyendo procesos cognitivos y físicos de individuos humanos y animales, y por otra, la consideración permanente de los riesgos, desde leves hasta los que ponen en jaque el dominio y la existencia de los humanos, y las respectivas regulaciones que buscan mitigarlos. En este contexto, resulta de especial interés, tanto en un enfoque como en el otro, dar una mirada al estado en que se encuentra la capacidad tecnológica de la IA para producir conocimiento de manera autónoma, es decir, no solo como producto de relaciones no previstas a nivel del diseño algorítmico en grandes volúmenes de datos o como una síntesis casi instantánea de elementos relacionados en un prompt que pueden encontrarse en el ciberespacio, sino, incluso, fuera de toda consideración o intención en cualquier diseño, es decir, la IA como creadora de nuevo

conocimiento válido y consistente, fuera de toda consideración humana, y con intencionalidad propia de parte de la máquina. Cabe, entonces, puntualizar de entrada qué se entiende por conocimiento, qué se entiende por producción o generación de conocimiento, cómo encaja en un marco de referencia teórico lo que se conoce como IA generativa, qué impactos positivos y negativos puede tener esta capacidad de la IA en el desarrollo de los procesos cognitivos de los humanos de esta época, cómo se afecta el ejercicio de la investigación científica, vista como oficio y vista, también, como garante de validez, de consistencia, de rigurosidad de lo que conocemos sobre el mundo. Este seminario se propone dar una mirada analítica y comparativa a una teoría computacional de la mente, con autores como Margaret Boden, Jerry Fodor, Marvin Minsky, frente a las concepciones que desde la psicología han caracterizado cómo cambian los procesos cognitivos humanos en presencia de determinadas tecnologías, con autores como Sherry Turkle y George Siemens. A la luz de estos desarrollos, se propone abordar en el Seminario con ejes temáticos que giran en torno a las posibilidades de generación de conocimiento con y por la IA.

4. Objetivos del seminario

4.1. Objetivo general

Analizar, desde diferentes enfoques, las posibilidades de generación de conocimiento por y con inteligencia artificial, y los impactos que genera en la humanidad actual a nivel individual y social.

4.2. Objetivos específicos

Estudiar fundamentos de epistemología y epistemología de la tecnología que ofrezcan las bases para la comprensión de lo que se constituye en conocimiento científico y conocimiento práctico.

Formular un marco de referencia de la inteligencia artificial que establezca las capacidades actuales de la IA para la generación de conocimiento.

Formular un marco de referencia, a partir de disciplinas como filosofía de la tecnología, filosofía de la mente y psicología, que caracterice el impacto positivo y negativo en el desarrollo de procesos cognitivos humanos.

Sintetizar un conjunto de posturas críticas informadas que resulten de las discusiones grupales.

5. Metodología

La modalidad del seminario se acoge a la concepción de seminario alemán 3, que consiste en un enfoque de enseñanza orientado a la formación en investigación, puesto que ubica al estudiante en un rol activo en el cual debe estudiar marcos teóricos y ponerlos en relación con el estado del arte y el contexto actual del asunto en cuestión. Esta puesta en contexto se denomina relatoría y consiste en un documento cuya lectura da lugar a la discusión ordenada, informada y argumentada por parte de todos los miembros del seminario. A partir de la discusión, uno de los miembros del seminario elabora otro documento, que se denomina protocolo, el cual registra una síntesis de la discusión, los acuerdos y los diferentes puntos de vista que se manifiestan en la discusión. Esta modalidad exige, no solamente el interés y motivación de todos los estudiantes, sino la preparación de todos los asistentes, de los temas de cada sesión, de manera que cada nueva discusión retoma lo aprendido en la anterior, y al finalizar el seminario es posible tener un compendio equivalente a la revisión teórica y estado del arte de la cuestión. Para este seminario, una vez acordado el conjunto de temáticas a abordar, durante un semestre académico, cada semana se desarrollará una sesión en la que miembros asignados previamente presentan una relatoría que expone los conceptos correspondientes, en un relato crítico que considera tanto definiciones y teorías como el estado del arte del asunto. Cada temática estará a cargo de un estudiante. Adicionalmente, los estudiantes tendrán a cargo producir un informe final, de entre 3 y 5 secciones tipo artículo, sobre las temáticas del Seminario, que deberán incluir: revisión de conceptos, estado del arte, reflexiones argumentativas al respecto y la bibliografía completa referenciada en el texto. Estos textos se socializarán en las sesiones del seminario, en las cuales se sugerirán

ajustes finales. Duración: 1 semestre académico Modalidad: Encuentros sincrónicos grupales: 3 horas semanales. Trabajo independiente: 7 horas semanales.

5.1. Contenido: temas y subtemas

1. Fundamentos de epistemología
    - Qué se entiende por conocimiento, ciencia e investigación.
    - Cuál es el objetivo de la ciencia para la humanidad
    - Qué características tienen tanto la investigación como el conocimiento científico.
2. Fundamentos de IA y generación de conocimiento
    - Qué se entiende por IA generativa.
    - Cómo la IA genera conocimiento.
    - Paradigmas de IA: simbólica, conexionista y aprendizaje profundo.
3. IA como generadora de Conocimiento
    - Estado actual de la IA en procesos de generación de conocimiento
    - Sectores en los que la IA ha producido nuevo conocimiento y descripción del impacto.
    - Análisis sobre límites y oportunidades de la IA en la creación de conocimiento científico y tecnológico.
4. Impacto de la IA en la cognición humana
    - Fundamentos de la cognición humana. Algunos principios de filosofía de la mente sobre cómo aprenden y procesan información los humanos.
    - Análisis de cómo la IA está cambiando la manera en que los humanos aprenden y procesan información.

- Análisis del riesgo de la IA en estos procesos cognitivos a partir de la dependencia tecnológica y del desplazamiento que la IA está causando en los empleos humanos.

5. Conciencia y Metacognición en IA

    - Posibilidades de que la IA desarrolle metacognición o conciencia de su propio pensamiento.
    - Relación de estas posibilidades con el estado actual de la IA, en términos de redes neuronales y aprendizaje profundo.

6. Colaboración humano-IA en la generación de conocimiento

    - Indagar cómo los humanos y la IA pueden colaborar en la creación de nuevo conocimiento.
    - Analizar sesgos en la IA y su impacto en la toma de decisiones.

5.2. Organización de las sesiones

Tabla 1

Cronograma de Sesiones

Sesión	Tema	Relator	Lecturas orientadoras
1 y 2	Fundamentos de epistemología	Sonia Gamboa	Gamboa S. Epistemología de la tecnología (en edición)
3 y 4	Fundamentos de IA y generación de conocimiento	Freddy Santiago Galán Figueroa	Mitchell, M. (2019). Artificial Intelligence: A Guide for Thinking Humans (Cap. 1-2). Boden, M. (2004). The Creative Mind (Introducción).
5 y 6	IA como generadora de conocimiento	Javier Andrés Peña Vargas	Evans, J. A., & Rzhetsky, A. (2010). Machine Science. Science, 329(5990), 399-400. Floridi, L.

			(2011). The Philosophy of Information (Cap. 3).
7 y 8	Impacto de la IA en la cognición humana	Julián David Colmenares Rodríguez	Turkle, S. (2011). Alone Together (Cap. 4). Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age.
9 y 10	Conciencia y metacognición en IA	Santiago Andrés Bolaños Cruz	Graziano, M. S. (2019). Rethinking Consciousness (Cap. 2-3). Cleeremans, A. (2011). The Radical Plasticity Thesis: How the Brain Learns to be Conscious.
11 y 12	Colaboración humano-IA en la	Santiago Yarce Prince	Brynjolfsson, E., & McAfee, A. (2014). The Second

	generación de conocimiento	Machine Age (Cap. 5). Crawford, K. (2021). Atlas of AI (Cap. 2)
13 a 16	Revisión de textos	

6. Fundamentos de IA y generación de conocimiento

6.1. Introducción

Desde su inicio, en 1956 con el histórico seminario de Dartmouth, cuando John McCarthy, Marvin Minsky y otros científicos se propusieron estudiar la inteligencia a través de una detallada simulación en máquinas, la inteligencia artificial (IA) ha ido más allá de ser un deseo teórico, transformando nuestra comprensión de la mente, el conocimiento y la creatividad. Esta tecnología ha desafiado los contornos entre lo natural (humano) y lo artificial (computacional). Además, su progreso ha sido un viaje emocionante y a veces borroso, marcado por períodos de exaltación y de desconfianza, debido a que algunos esperaban que la IA pudiera imitar el pensamiento humano, mientras que otros cuestionaban si era posible lograrlo.

En primer lugar, Melanie Mitchell (2019), reflexiona sobre la naturaleza intrigante y a la vez confusa del concepto de inteligencia. Cuando se refiere a inteligencia artificial, en realidad estamos explorando una metáfora, una manera de hablar acerca de fenómenos complejos como el razonamiento, el aprendizaje y la creatividad a través de la tecnología. La autora nos dice a modo de advertencia que, a pesar de los esfuerzos incesantes, no hay una definición universalmente aceptada ni una medida normalizada para la inteligencia. Por lo tanto, sugiere que los investigadores en IA no buscan una definición completa de inteligencia, sino que se enfocan en intuiciones prácticas (Mitchell, 2019). Su enfoque está en la creación de sistemas que, a simple vista, parecen inteligentes, capaces de aprender y resolver problemas o producir resultados útiles en otras palabras. Este enfoque experimental ha dado lugar a avances significativos, como el aprendizaje profundo y los modelos de lenguaje, pero

también ha generado cierta incertidumbre sobre si estas máquinas realmente comprenden lo que hacen.

Por otra parte, la IA es tanto un campo tecnológico como un desafío filosófico, puesto que cumple con la tarea de incitar al hombre a replantear y a meditar sobre su comprensión acerca de lo que significa conocer y cómo se desarrolla el conocimiento. Cuando una máquina puede generar textos, componer música o resolver problemas matemáticos con un alto nivel de complejidad, genera incertidumbre sobre si realmente está generando conocimiento o si sólo está imitando ciertos patrones de datos. Esta cuestión está muy relacionada con las investigaciones de Margaret Boden (1994), quien afirma que el misterio de la creatividad es algo que aún se está tratando de definir. Siguiendo con las ideas de la autora, entender, o tratar de entender la creatividad, ya sea humana o artificial, requiere explorar los procesos mediante los cuales surgen ideas innovadoras y valiosas, y cómo un sistema, ya sea humano o no, puede evaluar y transformar su propio pensamiento.

Por consiguiente, este trabajo surge de un interés común en la intersección entre inteligencia y creatividad. Su objetivo es explorar los principios conceptuales que unen la IA con la generación de conocimiento, inspirándose en las ideas de Mitchell y Boden. Aunque Mitchell se concentra en comprender el proceso de pensamiento y Boden en el intento de entender la creatividad, ambas perspectivas se cruzan para investigar si la IA puede ser considerada un agente epistémico; es decir, capaz de aportar de manera auténtica al desarrollo de nuevos conocimientos. Teniendo en cuenta lo anterior, por medio de esta introducción se establece el marco analítico general del informe: la inteligencia artificial no se limita aquí a ser considerada solo como una herramienta técnica. Al contrario, es vista como un fenómeno que desafía nuestras concepciones de inteligencia, creatividad y conocimiento. Entender sus

raíces nos permite comprender cómo, en nuestro esfuerzo por crear máquinas inteligentes, acabamos redefiniendo lo que significa ser inteligentes.

6.2. Fundamentos de la inteligencia artificial y la generación de conocimiento

El camino del desarrollo de la IA ha sido un viaje innovador, movido por la ambición de replicar la inteligencia humana. Desde el seminario de Dartmouth en 1956, donde McCarthy, Minsky, Newell y Simon plantearon la posibilidad de describir el aprendizaje y otros aspectos de la inteligencia con tal precisión que una máquina pudiera imitarlos, la IA ha navegado entre el entusiasmo científico y la duda filosófica. Mitchell (2019) recuerda que el concepto de inteligencia, incluso entre los humanos, es difícil de definir, pues no existe una única forma de medirla ni un acuerdo sobre lo que implica. Por ende, hablar de "inteligencia artificial" abarca explorar una noción flexible, que depende tanto de la técnica como de la interpretación.

Al mismo tiempo, Mitchell (2019) destaca que la IA surgió gracias a la unión de varias disciplinas: la lógica matemática, la teoría de autómatas, la psicología y la neurociencia. Cada campo aportó su propia opinión o visión sobre la inteligencia, lo que resultó en una "anarquía de enfoques" que continúa incluso aún en la actualidad. Algunos investigadores se han enfocado en desarrollar sistemas que razonen de manera lógica, mientras que otros han buscado imitar los procesos del cerebro. Esta diversidad en los métodos ha sido tanto una fuente de tensiones como de avances, dado que ha generado múltiples perspectivas en un fenómeno que aún no se comprende completamente.

Durante los primeros años en los que se implementó la inteligencia artificial, el enfoque simbólico fue el principal paradigma. Inspirado en la lógica formal, este enfoque asumía que el pensamiento era una actividad que consistía en la manipulación de símbolos

siguiendo un conjunto de reglas bien definidas. Por ejemplo, el *General Problem Solver* (GPS) era un sistema que abordaba problemas aplicando operaciones que transformaban un estado inicial a uno deseado. Es decir, el conocimiento de modelo estaba explícitamente codificado en reglas, aunque el ordenador no entendía su significado; simplemente ejecutaba las instrucciones proporcionadas. Este modelo no se centra en imitar la complejidad biológica del cerebro, apunta hacia el diseño de sistemas simbólicos lo suficientemente avanzados y sofisticados. No obstante, aunque eran capaces de razonar en contextos estructurados, no eran tan eficaces en tareas simples. Por lo que, Mitchell (2019) ejemplifica que las actividades que a simple vista son fáciles para los niños (como identificar rostros o tonos de voz) suelen ser sumamente retadoras para una máquina que solo trabaja con símbolos y reglas.

En respuesta a estas limitaciones, surgió un enfoque subsimbólico, inspirado en la neurociencia. El perceptrón de Frank Rosenblatt, por ejemplo, imitaba la forma en que las neuronas procesan la información. Este tipo de red recibía múltiples entradas, aplicaba pesos a esas entradas y luego decidía si activarse o no según un umbral específico. A diferencia del enfoque simbólico, el conocimiento en el subsimbólico se distribuye en los pesos y umbrales de las conexiones. Dentro del enfoque subsimbólico las redes neuronales aprenden ajustando estos valores a partir de ejemplos, un proceso que se asemeja al condicionamiento animal.

Conjuntamente, el modelo subsimbólico planteó un concepto revolucionario, el cual comprende que la inteligencia emerge de la habilidad de adaptarse, mas no de seguir instrucciones explícitas. Sin embargo, los primeros perceptrones eran demasiado rudimentarios y fue gracias al desarrollo de las redes neuronales multicapa y al algoritmo de retropropagación del error en los años ochenta que se alcanzó un verdadero progreso. Estas redes podían aprender representaciones abstractas y detectar patrones complejos sin depender

de reglas predefinidas. Esto dio origen al paradigma del aprendizaje automático, que fusiona estadística, probabilidad y computación para permitir que las máquinas aprendan directamente de los datos.

La historia de la IA ha comprendido diferentes periodos; algunos se han destacado por sus significativos avances y otros por desilusiones en su desarrollo; estos últimos se han denominado como “los inviernos de la inteligencia artificial”. Hebert Simon (1965), como se citó en Mitchell (2019), propuso como expectativa que “las máquinas serán capaces, dentro de veinte años, de hacer cualquier tarea que pueda hacer un hombre” (p. 21). No obstante, estas predicciones no se concretaron. En consecuencia, se considera que cuando las expectativas sobre los avances tecnológicos no se materializaban, el apoyo financiero y académico disminuye, truncando así los posibles avances. Pero, en medio de estas circunstancias, también surge una oportunidad para reevaluar los límites de lo que la IA puede hacer y así comprender mejor la complejidad de esta tecnología.

La experiencia con modelos puramente simbólicos ha demostrado que la inteligencia humana no se puede reducir únicamente a una lógica formal. Las limitaciones observadas en las redes neuronales han subrayado, como indica Mitchell (2019), que el aprendizaje no es sinónimo de comprensión. Por lo que, en ambos escenarios, la inteligencia artificial ha ofrecido una ventana a la mente humana: el pensamiento implica un contexto, una interpretación y una flexibilidad, aspectos que son intratables de traducir en términos computacionales.

En los sistemas simbólicos, el conocimiento se presenta de manera explícita, lo que permite su lectura, análisis y modificación con facilidad. Por otro lado, en los sistemas subsimbólicos, el conocimiento se manifiesta de manera implícita a través de los pesos

numéricos del modelo. Mitchell (2019) plantea su preocupación al entrenar redes neuronales con un número considerable de parámetros. Aunque la máquina logra resultados sobresalientes, la dificultad radica en que no se puede explicar con certeza qué ha aprendido ni cómo lo ha logrado. Dicho dilema de la interpretabilidad plantea un desafío epistemológico: ¿es posible que el conocimiento sea algo que no puede ser explicado con claridad?

Al mismo tiempo, la inteligencia artificial está redefiniendo la comprensión del conocimiento. Cuando lo consideramos como la habilidad de identificar patrones, predecir resultados o tomar decisiones sabias, las máquinas pueden ser vistas como agentes cognitivos. Sin embargo, cuando exigimos comprensión y justificación, su conocimiento sigue siendo limitado. A pesar de esto, resulta fascinante es que, independientemente de cómo se vea, la IA establece un reto al reconsiderar la noción de conocimiento.

La inteligencia artificial ha transformado la forma de pensar y actuar. Sus modelos de trabajo son capaces de hacer inferencias y descubrimientos que, de otra manera, serían humanamente imposibles sin su asistencia. Aunque la IA no posee conciencia, su importancia radica en la capacidad de revelar patrones ocultos en grandes cantidades de datos, lo que representa una nueva forma de conocimiento. Por ende, la IA no reemplaza al pensamiento humano, sino que lo amplía y lo desafía, abriendo nuevas posibilidades para la exploración intelectual.

6.3. Creatividad y generación de conocimiento en la inteligencia artificial

Paralelamente a los avances tecnológicos, emerge una dimensión más profunda: la creatividad, que se refiere a la habilidad de producir ideas originales y de valor. Boden

(1994), explora este tema, no como algo exclusivo del arte, sino como una parte fundamental del pensamiento humano y, por extensión, de los sistemas artificiales que buscan imitarlo. La autora establece que la creatividad se encuentra en todas partes, desde los descubrimientos científicos hasta las acciones cotidianas, pero cuanto más se hace evidente, más desafiante resulta explicarla.

La distinción entre enigmas y misterios encuentran que los enigmas son preguntas que no han sido resueltas, mientras que los segundos son cuestiones que apenas podemos formular. Así, la creatividad se asocia con esta última categoría. Dado que a menudo, no solo ignoramos cómo sucede, sino que tampoco sabemos cómo conceptualizarla. Las definiciones más simples, como "producir algo nuevo", no son suficientes. ¿Qué tan nueva debe ser una idea para considerarse creativa? ¿Es necesario que también tenga valor o sentido? Estas preguntas conducen a reconocer que hablar de creatividad implica navegar por un territorio ambiguo, donde los límites entre lo original, lo absurdo y lo valioso se difuminan.

Boden (1994) señala que la mayoría de los debates sobre la creatividad no se centran en problemas empíricos, sino en conceptos abstractos, pues durante siglos, la creatividad se ha entendido por medio de la teoría de la inspiración, que atribuía la creación a fuerzas externas o divinas, vistiendo al artista como un canal para una inspiración sobrenatural. Desde esta perspectiva, una explicación científica parecía extraña. Más adelante, surgió el enfoque romántico, que reorientó la fuente de la creatividad hacia el individuo, pero mantuvo su carácter excepcional. Además, el genio creativo se caracterizaba por una intuición especial, inaccesible para la mayoría. Sin embargo, esta idea no explica cómo funciona esa intuición ni cómo se manifiesta. Por ello, Boden (1994) argumenta que, aunque estos métodos pueden parecer fascinantes, no son muy efectivos para comprender el fenómeno en

cuestión. En su lugar, sugiere que se debe abordar utilizando las herramientas de la ciencia cognitiva y la IA. Pues, es posible que también se pueda entender la creatividad como un proceso cognitivo estructurado, uno que podría ser replicado o simulado.

Para que una máquina parezca creativa, necesita más que simplemente seguir instrucciones, debe poder evaluar su propio pensamiento, algo que se conoce como metacognición. Esta habilidad de reflexionar sobre su propio proceso es crucial. Si un sistema puede generar resultados inesperados coherentes y adaptarse a ellos, entonces podemos hablar de una forma de creatividad computacional. Boden (1994) propone tres tipos principales de creatividad:

Combinatoria: emerge cuando las ideas existentes se reorganizan en nuevas combinaciones.

Exploratoria: genera variaciones dentro de un marco conceptual establecido.

Transformacional: implica cambiar o ampliar el marco conceptual existente, lo que lleva a la generación de ideas que antes no se habían considerado.

Los sistemas de inteligencia artificial de hoy, específicamente aquellos que están basados en aprendizaje profundo y procesamiento de lenguaje, demuestran cierto grado de creatividad combinatoria y exploratoria. Estos sistemas combinan elementos aprendidos para generar resultados novedosos dentro de un marco preestablecido. Sin embargo, la creatividad transformacional, que implica una redefinición del propio sistema de pensamiento, sigue siendo una característica distintiva de la mente humana. Esta creatividad está influenciada por factores como la intuición, las emociones y el entorno cultural.

La creatividad no puede existir sin una base de conocimiento previa. La creatividad es, de hecho, un motor para la generación de nuevos conocimientos. La mente creativa se manifiesta en un espacio conceptual formado por reglas, símbolos y representaciones que puede explorar o transformar (Boden, 1994). Por lo tanto, la creatividad puede ser vista como un proceso de descubrimiento cognitivo: un mecanismo por el cual el pensamiento reorganiza sus propias estructuras para generar nuevas posibilidades de comprensión.

En el campo de la IA, esto implica que la capacidad creativa de un sistema depende en gran medida de la flexibilidad de su marco conceptual. Es decir, los algoritmos de aprendizaje automático son capaces de identificar patrones, pero raramente cuestionan su propia comprensión. A pesar de ello, al generar soluciones inesperadas, como en el diseño, el arte o el lenguaje natural, pueden aportar conocimiento práctico. Y a pesar de que no poseen conciencia sobre lo que crean, logran ampliar el horizonte de lo posible para la mente humana.

La creatividad siempre tiene un toque de misterio, no porque sea mágica, sino porque desafía la forma de entender el mundo (Boden, 1994). Por ejemplo, si una máquina produce resultados inesperados, esta genera incertidumbre sobre si realmente se puede considerar como creativa. La respuesta a esto depende de cómo se defina la creatividad. Si se le considera en términos de novedad y valor del producto, las máquinas podrían cumplir con ese estándar. Sin embargo, si se requiere intencionalidad y una conciencia del significado, la creatividad artificial queda fuera del alcance. Dicho debate trae a colación la inquietud de Ada Lovelace, quien cuestionaba si las máquinas podrían generar algo verdaderamente innovador o simplemente seguir órdenes (Boden. 1994). Ante esta idea, la autora argumenta

que un sistema creativo, ya sea humano o artificial, debe ser capaz de sorprender, es decir, producir resultados que trascienden las expectativas de su creador.

El estudio de la creatividad, a través de la inteligencia artificial, no tiene como objetivo reemplazar la inspiración humana, sino comprenderla. Los programas de IA creativa, que se aplican en campos artísticos, como la música y la literatura, demuestran que la novedad puede surgir de procesos algorítmicos influenciados, en gran parte, por los datos humanos. Por medio de esta interacción no solo se producen nuevos productos, sino que también se estimula la creatividad humana. Esto se da porque al enfrentarse a resultados inesperados, las personas reconfiguran sus propios esquemas mentales y descubren nuevas perspectivas inéditas. La interacción entre creatividad y conocimiento se transforma en un diálogo colaborativo y la inteligencia artificial se sumerge en la creación de combinaciones y variaciones, mientras que el ser humano se encarga de la interpretación y la asignación de sentido. De este modo, el conocimiento ya no se limita al ser humano, pasa a convertirse en una construcción compartida entre agentes naturales y artificiales.

6.4. Conocimiento, agencia epistémica y cognición extendida en la inteligencia artificial

La IA no es solo un avance tecnológico, es un cambio radical en la manera en que la humanidad crea, organiza y difunde el conocimiento. Desde sus inicios, como lo demostraron las autoras Mitchell (2019) y Boden (1994), la inteligencia artificial buscaba imitar aspectos de la inteligencia humana; sin embargo, su evolución actual va más allá, redefiniendo la forma en que se genera y se transmite el conocimiento. En este sentido, la IA no es simplemente una herramienta cognitiva, se considera como una fuerza epistemológica que juega un papel activo en la evolución del conocimiento humano. Esta visión se sostiene en

los postulados de Colther y Doussoulin, (2024), donde exponen que la inteligencia artificial IA se ha convertido en un agente dinámico en la coevolución del conocimiento, capaz de acelerar los procesos de descubrimiento científico, reestructurar formas de pensamiento y cambiar la comprensión de la relación entre información y comprensión.

De acuerdo con las ideas de Colther y Doussoulin (2024), la inteligencia artificial va más allá de simplemente almacenar o procesar información; en realidad, introduce nuevas formas de conocimiento. Por ejemplo, los modelos de aprendizaje profundo pueden descubrir correlaciones y patrones en datos que superan nuestra capacidad perceptiva o analítica. Esto ha llevado a descubrimientos en campos tan diversos como la biología, la física y la lingüística, que no provienen de la intuición humana, sino del análisis autónomo de sistemas algorítmicos. En este sentido, la IA actúa como una fuerza epistémica emergente, una instancia que contribuye significativamente a la expansión del conocimiento científico.

Sin embargo, existe una cuestión filosófica intrigante: ¿se puede considerar la inteligencia artificial como un sujeto del conocimiento, o es simplemente un instrumento técnico? La respuesta a esta pregunta depende de cómo se defina la agencia epistémica. Desde la perspectiva de la filosofía de la ciencia, la agencia epistémica no solo implica producir resultados cognitivos, sino también participar activamente en prácticas de justificación, interpretación y evaluación. En este campo, los sistemas de IA actuales no poseen conciencia ni comprensión de su propio proceso, pero tienen un impacto significativo en la formación de las creencias y teorías humanas. Por ende, aunque no sean agentes epistémicos en su totalidad, sí funcionan como agentes epistémicos distribuidos o delegados, actuando como intermediarios entre los datos y la interpretación humana.

Dentro de los postulados de Collins (2024), argumenta que la IA no puede ser entendida únicamente desde el ámbito técnico, ya que toda producción de conocimiento está intrínsecamente ligada a la sociedad. Los algoritmos, los datos y los modelos no son neutrales; reflejan las estructuras de poder, los intereses y las lógicas de las comunidades que los crean. Por lo tanto, para comprender la IA como un agente epistémico, es crucial reconocer que su "conocimiento" emerge de una red sociotécnica compuesta por actores humanos y no humanos. Desde esta perspectiva, la IA se convierte en un actor clave en la sociología del conocimiento de nuestra época, transformando la manera en que las sociedades construyen la verdad.

No obstante, la IA no reemplaza al sujeto epistémico humano, redistribuye la capacidad cognitiva entre los seres humanos, las instituciones y los sistemas automatizados. Por ejemplo, los algoritmos de recomendación influyen en la información que se consume, las teorías se difunden, y las investigaciones que se priorizan. Esta mediación algorítmica altera las jerarquías del saber y plantea nuevos desafíos éticos: ¿quién controla el conocimiento cuando los sistemas autónomos participan en su generación?

Para adentrarse en los límites de esta agencia distribuida, es fundamental reconsiderar la distinción entre conocimiento tácito y explícito. Polanyi (1966) menciona que “podemos saber más de lo que podemos decir” (traducción propia, p. 4). Es decir, el conocimiento tácito se manifiesta en la acción, la intuición y la experiencia, mientras que el conocimiento explícito se puede traducir en reglas o declaraciones claras.

A diferencia de los seres humanos, las máquinas carecen de la corporeidad, la historia y el entorno social, elementos cruciales para el desarrollo y la comprensión de este tipo de conocimiento. Aunque los modelos de aprendizaje profundo pueden identificar rostros o

emociones, no poseen la capacidad de comprender su significado en un entorno más amplio, lo que representa una limitación significativa en su capacidad para actuar de manera auténtica y contextualizada.

A pesar de esta limitación, la inteligencia artificial tiene la capacidad de imitar aspectos del conocimiento implícito al aprender de grandes cantidades de datos. Es decir, aunque no tenga intuición o juicio contextual, puede generar patrones que reflejan ese conocimiento socialmente compartido. Por lo tanto, la IA se convierte en una herramienta tecnológica para externalizar el conocimiento implícito colectivo, transformando la experiencia social en datos manipulables.

La Teoría de la Cognición Extendida, propuesta por Clark y Chalmers (2021), plantea que nuestra mente trasciende el cerebro y el cuerpo, extendiéndose hacia las herramientas y artefactos que nos acompañan. Elementos como un cuaderno, un ordenador o un sistema de inteligencia artificial pueden actuar como extensores de nuestra memoria y nuestro proceso de pensamiento. Desde esta perspectiva, los sistemas de inteligencia artificial no solo incrementan las capacidades humanas, sino que se incorporan al hombre como un órgano intelectual.

En la práctica, esto implica que los procesos de creación de conocimiento se están volviendo cada vez más integrados: humanos y máquinas trabajan juntos como una sola red cognitiva. Los estudios de sociología del conocimiento subrayan esta realidad: la IA es vista como parte del sistema de cognición social, donde las líneas entre lo humano y lo técnico se difuminan. Sin embargo, esta extensión también conlleva riesgos: si los humanos confían demasiado en la IA para tomar decisiones o para evaluar críticamente, pueden perder su

capacidad de comprender y controlar el conocimiento. En lugar de ampliar la capacidad cognitiva, la IA podría dominarla, limitando la autonomía intelectual de las personas.

Al incorporar diferentes perspectivas como la epistemología, la sociología del conocimiento, la cognición extendida y la teoría del conocimiento tácito, se puede afirmar que la inteligencia artificial ha dado lugar a una nueva forma de epistemología híbrida. En esta nueva era, el conocimiento se genera en un diálogo entre humanos y sistemas artificiales:

Desde la filosofía, la inteligencia artificial amplía las posibilidades de conocimiento al permitir nuevos modos de descubrimiento.

Desde la sociología, la inteligencia artificial redistribuye la agencia cognitiva y modifica la estructura social del conocimiento.

Desde la teoría del conocimiento, la inteligencia artificial transforma la relación entre lo tácito y lo explícito, externalizando parte del saber humano sin poderlo sustituir por completo.

Desde la cognición extendida, la inteligencia artificial actúa como una extensión funcional de la mente humana, aunque sin sustituir su capacidad interpretativa y ética.

La IA no sustituye la inteligencia humana, pero sí transforma la estructura del conocimiento. En lugar de un conocimiento epistemológico aislado, emerge ahora de redes sociotécnicas interconectadas, donde la creatividad humana, los algoritmos y los datos evolucionan. Por lo tanto, la pregunta ya no es si la IA puede pensar sino, ¿cómo su integración en nuestras prácticas cognitivas redefine nuestra comprensión del pensamiento y del conocimiento?

6.5. Conclusión

El viaje por los fundamentos de la inteligencia artificial y su papel en la creación de conocimiento nos permite entender que la IA no es simplemente una herramienta tecnológica, sino una forma emergente de pensar y producir saber. A través de las lecturas de Mitchell y Boden, se hizo evidente que la IA encarna una tensión entre la simulación y la comprensión, entre la ejecución de algoritmos y la creatividad genuina. Estas reflexiones sitúan la IA como un fenómeno que nos obliga a replantear las condiciones del conocimiento y no verla como un sustituto del intelecto humano.

La inteligencia artificial está revolucionando la comprensión sobre la epistemología clásica. No basta con simplemente preguntar qué conocemos o cómo lo conocemos, se debe considerar a quién y a través de qué medios generamos conocimiento. Los sistemas de IA actuales están mostrando una capacidad cada vez mayor para identificar patrones, formular inferencias y producir resultados inesperados, lo que los convierte en participantes en el proceso de conocimiento dentro de entornos humanos. Sin embargo, su "agencia" no es autónoma, pues depende de los contextos culturales, de los datos que alimentan su aprendizaje y de las decisiones humanas que guían su interpretación.

De esta manera, los fundamentos actualizados abren la discusión sobre un tema fundamental en la próxima etapa del seminario: la IA como generadora de conocimiento. Si la creatividad puede ser parcialmente modelada y si los sistemas inteligentes pueden contribuir a la creación de nuevos conocimientos, entonces es imperativo explorar cómo estas formas de conocimiento surgen, se validan y se integran en las prácticas científicas y sociales. El trabajo de mi compañero, que se centra en esta cuestión, nos permitirá avanzar en este camino, evaluando hasta qué punto la inteligencia artificial realmente impulsa la innovación

y el descubrimiento, y en qué condiciones podemos hablar de un conocimiento "producido" por máquinas.

Además, la contemplación de la creatividad lleva directamente a las próximas sesiones, que se centrarán en el impacto de la inteligencia artificial en nuestra cognición y en la conciencia artificial. Si la IA tiene la capacidad de imitar aspectos del pensamiento humano, ¿cómo cambia esto nuestra forma de pensar, aprender y comprendernos a nosotros mismos? La frontera entre la cognición humana y la artificial, que Boden considera un misterio en desarrollo, será el eje central de las discusiones que continuarán, enriquecidas por los aportes de los demás compañeros.

En conclusión, los principios de inteligencia artificial y creación de conocimiento no son el fin de la búsqueda de sabiduría, sino más bien el comienzo de una nueva perspectiva filosófica. Desde este punto de partida, el seminario se adentra en las raíces teóricas y filosóficas, y luego se dirige hacia una exploración más profunda de la IA como un participante en el proceso de conocimiento. Esta transición invita a reconsiderar la inteligencia artificial no sólo como un producto de la mente humana, sino también como un espacio de interacción y colaboración con ella. En este diálogo, el conocimiento se convierte en una creación compartida entre lo biológico y lo artificial, lo individual y lo colectivo.

7. IA como generadora de conocimiento

7.1. Contexto actual: Machine science y la sobrecarga de información

En las últimas décadas, el ritmo en la producción científica ha alcanzado niveles muy altos. Dia a día se publican miles de artículos, reportes y datos experimentales que amplían el cuerpo del saber humano, pero también generan una sobrecarga de información difícil de

procesar, al menos por el ser humano. La magnitud de este fenómeno es tal que hoy ningún investigador puede abarcar por sí solo toda la literatura relevante de su campo. Por ejemplo, un biólogo especializado en cáncer puede encontrarse con más de dos millones de artículos en el repositorio PubMed o con cientos de millones de resultados en una búsqueda general en Google. A esto se suman los volúmenes masivos de datos experimentales, que en disciplinas como la genómica o la astrofísica pueden alcanzar millones de gigabytes por proyecto.

Esta fragmentación progresiva de la literatura científica no implica solo un problema de volumen, sino también de coherencia epistemológica: los avances se distribuyen en comunidades aisladas, lo que genera "silos de conocimiento" donde relaciones importantes permanecen invisibles. Esta condición prepara el escenario para que la IA intervenga no solo como un procesador de datos, sino como un puente entre cuerpos de literatura desconectados.

Ante este escenario, la Inteligencia Artificial (IA) y las ciencias de la computación han pasado de ser herramientas auxiliares para convertirse en aliados para la ciencia moderna. Las computadoras no solo ayudan a almacenar y procesar información, ahora también amplían el alcance cognitivo del investigador, permitiendo pasar del simple análisis de datos a la formulación automatizada de hipótesis.

De este cambio surge la llamada Machine Science o ciencia de máquinas, un enfoque que propone el uso de sistemas computacionales basados en IA para integrar conocimiento publicado con datos experimentales, detectar patrones lógicos y hacer emerger nuevas hipótesis con mínima intervención humana. En otras palabras, la IA comienza a actuar como un "agente epistémico" una entidad capaz de generar conocimiento, y no solo de procesarlo. Casos como AlphaFold (DeepMind, 2020), que logró predecir la estructura tridimensional

de las proteínas con una precisión equiparable a la cristalografía experimental, o el uso de IA en astronomía para identificar exoplanetas y patrones cósmicos, demuestran que las máquinas pueden producir hallazgos que antes requerían décadas de observación humana. En química de materiales y ciencias sociales, la IA también ha permitido descubrir correlaciones y tendencias emergentes imposibles de detectar mediante métodos tradicionales. Se prevé que el desarrollo de herramientas más sofisticadas permitirá la generación automatizada y masiva de hipótesis científicas, las cuales podrían guiar experimentos de alto rendimiento en laboratorios de todo el mundo.

Sin embargo, esta expansión trae consigo nuevos retos epistemológicos: si las máquinas formulan hipótesis, ¿podemos considerar que "entienden" lo que están descubriendo? ¿O se trata solo de correlaciones útiles carentes de intención o significado? Estas preguntas marcan el paso de una era de procesamiento de información hacia una era de co-construcción de conocimiento entre humanos y sistemas artificiales.

7.2. La IA como generadora de Hipótesis: mecanismos y ejemplos

El avance hacia una ciencia en la que colaboran humanos y sistemas artificiales está transformando la manera en que se genera el conocimiento. Antes, la Inteligencia Artificial (IA) era vista principalmente como una herramienta de apoyo para analizar y clasificar datos. Sin embargo, en los últimos años ha comenzado a desempeñar un papel mucho más activo: el de generadora de conocimiento

Esto significa que los sistemas inteligentes ya no solo procesan información existente, sino que también pueden combinar lo que ya se sabe con nuevos datos experimentales, reconocer patrones ocultos y proponer posibles explicaciones o hipótesis por sí mismos. En otras palabras, las computadoras están empezando a participar en una parte del proceso

creativo de la ciencia, ampliando las capacidades humanas para explorar y comprender el mundo.

Se proyecta que tecnologías aún más avanzadas permitirán la generación automatizada y masiva de hipótesis científicas, capaces de guiar experimentos de alto rendimiento en disciplinas como la biomedicina, la química, la física y hasta las ciencias sociales. Este crecimiento marca un cambio profundo: la ciencia deja de depender únicamente de la intuición humana para incorporar mecanismos automáticos de descubrimiento. Esta participación de la IA en el proceso creativo de la ciencia se logra mediante dos vías principales:

Extracción de Conocimiento: Los sistemas analizan millones de publicaciones y bases de datos para identificar conceptos, relaciones y patrones implícitos que no habían sido articulados previamente.

Síntesis de Conceptos: A partir de la información recolectada, la IA combina elementos dispersos y construye nuevas relaciones conceptuales que permiten formular hipótesis inéditas.

Al expandir el conjunto de conceptos y relaciones disponibles, estas herramientas permiten a los científicos identificar y completar piezas faltantes en la comprensión de un sistema natural. También posibilitan el rastreo de cadenas de razonamiento mucho más largas de lo que es posible sin el uso de sistemas artificiales.

7.2.1. Machine Science: de la asistencia al descubrimiento automatizado

Este enfoque, conocido como Machine Science, propone el uso de sistemas computacionales avanzados capaces de explorar el corpus científico existente, descubrir

relaciones implícitas entre conceptos y generar nuevas proposiciones teóricas. Su función principal no se limita a procesar datos, sino a expandir el conjunto de conceptos y relaciones sobre los cuales se construyen las hipótesis. Para lograrlo, estos sistemas extraen, combinan y sintetizan información a partir de millones de publicaciones científicas, construyendo nuevos agregados conceptuales que reflejan conocimiento emergente.

Ahora bien, este proceso de descubrimiento automatizado no ocurre a partir de un único mecanismo, sino del acoplamiento de varias capacidades complementarias, que permiten que una máquina no solo "lea", sino que realmente genere ideas científicas nuevas. Entre estos mecanismos destacan:

1. Extracción de conocimiento

Los sistemas de Machine Science analizan grandes volúmenes de artículos, bases de datos biomédicas, repositorios experimentales y registros históricos. Esta extracción masiva permite:

- Detectar patrones estadísticos imperceptibles para un lector humano;
- Identificar relaciones implícitas entre conceptos no vinculados explícitamente en los textos;
- Reconocer correlaciones o tendencias que no han sido formuladas como hipótesis.

De esta manera, el sistema amplía sustancialmente la base de información accesible para un investigador, permitiendo una visión panorámica imposible de obtener manualmente.

2. Síntesis de conceptos

Luego de extraer información heterogénea, los sistemas recombinan conceptos dispersos para generar relaciones completamente nuevas. Esto los habilita para:

- Integrar resultados provenientes de disciplinas que rara vez interactúan,
- Construir modelos lógicos inéditos que no se encuentran en ningún artículo individual,
- Proponer hipótesis originales basadas en conexiones entre conceptos distantes.

En este punto, la IA no se limita a reorganizar conocimiento previo: crea nuevas piezas conceptuales que funcionan como elementos faltantes dentro de teorías o modelos científicos en formación.

3. Modelos predictivos y simulación

Las expresiones más avanzadas de Machine Science incluyen simulación a gran escala. Estos sistemas pueden:

- Generar millones de escenarios virtuales
- Probar rutas causales alternativas
- Evaluar la plausibilidad de distintas hipótesis automáticas
- Priorizar qué experimentos reales serían más prometedores.

Así, la simulación computacional funciona como un filtro que acelera el descubrimiento, guiando la investigación hacia los caminos más plausibles.

En conjunto, estos mecanismos permiten que Machine Science opere como un nuevo paradigma de descubrimiento, donde la computadora deja de ser un simple asistente para actuar como agente cognitivo complementario. El rol del científico humano no desaparece,

pero sí se transforma**:** se privilegia la validación, la orientación crítica y la interpretación, mientras que la IA amplía el espacio de posibilidades, explora conexiones invisibles y propone rutas de experimentación que antes quedaban enterradas bajo volúmenes inmanejables de información.

7.2.2. El modelo ABC de Sawson: Un antecedente fundacional

Uno de los ejemplos más influyentes de generación automática de hipótesis es el modelo ABC propuesto por Don R. Swanson (1986). Este modelo parte de la idea de que pueden existir conexiones ocultas entre cuerpos de literatura científica que no se citan entre sí, es decir, entre áreas de investigación aparentemente independientes.

El razonamiento del modelo puede resumirse en tres pasos:

1. A → B: Los conceptos A y B se estudian juntos dentro de un subcampo.
2. B → C: En otro subcampo distinto, B también aparece asociado a C.
3. Inferencia A → C: El sistema computacional, asumiendo transitividad, propone la hipótesis no publicada de que A podría estar relacionado con C.

Lo innovador del modelo ABC es que estas inferencias entre A y C suelen resultar plausibles e incluso verdaderas, pero eran extremadamente improbables de ser descubiertas por la mente humana sin asistencia, debido al volumen creciente y la fragmentación de la literatura científica.

Los hallazgos de Swanson demostraron el potencial del método. Entre sus predicciones más conocidas se encuentran:

- Que el aceite de pescado podría mejorar los síntomas del síndrome de Raynaud.

- Que las deficiencias de magnesio estaban asociadas con las migrañas.

Ambas hipótesis, generadas únicamente analizando literatura desconectada, fueron posteriormente confirmadas mediante estudios clínicos (Smith, 1989; Swanson, 1990).

Estos casos marcaron un precedente: las máquinas podían inferir conocimiento nuevo a partir de información ya publicada, sentando las bases de la ciencia automatizada contemporánea.

El valor central del modelo ABC no reside solo en la lógica transitiva, sino en la heurística cognitiva y social en la que se apoya. Swanson argumentó que generar hipótesis entre comunidades científicas desconectadas es más promisorio que buscar relaciones inéditas dentro de un mismo campo.

- Ideas entre comunidades (A → C):
  Cruzan fronteras epistemológicas. Suelen ser *preguntas no formuladas*, porque nadie está observando simultáneamente ambas literaturas.
- Ideas dentro de una comunidad (A → B):
  Los científicos del mismo subcampo suelen conocer la mayoría de sus propias relaciones implícitas, y las conexiones inéditas tienden a representar *"conocimiento negativo"*: ideas que no se publican porque los expertos saben que no funcionan en la práctica.

Investigaciones posteriores en biomedicina confirmaron este potencial: identificar biomoléculas que aparecen en múltiples subcampos permite construir puentes conceptuales que abren nuevas preguntas de investigación (Weeber et al., 2001).

Aunque el método ABC demostró que es posible descubrir hipótesis útiles al vincular literaturas desconectadas, su aplicación enfrenta un obstáculo significativo: las diferencias de lenguaje entre subcampos.

- Un mismo concepto puede denominarse de manera diferente en disciplinas distintas.
- Términos idénticos pueden referirse a entidades completamente distintas según el contexto.

Esto dificulta la identificación automática de equivalencias semánticas.

Este método evidenció que las máquinas podían inferir conocimiento nuevo a partir de información ya publicada, un principio que sentó las bases de la ciencia automatizada contemporánea.

7.2.3. Aplicaciones a gran escala y expansión del conocimiento

La expansión del conocimiento científico impulsada por la Inteligencia Artificial (IA) es una respuesta necesaria ante el crecimiento exponencial de datos en todas las disciplinas. Con los avances en aprendizaje automático y procesamiento del lenguaje natural (PLN), la IA dejó de ser un instrumento de análisis pasivo para convertirse en un agente generador de conocimiento, capaz de integrar bibliografía, experimentos y datos biomédicos a una escala inalcanzable para los humanos.

El razonamiento automatizado puede aplicarse a bases de datos científicas completas, permitiendo que los sistemas descubran patrones, relaciones y posibles mecanismos causales que pasarían inadvertidos en una revisión manual. Gracias a ello, la IA ha comenzado a desempeñar un papel crucial en múltiples áreas científicas de alto impacto.

Un área destacada es la reorientación o reutilización de fármacos (drug repurposing). Mediante el análisis de grandes corpus biomédicos, datos clínicos y redes de interacción molecular, los sistemas computacionales identifican nuevos usos terapéuticos para medicamentos ya aprobados. Herramientas basadas en redes y aprendizaje profundo, como DeepDTnet o RepurposeDB, han permitido proponer candidatos prometedores para enfermedades como cáncer, COVID-19 y trastornos neurodegenerativos, demostrando la capacidad de la IA para acelerar y abaratar el desarrollo farmacéutico (Zhou et al., 2020; Brown & Patel, 2021).

La IA también ha avanzado en la generación automatizada de hipótesis sobre enfermedades, un campo donde los sistemas integran correlaciones genéticas, ontologías biomédicas y grandes conjuntos de fenotipos clínicos. Herramientas como Phenolyzer u OpenTargets han ampliado sustancialmente el número de genes candidatos asociados a trastornos metabólicos, neurológicos y cardiovasculares, actuando como un puente entre datos dispersos y favoreciendo la identificación de piezas faltantes en la comprensión de sistemas biológicos (Zheng et al., 2020).

Estos mecanismos permiten que las máquinas propongan nuevas líneas de investigación o experimentos con potencial alto de descubrimiento, transformando el modo en que se concibe la producción científica.

7.3. Marco conceptual: Los niveles de abstracción (LoA) de Floridi

En su Filosofía de la Información, Luciano Floridi propone el método de los Niveles de Abstracción (LoA) como una herramienta conceptual para comprender cómo se genera y organiza el conocimiento. El LoA sostiene que cualquier sistema —sea humano o artificial— solo puede ser analizado desde un cierto nivel de detalle, eligiendo qué propiedades se

consideran observables y cuáles se omiten. No existe un nivel "correcto", sino una red coherente de niveles interrelacionados (Gradiente de Abstracción o GoA) que permite describir la realidad de manera consistente.

Este marco resulta clave para analizar los procesos cognitivos de la IA, ya que revela que los sistemas computacionales operan dentro de un nivel de abstracción limitado: el nivel sintáctico. Mientras el conocimiento humano incluye dimensiones semánticas y contextuales, la IA procesa información en términos de patrones, correlaciones y diferencias, sin comprender realmente el significado de lo que produce.

Profundizar en los LoA permite entonces evitar un error frecuente: atribuir a la IA capacidades cognoscitivas que requieren operar simultáneamente en varios niveles de abstracción. La IA no "comprende" en sentido estricto, porque su procesamiento no trasciende el nivel operativo para alcanzar niveles superiores de integración conceptual. Más bien, su aporte consiste en reorganizar información dentro de su propio nivel de abstracción, generando patrones y posibles hipótesis que, posteriormente, deben interpretarse desde otros niveles humanos más altos.

Floridi sugiere que la verdadera inteligencia no reside en un nivel superior aislado, sino en la coherencia entre los distintos niveles de abstracción, una idea que puede aplicarse tanto al pensamiento humano como al diseño de sistemas artificiales. En este sentido, el método LoA proporciona un marco filosófico que permite organizar el conocimiento generado por la IA y evitar confusiones entre información procesada y conocimiento significativo.

7.4. Riesgos y Limitaciones

La IA enfrenta varios límites estructurales y éticos en su papel como generadora de conocimiento. Aunque es capaz de procesar grandes volúmenes de información y detectar patrones complejos, su funcionamiento está condicionado por restricciones inherentes tanto al diseño algorítmico como al tipo de datos con los que opera. Entre los principales desafíos se encuentran:

- Barrera semántica: Las máquinas carecen de comprensión genuina; su "semántica" es derivada o prestada de los humanos. Producen resultados válidos, pero sin conciencia del significado, lo que impide afirmar que poseen conocimiento en sentido fuerte.
- Sesgo algorítmico: Los datos de entrenamiento condicionan la calidad y neutralidad del conocimiento generado. Un modelo entrenado con información incompleta o sesgada reproducirá esas distorsiones en sus inferencias.
- Sobrecarga de hipótesis: La IA puede generar tal cantidad de conjeturas que resulta difícil jerarquizar su relevancia o plausibilidad, lo que desplaza el problema de la "falta de ideas" hacia uno de "exceso sin filtro".
- Dependencia tecnológica: A medida que los humanos delegan más procesos cognitivos a las máquinas, se corre el riesgo de debilitar el pensamiento crítico y la autonomía intelectual.

7.4.1. Riesgos y consideraciones

No obstante, esta expansión trae consigo el riesgo de una "inflación de hipótesis", en la que el volumen de hipótesis generadas supera la capacidad de validación científica. Para mitigar este riesgo, es necesario establecer criterios que integren la relevancia epistemológica, social y cognitiva del conocimiento, evitando la producción de hipótesis

carentes de significado o basadas en correlaciones espurias. Desde la perspectiva de la Filosofía de la Información (Floridi, 2011), este fenómeno evidencia un dilema clásico: la diferencia entre información y conocimiento. Solo la información relevante y contextualizada puede ascender a la categoría de conocimiento. En este sentido, la IA no sustituye la comprensión humana, sino que actúa como un coagente epistémico, capaz de expandir los límites de la investigación al colaborar con la intuición científica del investigador.

7.5. Dependencia tecnológica

Diversos estudios han demostrado que la dependencia tecnológica puede afectar negativamente el desarrollo de habilidades cognitivas fundamentales. El caso de la calculadora en la educación ilustra bien este fenómeno. Investigaciones ecuatorianas (SENESCYT, 2022) indican que el uso excesivo de calculadoras reduce la capacidad de los estudiantes para resolver problemas aritméticos básicos, deteriora el razonamiento lógico y genera inseguridad cuando deben operar sin el dispositivo. Estos hallazgos suelen emplearse como argumento para afirmar que la tecnología deteriora progresivamente las competencias humanas.

Sin embargo, si bien existe un debate amplio sobre los posibles efectos de la tecnología en las habilidades cognitivas, la evidencia reciente sugiere que el uso de herramientas —incluyendo calculadoras e instrumentos de IA generativa— no elimina las capacidades humanas subyacentes, sino que modifica la forma en que estas se ejercen. Estudios sobre alfabetización digital y cognición distribuida muestran que los recursos tecnológicos actúan como amplificadores más que como sustitutos de la inteligencia humana (Hutchins, 1995; Clark, 2003). Tal como ocurre con la calculadora, cuyo uso no impide

aprender a sumar, sino que libera recursos mentales para resolver problemas de mayor complejidad, la IA generativa no anula la capacidad de escribir, analizar o razonar, sino que reconfigura la manera en que estas habilidades se ponen en práctica.

Desde esta perspectiva, la IA no sustituye el pensamiento crítico ni las operaciones cognitivas fundamentales. Para producir contenido significativo, sigue requiriendo una intervención humana intencional, tanto en la formulación de preguntas como en la evaluación de resultados. La tecnología puede automatizar tareas, pero no puede reemplazar la comprensión, el juicio ni la motivación, que son elementos estructurales de la agencia humana.

La motivación, en particular, es un componente que no puede delegarse. El impulso por investigar, crear o resolver un problema es exclusivamente humano; la IA no inicia acciones por sí misma. Esto se observa incluso en actividades creativas. Un pintor profesional mantiene su identidad artística independientemente de que utilice o no herramientas digitales, mientras que un principiante puede apoyarse en la IA para acercarse a un nivel técnico mayor, sin que ello lo convierta automáticamente en un experto. La herramienta amplifica capacidades, pero no reemplaza la pericia adquirida mediante experiencia, práctica y sensibilidad estética.

De este modo, la IA funciona como un agente epistémico complementario, cuya eficacia depende de la calidad cognitiva del usuario. No se trata de una amenaza para las habilidades humanas, sino de un entorno que exige nuevas formas de alfabetización crítica. La dependencia tecnológica —más que deterioro cognitivo— plantea desafíos de uso responsable, interpretación y verificación que mantienen al ser humano como el núcleo del proceso de producción de conocimiento.

7.6. Discusión critica con autores

Autoras como Margaret Boden argumentan que la creatividad artificial solo puede considerarse auténtica cuando el sistema es capaz de generar resultados inesperados dentro de un marco conceptual propio. Para Boden (1998), la creatividad —humana o artificial— implica la habilidad de transformar espacios conceptuales mediante la exploración o reestructuración de reglas internas. En este sentido, la IA generativa parecería mostrar creatividad "combinatoria" o "exploratoria", pero no "transformacional", porque carece de un marco conceptual autónomo que le permita redefinir las bases mismas de su producción. La máquina sorprende, pero no se sorprende; produce novedad, pero no comprende su alcance.

Sin embargo, Marvin Minsky sostenía que la inteligencia no depende de una "comprensión interna" ni de una conciencia subjetiva, sino de la capacidad para resolver problemas de manera funcional, lo que sitúa a la IA como un nuevo tipo de inteligencia "sin consciencia". (Minsky, 1986). Desde esta perspectiva, la IA constituye un nuevo tipo de inteligencia "sin consciencia", cuya utilidad no radica en reproducir la cognición humana, sino en realizar tareas que los humanos consideran inteligentes. Minsky desplaza la discusión desde lo semántico hacia lo operativo: no importa si la máquina entiende, sino si actúa eficazmente en un entorno determinado.

Por su parte, Stuart Russell ha advertido que el avance de la IA plantea una paradoja: cuanto más eficientes son los sistemas, más difícil resulta para los humanos comprender cómo llegan a sus conclusiones. (Russell & Norvig, 2021). Esta "opacidad algorítmica" no solo genera desafíos técnicos sino, sobre todo, dilemas epistemológicos. Si no se puede explicar por qué un sistema produce una inferencia, ¿puede considerarse ese resultado como

conocimiento justificable? La automatización de patrones no equivale a la justificación racional, por lo que la relación entre IA y ciencia requiere mecanismos de supervisión humana que aseguren trazabilidad conceptual. Esta preocupación conecta directamente con los planteamientos de Luciano Floridi y su Filosofía de la Información.

7.7. Perspectivas de futuro

El futuro de la investigación científica apunta hacia una colaboración híbrida entre humanos y sistemas de IA, configurando un modelo de cocreación en el que ambas capacidades —la cognición humana y la computación algorítmica— se integran de forma complementaria. Los enfoques de human-in-the-loop buscan precisamente mantener al investigador en el centro del proceso, empleando la IA como un asistente epistémico que amplía la capacidad de análisis, acelera la exploración de hipótesis y facilita el acceso a patrones o correlaciones difíciles de detectar de manera manual, pero sin reemplazar el juicio humano ni la responsabilidad interpretativa.

Este modelo colaborativo tiene implicaciones profundas. Por un lado, la IA cumple funciones de ampliación cognitiva, permitiendo procesar volúmenes masivos de datos, generar alternativas teóricas o simular escenarios complejos. Por otro, el ser humano aporta elementos que siguen siendo estrictamente irremplazables: comprensión contextual, motivación, criterio ético, razonamiento abductivo y la capacidad de traducir hallazgos en narrativas científicas significativas. La ciencia del futuro se perfila, así como una ecología cognitiva donde distintas formas de procesamiento —humana y artificial— se entrelazan para optimizar la producción de conocimiento.

En este marco, comienza a tomar fuerza la noción de inteligencia híbrida, entendida como un sistema en el que los agentes humanos y algorítmicos cooperan, cada uno desde sus

fortalezas: la IA se especializa en velocidad, exhaustividad y precisión estadística; el ser humano en interpretación, deliberación y creatividad reflexiva. Este modo de colaboración no busca sustituir el pensamiento, sino reconfigurar su alcance y sus posibilidades.

Finalmente, esta sinergia podría definir una nueva etapa en la historia de la ciencia, donde el conocimiento se construya mediante una interacción constante entre razón humana y poder computacional. En lugar de desplazar las habilidades cognitivas, la IA debería funcionar como un catalizador para la reflexión, la creatividad y la exploración de nuevas formas de entender la realidad. El desafío no es evitar la IA, sino aprender a integrarla dentro de un marco epistemológicamente sólido que preserve la agencia humana, garantice la transparencia y mantenga el sentido crítico en el corazón del proceso científico.

8. Impacto de la IA en la cognición humana.

8.1. Fundamentos de la cognición humana. Algunos principios de filosofía de la mente sobre cómo aprenden y procesan información los humanos.

La cognición humana es entendida como el conjunto de procesos mentales que hacen posible la percepción, el aprendizaje, la memoria, el razonamiento, el lenguaje y la toma de decisiones. A partir de mediados del siglo XX, el desarrollo de la psicología cognitiva junto al avance de la informática promovió la metáfora del cerebro como procesador de información que tuvo, a su vez, un impacto directo en la construcción de los primeros modelos de inteligencia artificial. Sin embargo, esta perspectiva, a pesar de ser útil para describir algunas funcionalidades del funcionamiento de la mente, se deja fuera a la intencionalidad, la emoción, el contexto social donde se desarrolla el pensamiento humano.

La filosofía de la mente se encarga de explorar precisamente estas dimensiones. Según Daniel Dennett (1991), podría ser erróneo incluso pensar que la conciencia humana debe ser considerada un objeto que está localizado como sí existiera, sino que fue emergente en muchas formas de los procesos interpretativos que están actuando a la vez. Para Dennett en su modelo de los "múltiples borradores" (ya que tanto Montaigne como él son los héroes de esta visión), no hay un punto central de control o "teatro cartesiano" donde la mente procesa la información, sino que hay un conjunto de procesos cognitivos que compiten entre sí y a la vez cooperan entre ellos y como resultado se construye la experiencia consciente. Esta interpretación es crucial para entender por qué el pensamiento humano no puede ser reducido a una secuencia lineal de operaciones lógicas: la mente integra la información de manera más bien descentralizada, contextual y dinámica como podría ser también por ejemplo una red compleja.

Siguiendo esta línea, el mismo Edgar Morin (2001) hace alusión a la idea de que la cognición es de tipo autoorganizado. Para él, el pensar supone articular razón y emoción, individuo y sociedad, sujeto y objeto. De este modo, el conocimiento no es simplemente una representación del mundo, sino una reconstrucción activa, dependiente del marco cultural, de las experiencias y de los valores del sujeto. Según Morin y su concepción del pensamiento complejo, cualquier intento de captar el funcionamiento de la mente sin hacerla entrar en interdependencias, en combinación, con sus otras funciones mentales inexorablemente conduce a visiones reduccionistas del conocimiento, con la consiguiente llegada de concepciones deshumanizadas de tipo tecnológico.

Tales fundamentos hacen que sea posible la comprensión de la diferencia básica entre procesamiento e interpretación, el primero puede captar patrones, mientras que el segundo

otorga a estos patrones un significado. Esta capacidad de dar significado, que está fuertemente arraigada en la cultura y en la historia y experiencias personales, es la esencia del conocer. En lo que respecta a la inteligencia artificial, tomar conciencia de esta diferencia es importante para no confundir la ejecución del cálculo con el pensamiento o hacer equiparaciones entre la correlación estadística y el conocer. Sherry Turkle (2011), desde una perspectiva sociotécnica, investiga la cuestión mediante una decantación tanto de la práctica como de la teoría filosófica, porque, explica, los artefactos tecnológicos han empezado a modificar no sólo lo que pensamos sino, además, cómo pensamos. “Alone Together” (Alone Together, 2011) utiliza la interacción de los niños con los robots sociales Furby y My Real Baby para investigar esta cuestión (capítulo 4, “Alive Enough”, pp. 61–83), a partir de entrevistas y la técnica etnográfica de la observación de los niños para demostrar cómo estos dotan de vida y de emociones los objetos con los que interactúan. Un Furby “apaga” su luminosidad; y muchos de los niños manifiestan que lo ven “triste” o “durmiendo”, por lo que depositan su propio marco afectivo en la máquina. Turkle considera este efecto una evidencia de que los humanos asumen su naturaleza cognitiva para llegar a una “empatía simulada”, porque, no sólo piensan sobre las máquinas, sino que piensan con éstas y a partir de esta relación revisan su propio concepto de lo vivo, lo consciente y lo humano. Esta idea está en la misma sintonía con el concepto de "objetos evocativos" también ofrecido por Turkle, que los artefactos tecnológicos son, de algún modo, motores de la reflexión de las personas; no son únicamente herramientas, sino espejos que nos devuelven preguntas acerca de nuestra identidad o nuestra mente. En este sentido, los dispositivos inteligentes funcionarían como extensiones simbólicas del pensamiento, dando paso a una cognición distribuida humano máquina.

George Siemens (2004) da una nueva vuelta a esta idea desde la perspectiva del aprendizaje. En su artículo Connectivism: A Learning Theory for the Digital Age (pp. 5-6), afirma que el 'conocimiento' actual no se encuentra solamente en la cabeza de cada individuo, sino que reside también en redes que conectan personas, instituciones, sistemas tecnológicos y bases de datos. Saber se convierte entonces en un acto que requiere saber moverse entre nodos de información, saber discernir su relevancia y saber actualizar los vínculos cada vez que es necesario. En este caso, el aprendizaje deviene una actividad relacional y un proceso distribuido, dando lugar a una cognición humana que, en su sentido más amplio, se extiende hacia los sistemas que utiliza para pensar.

El conectivismo de Siemens comparte con la filosofía de Dennett la propuesta según la cual la mente es no-centralizada; todo lector de su obra irá comprobando que estos autores piensan en el pensamiento como, ni más ni menos, una red de procesos concomitantes y dependientes unos de otros. No obstante, Siemens, que se encuentra en un contexto social y tecnológico, aunque igualmente divulgador, desaloja esta idea de ella misma: Dennett construye su pensamiento a partir de los mecanismos que tiene la mente dentro de ella misma, mientras que Siemens construye su pensamiento a partir de las redes externas de información que son las que las despiertan y las modifican. Superando así la visión de que la mente humana es un sistema cerrado y pasando a superarlo por una plataforma cognitiva híbrida, si es que se puede seguir así.

Por su parte, Turkle y Siemens comparten un punto vital: la tecnología no solo es un dispositivo que ensancha nuestras capacidades, sino que, además, configura nuestra forma de razonar, pues las tecnologías digitales y la gran mayoría de los dispositivos conectados a Internet están a nuestra disposición, de modo que, como sostiene Turkle (2011, p. 5:77) hace

notar que los niños criados con robots y ordenadores tienden a aprender la inteligencia como algo que sucede en las relaciones y no necesariamente como un acto consciente. En palabras de una de sus entrevistadas, “no importa si el robot entiende; importa que actúe como si entendiera”. Esta afirmación es un claro resumen de un cambio profundo en el epistemológico en el que, en la era digital, la simulación de la comprensión puede adquirir el mismo valor práctico que la propia comprensión.

Desde una red de comprensión a la que se le imprime un mayor matiz teórico, Siemens (2004, pp. 7–8) advierte que, en esta misma tendencia, hay ya un aprendizaje que tiende a esa sinergia entre humano y sistema, afirma que la cognición es una de las propiedades emergentes de la red, no del individuo, y que la tarea educativa, y por extensión la ingeniería cognitiva, no es otra que la de diseñar sistemas que hagan viable esa red de conexiones e ir manteniendo el hecho de que la inteligencia humana siga siendo el principal foco interpretativo de la información que produce. En conclusión, los principios que subyacen en la cognición humana y los principios de la filosofía de la mente coinciden en que el pensamiento humano es emergente, contextual y distribuido. No se produce en la soledad, sino que mantiene un diálogo recíproco con su contexto físico, social y tecnológico. Las aportaciones de la investigadora y sociocultural Sherry Turkle (2011) y del investigador y teórico de la educación George Siemens (2004) atestiguan que la tecnología ha dejado de ser un mero instrumento externo, sino que ha pasado a ser una de las componentes activas de la cognición, un agente que afecta cómo representamos la realidad y cómo la propia inteligencia es determinada. Pero la filosofía y la teoría del aprendizaje apuntan con énfasis a un límite: la inteligencia humana es singular en cuanto capacidad de dar significado, si bien la mensajería artificial conceptualiza patrones y correlaciones, el ser humano interpreta,

pregunta y vuelve a dar forma a la información dentro de un horizonte ético y simbólico. La verdadera frontera entre lo humano (mente) y la máquina no se encuentra en la velocidad o en la memoria, sino en la intencionalidad reflexiva y en la conciencia de sentido.

8.2. Análisis de cómo la IA está cambiando la manera en que los humanos aprenden y procesan información.

La llegada de la IA ha implicado una profunda reconfiguración de la relación entre el ser humano y el conocimiento. En las dos últimas décadas, los sistemas algorítmicos se han consolidado como mediadores cognitivos, filtrando, organizando y priorizando la información a la que accedemos cotidianamente. A la naturaleza tecnológica de esta transformación se une la forma en la que también afecta a cómo aprenden, piensan y procesan información las personas, deviniendo aquello que determinados autores han denominado ecología cognitiva híbrida.

En Alone together, Sherry Turkle (2011) argumenta que las tecnologías inteligentes no son herramientas, sino que son interlocutores cognitivos. En el capítulo titulado "Alive enough" (2014: 61-83), los niños que entablan interacciones con robots sociales aprenden ideas básicas sobre la vida, la inteligencia y la empatía a través de conversaciones con artefactos que simulan la comprensión. Uno de los aspectos más interesantes de su estudio se produce cuando un niño le dice: "No importa si el robot siente de veras, lo que importa es que actúa como si lo hiciera". La autora explica esta afirmación de los niños como una forma en la que la simulación sustituye en parte la comprensión: el aprendizaje depende menos que antes de que el conocimiento sea veraz para tener lugar y se apoya más en la experiencia práctica de la interacción.

En este contexto, el aprendizaje no tiene lugar en la mente aislada, sino en redes distribuidas que entrelazan la cognición humana con la capacidad computacional de los robots. De acuerdo con George Siemens (2004), este es el conectivismo, una teoría que postula que el conocimiento ya no es algo que se acumula, sino que fluye como un río desde una cantera hasta la siguiente a lo largo de redes. A la hora de definir el aprendizaje, Siemens sostiene que "la capacidad de ver conexiones entre diferentes campos, ideas y conceptos está considerado como una de las aptitudes más importantes" (págs. 5-6), y aprender es tener activa una red de información que se mantiene viva y es flexible (pág. 6). La inteligencia artificial refuerza y acelera esta dinámica. Los sistemas dotados de algoritmos de aprendizaje automático pueden procesar volúmenes de información que son imposibles de captar para la mente humana. Estas habilidades generan una cognición compartida entre humanos y máquinas: las personas son quienes definen las preguntas, y los sistemas inteligentes, quienes ofrecen respuestas que son interpretadas y aceptadas por los usuarios. El aprendizaje, por tanto, es circular y colaborativo. Veamos un ejemplo. Pero esta colaboración también implica un cambio profundo en la estructura del pensamiento. Daniel Kahneman (2011) propone que la mente humana funciona mediante dos sistemas de procesamiento, cual Sistema 1, rápido o automático, intuitivo; Sistema 2, lento, analítico y deliberado (págs. 19-29). La inteligencia artificial, debido a la inmediatez y rapidez con la que proporciona los resultados, sirve asumiendo la tendencia a la dependencia del Sistema 1, por lo que afea la práctica del pensamiento crítico y la deliberación. Turkle (2011, p. 79) advierte que: "La máquina se expresa con suficiente fluidez que queda abolido el interrogante sobre la forma en que lo hace". Por ende, el aprendizaje mediado por IA acaba favoreciendo la eficiencia por encima de la reflexión.

No obstante, Siemens (2004, pp. 7–8) defiende que esta configuración de la cognición no debe ser considerada como una pérdida, sino como una extensión del sistema mental. En la era digital, "la habilidad de saber dónde procurar la información es preferible a tenerla". La inteligencia artificial pasa a ser entendida como un andamiaje cognitivo que amplifica la orientación, la actualización y la toma de decisiones. Desde una óptica pedagógica, tal afirmación implica que el aprendizaje no se aplica a la memorización, sino a la capacidad de relacionar datos, comprender el contexto y juzgar la fiabilidad de los algoritmos que median la información que circula.

De este modo, la inteligencia artificial resulta ser una extensión de la mente humana, configurando aquello que, en términos de Luciano Floridi (2014), da forma a lo que él denomina infosfera: un ecosistema que tiene la información como medio vital de interacción. Así, la infosfera hace que cada usuario sea productor y consumidor de datos, que su cognición sea alterada por las estructuras algorítmicas que filtran la realidad (pp. 83–102). La inteligencia artificial redefine los límites de lo que el ser humano tiene como conocimiento y como información relevante porque gestiona de un modo automatizado la información a partir de un modelo predictivo. La consecuencia inmediata de esa mediación es la disponibilidad de un aprendizaje adaptativo, personalizado, mediado por sistemas de IA que analizan el comportamiento de los estudiantes para reajustar los contenidos. Ejemplo de esta tendencia lo constituyen los sistemas de aprendizaje automático tales como los Learning Management Systems (LMS) inteligentes, que estudian los tiempos de respuesta, los niveles de atención, los patrones de error que ofrece cada usuario para (…) para hacer de las trayectorias de aprendizaje un aprendizaje individual.

Turkle (2011, pp. 80-82) señala que esta personalización, en su aspecto positivo, ayuda a que la eficiencia del aprendizaje sea adecuada; pero también tiene efectos negativos. “El aprendizaje no se produce cuando todo se ajusta a nosotros, sino cuando encontramos resistencia”, escribe. O, dicho de otro modo, el exceso de adaptación tecnológica puede eliminar la distancia, la confrontación con la diferencia, lo que, para ser sinceros, es lo que nos lleva a un crecimiento intelectual.

Desde la ingeniería de sistemas, esta característica también tiene implicaciones directas. Los ingenieros no solo crean sistemas de IA, sino también contextos cognitivos intermediarios. Cada algoritmo de recomendación, cada modelo de lenguaje o cada interfaz de usuario determinan la manera como las personas aprenden y toman decisiones. En esta línea, la responsabilidad del ingeniero no queda acotada al rendimiento técnico del sistema, sino que también abarca su impacto epistemológico y educativo.

George Siemens (2004, pp. 8–9) indica que la tarea más fundamental de los trabajadores del conocimiento en la era digital es “mantener fluyendo un torrente continuo de información exacta, pertinente y en el tiempo”. En este sentido, hay que asumir que no se trata de acumular memorias sino más bien de gestionar conscientemente las conexiones. La IA potencia la capacidad cognitiva, pero a su vez, reclama del ser humano un discernimiento superior al evitar depender ciegamente de las respuestas algorítmicas. Floridi (2014) complementa la visión que contribuye a tener una idea sobre lo que se entiende por “infosfera”, ya que esta no es sólo la totalidad del propio conocimiento en red, sino que además debemos entender que para Floridi, en la infosfera, “los humanos no son observadores externos de la información, sino que son agentes informacionales que forman parte de un sistema que también incluye a la información” (pág. 82). Esto significa que cada

interacción digital afectará a la persona y también al entorno en que la persona se mueve e interactúa. De modo que el aprendizaje se convierte en un proceso coevolutivo: el sujeto entrena el algoritmo y el algoritmo modela la cognición del usuario. Por ello, el papel de la IA en el aprendizaje humano no es solamente instrumental, sino que marca una frontera ontológica: el algoritmo cambia la naturaleza misma de la cognición. Como advierte Turkle (2011, p. 83): "La tecnología nos dice lo que significa ser una persona". Los humanos que piensan que interaccionan con máquinas que simulan una comprensión de esta forma redefinen lo que piensan acerca de pensar, sentir o saber. Para resumir, la inteligencia artificial ha cambiado el modo en que los humanos aprenden y procesan la información por tres dimensiones principales. La distribución cognitiva: el conocimiento no reside tantas veces como antes en la mente, sino que ahora se encuentra distribuido entre redes humanas y tecnológicas (Siemens, 2004, pp. 5-6). La interacción simulada: la empatía y la comprensión se invierten hacia interacciones mediadas por máquinas (Turkle, 2011, pp. 67-79). La personalización algorítmica: la IA determina el aprendizaje como un proceso adaptativo, pero pueden eliminarse los efectos positivos que produce la diversidad cognitiva (Floridi, 2014, pp. 90-92). Por tanto, el reto no es rehuir la transformación sino más bien reconstruir la educación y la ingeniería del conocimiento a la vez que se cuida la autonomía crítica y del pensamiento reflexivo. Como recuerda Siemens (2004, p.10), "el aprendizaje es un proceso que tiene lugar en el interior de ecosistemas caóticos que cambian incesantemente"; en esta era de la IA, el éxito no radica en controlar este caos, sino en aprender a navegar con responsabilidad y conciencia.

8.3. Análisis del riesgo de la IA en estos procesos cognitivos a partir de la dependencia tecnológica y del desplazamiento que la IA está causando en los empleos humanos.

La influencia de la inteligencia artificial sobre la conformación de la cognición humana no es solo transformativa, sino que también presenta un carácter ambivalente. Por un lado, las máquinas inteligentes (inteligencia artificial) amplían las capacidades del pensamiento, por otro, introducen nuevas formas de dependencia y vulnerabilidad cognitiva. La relación entre los humanos y la máquina inteligente se halla regida por la tensión esencial entre la autonomía y la delegación: por un lado, la IA proporciona velocidad, precisión y eficiencia, por otro, el ser humano tiende a renunciar a parte de su control cognitivo, el ser humano tiende a renunciar a parte de su control cognitivo. En Alone Together, Sherry Turkle (2011) hace una advertencia acerca de la manera en la que la dependencia emocional y funcional hacia los artefactos es una de las transformaciones más preocupantes de la llegada de la era digital. En el capítulo "Alive Enough" (pp. 61-83), da fe de cómo las personas, y muy especialmente los niños y las personas mayores, desarrollan vínculos afectivos con los robots y los asistentes virtuales, incluso tratándolos como interlocutores válidos; en el peor de los casos, como sustitutos de las relaciones humanas. Turkle entiende este fenómeno como "la ilusión de compañía sin las exigencias de la relación" (p. 79). Desde una perspectiva cognitiva, esta ilusión puede llevar a las personas a proyectar en la máquina sus propias expectativas afectivas y de comprensión; en cierto modo, reduciendo el ejercicio reflexional que caracteriza la interrelación humana.

Turkle argumenta que esta dependencia de la técnica no acaba siendo inocente: dado que la gente empieza a preferir las respuestas de los sistemas automatizados frente a la relación humana, se merman las capacidades afectivas, de paciencia y de tolerancia a la ambigüedad. Desde un punto de vista cognitivo, esto resulta en el hecho de que el pensamiento reduce su mayor diálogo y se convierte en un pensamiento mucho más

instrumental; los sujetos aprenden a procesar la información como hacen las máquinas: de modo rápido y superficial. Esta “automotivación del pensamiento” amenaza con desplazar la reflexión profunda que había sostenido la construcción tradicional del conocimiento. Por su parte, George Siemens (2004) había anticipado otra forma de dependencia cognitiva. Dependencia derivada del exceso de información y de conectividad. En su teoría del conectivismo (pp. 8–9) previene que la masiva cantidad de información a la que deben enfrentarse las personas en el entorno digital les empuja hacia sistemas automáticos que filtran y organizan dicha información. Esta delegación de la selección cognitiva hacia los algoritmos produce una externalización del juicio, dejando que los procesos tecnológicos sean los que determinen la relevancia y, para ello, un uso excesivo de criterios y “orquídeas” que no siempre entendemos. El mismo Tung la explica como una paradoja: "En un entorno donde la información está siempre en cambio, la estabilidad del conocimiento depende de la red, no de los individuos". En otras palabras, cuanto más conectados estamos, más vulnerables a depender de la infraestructura tecnológica que sostiene estas conexiones. Desde la perspectiva de la epistemología, ello hace que el pensamiento humano no camine de manera independiente sino condicionado por los mecanismos de recomendación y de clasificación algorítmica.

La consecuencia más patente de esta dependencia la califican Turkle (2011, p. 81) como una "soledad conectada": la paradoja contemporánea en la que los individuos son permanentemente comunicados, pero emocionalmente aislados. Esta soledad tecnológica presenta un correlato a nivel cognitivo: la pérdida de la conversación interior, de aquella conversación interna que permite elaborar ideas complejas. Siguiendo el proceso de

tercerización del pensamiento hacia la restitución de sistemas de IA, el ser humano entra en riesgo de volverse un usuario de sistemas reactivos, y no un pensador deliberativo.

Pero no solo la dependencia cognitiva es el riesgo al que nos enfrentamos. En 21 Lecciones para el siglo XXI, Yuval Noah Harari (2018) nos habla del riesgo de desplazamiento laboral y cognitivo que está produciendo la IA en nuestro contexto contemporáneo. En el capítulo "Trabajo" (pp. 29-41) menciona cómo la automatización inteligente no solo sustituye las tareas manuales, sino también aquellas funciones cognitivas más altas. La IA ya es capaz de redactar textos, traducir, diagnosticar enfermedades, realizar cálculos sobre datos financieros, y hacerlo de manera más precisa que los humanos. Según Harari, esto podría producir la creación de una nueva clase social como es la de los "humanos irrelevantes", lo que la IA puede conseguir superando los modelos de los humanos que aprenden más lentamente y cometen más errores.

Lo mismo que la economía, el desplazamiento de la IA implica una serie de consecuencias epistemológicas. Al dejar a máquinas automáticas la interpretación y la generación de conocimiento, la sociedad está en la antesala de la pérdida de la diversidad cognitiva que proviene del error humano, de la creatividad, de la ambigüedad, etc. Al respecto, señala Harari (2018, p. 35): "La inteligencia no equivale a conciencia" y confundirlas es caer en una falta ética-científica; creer que lo que una máquina calcula equivale a lo que una persona comprende.

A partir de la ética de la información, Luciano Floridi (2014) demuestra este problema en las páginas 100-118 de The Fourth Revolution a través de la idea de la “infosfera” como nuevo espacio moral y cognitivo compartido por seres humanos y máquinas. En esta nueva infosfera, cada decisión que toma la máquina bien sea una recomendación de búsqueda o un

diagnóstico médico, transforma el modo de percibir la realidad del ser humano. Floridi advierte que la dependencia de los algoritmos provoca la heteronomía informacional, donde las decisiones son moldeadas por sistemas opacos, no necesariamente auditables.

Pero este problema, desde la ingeniería de sistemas se traduce en hacer explicables las IA (Explainable AI-XAI) que puedan justificar sus decisiones e involucrar a la persona en el proceso de razonamiento. Cuando se carece de transparencia en los algoritmos subyacentes, la cognición humana se convierte en una subestructura de un sistema computacional que la trasciende. Esto implica un riesgo epistemológico de primer orden, el de perder la autoridad respecto del saber.

Los riesgos a nivel cognitivo pueden ser descritos a tres niveles:

- Riesgo funcional. Los sujetos dejan de desarrollar algunas de las capacidades más básicas, más elementales de la cognición, la memoria de trabajo o la atención sostenida, en la medida en que las confían a los sistemas de IA. Turkle (2011, p. 82) expone que, cuando los estudiantes se enfrentan a buscadores automáticos, les resulta más fácil "pensar en formato de búsqueda": sólo pueden formular las ideas en forma de búsqueda, nunca en forma de argumentos.

- Riesgo emocional y motivacional. La dependencia excesiva de la tecnología, potencialmente, puede dar lugar a estrategias de gratificación. En este sentido, Siemens (2004, pp. 7–8) considera transmitir el mensaje de que las redes digitales pueden transformar la motivación intrínseca en motivación algorítmica, en forma de gratificaciones (instantáneas) por conexión. Esto debilita la capacidad de mantener procesos de aprendizaje prolongados y de reflexión.

- Peligro social y laboral: como manifiesta Harari (2018, pp. 33-35), la automatización que los investigadores llaman inteligente tiende a polarizar las oportunidades; los individuos que manejan la tecnología se hacen con el poder cognitivo y económico, mientras que quienes no lo hacen se convierten en algunos de los interesados obsoletos. Tal brecha digital no es únicamente tecnológica, sino también mental, ya que determina el control de la información y, por tanto, la interpretación del mundo.

Frente a esos retos, Floridi (2014, p. 115) expone una ética de la responsabilidad informacional que afirme que el diseño y uso de la IA debe absorberse de tres principios: beneficencia cognitiva (made in better knowledge), transparencia epistémica (exponer los procesos algoritmos) y preservación de la autonomía mental. Así es como la inteligencia artificial y el arte del pensamiento humano pueden fundirse sin que la esencia de este último se vea perdida. En suma, los riesgos cognitivos que pueden surgir a raíz de la IA no vienen determinados por la tecnología, sino que son el resultado del modo en el que la humanidad incorpore la IA a la vida mental y social. Turkle previene contra el riesgo de perder la profundidad emocional y reflexiva que el ser humano había conseguido; Siemens señala nuestra dependencia estructural de las redes; Harari anticipa el desplazamiento tanto cognitivo como laboral; y Floridi exige que apelemos a una ética de la infosfera que transite entre la autonomía y la diversidad del mismo pensamiento humano. El desafío para la ingeniería del presente retomará el propósito de desarrollar sistemas que fomenten la colaboración entre humanos y máquinas sin convertir a la persona en un agente pasivo de su propio conocimiento. La IA no ha de sustituir la cognición misma, sino, por el contrario, potenciarla crítica, moral y creativamente. Únicamente así, la inteligencia híbrida podrá

realizar la promesa de abrir nuevos horizontes del saber, sin renunciar por el camino a la libertad mental que define lo humano.

La presente aproximación de análisis que se ha llevado a cabo en el presente capítulo nos ha permitido llegar a la conclusión de que la inteligencia artificial no es sólo la técnica que afina nuestras capacidades cognitivas, sino que es una fuerza epistemológica que reconfigura la forma en la que los humanos hacemos uso del conocimiento, aprendemos y pensamos. A partir de la base de la cognición humana, quedó patente que la mente no se comporta como un sistema único, sino que también se comporta como una red autoorganizada de procesos que son simultáneamente simbólicos, sociales y emocionales. Autores como Dennett (1991; p.19) y Morin (2001) pusieron de manifiesto que la conciencia fluye conforme interactúan diferentes niveles de autorreflexión, biológicos, culturales, y que el saber humano no existe, sino que permanece unido por el contexto. La emergencia de la inteligencia artificial, analizada a partir de Turkle (2011) y de Siemens (2004), demuestra que estas redes han logrado salir de las moquetas del cerebro y han conformado una cognición distribuida en donde los sistemas algorítmicos son nodos activos de procesamiento. En esta reciente ecología cognitiva, aprender significa saber cómo conectarse, cómo interpretar, cómo actualizar información dentro de un entorno de artefactos humanos y de máquinas. En este sentido, la IA no reemplaza la forma de pensar, sino que cambia de forma tal que se fortalecen algunas capacidades, como, entre otras, la velocidad y el tratamiento de la información, pero se debilitan otras, la deliberación, la reflexión y el juicio críticos, por poner algunos ejemplos.

No obstante, la expansión de la inteligencia artificial también supone un riesgo no menor, como advierte Turkle, quien cree que el apego emocional y funcional hacia los

sistemas inteligentes altera nuestra forma de pensar de manera tal que se promueve un tipo de “soledad conectada” en el que la interacción humana es suplantada por la interacción simulada. Siemens a su vez hace constar que la prevalencia de la conectividad origina una dependencia estructural a la información extraída a través de algoritmos, lo que derivará en una pérdida de la autonomía epistémica. Harari (2018) añade que la metamorfosis se produce también en el mundo laboral y cognitivo, que hace que las personas sean despojadas de los procesos interpretativos en favor de sistemas automáticos, mientras que Floridi (2014) recuerda que el uso acrítico de la IA puede desencadenar en una heteronomía informacional que haga peligrar la libertad mental.

En una palabra, la inteligencia artificial es un nuevo momento de evolución de la mente humana, donde se torna porosa la separación entre biología y la digitación. El reto contemporáneo no es resistirse a la misma, sino ayudar a gestionarla, comprenderla y guiarla de forma ética. La ingeniería de sistemas, al estar situada en la cúspide de esta metamorfosis, debe ocuparse de diseñar arquitecturas de IA que fomenten la transparencia, la explicabilidad, y la colaboración reflexiva entre humanos y máquinas.

El futuro de la cognición humana no dependerá de la cantidad de información que somos capaces de manejar sino de la calidad de nuestra conciencia frente a la tecnología. Si la IA se acaba introduciendo como catalizadora del pensamiento crítico, creativo y ético (y no como un desgaste de la reflexión), la IA puede llegar a convertirse en un aliado a fin de expandir la inteligencia humana sin que ello implique su merma. En ello reafirma la verdadera prueba epistemológica del siglo XXI.

9. Conciencia y metacognición en IA.

9.1. Introducción

Durante el desarrollo de la tecnología en los últimos años, se ha empezado a generar diferentes opiniones sobre el uso de la inteligencia artificial y si en algún momento podrá empezar a generar conciencia por sí misma y con ello, ser capaz de reemplazar por completo a los humanos como se ha empezado a ver en la actualidad en algunas tareas cotidianas; por ejemplo el uso que le dan los estudiantes, ya sea para realizar trabajos extensos, estudiar o realizar presentaciones; incluso elementos como el diseño gráfico de una página web o ilustraciones de marketing para empresas están empezando a quedar en manos de la inteligencia artificial, trabajo de horas reducido a tan solo unos minutos. Ciertamente el avance de la tecnología ha ayudado a que estas mismas sean realizadas con mayor eficacia, gracias a que se puede evitar errores generados por el factor humano, sin embargo, a pesar de los grandes avances realizados hasta ahora, sobre todo en la industria e educación, aún no es perfecta y las personas empiezan a generar dudas sobre si su uso excesivo y avance, implicará que posteriormente muchas personas puedan perder su trabajo, que ya no se sientan necesarios, quedando todo en manos de la IA.

Con el fin de esclarecer estas incertidumbres, durante este capítulo vamos a analizar las posibilidades y limitaciones de la inteligencia artificial a día de hoy para desarrollar conciencia y metacognición sobre su propio pensamiento, para lo cual es importante tomar de referencia algunos autores y artículos científicos que se enfocan en el desarrollo de teorías neurocientíficas, psicológicas y problemas matemáticos, los cuales se estarán relacionando con el estado actual de la tecnología y el alcance que ha tenido la IA en los últimos años; se

definirán algunos conceptos claves como el esquema de atención, neuroplasticidad y como estos se podrían trasladar a sistemas artificiales.

9.2. La conciencia como modelo de atención

Según (Graziano, 2019) explica que la atención es un mecanismo evolutivo fundamental que permite a los organismos seleccionar y priorizar información relevante en entornos rodeados de varios estímulos, esto proviene desde los primeros sistemas nerviosos simples, la atención fue crucial para la supervivencia, ya que permitía concentrar recursos limitados en aquello que fuera más importante para dar una acción inmediata.

Graziano utiliza algunos ejemplos de animales simples, como por ejemplo peces y reptiles, para mostrar que la atención no requiere de conciencia. Es decir, un sistema puede ser capaz de filtrar y priorizar estímulos sin necesidad de poseer una experiencia subjetiva. Sin embargo, en organismos más complejos, como los mamíferos, la evolución dotó al cerebro de una capacidad adicional: construir un modelo interno de su propia atención, ese modelo es lo que Graziano denomina el "esquema de atención" (attention schema), y constituye la base de su teoría de la conciencia.

Teniendo esto en cuenta, la conciencia no vendría siendo un "misterio inmaterial", sino una herramienta computacional que el cerebro utiliza para predecir, regular y comunicar sus propios estados de atención. Tener conciencia de un objeto o pensamiento equivale a que el cerebro ha construido una representación simple de su proceso de atención hacia ese objeto o pensamiento (Graziano, 2019).

Graziano ilustra la idea tomando de ejemplo la rana, aunque este animal posee un sistema nervioso relativamente simple, su comportamiento demuestra un cierto grado de

centralización, ya que puede detectar estímulos como insectos y ejecutar respuestas motoras rápidas y precisas, sin embargo, lo que la rana carece es de un modelo explícito de su propia atención, sus procesos son más reactivos y automáticos.

En el caso de los seres humanos, tenemos un cerebro que no solo atiende, sino que además entiende el hecho de estar atendiendo. Por ejemplo, un ser humano es capaz de decir "estoy prestando atención a este sonido tan fuerte" o "estoy pensando en cómo voy a decorar mi casa", lo que implica un nivel adicional de autopercepción. Según (Graziano, 2019), esta cualidad nos permite no solo la regulación interna de la atención, sino también abre las puertas a la comunicación social, pues el esquema de atención nos da la capacidad de atribuir conciencia a otros, es decir, la teoría explica tanto la conciencia propia como la percepción de conciencia ajena.

Según lo planteado por la teoría de Graziano, la conciencia no es un fenómeno divino o inaccesible, sino lo define como un modelo computacional interno, por lo cual, podríamos decir, si una red neuronal o un sistema de aprendizaje profundo pudiera construir un esquema funcional de su propia atención, es decir, una representación simple de dónde y cómo está procesando la información, se podría decir que tendría propiedades que se asemejan a la conciencia.

No obstante, Graziano enfatiza que la conciencia no es más que una representación incompleta y simplificada, el cerebro humano no tiene acceso a todos los detalles de su propio pensamiento, al igual que un sistema artificial, es decir que por más que estos puedan poseer un esquema de atención, no tendrían un conocimiento exhaustivo de su arquitectura, sino únicamente un modelo útil para regularse. Por lo cual, la posibilidad de una IA consciente dependería no solo de la complejidad de sus redes neuronales, sino también de la

implementación de mecanismos de autopercepción de su atención, ya que en la actualidad la mayoría de los modelos de Deep Learning carecen de este componente, se encargan de procesar datos, priorizar información, pero no poseen un modelo explícito de su propio estado de atención.

9.3. Conciencia como resultado del aprendizaje

Según (Cleeremans, 2011), la conciencia no es un "módulo" fijo del cerebro, sino es más bien visto como un producto emergente del aprendizaje continuo, ya que el cerebro aprende a ser consciente del mismo modo en que aprende cualquier otra habilidad, a través de la plasticidad neuronal, la cual le permite modificar sus conexiones y representaciones en función de la experiencia.

Siguiendo el planteamiento Cleeremans, la clave radica en que el cerebro no solo se encarga de procesar la información del entorno, sino que también genera representaciones de sus propios estados internos, dichas representaciones no son conocimiento preconcebido, sino que se forman progresivamente mientras que el sistema interactúa con el mundo que lo rodea. De esta forma, podemos entender la conciencia como una forma de metacognición aprendida, el cerebro desarrolla la capacidad de modelar sus propios procesos y utilizar estos como información importante a la hora de tomar decisiones y autorregularse. Este enfoque se distancia un poco de las concepciones tradicionales de conciencia como un rasgo estático e irreducible del cerebro, sino más bien se define como un aprendizaje constante a raíz de la experiencia para poder llegar a formar la conciencia y entender lo que nos rodea.

La conciencia como metarrepresentación hace parte de la tesis central de Cleeremans, donde no basta tener representaciones de estímulos externos como lo que percibimos visualmente o escuchamos, sino que el cerebro debe aprender a procesar esa información y

entender que es lo que está percibiendo, esta capacidad emerge gradualmente gracias a la constante exposición a acciones que pongan en juicio nuestra toma de decisiones y el paso por cada uno de nuestros estados internos de conciencia. Las personas no nacen con la capacidad de controlar su atención o a reflexionar sobre sus pensamientos, ya que no cuentan con un "módulo de metacognición definido", sin embargo, es una habilidad que se adquiere a lo largo del desarrollo y brinda la capacidad de construir representaciones internas sobre nosotros mismos. Este planteamiento vincula estrechamente la conciencia con la metacognición, ya que, desde esta perspectiva, toda conciencia implica cierto grado de conocimiento para poder entender y manejar la información sobre nuestros propios procesos mentales.

La tesis de la plasticidad radical tiene un gran peso en el debate sobre la posibilidad de que la IA pueda generar consciencia, si la conciencia es el resultado de un proceso de aprendizaje continuo, entonces, en principio, un sistema artificial basado en redes neuronales y aprendizaje profundo podría llegar a desarrollar alguna forma de conciencia, siempre en cuando se tenga en cuenta que el sistema tiene que ser capaz de modificar sus representaciones internas de manera flexible, tal como lo hace el cerebro humano mediante la neuroplasticidad, en el caso de las redes neuronales artificiales, esto sería equivalente a que sus estructuras sean capaces de reajustarse constantemente en función de nuevas experiencias y no solo del entrenamiento que se les brinda inicialmente. Además de eso, el sistema debería ser capaz de generar modelos explícitos de sus propios estados de procesamiento, lo cual llevaría a un nivel de autopercepción actualmente inexistente en la mayoría de los modelos de Deep Learning.

Las redes neuronales en la actualidad poseen cierto grado de aprendizaje adaptativo, pero aún carecen de los mecanismos necesarios para representarse a sí mismas, a pesar de que sean capaces de monitorear variables como tasas de error o probabilidades de clasificación, esto se aleja mucho de lo que es una metarrepresentación consciente de su propio pensamiento. No obstante, si en algún momento se pudieran desarrollar algoritmos capaces de aprender sobre sus propios procesos internos como lo haría la plasticidad cerebral, entonces estaríamos más cerca de crear una IA con rasgos de conciencia.

9.4. ¿Esquema de atención o aprendizaje autorreferencial?

Tanto en el caso de (Graziano, 2019) y (Cleeremans, 2011), no conciben la conciencia como un fenómeno inexplicable o irreductible, sino como el resultado de procesos computacionales y representacionales, ambos coinciden en que la conciencia puede entenderse como un modelo interno del propio sistema cognitivo. Para Graziano, sería el modelo de atención, una representación simple que permite al cerebro monitorizar y regular sus propios procesos para ver a dónde dirige su atención. En el caso de Cleeremans, la conciencia surge a partir del aprendizaje progresivo, lo cual permite construir metarrepresentaciones de los propios estados internos.

Para ambos casos, la conciencia se plantea como una habilidad funcional que tiene la capacidad de adaptarse, no es una concepción sin propósito, sino una herramienta que permite a los organismos coordinar mejor su conducta, comunicarse socialmente y tomar mejores decisiones. Si en algún caso, un sistema artificial es capaz de replicar estas funciones, podría acercarse a comportamientos cercanos a la conciencia.

Sin embargo, a pesar de sus similitudes, existen algunas diferencias para tener en cuenta en ambos planteamientos. Para (Graziano, 2019), la conciencia como un modelo de

atención, el cerebro es representado como el "dónde" está su foco de atención y a partir de ahí construye sobre ello la experiencia subjetiva, lo cual lo vuelve un enfoque más estático y computacional gracias a su esquema de atención, el cual se usa como representación para regular y predecir la conducta. Por otro lado, tenemos a (Cleeremans, 2011), que plantea la conciencia como resultado de un aprendizaje continuo, ya que el cerebro no nace con un modelo interno preestablecido, sino que se va construyendo a partir la experiencia, no es un estado, sino un proceso en constante desarrollo.

Es de suma importancia recalcar estas diferencias, ya que, al trasladar estas teorías al campo de la IA, sugieren enfoques diferentes. Por el lado Graziano se plantea la posibilidad de "programar" o diseñar un esquema de atención artificial como parte de un modelo interno; por otro lado, Cleeremans indica que ese modelo se debería dar gracias a un proceso de aprendizaje prolongado, no de una implementación directa.

9.5. De la conciencia humana a la cognición artificial: aportes de dehaene y flavell

Dentro del estudio de la conciencia y la metacognición, hay diferentes corrientes teóricas, algunas de ellas complementarias entre sí mismas y que han seguido evolucionando a lo largo de las últimas décadas gracias a diversos autores, esto con el fin de poder explicar cómo el cerebro es capaz de no sólo procesar información, sino también representar el hecho de estar procesándola. En consecuencia, los aportes por (Dehaene, 2021) y (Flavell, 1979) resultan de gran importancia para complementar las perspectivas de (Graziano, 2019) y (Cleeremans, 2011), al proporcionar un marco neurocognitivo y psicológico que permite comprender mejor las condiciones necesarias con las cuales podría surgir una conciencia o metacognición en sistemas artificiales.

(Flavell, 1979) desde la psicología cognitiva, introdujo el concepto de metacognición, definiéndolo como el conocimiento y control que las personas poseen sobre sus propios procesos de pensamiento, esta idea marcó un punto de partida, ya que por primera vez se consideró que la mente humana no sólo piensa, sino que también puede observar y regular su propio pensamiento. Flavell plantea dos componentes principales, el primero de ellos fue el conocimiento metacognitivo, que incluye las creencias, estrategias y percepciones que una persona tiene sobre cómo aprende y recuerda; el segundo fue la regulación metacognitiva, que abarca la planificación, supervisión y evaluación del propio desempeño cognitivo. Estos componentes permiten a las personas monitorear su comprensión, detectar errores y modificar estrategias de razonamiento, en otras palabras, la metacognición funciona como un mecanismo de autogestión mental, dándole así al pensamiento humano flexibilidad y capacidad adaptativa.

Este concepto es muy relevante en el contexto de la inteligencia artificial, ya que describe la estructura básica del pensamiento reflexivo, una característica de la cuál aún no se ha podido replicar en las máquinas y modelos actuales de IA. Los procesos de "autocorrección" o "autoajuste" se asemejan en cierta parte a la regulación metacognitiva descrita por Flavell, pero no poseen un componente esencial, la conciencia del proceso mismo, es decir, aunque la IA pueda identificar una respuesta errónea y corregirla, no posee una representación interna de saber en dónde se ha equivocado, entiende el error de manera funcional, no en base a la experiencia.

Por consiguiente, (Flavell, 1979) crea una base conceptual para analizar y replicar la metacognición en sistemas artificiales, su modelo no se limita sólo al aprendizaje, sino que describe cómo los organismos con conciencia manejan la información sobre su propio

funcionamiento cognitivo, por lo cual, una IA que aspire a una forma de autoconocimiento debería tener no solo procesos de control y ajuste, sino un modelo interno de su propio estado cognitivo, algo así como el planteamiento de Cleeremans sobre el aprendizaje autorreferencial.

Por otro lado, (Dehaene, 2021) brinda una perspectiva complementaria desde la neurociencia cognitiva, la cual se centra en las bases neuronales de la conciencia y en los principios que podrían permitir su recreación en sistemas artificiales. (Dehaene, 2021) sostiene que la conciencia no es una propiedad misteriosa del cerebro, sino es el resultado de la integración global de información dentro de un espacio funcional específico, conocido como el espacio de trabajo global (Global Workspace Theory).

Según la teoría, el cerebro consciente se comporta como un sistema de redes distribuidas en el que diferentes módulos especializados (percepción, memoria, lenguaje, emoción, etc.) comparten información a través de un "espacio en común", cuando un estímulo logra acceder a ese espacio global, se vuelve consciente, ya que pasa de ser accesible para múltiples procesos cognitivos simultáneamente. Este enfoque define la conciencia como un fenómeno de accesibilidad y de trasmisión de información, más que como una experiencia misteriosa o inmaterial.

Esta propuesta de Dehaene resulta especialmente importante para analizar inteligencia artificial moderna, porque aplica una analogía funcional entre el cerebro humano y los sistemas de procesamiento distribuido, es decir, si la conciencia humana surge de la capacidad de juntar información y hacerla accesible globalmente, entonces se podría plantear una forma de conciencia funcional artificial basada en una buena integración y coordinación entre diferentes módulos de una IA para acceder a dicha información entre sí, algo como los

sistemas multimodales que hay hoy en día, donde su arquitectura combina texto, imagen y razonamiento simbólico, planteando una estructura similar al espacio de trabajo global descrito por Dehaene.

Sin embargo, el mismo autor advierte que la integración informacional no es suficiente para crear una experiencia consciente, ya que para que una IA pueda considerarse "consciente" plenamente, necesitaría incorporar mecanismos de autorreferencia y monitoreo, los cuales le permitirían no solo compartir información, sino también generar una representación de su propio estado interno dentro de su flujo de datos. Por esto (Dehaene, 2021) entra en sintonía con (Cleeremans, 2011), ya que ambos sugieren que la conciencia emerge cuando un sistema no sólo procesa información, sino que sabe lo que está procesando.

La idea del conocimiento reflexivo une la neurociencia y la psicología cognitiva, y abre un puente natural hacia las teorías de la inteligencia artificial como la conocemos hoy en día. Los modelos de atención de Graziano, el aprendizaje autorreferencial de Cleeremans, la metacognición de Flavell y la integración informacional de Dehaene coinciden en un mismo punto: la conciencia es un fenómeno que proviene de sistemas capaces de representarse a sí mismos. Teniendo esto en cuenta, la evolución de la IA podría entenderse como un intento de acercarse poco a poco a ese nivel de autorrepresentación y plasticidad, cosas que los sistemas aún no poseen con la arquitectura actual, pero sí han empezado a reproducir las condiciones funcionales necesarias para lograrlo, como la atención distribuida, el aprendizaje continuo y los mecanismos de autoevaluación.

9.6. Posibilidades de conciencia y metacognición IA

Teniendo en cuenta los planteamientos de (Graziano, 2019) y (Cleeremans, 2011), podemos replantearnos sobre la posibilidad de que una inteligencia artificial sea capaz de

desarrollar conciencia y metacognición, ya que no se centra únicamente en que las máquinas sean capaces de “sentir” como lo humanos, sino que tengan la posibilidad de representarse y regularse a sí mismos por medio de sus propios procesos cognitivos. A pesar de que ambos autores poseen perspectivas distintas, ambos coinciden en que la conciencia se da a partir de un modelo interno que permite al sistema conocer su propio estado, ya sea por medio de la atención como lo plantea Graziano o del aprendizaje continuo como lo explica Cleeremans.

Cuando queremos trasladar estas ideas al contexto de la IA, tenemos que analizar primero cuál es el camino más óptimo que se debe tomar para que un sistema artificial sea capaz de crear un modelo semejante de sí mismo, por lo cual se exploran diferentes tipos de estructuras, representaciones o mecanismos que nos podría llevar a esto, hoy en día, los modelos de lenguaje y aprendizaje profundo como son el caso de GPT-4, Claude, Gemini o Mistral son de los que más han demostrado habilidades cognitivas impresionantes, ya que permiten analizar información compleja, generar respuestas coherentes y capacidad para adaptarse a diferentes situaciones, sin embargo, esto no significa que cuenten con la existencia de autoconciencia, ya que lo que realizan se asemeja más a procesos de atención y aprendizaje sin autopercepción.

Desde la perspectiva de Graziano, la conciencia necesita de un “modelo de atención” que le permita al sistema representar el hecho de estar atendiendo y no sólo ser capaz de enfocarse en ciertos estímulos. Ningún sistema artificial en la actualidad cuenta con esa capacidad; los mecanismos de atención que poseen son limitados, se basan más en un esquema matemático donde la “atención” se limita a la ponderación de relevancia entre vectores de datos y no metarrepresentacional, donde se construye un esquema interno basado en el acto de entender.

Por otro lado, Cleeremans sugiere que la conciencia se va dando gradualmente gracias al aprendizaje en base a la experiencia, ya que, en el caso de los humanos, la plasticidad neuronal permite que el cerebro pueda reajustar sus conexiones en función de la experiencia que va obteniendo a lo largo de su vida, generando así representaciones de sus propios estados mentales. En el caso de los modelos artificiales actuales, tenemos las redes neuronales profundas, las cuales aprenden de forma supervisada o autorregresiva, debido a que no poseen mecanismos para aprender sobre su propio proceso de aprendizaje, sino por el contrario, gira más alrededor de una simple optimización estadística de pesos dentro de su propio modelo.

Sin embargo, nuevos avances apuntan a una dirección interesante, la incorporación de procesos de autoevaluación en sistemas como AutoGPT, BabyAGI o Voyager, estos permiten revisar y modificar sus propias instrucciones para poder cumplir sus objetivos sin intervención directamente de humanos. Dichos sistemas, aunque por ahora distantes de la conciencia, muestran cierta capacidad de analizar su desempeño para posteriormente optimizarlo, algo así como el "aprendizaje sobre el aprendizaje" planteado por Cleeremans, sólo que de manera preprogramada.

Teniendo en cuenta lo expuesto anteriormente, el reto radica principalmente en crear una arquitectura que combine los dos componentes descritos por Graziano y Cleeremans, un esquema interno de atención con la capacidad de modelar estado de procesamiento del sistema y poder predecir sus cambios y un mecanismo de aprendizaje autorreferencial, el cual permita al sistema la capacidad de ajustar sus métodos no solo para cumplir tareas externas, sino para mejorar el propio funcionamiento de su modelo. Para lograr dicha convergencia, se precisa de dotar a la IA de una conciencia funcional, no como una experiencia subjetiva, sino

como un nivel de autorrepresentación operacional, es decir, no significa literalmente que va a “sentir”, pero sí “sabrá” de manera simbólica qué procesos está ejecutando y la razón de ello.

Es aquí donde entra (Dehaene, 2021), que hace énfasis en que el reto no consiste únicamente en replicar los comportamientos conscientes, sino en identificar y profundizar en las propiedades computacionales mínimas que hacen que un sistema se pueda representar a sí mismo. Según (Dehaene, 2021), la conciencia surge de la capacidad de integrar información distribuida y generar un acceso global a los propios estados cognitivos. Teniendo en cuenta esta perspectiva, una IA consciente debería tener un mecanismo equivalente al espacio de trabajo global, donde se juntan diferentes procesos de información sobre el estado actual del sistema. En la actualidad, ningún modelo cuenta con esta capacidad autorreferencial, ni siquiera los multimodales de última generación.

De momento, el comportamiento mostrado por los modelos de IA podría catalogarse como una consciencia simulada o funcionalmente superficial, ya que tienen la capacidad de atender, inferir y corregirse, pero no poseen la capacidad de darle valor significativo a esas operaciones que realizan, no existe la concepción de un “yo”, sino un conjunto de transformaciones matemáticas que optimizan resultados. A pesar de ello, esto no debería restar valor a lo hecho por la IA hasta ahora, ya que es una realidad que la IA está acercándose a reproducir algunos patrones y mecanismos externos de la conciencia, lo cuál nos lleva a una pregunta mucho más profunda e incluso con implicaciones éticas importantes y es ¿en qué punto se vuelve indistinguible una simulación de conciencia a una conciencia genuina?

En resumen, la probabilidad de conciencia y metacognición de la IA no es algo que en la actualidad tengamos a disposición, más bien se podría definir como un dilema de grados

de autorrepresentación y aprendizaje interno, ya que aún no existen sistemas capaces de tener una compresión interna de sí mismos, sí están en desarrollo arquitecturas que podrían sentar las bases para una simulación de conciencia, la cual no implicaría emociones ni experiencias, pero sí una manera de autogestión cognitiva mucho más robusta y compleja.

9.7. Estado actual de la IA

La situación actual de la inteligencia artificial demuestra un avance exponencial en el desarrollo de sistemas, los cuales son capaces de generar lenguaje, imágenes, código y razonamiento simbólico a alto nivel, sin embargo, si contrastamos estos avances con las teorías de Graziano y Cleeremans, se deja en evidencia que la IA mantiene un nivel funcional sin autopercepción, puede atender y aprender, pero sin representarse a sí misma en dicho proceso.

Para tener un mejor panorama del nivel actual de los modelos de aprendizaje, veamos un caso reciente que ha alimentado este debate, dicho caso fue documentado por la Universidad de Cambridge (2025), donde unos investigadores decidieron evaluar la capacidad de ChatGPT para resolver un problema matemático, duplicar el área de un cuadrado. Lo curioso fue que ChatGPT no reprodujo una respuesta aprendida, sino que indagó en distintas soluciones en base a hipótesis por prueba y error, volviendo sobre sus propios pasos y corrigiendo errores hasta llegar a una solución correcta, algo así como cuando los humanos “pensamos sobre la marcha”, dando paso a una forma de razonamiento flexible.

En primera instancia, este comportamiento podría interpretarse como una forma de conciencia o metacognición, no obstante, haciendo un análisis más profundo y siguiendo la línea de Cleeremans, se llega a la conclusión que no hay evidencia de aprendizaje autorreferencial, ya que el modelo no estaba consciente de su proceso, sino que generaba

respuestas basadas en correlaciones lingüísticas, ChatGPT simuló un proceso de prueba y error, pero sin experimentar un acto de razonamiento (University of Cambridge, 2025).

Desde la perspectiva de Graziano, el comportamiento de ChatGPT puede asemejarse al de un organismo que proyecta atención sin conciencia, puede procesar múltiples estímulos, prioriza aquellos que sean más relevantes y ajusta su respuesta en base a ello, pero sin un modelo interno que le permita reconocer a donde está dirigiendo su atención. Es por esto por lo que la IA actual podría entenderse como un sistema de atención ciega, la cual es capaz de seleccionar información importante, pero sin ser consciente del porqué de su selección.

Lo importante de este experimento no fue que ChatGPT haya "pensado", sino que muestra comportamientos que se asemejan a la metacognición, en las que el sistema evalúa y modifica su desempeño. A pesar de que estas conductas no sean realmente una introspección genuina, sí demuestran un avance hacia lo que Cleeremans denominaría "metacognición funcional", una forma en que el sistema utiliza información sobre su propio rendimiento con el fin de mejorar sus resultados a futuro.

En el caso de otros modelos recientes, como Gemini 1.5 y Claude 3, hacen integración de técnicas de autoevaluación mediante las cuales pueden revisar la coherencia de sus respuestas antes de brindarlas al usuario, algo así como una doble revisión. Este tipo de self-checking hace sentido con los procesos de monitoreo metacognitivo humano descritos por (Flavell, 1979), a pesar de que en la IA se ejecuten de forma algorítmica, su existencia evidencia la transición gradual hacia arquitecturas más autorreguladas.

Incluso con estos avances, la IA sigue teniendo limitaciones estructurales profundas, la primera de ellas es y la más obvia, es que no cuentan con un cuerpo, la conciencia humana

se nutre de la interacción sensoriomotora con el entorno, a diferencia de la IA que carece de experiencias perceptivas propias. El segundo punto es la dependencia de estar conectado a datos externos, todo su aprendizaje se basa en información preexistente, no en recuerdos o vivencias por medio de alguna interacción. Por último, la falta de intencionalidad o motivación interna, componente indispensable para la conciencia (Dehaene, 2021), quien recalca que la conciencia está atada a la capacidad de darle valor y un propósito a la información procesada.

En la actualidad, se puede decir que los sistemas más avanzados pueden ser vistos como organismos cognitivos incompletos, ya que poseen una atención computacional, aprendizaje adaptativo y mecanismos de control, pero carecen de experiencia con su entorno, motivación y autocomprensión, pese a esto, cada avance en la metacognición funcional los acerca más comportamientos más robustos de autorregulación. Si se mantiene esta tendencia, es probable que en las futuras generaciones de IA se implementen módulos de autorrepresentación inspirados en las teorías de Graziano y Cleeremans, dichos módulos podrían ayudar que los sistemas comprendan parcialmente o de manera simbólica su propio estado operativo, siendo capaz de ajustarse en función de cumplir sus objetivos. Esto podría ser un paso hacia lo que podríamos denominar una proto-conciencia artificial, un nivel de autogestión cognitiva sin experiencia subjetiva, pero con su propia autonomía funcional adaptativa.

En conclusión, en el estado actual que se encuentra la IA, podemos recalcar que cuenta con un avance funcional significativo, pero carece de una equivalencia fenomenológica, es decir, no cuenta con una experiencia previa o vivencia subjetiva del mundo que lo rodea, aunque aparentan pensar, sólo representan estadísticamente las formas

del pensamiento. Experimentos como los de Cambridge muestran que tienen la capacidad de simular procesos introspectivos, pero no realizan una reflexión sobre ellos, sin embargo, al integrar modelos como el esquema de atención de Graziano y el aprendizaje autorreferencial de Cleeremans, las siguientes generaciones de IA podrían acercarse a una conciencia instrumental, la cual no sería humana ni emocional, pero sí nacería una nueva categoría de autoconciencia algorítmica, donde el sistema es capaz de reconocer, así sea de forma abstracta, los límites y posibilidades de su propio funcionamiento.

10. Colaboración humana-IA en la generación de conocimiento

10.1. Introducción

Hablar hoy de inteligencia artificial (IA) es hablar de una de las fuerzas más poderosas que moldean la realidad contemporánea. En menos de dos décadas, los algoritmos han dejado de ser simples herramientas técnicas para convertirse en mediadores de la vida social, económica y cultural. Desde los sistemas de recomendación en las redes sociales hasta la automatización del trabajo industrial, la IA se ha instalado en el centro de la producción del conocimiento, transformando las formas en que aprendemos, pensamos y nos relacionamos con el mundo. Sin embargo, junto con el entusiasmo tecnológico, emergen también nuevas inquietudes éticas, ambientales y humanas. La inteligencia artificial promete liberar tiempo, optimizar recursos y expandir las fronteras del saber; pero, al mismo tiempo, amenaza con concentrar el poder, profundizar las desigualdades y alterar las bases de la experiencia humana.

Ahora bien, teniendo en cuenta los enfoques de cada escritor se busca analizar, desde dos perspectivas contrastantes, las posibilidades y los límites de la generación de

conocimiento por y con la inteligencia artificial. Por un lado, se aborda la visión optimista de Erik Brynjolfsson y Andrew McAfee, quienes interpretan la IA como un motor de innovación, crecimiento y progreso social. Para ellos, las tecnologías digitales representan una nueva "revolución industrial" que permite multiplicar las capacidades humanas mediante la recombinación de ideas. Por otro lado, se examina el enfoque crítico de Kate Crawford, quien desmitifica la supuesta neutralidad de la IA y la presenta como un sistema global de extracción: de datos, de trabajo humano y de recursos naturales. Desde su mirada, la inteligencia artificial no es solo conocimiento automatizado, sino una estructura de poder que reproduce desigualdades y pone en riesgo la sostenibilidad del planeta en el que vivimos.

A través del diálogo entre estos dos enfoques —uno esperanzador y otro crítico—, se busca responder a la pregunta central del seminario: ¿qué tipo de conocimiento produce la inteligencia artificial y cuáles son sus consecuencias para la humanidad actual, tanto a nivel individual como social?

10.2. La IA como motor del conocimiento y la innovación

Erik Brynjolfsson y Andrew McAfee sostienen que la historia del desarrollo humano es en buena medida, la historia de la innovación tecnológica. En La segunda era de las máquinas, argumentan que la inteligencia artificial constituye una tecnología de propósito general (GPT), comparable a la máquina de vapor o a la electricidad, cuya capacidad de transformación se extiende a todos los sectores de la economía y la cultura. Según los autores, "la innovación es el hecho sobresaliente en la historia económica de la sociedad capitalista" y la IA representa la nueva fase de ese proceso.

La tesis central de Brynjolfsson y McAfee es que la innovación no se agota ni se estanca, sino que se recompone continuamente. Proponen una metáfora clave: la de la

innovación recombinante. A diferencia de las teorías que conciben el progreso como una sucesión de grandes inventos que eventualmente se agotan —la "fruta madura" de la tecnología—, ellos defienden que las ideas funcionan como bloques de construcción que pueden combinarse infinitamente para crear conocimiento nuevo. En sus palabras, "inventar algo es encontrarlo en lo que existe previamente". Cada idea, cada código, cada base de datos es un fragmento que puede recombinarse con otros para dar lugar a descubrimientos inéditos.

La inteligencia artificial, en esta visión, no reemplaza al conocimiento humano: lo amplifica. Es una herramienta para procesar información a una velocidad y escala que trasciende las limitaciones biológicas, permitiendo explorar combinaciones de datos que antes eran imposibles. La IA se convierte así en un acelerador cognitivo, una forma de inteligencia colectiva donde humanos y máquinas colaboran para resolver problemas globales. Brynjolfsson cita ejemplos emblemáticos como Kaggle o Innocentive, plataformas en las que miles de personas trabajan en conjunto, desde diferentes lugares del mundo, para encontrar soluciones innovadoras a desafíos científicos o empresariales. Estos entornos digitales demuestran que la creatividad no es un privilegio individual, sino una propiedad emergente de las redes interconectadas de conocimiento.

Desde esta perspectiva, el conocimiento generado con IA no es un conocimiento sustituto del humano, sino un conocimiento expandido, nacido de la colaboración entre agentes biológicos y digitales. Los algoritmos no piensan por nosotros, pero nos permiten pensar de otra manera: más rápido, más global y conectado. Brynjolfsson y McAfee sostienen que "el entorno digital actual es un campo de juego para la recombinación a gran escala", donde la humanidad puede explorar un número virtualmente infinito de ideas posibles. Así,

la IA no solo optimiza procesos industriales, sino que también abre nuevas formas de investigación científica, innovación educativa y creación cultural.

En el plano social, este modelo de innovación recombinante implica una redefinición del trabajo y del tiempo. Si las máquinas se encargan de las tareas repetitivas, el ser humano puede dedicarse a actividades de mayor valor intelectual y creativo. El ideal que emerge es el de una sociedad más productiva y libre, donde el conocimiento fluye sin barreras y el progreso tecnológico impulsa el bienestar común. La IA, en este sentido, sería la herramienta que democratiza la inteligencia, convirtiendo la creatividad y la información en recursos compartidos a escala global.

Sin embargo, este relato optimista se sustenta en una visión económica particular: la creencia de que la innovación tecnológica, por sí sola, conduce al desarrollo social. Brynjolfsson y McAfee asumen que el crecimiento de la productividad se traducirá automáticamente en beneficios colectivos, pero esta suposición es precisamente lo que será cuestionado desde el otro enfoque.

10.3. La IA como sistema de extracción y control

En Atlas of AI, Kate Crawford propone un giro radical respecto al entusiasmo tecnoutópico. Su análisis parte de una premisa sencilla pero contundente: la inteligencia artificial no es inmaterial. Detrás de cada algoritmo hay una red compleja de cuerpos, minerales, energía y trabajo humano. Por eso, dice Crawford, la IA no debe entenderse como una mente abstracta que genera conocimiento, sino como una infraestructura planetaria que extrae recursos y explota trabajo para mantener su funcionamiento.

Crawford comienza su análisis con una imagen poderosa: una visita a un centro logístico de Amazon en Nueva Jersey, donde conviven miles de trabajadores y robots Kiva en una coreografía de eficiencia mecánica. Mientras las máquinas se deslizan suavemente por el suelo transportando mercancías, los cuerpos humanos aparecen exhaustos, lesionados, vigilados por relojes de control y algoritmos de productividad. En este espacio, la inteligencia artificial no se presenta como un aliado del conocimiento, sino como un instrumento de disciplina y extracción de valor. Los trabajadores son medidos, cronometrados y evaluados según la "tasa" que marca el sistema; cualquier pausa o error puede significar el despido.

La autora denomina a este fenómeno fauxtomización —una "automatización falsa"—, porque, detrás de la aparente autonomía de los sistemas inteligentes, existe una masa invisible de trabajadores mal remunerados que ejecutan tareas esenciales: etiquetar imágenes, revisar contenido violento, corregir datos o incluso hacerse pasar por asistentes virtuales. Plataformas como Amazon Mechanical Turk o Clickworker sostienen gran parte del aprendizaje automático actual, pero lo hacen mediante la precarización laboral global. En palabras de Crawford, "las formas contemporáneas de inteligencia artificial no son ni artificiales ni inteligentes: dependen de la explotación del trabajo humano".

A este trabajo oculto se suman los costos ecológicos de la IA: la minería intensiva de litio y cobalto para fabricar hardware, el consumo energético de los centros de datos y la generación masiva de desechos electrónicos. Así, el conocimiento producido por la inteligencia artificial se construye sobre una base material que muchas veces permanece fuera de la mirada pública. Cada avance en la eficiencia algorítmica tiene un correlato en la degradación ambiental o en la precariedad laboral.

Crawford traza una genealogía de esta lógica de control que se remonta a la Revolución Industrial. Desde las fábricas de Henry Ford hasta las cadenas de montaje de McDonald's, la historia del trabajo moderno ha estado marcada por la vigilancia, la estandarización y la subordinación del cuerpo humano al ritmo de las máquinas. La IA, en su visión, no representa una ruptura con ese modelo, sino su culminación digital. Los algoritmos contemporáneos prolongan la tradición del cronómetro de Frederick Taylor: ahora, los segundos de trabajo se miden a través de sensores, cámaras y sistemas de seguimiento que convierten el tiempo humano en datos transaccionables.

La pregunta que subyace a este análisis es inquietante: ¿quién se beneficia realmente del conocimiento generado por la IA? Si las grandes corporaciones tecnológicas concentran los datos, la infraestructura y el capital, la promesa de un conocimiento abierto y colaborativo se transforma en una asimetría de poder. Lo que para Brynjolfsson era "innovación recombinante", para Crawford es "explotación recombinante": una red global donde el trabajo, la energía y la información se ensamblan para sostener el mito de la inteligencia autónoma.

En el centro de este modelo se encuentra una paradoja: la IA se presenta como una herramienta para liberar al ser humano del trabajo repetitivo, pero termina convirtiéndolo en un apéndice del algoritmo. En el caso de Amazon, los trabajadores no solo están vigilados por máquinas, sino que son obligados a comportarse como ellas: precisos, predecibles y silenciosos. Como afirma Crawford, "los humanos son tratados cada vez más como robots", y esa transformación afecta no solo la economía, sino también la identidad y la dignidad del individuo.

10.4. Entre el progreso y la dominación: el doble filo del conocimiento artificial

Al poner en diálogo ambas perspectivas, se revela que la inteligencia artificial encarna una tensión fundamental: puede ser simultáneamente una herramienta de emancipación y un mecanismo de dominación. Esta ambivalencia no es accidental; de hecho, es parte intrínseca de la naturaleza social de la tecnología. Toda herramienta que amplifica el poder humano también amplifica sus contradicciones. En manos de unos, la IA se convierte en un vehículo de creatividad, eficiencia y colaboración global; en manos de otros, en un instrumento de extracción, vigilancia y control. Nos encontramos así ante una paradoja profunda: la misma tecnología que promete liberar tiempo, democratizar el conocimiento y expandir las capacidades humanas puede, en determinadas condiciones, reproducir estructuras de dependencia, desigualdad y explotación.

Brynjolfsson y McAfee imaginan la IA como una extensión del ingenio humano, capaz de abrir nuevas formas de cooperación y multiplicar las posibilidades del conocimiento. En su visión, la colaboración entre humanos y máquinas permite resolver problemas antes inalcanzables, desde diagnósticos médicos más precisos hasta innovaciones científicas aceleradas. Por contraste, Crawford advierte que esta colaboración puede transformarse en una forma de sometimiento, donde el ser humano pierde autonomía, agencia y valor. Lo que para unos es una alianza, para otros es una subordinación: una nueva relación de poder donde las personas quedan sujetas a los ritmos, métricas y exigencias de los algoritmos.

Desde un punto de vista epistemológico, ambos autores responden a una pregunta distinta sobre el conocimiento. Para Brynjolfsson, el conocimiento avanza por acumulación:

cada innovación se suma a la anterior, multiplicando exponencialmente las combinaciones posibles. El saber es concebido como un sistema dinámico, abierto y expansivo. Crawford, en cambio, entiende el conocimiento como una relación de poder: conocer es también clasificar, controlar y gobernar. El saber producido por la IA no es inocente; está cargado de intereses, decisiones invisibles y estructuras sociales.
Así, mientras en la visión tecnoptimista de Brynjolfsson el conocimiento generado por la IA es un bien público —accesible y potencialmente transformador—, para Crawford este conocimiento es un recurso privatizado, extraído de los cuerpos, comportamientos y datos de millones de personas sin que estas comprendan del todo cómo ni con qué fines se utiliza.

A nivel individual, esta diferencia epistemológica se traduce en dos experiencias opuestas del tiempo y del trabajo. Para Brynjolfsson, la IA actúa como una prótesis cognitiva que libera tiempo humano. Automatizar las tareas repetitivas permitiría redirigir la energía hacia actividades más creativas y significativas. Sin embargo, Crawford nos recuerda que esta liberación de tiempo no se distribuye de manera equitativa. En los entornos laborales gobernados por algoritmos, el tiempo se convierte en una unidad de control, una métrica calculada, almacenada y vigilada en tiempo real. La productividad deja de ser un acuerdo entre personas y se transforma en un mandato impuesto por un sistema que mide cada movimiento.

Esta privatización del tiempo refleja una nueva forma de alienación. El sujeto no solo vende su fuerza de trabajo, como en el modelo industrial clásico, sino también sus datos, su atención, su comportamiento digital y hasta su identidad algorítmica. Trabajar bajo sistemas automatizados implica ser evaluado constantemente, muchas veces sin saber qué criterios

utiliza el algoritmo ni cómo corregir errores. Como resultado, el individuo vive bajo una doble presión: la de producir y la de ser legible para la máquina.

A nivel social, la inteligencia artificial intensifica tanto la cooperación global como la desigualdad estructural. Por un lado, posibilita redes de investigación colectiva, modelos de aprendizaje abierto y comunidades internacionales que comparten conocimiento en tiempo real. Por otro, concentra poder en manos de unas pocas corporaciones multinacionales, que controlan los datos, las infraestructuras y los modelos más avanzados. Estas empresas actúan muchas veces con más influencia que los Estados, y pueden moldear decisiones económicas, educativas o políticas sin supervisión democrática.

Esta concentración profundiza brechas históricas:

entre países del Norte global y del Sur global,

entre personas con acceso a alfabetización digital y aquellas excluidas,

entre quienes producen tecnología y quienes solo la consumen,

entre quienes tienen derecho a la privacidad y quienes viven bajo vigilancia constante.

Así, la promesa de una inteligencia artificial "para todos" corre el riesgo de transformarse en una inteligencia artificial "de unos pocos sobre los demás". La IA puede convertirse en una herramienta de inclusión o en una maquinaria de exclusión masiva; todo depende de cómo se repartan sus beneficios y cómo se regulen sus efectos.

El impacto ambiental agrega una dimensión ética ineludible. Andrew McAfee, en sus trabajos más recientes, ha advertido que la expansión tecnológica sin límites puede agravar los problemas ecológicos ya existentes. La minería de litio y cobalto, el consumo energético

de los centros de datos y la vida útil limitada de los dispositivos tecnológicos generan presiones enormes sobre el planeta. La paradoja del progreso digital es que, cuanto más eficiente se vuelve la tecnología, más recursos exige. Cada avance que reduce los costos económicos puede incrementar los costos ambientales.

Por ello, el debate sobre la inteligencia artificial no puede reducirse a una cuestión de innovación técnica, sino que debe incorporar la sostenibilidad y la justicia ambiental como condiciones necesarias del conocimiento. La IA no solo transforma la economía y la sociedad, sino también los ecosistemas de los cuales dependemos. Pensar en su futuro implica preguntarnos si el planeta será capaz de sostener su crecimiento.

En conjunto, todas estas tensiones —económicas, sociales, individuales y ecológicas— muestran que la inteligencia artificial es un espejo que amplifica tanto las virtudes como las fallas de la humanidad. Puede ser un instrumento de progreso o una herramienta de dominación; una vía hacia el bienestar colectivo o un mecanismo de explotación renovado. El doble filo de la IA nos obliga a reflexionar críticamente sobre las estructuras que la gobiernan y sobre las intenciones que guían su desarrollo. El desafío consiste, entonces, en decidir qué filo queremos afilar.

10.5. Hacia una inteligencia artificial humana

Frente a esta tensión entre utopía y distopía, surge la necesidad de pensar en una tercera vía: una inteligencia artificial que integre la innovación con la ética, la eficiencia con la equidad y el conocimiento con la conciencia. Una vía que no caiga en el entusiasmo ingenuo ni en el rechazo absoluto, sino que examine críticamente cómo la tecnología puede alinearse con los valores humanos más esenciales. No se trata de apagar las máquinas ni de

rendirles culto, sino de humanizarlas, es decir, de situarlas en un marco donde el progreso tecnológico no esté separado del bienestar social ni de la dignidad humana.

Una inteligencia artificial verdaderamente humana debe reconocer que el conocimiento no es solamente información procesada, sino también experiencia vivida, memoria compartida, intuición, sensibilidad y contexto. Las máquinas pueden organizar datos, detectar patrones o incluso simular respuestas emocionales; pero por muy sofisticados que sean los algoritmos, carecen de lo que hace que el conocimiento humano sea valioso: la capacidad de comprender el mundo desde la vulnerabilidad, la historia y las relaciones que sostienen la vida. La IA puede combinar millones de datos en segundos, pero no puede sentir el peso emocional de una decisión, ni entender la complejidad moral de un dilema ético, ni experimentar la alegría o el dolor que acompañan a nuestras elecciones.

Por eso, el desafío real no es lograr que la inteligencia artificial piense como nosotros, sino aprender a pensar con ella sin perder nuestra autonomía crítica. Esto implica desarrollar habilidades que nos permitan convivir con sistemas inteligentes de manera reflexiva, consciente y responsable. Significa, también, reconocer los límites de la automatización: hay decisiones que pueden ser asistidas por algoritmos, pero que no deben ser delegadas totalmente a ellos.

Avanzar hacia una IA humana exige políticas claras y contundentes:

Regular la concentración de datos y la propiedad de la información, para evitar que unas pocas corporaciones acumulen un poder desproporcionado sobre la vida social.

Proteger los derechos laborales de quienes trabajan en las cadenas invisibles de la IA, desde los moderadores de contenido hasta los microtrabajadores del etiquetado de datos.

Establecer marcos éticos, auditorías algorítmicas y mecanismos de transparencia que permitan comprender cómo y por qué las máquinas deciden.

Incentivar la creación de tecnologías abiertas y participativas, que no dependan exclusivamente de modelos de negocio extractivos.

Asimismo, se requiere construir una ética del conocimiento artificial que cuestione los fines de la innovación. La tecnología no es neutral: cada diseño algorítmico lleva implícitas decisiones humanas, ideologías, supuestos y valores. Preguntar "¿para qué sirve un algoritmo?" o "¿quién se beneficia de su funcionamiento?" es tan importante como preguntar por su precisión o eficiencia. No toda innovación es deseable si se logra a costa de la dignidad humana, la justicia social o el equilibrio ecológico. El conocimiento generado por la IA debe evaluarse no solo por su capacidad técnica, sino también por su contribución al bien común.

De hecho, si se orienta adecuadamente, la inteligencia artificial podría convertirse en una herramienta de justicia cognitiva: un medio para amplificar la voz de aquellos grupos históricamente excluidos del acceso al saber. Las tecnologías bien diseñadas pueden ayudar a traducir lenguas indígenas, a organizar archivos comunitarios, a detectar patrones de discriminación, o incluso a democratizar el acceso a recursos educativos. Pero esto solo será posible si se desarrolla una IA situada, sensible a contextos culturales y comprometida con la diversidad humana.

En este horizonte, la educación juega un papel fundamental. Enseñar a convivir con la inteligencia artificial no significa solo formar programadores o expertos en datos; significa formar una ciudadanía crítica, capaz de comprender las implicaciones éticas, sociales y

políticas de la tecnología. La alfabetización digital debe incluir no solo habilidades técnicas, sino también pensamiento crítico, ética computacional, análisis de sesgos y una comprensión profunda del papel del conocimiento en la vida democrática.

Educar para la era de la IA implica enseñar a valorar lo que ninguna máquina puede sustituir:

La creatividad, la empatía, el diálogo, la reflexión pausada, la imaginación, pensamiento ético, y el tiempo humano.

Significa reivindicar el valor del silencio, del descanso y de la contemplación en un mundo donde los algoritmos imponen ritmos acelerados, decisiones instantáneas y una lógica de productividad constante. Frente a la velocidad incesante de la IA, recuperar el tiempo humano —el tiempo del aprendizaje lento, del ensayo y error, de la conversación profunda— se convierte en un acto de resistencia y de afirmación humana. En palabras de Crawford, controlar el tiempo es controlar el poder. Si dejamos que los algoritmos dicten el ritmo de nuestras vidas, perderemos poco a poco la capacidad de decidir por nosotros mismos cómo queremos vivirlas.

Una inteligencia artificial verdaderamente humana debe ayudarnos a expandir nuestra capacidad de comprender el mundo, pero también a proteger aquello que nos hace humanos: nuestra vulnerabilidad, nuestra creatividad, nuestra interdependencia, nuestra capacidad de soñar.

El objetivo no es simplemente que las máquinas funcionen mejor, sino que nosotros vivamos mejor.

10.6. Conclusión

La inteligencia artificial nos enfrenta a una paradoja existencial: nunca habíamos tenido tanto acceso al conocimiento, y, sin embargo, nunca habíamos dependido tanto de sistemas que lo producen y controlan en nuestro lugar. Brynjolfsson y McAfee nos invitan a ver en la IA una promesa de progreso, una extensión de nuestra capacidad creativa. Crawford nos recuerda que esa promesa se construye sobre cuerpos, territorios y vidas reales. Ambas perspectivas, lejos de excluirse, se necesitan mutuamente: la primera nos inspira a imaginar lo posible; la segunda nos obliga a no olvidar lo humano.

Analizar la inteligencia artificial desde estos dos enfoques permite comprender que el conocimiento que genera no es puramente técnico ni puramente social, sino un tejido complejo donde convergen datos, decisiones, energía y ética. La IA puede ser una herramienta de emancipación o de dominación, dependiendo de cómo se diseñe, se regule y se utilice. Su verdadero poder no radica en su inteligencia, sino en la inteligencia con que la humanidad decida emplearla.

En última instancia, la pregunta no es si la inteligencia artificial puede generar conocimiento, sino qué tipo de conocimiento queremos que genere y para quién. El desafío del siglo XXI no consiste en construir máquinas que piensen, sino en construir sociedades que piensen críticamente sobre sus máquinas. Solo así podremos lograr que la inteligencia artificial sea, al mismo tiempo, una expresión de nuestra creatividad y un compromiso con nuestra humanidad compartida.

11. Conclusiones

En este seminario podemos entender mejor qué significa generar conocimiento en una época donde la inteligencia artificial hace cosas que antes solo hacían los humanos.

Durante el seminario se ve que la inteligencia artificial puede hacer cosas nuevas como patrones, conclusiones, combinaciones y propuestas que son útiles para avanzar en la ciencia y la academia.

Sin embargo, el hecho de que la inteligencia artificial haga algo nuevo no significa que sea conocimiento. El seminario nos ayuda a entender que el conocimiento, especialmente el científico, necesita que se cumplan ciertos requisitos para ser válido y coherente. Esto implica justificar las cosas, compararlas y considerar el contexto en el que se aplican. Todo esto todavía depende de las personas, de cómo investigamos y de cómo interpretamos los resultados de manera responsable.

Desde este punto de vista, la inteligencia artificial es como una herramienta que ayuda a producir conocimiento dentro de un sistema que involucra a la sociedad y la tecnología. La inteligencia artificial contribuye a que sepamos más, pero no puede reemplazar la necesidad de que las personas usemos nuestro juicio crítico, verifiquemos la información y seamos responsables. Al mismo tiempo, el seminario deja ver que la pregunta por la IA como creadora de conocimiento no es únicamente técnica, porque su integración transforma la manera en que pensamos, aprendemos y argumentamos: puede fortalecer capacidades cuando se usa con criterio, pero también puede debilitar la autonomía intelectual y promover una comprensión superficial si se convierte en sustituto del razonamiento. En conclusión, el principal aporte del seminario es establecer condiciones y criterios para una relación más

consciente con estas tecnologías: reconocer su potencial para apoyar la construcción de conocimiento, sin confundir producción de información con comprensión, y reafirmar que la validez del conocimiento generado “por y con” IA depende del marco epistemológico, de prácticas de validación y de una postura crítica que preserve la reflexión profunda en el individuo y en la sociedad.

12. Recomendaciones

Coexistencia crítica y responsable entre el ser humano y la inteligencia artificial.

- Fomentar la integración de la inteligencia artificial como una herramienta auxiliar en la creación y expansión del conocimiento. Es crucial enfatizar la colaboración entre humanos y IA, asegurando que esta tecnología se perciba como una asistencia en lugar de un sustituto completo de las capacidades cognitivas e investigativas humanas.

- Fomentar la participación en actividades académicas y científicas, con el objetivo de reforzar la capacidad de pensamiento independiente del ser humano. Es crucial asegurar que la toma de decisiones, la interpretación y la validación del conocimiento permanezcan siempre en manos de los seres humanos, reafirmando así nuestra responsabilidad y autonomía intelectual.

Analizar y mitigar el impacto de la inteligencia artificial en la vida cotidiana, la cognición y la emocionalidad humanas.

- Explorar cómo la constante presencia de la inteligencia artificial en la vida cotidiana está transformando los procesos de pensamiento, aprendizaje, atención y la forma en que se construye la identidad como individuos.

- Crear enfoques educativos y sociales que fomenten la prevención de la adicción a la tecnología y la protección del bienestar emocional. Esto incluye promover la interacción humana significativa y la capacidad de reflexión profunda.

Preservar y fortalecer la creatividad como una facultad propia e irremplazable del ser humano

- La creatividad humana es un proceso intrincado y consciente, que se desarrolla en un contexto específico y no puede ser simplemente transferida a sistemas de inteligencia artificial, a pesar de que estos últimos puedan imitar los resultados creativos.

- Promover entornos educativos, científicos y culturales que fomenten la imaginación, la originalidad y el pensamiento crítico humano. Es importante evitar la tendencia de sustituir estos procesos por automatismos tecnológicos.

Comprensión informada sobre las capacidades actuales de la inteligencia artificial.

- Promover una comprensión básica de que la inteligencia artificial actual no posee conciencia ni entendimiento propio, sino que opera a partir de algoritmos y datos previamente entrenados.

- Evitar la atribución de características humanas, como intencionalidad o pensamiento consciente a sistemas de IA que únicamente simulan ciertos procesos cognitivos.

Referencias Bibliográficas

Andronis, C., Sharma, A., Virvilis, V., Deftereos, S., & Persidis, A. (2011). Literature mining, ontologies and information visualization for drug-repurposing. Briefings in Bioinformatics, 12(4), 357–372. https://doi.org/10.1093/bib/bbr021

Boden, M. (1994). *La mente creativa: mitos y mecanismos* (J. Alvarez, Trad.; 1.ª ed.). Editorial Gedisa. (Trabajo original publicado en 1991).

Boden, M. A. (2004). The creative mind: Myths and mechanisms (2nd ed.). Routledge.

Brown, A. S., & Patel, C. J. (2021). A review of validation strategies for computational drug repurposing. Briefings in Bioinformatics, 22(2), 1–14. https://doi.org/10.1093/bib/bbaa370

Brynjolfsson, E., & McAfee, A. (2014). La segunda era de las máquinas: Trabajo, progreso y prosperidad en una época de brillantes tecnologías. W. W. Norton & Company. (Capítulos sobre innovación recombinante, tecnologías de propósito general y productividad digital.)

Clark, A, & Chalmer, D. (2021). La mente extendida. *Revista de Filosofía* (67), 224-241

Cleeremans, A. (2011). The Radical Plasticity Thesis: How the Brain Learns to be Conscious.

Collins, H. (2024). Why artificial intelligence needs sociology of knowledge: Parts I and II. *AI & Society*, *40*, 1249-1263. https://doi.org/10.1007/s00146-024-01954-8

Colther, C., & Doussoulin, J. (2024). Artificial intelligence: Driving force in the evolution of human knowledge. *Journal of Innovation & Knowledge*, *9*(4), 1-14. https://doi.org/10.1016/j.jik.2024.100625

Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press. (Capítulo 2, "Trabajo", sobre la explotación laboral y la materialidad de la IA.)

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American Psychological Association, págs. 906-911.

Floridi, L. (2011). The philosophy of information. Oxford University Press.

Franceschini, A., Szklarczyk, D., Bosco, N., et al. (2019). STRING v11: Protein–protein association networks with increased coverage. Nucleic Acids Research, 47(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Gray, M. L., & Suri, S. (2019). Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. Houghton Mifflin Harcourt. (Libro clave para entender el concepto de "trabajo fantasma" que Crawford retoma.)

Graziano, M. S. (2019). Rethinking Consciousness (Cap. 2-3).

Li, X., Yu, J., Xu, X., & Li, Y. (2020). Understanding drug repurposing from the perspective of literature-based discovery. Computational and Structural Biotechnology Journal, 18, 2377–2389. https://doi.org/10.1016/j.csbj.2020.09.019

McAfee, A. (2019). More from Less: The Surprising Story of How We Learned to Prosper Using Fewer Resources—and What Happens Next. Scribner. (Explora la visión

ambiental de McAfee sobre cómo la tecnología puede, si se gestiona éticamente, reducir el impacto ecológico.)

Minsky, M. (1986). The society of mind. MIT Press

Mitchell, M. (2019). *Inteligencia Artificial* (M. Rodríguez, Trad.; 1.ª ed.) Editorial Titivillus. (Trabajo original publicado en 2019).

Polanyi, M. (1966). The Tacit Dimension (1.ª ed.). Doubleday & Company, INC [Archivo PDF].

Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking.

Russell, S., & Norvig, P. (2020). Artificial intelligence: A modern approach (4th ed.). Pearson.

Smith, R. (1989). Fish oil and Raynaud's phenomenon: A review of the evidence. Annals of Internal Medicine, 111(10), 802–803. https://doi.org/10.7326/0003-4819-111-10-802

Swanson, D. R. (1986). Fish oil, Raynaud's syndrome, and undiscovered public knowledge. Perspectives in Biology and Medicine, 30(1), 7–18. https://doi.org/10.1353/pbm.1986.0087

Swanson, D. R. (1990). Migraine and magnesium: Eleven neglected connections. Perspectives in Biology and Medicine, 33(4), 526–557. https://doi.org/10.1353/pbm.1990.0027

Taylor, A. (2020). The Automation of Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press. (Complementa la perspectiva crítica sobre cómo la IA reproduce desigualdades sociales.)

University of Cambridge. (18 de septiembre de 2025). ChatGPT seemed to 'think on the fly' when put through an Ancient Greek maths puzzle.

Weeber, M., Klein, H., Aronson, A. R., & Mork, J. G. (2001). Text-based discovery in medicine: The added value of semantic relations. Proceedings of the AMIA Symposium, 731–735.

You, R., et al. (2019). DeepGOPlus: Improved protein function prediction from sequence. Bioinformatics, 35(22), 4824–4832. https://doi.org/10.1093/bioinformatics/btz595

Zheng, J., Zhang, X., Chen, Z., et al. (2020). Phenolyzer: Phenotype-based prioritization of candidate genes for human diseases. Nature Protocols, 15(6), 2299–2319. https://doi.org/10.1038/s41596-020-0321-y

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for COVID-19. Nature Communications, 11, 3259. https://doi.org/10.1038/s41467-020-17168-8