

## Capítulo 1

# Elementos clave de la IA en neuroética: un análisis interdisciplinar para seguridad y defensa\*

---

DOI: <https://doi.org/10.25062/9786287818224.01>

**German Darío Corzo-Ussa**

Escuela Superior de Guerra "General Rafael Reyes Prieto"

**Angélica María Patiño Zuluaga**

Escuela Superior de Guerra "General Rafael Reyes Prieto"

**Resumen:** Esta investigación se propone analizar desde un enfoque interdisciplinar los elementos clave de la IA explicable aplicables a la neuroética para su uso en los sectores de seguridad y defensa. Se parte del reconocimiento de los riesgos éticos, jurídicos y sociales asociados al uso de IA en contextos de alto impacto, especialmente cuando interfieren con la autonomía y la integridad mental de las personas. Se adopta un enfoque cualitativo exploratorio, basado en análisis bibliométrico con Bibliometrix y validación teórica axial, así como un marco teórico que articula tres ejes: inteligencia artificial explicable; neuroética y derechos emergentes, y aplicaciones interdisciplinares. A partir del análisis de autores, conceptos y documentos clave, se consolidó un marco conceptual para orientar el desarrollo de tecnologías responsables, transparentes y centradas en la persona, aportando insumos útiles para la formulación de políticas éticas y regulaciones en entornos de defensa y seguridad.

**Palabras clave:** derechos humanos; ética de la tecnología; gobernanza; integridad mental; inteligencia artificial; neurociencia; neuroética

---

\* Capítulo de libro resultado del proyecto de investigación "Desafíos contemporáneos en la investigación para la formación y la doctrina en seguridad y defensa de la Escuela Superior de Guerra: Reingeniería VINVE Fase I" del grupo de investigación Centro de Gravedad de la Escuela Superior de Guerra "General Rafael Reyes Prieto", categorizado en A1 por MinCiencias (código COL0104976). Los puntos de vista y los resultados de este capítulo son responsabilidad de sus autores y no reflejan necesariamente los de las instituciones participantes.

### German Darío Corzo-Ussa

Mayor del Ejército Nacional de Colombia. Doctor en Planeación Estratégica y Dirección de Tecnología, Universidad Popular Autónoma del Estado de Puebla, México. Magister en Ingeniería Electrónica, Pontificia Universidad Javeriana, Colombia. Ingeniero Electrónico, Universidad Distrital Francisco José de Caldas, Colombia. Docente universitario e Investigador Senior del Sistema Nacional de Ciencia Tecnología e Innovación, Colombia. Estudiante del Curso de Información Militar, Escuela Superior de Guerra "General Rafael Reyes Prieto", Colombia.

<https://orcid.org/0000-0001-7603-0896> - Contacto: [german.corzo@esdeg.edu.co](mailto:german.corzo@esdeg.edu.co)

### Angélica María Patiño Zuluaga

Mayor del Ejército Nacional de Colombia. Magister en Derecho Disciplinario, Universidad Libre, Colombia. Especialista en Derecho Probatorio de la Universidad Sergio Arboleda, Colombia. Especialista en Derecho Sancionatorio, Universidad Militar Nueva Granada, Colombia. Abogada, Universidad Libre Seccional Pereira, Colombia. Estudiante del Curso de Información Militar, Escuela Superior de Guerra "General Rafael Reyes Prieto", Colombia.

<https://orcid.org/0009-0003-8345-9801> - Contacto: [angelica.patino@esdeg.edu.co](mailto:angelica.patino@esdeg.edu.co)

**Citación APA:** Corzo-Ussa, G. D., & Patiño Zuluaga, A. M. (2025). Elementos clave de la IA en neuroética: un análisis interdisciplinar para seguridad y defensa. En A. Serrano Cuervo & D. F. Monroy Anaya (Eds), *Aplicación de la IA y ética en el diseño de soluciones interdisciplinarias para la defensa* (pp. 17-44). Sello Editorial ESDEG. <https://doi.org/10.25062/9786287818224.01>

## APLICACIÓN DE LA IA Y ÉTICA EN EL DISEÑO DE SOLUCIONES INTERDISCIPLINARIAS PARA LA DEFENSA

ISBN impreso: 978-628-7818-21-7

ISBN digital: 978-628-7818-22-4

DOI: <https://doi.org/10.25062/9786287818224>

### Colección Seguridad y Defensa

Sello Editorial ESDEG

Escuela Superior de Guerra "General Rafael Reyes prieto"

Bogotá D.C., Colombia

2025



## Introducción

El avance vertiginoso de la inteligencia artificial (IA) ha generado un profundo impacto en sectores estratégicos como la defensa, la justicia, la seguridad y en general el bienestar social. Su capacidad para automatizar procesos, analizar grandes volúmenes de información y asistir en la toma de decisiones en tiempo real, la ha convertido en una herramienta indispensable. Sin embargo, esta misma capacidad plantea riesgos éticos, jurídicos y sociales, especialmente en contextos de alto impacto, donde sus aplicaciones pueden comprometer principios fundamentales como la autonomía, la integridad mental y los derechos humanos (Floridi et al., 2018; Yuste et al., 2017).

Particularmente preocupante es su integración con tecnologías neurocognitivas. Esta convergencia, cada vez más frecuente en sectores como la seguridad y la defensa, plantea el riesgo de influencias tecnológicas no consentidas sobre los procesos mentales. Desde esta perspectiva, resulta urgente reflexionar sobre la gobernanza y regulación de estas tecnologías, pues su aplicación sin un marco ético y normativo adecuado puede conducir a prácticas discriminatorias, sesgos estructurales y vulneraciones a la dignidad humana (Ienca & Andorno, 2017; Lighthart et al., 2023).

En respuesta a estos desafíos, se han desarrollado propuestas como la IA explicable (XAI), que promueve la transparencia, trazabilidad y supervisabilidad de los procesos algorítmicos. La Comisión Europea (2020) ha planteado que los sistemas de IA deben ser comprensibles por los seres humanos, especialmente cuando se aplican en sectores de alto riesgo como la defensa. Este enfoque busca garantizar la confianza institucional y la rendición de cuentas, elementos clave para evitar abusos o errores automatizados.

Complementariamente, el campo de la neuroética ha adquirido relevancia como disciplina encargada de estudiar las implicaciones éticas, legales y sociales de las tecnologías que interactúan con el cerebro. Autores como Yuste et al. (2017) han propuesto el reconocimiento de los llamados “*neurorights*” o derechos neurocognitivos, que incluyen la privacidad mental, la libertad cognitiva y la integridad psicológica. Estos derechos emergentes buscan limitar los riesgos de una tecnología que, sin control, puede invadir los dominios más íntimos del pensamiento y la voluntad.

Desde un enfoque aplicado, experiencias como la documentada por Corzo-Ussa et al. (2023) evidencian que las herramientas de IA diseñadas con fines militares pueden adaptarse exitosamente a propósitos civiles, como la sostenibilidad ambiental. Este enfoque dual, militar-civil, refuerza la viabilidad de modelos éticos y técnicamente explicables en proyectos de desarrollo. Sin embargo, la literatura también advierte sobre los riesgos de un uso no regulado. Por ejemplo, Eke (2024) alerta sobre el “*dumping ético*”, fenómeno que describe la implementación de tecnologías sin estándares éticos en regiones con escasa regulación, lo que genera impactos negativos en poblaciones vulnerables y debilita la legitimidad institucional.

Este panorama evidencia una necesidad urgente: identificar los elementos clave de la XAI desde una perspectiva neuroética, de modo que sea posible diseñar soluciones interdisciplinarias aplicadas a la defensa y la seguridad que sean éticas, transparentes y centradas en la protección de la autonomía y la integridad mental de las personas. El presente capítulo se propone abordar esta necesidad mediante una investigación que parte de la siguiente pregunta: ¿Cuáles son los elementos clave de la XAI que deben integrarse desde una perspectiva neuroética en el diseño de soluciones interdisciplinarias aplicadas a contextos de seguridad y defensa?

Para dar respuesta a esta pregunta, se definió como objetivo general determinar los elementos clave de la XAI aplicables a la neuroética, con el fin de orientar el diseño de soluciones interdisciplinarias en los sectores de seguridad y defensa. Dichas soluciones deben garantizar el desarrollo de tecnologías transparentes, responsables y centradas en la protección de la autonomía y la integridad mental de las personas, considerando la sensibilidad de los contextos donde se implementan. De este objetivo general se desprenden tres propósitos específicos: 1) recolectar y sistematizar información bibliográfica relevante sobre XAI y neuroética aplicadas a los sectores de seguridad y defensa; 2) identificar las principales tendencias, autores clave, conceptos emergentes y redes temáticas que estructuran

la producción académica y científica en estas áreas estratégicas; y 3) clasificar los constructos teóricos y conceptuales que sirvan de base para el diseño de soluciones interdisciplinarias orientadas a salvaguardar la autonomía y la integridad mental en escenarios críticos.

Se adoptó un enfoque teórico organizado en torno a tres ejes complementarios y convergentes. El primero es la XAI, entendida como un conjunto de tecnologías orientadas a la transparencia, la auditabilidad y la rendición de cuentas de los sistemas algorítmicos (Comisión Europea, 2020; Floridi et al., 2018). El segundo eje corresponde a la neuroética y los derechos emergentes, que buscan proteger la privacidad mental, la libertad cognitiva y la integridad psicológica ante los riesgos potenciales derivados de la convergencia entre neurotecnología e IA (Ienca & Andorno, 2017; Lighthart et al., 2023). Finalmente, el tercer eje contempla las aplicaciones interdisciplinarias, que integran aportes de la ingeniería, el derecho, la neurociencia y la política pública para desarrollar soluciones tecnológicas sólidamente fundamentadas en principios éticos y orientadas a responder de manera responsable a los desafíos sociales y regulatorios contemporáneos (Eke, 2024; Francisco, 2023).

Metodológicamente, la investigación adopta un enfoque cualitativo de tipo exploratorio, respaldado por análisis bibliométrico y validación teórica interpretativa (Hernández Sampieri et al., 2014). El estudio se basa en el uso del paquete Bibliometrix (Aria & Cuccurullo, 2017), herramienta desarrollada en R que permite el análisis exhaustivo de tendencias, redes de coautoría, documentos más citados y teorías emergentes a partir de bases de datos como Scopus y Web of Science. Esta técnica se complementa con un análisis temático cualitativo que permite interpretar los hallazgos desde una perspectiva normativa y conceptual, siguiendo la propuesta de Díez-Gómez et al. (2019).

Entre los principales resultados se encuentra la sistematización de las teorías y marcos conceptuales más relevantes sobre XAI y neuroética, así como la identificación de constructos clave que orienten el diseño de soluciones interdisciplinarias en defensa y seguridad. Asimismo, se espera aportar un marco conceptual útil para la formulación de políticas públicas, protocolos de diseño tecnológico y recomendaciones de gobernanza ética centrada en las personas.

En definitiva, este capítulo busca ser un aporte riguroso y crítico al debate contemporáneo sobre la convergencia entre IA y neurotecnología, y su regulación ética en sectores estratégicos. La integración de enfoques interdisciplinarios permitirá

avanzar hacia un desarrollo tecnológico que no solo sea eficiente, sino también justo, transparente y profundamente humano.

## Fundamentos de la XAI y la neuroética

La XAI se refiere al diseño de sistemas cuyos procesos de decisión puedan ser comprendidos, auditados y justificados por seres humanos. Este principio es fundamental en contextos críticos como la defensa, la justicia o la salud, donde las decisiones automatizadas pueden tener consecuencias significativas para los derechos y la integridad de las personas. La Comisión Europea (2020) ha subrayado que la IA en entornos de alto riesgo debe ser transparente, trazable y estar sujeta a supervisión humana. Estas condiciones no solo promueven la confianza institucional, sino que permiten una rendición de cuentas efectiva frente a errores, sesgos o resultados discriminatorios.

En consonancia con este enfoque, Floridi et al. (2018) proponen una ética para la IA basada en los principios clásicos de la bioética: beneficencia, no maleficencia, autonomía y justicia. A estos principios añaden la explicabilidad como componente clave para garantizar que los sistemas algorítmicos respeten los derechos fundamentales, en especial en situaciones donde las decisiones automatizadas afectan la libertad, la igualdad o la dignidad humanas.

Desde una perspectiva complementaria, Binns (2018) sostiene que la equidad algorítmica no puede definirse exclusivamente desde criterios técnicos. Mediante un análisis anclado en la filosofía política, argumenta que distintas concepciones de justicia pueden producir resultados diversos en el diseño e implementación de algoritmos. Por tanto, cualquier arquitectura de IA representa también una elección normativa que requiere una reflexión desde el punto de vista de su impacto social.

Estas dimensiones éticas se conectan directamente con el campo de la neuroética, disciplina que estudia las implicaciones éticas, legales y sociales de las tecnologías que interactúan con el cerebro y la mente humana. En este marco, Lenca y Andorno (2017) advierten que la IA, cuando incide sobre procesos neurocognitivos, puede comprometer la autonomía, la privacidad mental o la libertad decisional. Esta preocupación ha dado lugar al reconocimiento de una nueva categoría de derechos emergentes: los *neurorights*, que buscan proteger la integridad psicológica frente a posibles interferencias tecnológicas no consentidas.

## Exploración teórica de la IA y la neuroética en seguridad y defensa

Una exploración teórica que articule la XAI con la neuroética resulta fundamental para sustentar el desarrollo de soluciones tecnológicas responsables en sectores estratégicos. Este marco no solo permite comprender los desafíos éticos que implica el uso de IA en escenarios complejos, sino que también ofrece una base conceptual sólida para orientar políticas públicas, procesos normativos y diseños tecnológicos centrados en las personas.

### Búsqueda de autores y fuentes

Con el propósito de recolectar información bibliográfica relevante sobre XAI y neuroética, aplicada específicamente a los sectores de seguridad y defensa, se efectuó una búsqueda sistemática en la base de datos Scopus. Esta búsqueda se diseñó a partir de las bases teóricas previamente exploradas en este estudio y se estructuró mediante la selección de palabras clave estratégicas y el uso de algoritmos booleanos. El objetivo fue obtener una muestra representativa de publicaciones científicas que aportaran perspectivas conceptuales y empíricas pertinentes al campo de investigación. A continuación, se presentan las combinaciones de términos y operadores utilizados en la estrategia de búsqueda:

1. ("Explainable Artificial Intelligence" OR XAI) AND (neuroethics OR neurorights) AND (security OR defense)
2. ("AI governance" OR "algorithmic transparency") AND ("human rights" OR "cognitive autonomy") AND (military OR defense)
3. ("Explainable AI" OR "ethical AI") AND (surveillance OR decision-making) AND (privacy OR mental integrity)
4. ("Artificial Intelligence" AND "neuroethics") AND ("autonomy" OR "privacy") AND (security OR "policy framework")
5. ("Neurotechnology" OR "AI applications") AND (ethics OR "human rights") AND (defense OR intelligence)
6. Con base en este listado, se obtuvo un archivo bibliométrico para procesar con el uso de Bibliometrix. Los resultados se presentan en la Tabla 1.

**Tabla 1.** Resultados de búsqueda de documentos en Scopus

Descripción	Resultados
Periodo de tiempo	2012:2025
Fuentes (revistas, libros, etc)	432
Documentos	541
Tasa de crecimiento anual %	44,15
Promedio de citas por documento	18,29
Palabras clave plus	2721
Palabras clave del autor	1563
Autores	2126

**Fuente:** Elaboración propia a partir de Bibliometrix

De la misma forma, con el uso de Bibliometrix se obtuvo un listado de los diez autores más citados. Los resultados se muestran en la Tabla 2.

**Tabla 2.** Autores más citados

Autores	Artículos	Artículos fraccionados
Ienca Marcello	10	3,45
Joshi Sameer	4	1,01
Buyx Alena	3	0,78
Chen Tian	3	0,58
Eaton Sara Elaine	3	3,00
Fiske Amelia	3	1,08
Friedrich Orsolya	3	0,48
Ho Manh Toan	3	0,63
Holzinger Andreas	3	0,66
Liu Jian	3	1,66

**Fuente:** Elaboración propia a partir de Bibliometrix

Finalmente, con el uso de Bibliometrix se obtuvo un listado de los diez documentos más citados (Tabla 3).

**Tabla 3.** Documentos más citados

Título del documento	Citaciones	Citaciones por año
Artificial intelligence in dentistry: Chances and challenges (Schwendicke et al., 2020)	547	91,17
A governance model for the application of ai in health care (Reddy et al., 2020)	451	75,17
Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy (Fiske et al., 2019)	405	57,86
A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion (Albahri et al., 2023)	353	117,67
Global evolution of research in artificial intelligence in health and medicine: A bibliometric study (Tran et al., 2019)	306	43,71
Human- versus artificial intelligence (Korteling et al., 2021)	259	51,80
A review of the role of artificial intelligence in healthcare (Al Kuwaiti et al., 2023)	257	85,67
Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine (Harrer, 2023)	228	76,00
The ugly truth about ourselves and our robot creations: The problem of bias and social inequity (Howard & Borenstein, 2018)	220	27,50
Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale effects of fairness and explanation on trust in ethical AI (Wang et al., 2023)	204	68,00

**Fuente:** Elaboración propia a partir de Bibliometrix

A partir de esta base documental se relacionan las teorías con los conceptos previamente definidos con el fin de identificar las tendencias y conceptos emergentes que permitan definir las redes temáticas y conceptuales sobre XAI y neuroética.

## Identificación de tendencias, conceptos y redes temáticas

La IA ha evolucionado aceleradamente, consolidándose como una tecnología estratégica con amplias aplicaciones en sectores sensibles como la salud, la seguridad y la educación. El estudio de Tran et al. (2019) muestra un crecimiento exponencial en publicaciones científicas sobre IA, especialmente en áreas como el aprendizaje automático, la XAI y la ética, subrayando la necesidad de enfoques multidisciplinares que articulen el desarrollo tecnológico con principios normativos y sociales. En esa línea, Korteling et al. (2021) reflexionan sobre los límites de comparar la inteligencia humana con la artificial, proponiendo la colaboración entre ambas y enfatizando la necesidad de que los humanos desarrollen conciencia crítica sobre la IA para utilizarla de manera efectiva y ética.

Por su parte, Howard y Borenstein (2018) advierten cómo los sesgos sociales y cognitivos humanos pueden ser replicados y amplificados por los algoritmos, generando desigualdades estructurales. Esta dimensión es especialmente crítica en contextos de seguridad, donde las decisiones automatizadas pueden impactar derechos fundamentales. Por otro lado, Wang et al. (2023) desarrollan un modelo para evaluar la competencia en IA donde la dimensión ética es un componente esencial. Este modelo es clave para fortalecer la alfabetización crítica y la toma de decisiones responsables frente a tecnologías cognitivas.

Estos trabajos convergen en la necesidad de integrar la ética, la transparencia y la regulación social en el diseño y uso de la IA en entornos de alto impacto. El análisis con base en términos clave que se hace más adelante muestra cómo algunos de estos temas, anunciados por estos autores, han evolucionado en el tiempo. Allí se evidencia que solo a partir de 2021 comienza a nombrarse la IA junto con la tecnología ética.

### Neuroética y derechos emergentes

La neuroética examina las implicaciones éticas y jurídicas del uso de tecnologías que interactúan con el cerebro, especialmente en contextos sensibles como la seguridad y la defensa. Estas tecnologías, al intervenir o extraer datos del sistema nervioso, plantean riesgos para la autonomía, la identidad y la privacidad mental de las personas. En este contexto, Lighthart et al. (2023) proponen una base ética y jurídica para los llamados *neurorights*, que comprenden el derecho a la privacidad mental, la integridad mental y la libertad cognitiva. Estos derechos buscan

proteger el dominio mental frente a accesos no consentidos, manipulaciones o interferencias tecnológicas.

Yuste et al. (2017) también destacan la necesidad urgente de reconocer estos derechos emergentes, dado el avance de la IA en conjunto con neurotecnologías. La posibilidad de leer, modificar o condicionar estados mentales requiere marcos normativos que garanticen la autodeterminación y la dignidad humana. La Comisión Europea (2020) respalda estos enfoques al exigir que la IA sea explicable, supervisable y centrada en el ser humano.

La integración de la neuroética en el diseño de tecnologías de defensa e IA permite establecer límites éticos que aseguren el respeto a la mente como espacio inviolable y fundamento de los derechos humanos.

## Aplicaciones interdisciplinarias en seguridad y defensa

La IA ha transformado los sectores de seguridad y defensa mediante la automatización de procesos críticos como la vigilancia, el reconocimiento y la toma de decisiones operativas. Estas tecnologías plantean oportunidades relevantes, pero también desafíos éticos y jurídicos, especialmente cuando interfieren con la autonomía cognitiva y los derechos fundamentales. Casos como el uso de IA militar con fines ambientales, como el expuesto por Corzo-Ussa et al. (2023), evidencian la posibilidad de adaptar herramientas diseñadas para conflictos a otros contextos, como el de la sostenibilidad ambiental, siempre que se respeten criterios de transparencia, trazabilidad y supervisión humana.

Francisco (2023) alerta que la integración de la IA en el discurso de seguridad puede derivar en prácticas de control político, vigilancia masiva o instrumentalización del medio ambiente, si no se desarrolla una gobernanza inclusiva y centrada en el ser humano. En esta línea, Reep (2024) propone el uso de sistemas híbridos que combinen inteligencia humana y artificial para anticipar necesidades logísticas y estratégicas, tanto en entornos comerciales como militares. Esta visión de la IA como herramienta dual permite maximizar su eficacia mientras se minimizan riesgos operacionales y éticos.

Finalmente, Eke (2024) subraya la importancia de considerar el contexto sociopolítico y cultural donde se implementan estas tecnologías, para evitar el “*dumping* ético” y garantizar soluciones que respeten la autonomía, la justicia y la seguridad integral. De allí que se requiera un enfoque verdaderamente interdisciplinar, capaz de integrar neuroética, derecho, ingeniería y política pública.



perspectivas. La investigación de Al Kuwaiti et al. (2023) destaca la expansión de la IA en salud, su potencial transformador y los retos éticos y regulatorios, mostrando la necesidad de una gobernanza robusta para equilibrar innovación y protección de derechos. Eaton (2025) explora la transformación educativa mediante IA, neurotecnologías y aprendizajes orientados al futuro, enfatizando la integridad académica y la ética en entornos tecnológicos disruptivos. Por su parte, Binns (2018) conecta la equidad algorítmica con teorías filosóficas de la justicia, aportando un marco normativo para comprender la "justicia algorítmica" y sus implicaciones políticas. Finalmente, Liu et al. (2024) expone su tesis sobre la confianza y estrategias éticas en IA, y destaca la importancia de soluciones explicables, auditables y socialmente responsables.

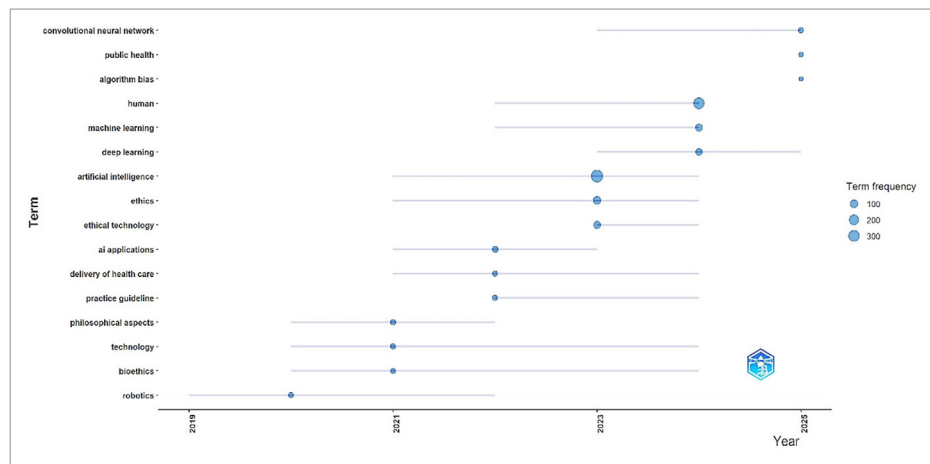
Estos aportes revelan tendencias clave: la expansión interdisciplinar de la IA, el énfasis en la ética y la gobernanza, la preocupación por la equidad y el rol central de la educación y la salud en su adopción. Estas conexiones reafirman un campo en consolidación, con una orientación hacia la responsabilidad, la equidad y la sostenibilidad tecnológica.

## Evolución y tendencia de términos

La evolución temporal de términos permite contextualizar las tendencias identificadas en la nube de palabras, para mostrar no solo qué conceptos son relevantes, sino también cómo han ganado o perdido protagonismo en la literatura científica a lo largo del tiempo, como se muestra en la Figura 2. La persistencia de términos como *inteligencia artificial*, *ética* y *tecnología ética* refleja un interés sostenido por la relación entre innovación tecnológica y valores humanos, mientras que la aparición reciente de expresiones como *sesgo algorítmico*, *salud pública* y *aspectos filosóficos* indica un viraje hacia desafíos sociales y regulatorios. Esta dinámica confirma que la investigación actual no se limita al desarrollo técnico, sino que integra la reflexión sobre impactos éticos, equidad y supervisión humana en aplicaciones críticas (Binns, 2018).

La Figura 2 evidencia cómo la investigación sobre IA ha transitado desde un enfoque técnico hacia una visión más integral, que incorpora aspectos éticos, sociales y humanos. Conceptos como *inteligencia artificial*, *ética*, *sesgo algorítmico* y *salud pública* presentan un crecimiento sostenido, lo que indica que la comunidad científica ha ampliado sus preocupaciones más allá del rendimiento tecnológico, hacia el impacto social de la IA (Binns, 2018; Floridi et al., 2018).

Figura 2. Evolución y tendencia de términos clave sobre IA y neuroética



Fuente: Elaboración propia a partir de Bibliometrix

El aumento de términos asociados con *supervisión humana*, *privacidad de datos* y *toma de decisiones* refleja la importancia de mantener el control humano en sistemas complejos y automatizados, un desafío central identificado en estudios recientes sobre gobernanza tecnológica y riesgos emergentes (Yuste et al., 2017). Asimismo, el crecimiento de temas vinculados con *educación tecnológica* y *alfabetización en IA* confirma la tendencia a preparar a ciudadanos y profesionales para interactuar con estas tecnologías de manera ética y segura (Huang et al., 2024).

Otro hallazgo relevante es el surgimiento de términos relacionados con aplicaciones en *salud mental* y *autogestión tecnológica*, lo cual evidencia que el uso de IA ha trascendido al ámbito clínico y terapéutico. Estas aplicaciones, aunque ofrecen beneficios en diagnóstico y prevención, también plantean dilemas éticos sobre *autonomía*, *privacidad* y *equidad* en el acceso a soluciones tecnológicas (Friedrich et al., 2021).

En síntesis, esta evolución temporal de los conceptos refleja un proceso de maduración de la investigación, orientado a la construcción de una IA confiable, explicable y socialmente responsable. La consolidación de conceptos clave evidencia que la investigación en IA avanza hacia una mayor complejidad interdisciplinaria. Este cambio no solo refleja el crecimiento del interés por la *ética*, la *salud* y la *educación*, sino que también señala la necesidad de comprender cómo los actores, instituciones y áreas temáticas se relacionan entre sí en el ecosistema científico. Analizar las redes temáticas permitirá identificar los núcleos de conocimiento más

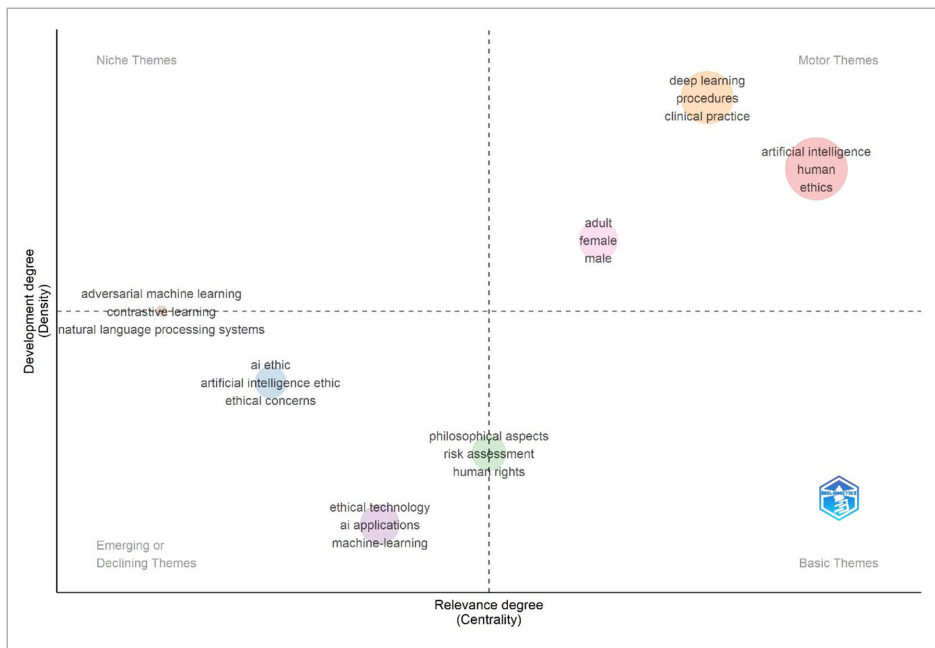
influyentes, las conexiones entre campos emergentes y los autores que lideran estas discusiones, facilitando así una visión integral del desarrollo de la IA y su impacto social.

## Redes temáticas

El mapa temático del *software* Bibliometrix permite identificar y clasificar los temas de investigación según su relevancia y grado de desarrollo, organizándolos en cuatro cuadrantes: *motores*, *básicos*, *especializados* y *emergentes*. Su análisis, basado en indicadores de centralidad y densidad, ofrece una visión estratégica de la estructura temática del campo, mostrando áreas consolidadas, en desarrollo o con potencial de crecimiento. De este modo, complementa otros análisis bibliométricos y facilita la comprensión de la dinámica de la investigación (Aria & Cuccurullo, 2017).

La Figura 3 muestra el mapa temático construido a partir de las palabras clave extraídas de los documentos revisados en la base de datos correspondientes a los últimos diez años. Revela cuatro áreas diferenciadas en la literatura analizada.

**Figura 3.** Mapa temático sobre IA y neuroética



Fuente: Elaboración propia a partir de Bibliometrix

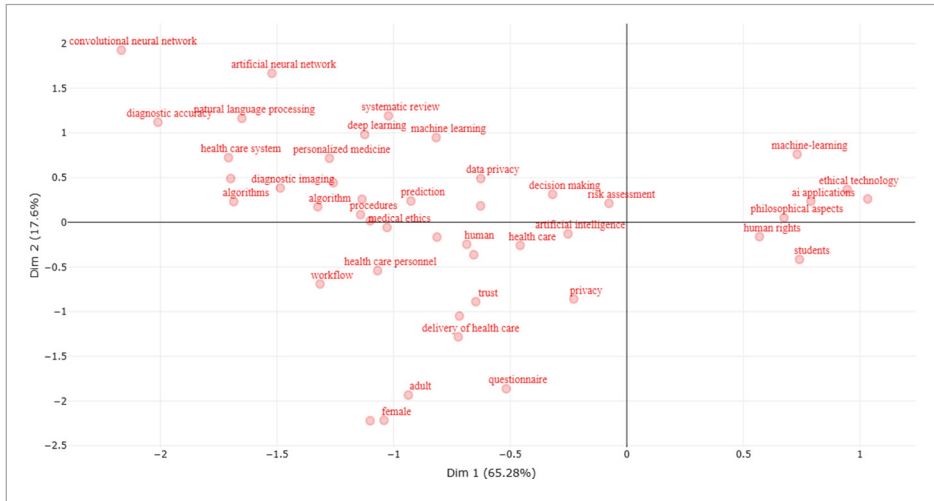
En el cuadrante de *temas motores* se ubican *human, humans y ethics*, junto con *artificial intelligence, machine learning y deep learning*. Estos conceptos presentan alta relevancia y un desarrollo consolidado, lo que refleja que la investigación actual se centra tanto en la aplicación técnica de la IA como en sus implicaciones éticas. Esta relación se manifiesta en la creciente adopción de la IA en la atención sanitaria y la medicina personalizada, donde los modelos de aprendizaje profundo mejoran el diagnóstico temprano y la gestión de enfermedades complejas (Sobhi et al., 2025), al tiempo que plantean la necesidad de salvaguardar la autonomía cognitiva y la integridad mental de los individuos (Herrero, 2025).

En el cuadrante de *temas especializados* se destacan *adversarial machine learning, contrastive learning y computational education*, que muestran alta cohesión interna pero baja conexión con otros temas. Estos representan nichos de investigación centrados en la robustez algorítmica y la formación de talento en IA, relevantes para la seguridad informática y la detección de amenazas avanzadas mediante modelos híbridos y técnicas explicables (Desidi et al., 2025).

Finalmente, en la zona de *temas emergentes o en declive* aparecen *ethical concerns, human rights y philosophical aspects*, que sugieren un desarrollo incipiente. Estos temas están relacionados con la necesidad de establecer marcos regulatorios claros, reconocer los neuroderechos y garantizar principios de equidad en la implementación de IA (Ghanem et al., 2025; Kamila & Jasrotia, 2025).

En la Figura 4 se presenta el mapa de palabras (*word map*), que complementa este análisis al mostrar la relación conceptual entre los términos clave y la estructura temática identificada. Generado mediante análisis factorial, este mapa representa la relación conceptual entre términos clave presentes en la literatura analizada, mostrando su disposición en un espacio bidimensional definido por dos ejes principales. El eje 1 (65,28 %) diferencia claramente los términos técnicos y clínicos, ubicados hacia la izquierda, de aquellos relacionados con consideraciones éticas, sociales y filosóficas, situados hacia la derecha. Por su parte, el eje 2 (17,66 %) distingue los enfoques orientados a metodologías y rendimiento, posicionados en la parte superior del gráfico (*convolutional neural network, artificial neural network, diagnostic accuracy, systematic review, deep learning, machine learning*), de los términos vinculados con la implementación práctica y aspectos operativos en el ámbito clínico, ubicados en la parte inferior (*workflow, delivery of health care, trust, health care personnel*).

Figura 4. Mapa de palabras sobre IA y neuroética



Fuente: Elaboración propia a partir de Bibliometrix

A partir de esta distribución se identifican tres agrupaciones conceptuales principales. En primer lugar, un *clúster técnico*, conformado por términos como *deep learning*, *machine learning*, *artificial neural network*, *convolutional neural network* y *prediction*, que refleja el núcleo tecnológico del campo, con aplicaciones consolidadas en el diagnóstico temprano y la medicina personalizada (Sobhi et al., 2025).

En segundo lugar, un *clúster clínico-operacional*, que incluye conceptos asociados a la práctica médica y la integración de la IA en sistemas de salud, tales como *health care system*, *diagnostic imaging*, *personalized medicine*, *procedures*, *workflow* y *delivery of health care*, evidenciando su rol en la optimización de procesos y la mejora de la calidad asistencial (Ghanem et al., 2025).

Finalmente, se observa un *clúster ético-social*, conceptualmente diferenciado, con términos como *ethical technology*, *philosophical aspects*, *human rights*, *students* y *data privacy*, que reflejan la creciente preocupación por la privacidad de la información, la gobernanza de datos y la protección de la autonomía cognitiva en escenarios críticos (Herrero, 2025; Kamila & Jasrotia, 2025).

La disposición de estos clústeres y la proporción de varianza explicada por los ejes evidencian una estructura conceptual robusta, donde el desarrollo técnico y sus aplicaciones clínicas dominan el campo, pero acompañados de un creciente debate ético y social que se proyecta como un componente emergente en expansión (Desidi et al., 2025). Estos hallazgos complementan el mapa temático, en el

que los desarrollos tecnológicos aparecen como temas motores consolidados, mientras que los aspectos éticos y de derechos humanos se perfilan como áreas de investigación en evolución.

## Constructos teóricos de IA y neuroética para seguridad y defensa

El análisis combinado de la nube y la tendencia temporal de palabras, el mapa temático y el análisis factorial de palabras resulta útil para clasificar los constructos teóricos y conceptuales sobre IA y neuroética que sirvan de base para el diseño de soluciones interdisciplinarias orientadas a salvaguardar la autonomía y la integridad mental en escenarios críticos.

### Análisis temático

En la base documental se evidencia una estructura conceptual en la que convergen desarrollos tecnológicos, aplicaciones clínicas y consideraciones éticas vinculadas a la IA. Los temas motores destacan la consolidación de *artificial intelligence*, *machine learning* y *deep learning* aplicados a la práctica clínica y procedimientos médicos, integrando la dimensión ética, humana y normativa. Estos constructos reflejan el núcleo técnico-científico del campo, pero también su anclaje en principios fundamentales de la neuroética, al requerir la protección de la autonomía cognitiva y la integridad de los datos personales, especialmente en poblaciones vulnerables (Friedrich et al., 2021; Yuste et al., 2017).

Por otra parte, los temas especializados, como *adversarial machine learning*, *explainable AI* y *contrastive learning*, muestran la creciente relevancia de la robustez algorítmica y la seguridad de los sistemas frente a ataques o fallos, aspectos cruciales para la ciberdefensa, la confianza operativa y la protección de infraestructuras críticas en contextos de seguridad y defensa (Holzinger et al., 2025). Estas dimensiones técnicas, sin embargo, no pueden analizarse de forma aislada, ya que conllevan implicaciones éticas, regulatorias y de gobernanza.

Asimismo, los temas transversales relacionados con variables demográficas evidencian la importancia de la equidad y la mitigación de sesgos algorítmicos en la toma de decisiones automatizadas, especialmente en sistemas aplicados a salud, justicia o seguridad (Binns, 2018). La inclusión de estas variables permite avanzar hacia una IA más justa, consciente de las asimetrías sociales que puede amplificar si no se diseñan mecanismos adecuados de corrección e intervención.

Finalmente, los temas emergentes vinculados con *derechos humanos*, *supervisión humana* y *evaluación de riesgos* reflejan un interés creciente por el desarrollo de marcos normativos y de gobernanza tecnológica que aseguren transparencia, explicabilidad y responsabilidad social. Este enfoque se articula con propuestas como el marco *AI4People* (Floridi et al., 2018), que promueve una IA “buena” basada en cinco principios éticos, y con el llamado a fortalecer la regulación y supervisión en entornos clínicos (Al Kuwaiti et al., 2023).

En conjunto, esta estructura temática refuerza la necesidad de un enfoque interdisciplinar que integre tecnología, neuroética y políticas de seguridad, permitiendo avanzar hacia soluciones tecnológicas que no solo sean eficientes, sino también confiables, humanas y justas.

La Tabla 4 presenta un resumen de los resultados, identificando los temas clave y su interpretación, para conectar las teorías encontradas sobre IA y neuroética con sus aplicaciones interdisciplinares en seguridad y defensa.

**Tabla 4.** *Análisis temático documental*

Indicador	Temas clave	Interpretación
Temas motores	Inteligencia artificial, aprendizaje automático, aprendizaje profundo, práctica clínica, procedimientos, humanos, ética.	Representan el núcleo tecnológico y ético consolidado del campo. Constructos clave: núcleo tecnológico y ético. Conexión con neuroética: desarrollo de IA aplicada a la salud con principios éticos explícitos. Implicación en seguridad y defensa: soporte de decisiones médicas en contextos militares, integridad de datos y derechos humanos.
Temas altamente desarrollados o especializados	Aprendizaje automático adversarial, aprendizaje contrastivo, sistemas de procesamiento del lenguaje natural.	Temas especializados con bajo nivel de conexión externa. Constructos clave: técnicas avanzadas de IA (robustez y procesamiento avanzado). Conexión con neuroética: seguridad algorítmica y resiliencia frente a ataques adversarios. Implicaciones en seguridad y defensa: aplicaciones en ciberseguridad y sistemas de información crítica.
<b>Temas transversales de datos demográficos</b>		
Temas básicos y transversales	Adulto, mujer, hombre (datos demográficos asociados a aplicaciones clínicas de IA)	Constructos clave: procesos transversales y equidad. Conexión con neuroética: análisis de equidad y representación justas. Implicación en seguridad y defensa: políticas de datos éticos y protección de poblaciones vulnerables.

Continúa tabla...

Indicador	Temas clave	Interpretación
Temas emergentes	Ética de la inteligencia artificial, preocupaciones éticas, tecnología ética, aplicaciones de la IA, derechos humanos, aspectos filosóficos, evaluación de riesgos	Temas emergentes relacionados con ética, derechos humanos y evaluación de riesgos. Constructos clave: gobernanza emergente, regulación, privacidad y derechos. Conexión con neuroética: marcos normativos y gobernanza de la IA. Implicación en seguridad y defensa: evaluación ética en el desarrollo de IA y cumplimiento de normas internacionales.

Fuente: Elaboración propia

## Clasificación de constructos

El análisis de la Tabla 4 permite clasificar los constructos teóricos y conceptuales sobre IA y neuroética, aportando una base sólida para el diseño de soluciones interdisciplinarias orientadas a la protección de la autonomía y la integridad mental en escenarios críticos de seguridad y defensa.

En el *núcleo tecnológico y ético* se identifican desarrollos consolidados en *machine learning*, *deep learning* y *clinical practice*, los cuales han demostrado su potencial para mejorar diagnósticos, optimizar tratamientos y apoyar procesos de toma de decisiones en entornos complejos (Sobhi et al., 2025). Estos avances se acompañan de principios éticos y humanos que, como señala Plá Herrero (2025), son esenciales para garantizar el respeto a los *neuroderechos* y la *autonomía cognitiva*.

En el ámbito de la *robustez y el procesamiento avanzado* destacan *adversarial machine learning* y *natural language processing systems*, herramientas fundamentales para garantizar la seguridad algorítmica frente a ataques que puedan manipular datos o alterar procesos decisionales. Desidi et al. (2025) evidencian que el uso de arquitecturas híbridas, como *CNN* y *LSTM*, combinado con *blockchain*, fortalece la resiliencia de los sistemas críticos.

Los *procesos transversales*, representados por *datos demográficos* y *flujos de trabajo clínico*, ponen de manifiesto la necesidad de reducir sesgos y asegurar equidad en el uso de datos. Ghanem et al. (2025) sostienen que integrar criterios de equidad en la IA es clave para evitar discriminaciones que afecten la diversidad cognitiva, especialmente en contextos militares y de seguridad. En esta línea, Binns (2018) argumenta que la *justicia algorítmica* debe basarse en principios normativos sólidos para garantizar legitimidad.

Finalmente, emergen constructos vinculados a *ética, derechos humanos* y *evaluación de riesgos*, que refuerzan la importancia de marcos regulatorios y de gobernanza tecnológica. Reddy et al. (2020) proponen modelos de gobernanza que promueven *transparencia, responsabilidad* y *confianza* en la implementación de IA, mientras que Kamila y Jasrotia (2025) enfatizan que la anticipación de riesgos éticos es esencial para prevenir daños y salvaguardar la integridad mental.

La Tabla 5 presenta la clasificación de constructos con los conceptos clave de IA y neuroética relevantes para soluciones interdisciplinarias en seguridad y defensa.

**Tabla 5.** Clasificación de constructos y conceptos clave

Constructo	Conceptos clave	Relación con neuroética	Aplicaciones en seguridad y defensa
Núcleo tecnológico y ético	Inteligencia artificial, aprendizaje automático, aprendizaje profundo, práctica clínica, procedimientos, ética, humanos	Soporte al diagnóstico y monitoreo cognitivo con principios éticos explícitos, protegiendo la autonomía y la integridad mental.	Sistemas de soporte a la salud mental del personal, IA aplicada a la evaluación cognitiva en escenarios críticos.
Robustez y procesamiento avanzado	Aprendizaje automático adversarial, aprendizaje contrastivo, sistemas de procesamiento del lenguaje natural	Resiliencia frente a ataques adversarios y manipulación de información, garantizando procesos cognitivos seguros.	Defensa cibernética de sistemas cognitivos, seguridad en sistemas de información críticos.
Procesos transversales y equidad	Datos demográficos, flujo de trabajo, prestación de servicios de salud	Mitigación de sesgos y equidad en datos, protegiendo la diversidad cognitiva y la autonomía en entornos operativos.	Protocolos inclusivos de integración de IA, minimizando sesgos que puedan afectar decisiones estratégicas.
Gobernanza emergente, regulación, privacidad y derechos.	Ética de la IA, tecnología ética, derechos humanos, aspectos filosóficos, evaluación de riesgos	Desarrollo de marcos regulatorios y principios de gobernanza que protejan derechos cognitivos y privacidad mental.	Políticas y normas para el uso ético de IA en operaciones militares y de seguridad, protegiendo la autonomía cognitiva.

Fuente: Elaboración propia

## Identificación de elementos clave

La Tabla 6 muestra los elementos clave, identificados a partir del análisis temático presentado en la Tabla 4 y de los constructos y conceptos clave clasificados en la Tabla 5. Se presenta, además, una propuesta de variables de estudio para cada elemento y las dimensiones con las que pueden medirse.

**Tabla 6.** Clasificación de constructos y conceptos clave

Elemento clave	Descripción	Variables de estudio	Dimensiones de medición
Autonomía cognitiva y toma de decisiones asistida por IA	Evalúa cómo las tecnologías de IA influyen en la capacidad de individuos (militares o civiles) para tomar decisiones libres y no manipuladas en entornos de alta presión.	Grado de dependencia tecnológica, impacto en la percepción de control, efectos en la carga cognitiva.	Percepción de control, nivel de autonomía percibida, precisión de decisiones, tiempos de respuesta cognitiva.
Integridad mental y protección frente a interferencias cognitivas	Considera el riesgo de manipulación de percepciones, emociones o procesos cognitivos por sistemas de IA, especialmente en contextos de guerra cognitiva o entrenamiento militar.	Exposición a sistemas autónomos, vulnerabilidad psicológica, resiliencia cognitiva, respuestas fisiológicas al estímulo de la IA.	Indicadores de estrés, estabilidad emocional, desempeño en tareas cognitivas críticas, resistencia a sesgos inducidos por IA.
Gobernanza ética y regulatoria de sistemas autónomos	Analiza la existencia y efectividad de normas que regulen el uso de IA con potencial de afectar procesos mentales.	Cumplimiento normativo, marcos legales aplicables, políticas de gobernanza tecnológica, auditorías éticas.	Grado de cumplimiento normativo, existencia de marcos regulatorios efectivos, confianza institucional en sistemas autónomos.
Equidad y mitigación de sesgos algorítmicos	Evalúa cómo los sesgos en datos y algoritmos pueden afectar poblaciones diversas (género, cultura, nivel educativo) y su relación con la justicia cognitiva.	Diversidad de datos, errores de predicción por grupo poblacional, percepciones de justicia algorítmica.	Índices de equidad algorítmica, diversidad de datos representados, tasas de error diferenciadas por grupo demográfico.
Resiliencia cibernética y seguridad algorítmica	Considera la protección de sistemas de IA frente a ataques adversarios que puedan alterar su funcionamiento o inducir errores cognitivos.	Vulnerabilidad algorítmica, efectividad de defensas, impacto de ataques adversarios en contextos críticos.	Robustez frente a ataques adversarios, continuidad operativa, tiempo de recuperación ante incidentes, eficacia de medidas de seguridad.

Fuente: Elaboración propia

Las dimensiones propuestas en la Tabla 6 ofrecen un marco metodológico sólido para evaluar cualitativa y cuantitativamente el impacto de la IA en la neuroética, con un enfoque orientado a la seguridad y la defensa. Este marco permite operacionalizar variables complejas y transformarlas en indicadores medibles, lo que posibilita un análisis riguroso y comparable entre distintos contextos. Por ejemplo, la autonomía cognitiva puede analizarse a través de indicadores como la percepción de control, la precisión de las decisiones y el tiempo de respuesta cognitiva, aspectos que son influenciados por el diseño y la supervisión de los sistemas autónomos (Holzinger et al., 2025).

La integridad mental, por su parte, puede evaluarse considerando niveles de estrés, estabilidad emocional y resiliencia frente a sesgos inducidos por IA. Estudios como el de Kuhn et al. (2021) demuestran la relevancia de proteger los procesos cognitivos frente a posibles influencias no deseadas en interacciones prolongadas con sistemas autónomos. En paralelo, los neuroderechos adquieren un papel central en la preservación de la privacidad mental y la identidad personal, proponiendo marcos regulatorios que prioricen la autonomía y la integridad cognitiva (Plá Herrero, 2025).

En el ámbito de la gobernanza ética, el grado de cumplimiento normativo y la confianza institucional actúan como indicadores clave para la sostenibilidad de los sistemas de IA. Modelos como el de Reddy et al. (2020) destacan la necesidad de estructuras regulatorias transparentes que integren evaluaciones de riesgo y mecanismos de supervisión. Del mismo modo, Kamila y Jasrotia (2025) advierten que anticipar y mitigar riesgos éticos es esencial para prevenir daños potenciales y preservar la legitimidad de la tecnología.

La equidad algorítmica se mide mediante índices de representación de datos y tasas de error diferenciadas, buscando evitar sesgos que comprometan la justicia cognitiva. Ghanem et al. (2025) subrayan que la integración de criterios de equidad es un requisito esencial para garantizar que la IA no reproduzca desigualdades estructurales, mientras que Binns (2018) aporta un marco normativo para orientar las decisiones algorítmicas hacia principios de justicia.

Finalmente, la resiliencia cibernética implica métricas de robustez y continuidad operativa, esenciales para proteger sistemas críticos frente a ciberataques o manipulación adversaria. Desidi et al. (2025) evidencian que la combinación de arquitecturas de redes neuronales con *blockchain* fortalece la capacidad de defensa de sistemas en entornos de alta amenaza. Estos elementos, integrados de

manera interdisciplinaria, facilitan el diseño de soluciones orientadas a salvaguardar la autonomía y la integridad mental en escenarios críticos de seguridad y defensa.

## Conclusiones

El análisis bibliométrico realizado con Bibliometrix revela que la investigación sobre IA aplicada a la neuroética presenta una estructura temática consolidada, en la que convergen desarrollos tecnológicos avanzados y un marco ético en crecimiento. Por un lado, se identifican algoritmos como *machine learning* y *deep learning* aplicados en contextos clínicos y de apoyo a la toma de decisiones, lo que evidencia un campo robusto en el diseño de soluciones técnicas orientadas a la salud y al procesamiento cognitivo. Por otro lado, emergen conceptos relacionados con la ética, la autonomía y los derechos humanos, que subrayan la preocupación por el impacto de estas tecnologías en la integridad mental y la autonomía cognitiva de las personas (Plá Herrero, 2025).

Los resultados muestran cómo la neuroética se consolida como un eje interdisciplinario clave para escenarios de seguridad y defensa. El protagonismo creciente de temas vinculados a los derechos humanos, la privacidad de datos, la evaluación de riesgos y la filosofía de la tecnología demuestra que el debate ético ya no es marginal, sino central en el desarrollo y la implementación de IA. En entornos donde la interacción entre sistemas autónomos y procesos humanos de toma de decisiones es crítica, garantizar la integridad mental frente a posibles interferencias se convierte en una prioridad (Ghanem et al., 2025).

Asimismo, la presencia de clústeres especializados en áreas como *adversarial machine learning* y procesamiento avanzado del lenguaje natural evidencia la necesidad de contar con sistemas robustos y resilientes ante amenazas externas. En contextos militares o de defensa, la manipulación algorítmica podría comprometer la estabilidad cognitiva de los operadores y la fiabilidad de las operaciones, situando la seguridad algorítmica como un componente estratégico. Del mismo modo, la equidad y la mitigación de sesgos emergen como factores determinantes para garantizar justicia cognitiva, dado que los sistemas de IA pueden replicar o amplificar sesgos presentes en los datos, afectando a poblaciones específicas y generando consecuencias éticas y operativas relevantes.

La identificación de elementos clave como la autonomía cognitiva, la integridad mental, la gobernanza ética, la equidad algorítmica y la resiliencia cibernética,

junto con sus dimensiones de medición, constituye un marco metodológico robusto para evaluar y diseñar soluciones interdisciplinarias que integren tecnología, psicología, derecho y ciberseguridad. Este marco no solo fortalece la capacidad de análisis y supervisión, sino que también proporciona herramientas prácticas para su aplicación en entornos reales (Desidi et al., 2025).

Finalmente, a pesar de la solidez de los hallazgos, la investigación presenta limitaciones derivadas de la dependencia de bases de datos específicas, lo que puede restringir la representatividad global. Asimismo, el enfoque bibliométrico no sustituye la validación empírica en escenarios operativos, lo que limita la generalización de los resultados. Futuras investigaciones deberían profundizar en estudios de campo que midan la efectividad de estos elementos en entornos militares, de ciberdefensa y de gestión de crisis, así como explorar métricas de neuroderechos y análisis comparativos de marcos regulatorios internacionales, con el fin de buscar la armonía entre la innovación tecnológica y la protección de la mente humana.

## Referencias

- Al Kuwaiti, A., Nazer, K., Al-Reedy, A., Al-Shehri, S., Al-Muhanna, A., Subbarayalu, A. V., Al Muhanna, D., & Al-Muhanna, F. A. (2023). A review of the role of artificial intelligence in healthcare. *Journal of Personalized Medicine, 13*(6). <https://doi.org/10.3390/jpm13060951>
- Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., Albahri, O. S., Alamoodi, A. H., Bai, J., Salhi, A., Santamaría, J., Ouyang, C., Gupta, A., Gu, Y., & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion, 96*, 156-191. <https://doi.org/10.1016/J.INFFUS.2023.03.008>
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research, 81*, 1-11.
- Comisión Europea. (2020). *White paper on artificial intelligence: A European approach to excellence and trust*. <https://tinyurl.com/46dj5up3>
- Corzo-Ussa, G. D., Álvarez-Aros, E. L., Mariño, J. P., & Amézquita-Gómez, N. (2023). Military artificial intelligence applied to sustainable development projects: Sound environmental scenarios. *DYNA, 90*(228), 115-122. <https://doi.org/10.15446/dyna.v90n228.108639>

- Desidi, N. R., Jyothi, D., Vijay, P. J., Kumar, M. K., & Lakshmi, R. V. (2025). Design of an improved method for intrusion detection using CNN, LSTM, and blockchain. *Journal of Theoretical and Applied Information Technology*, 103(1). <https://www.researchgate.net/publication/389314306>
- Díez-Gómez, D. A., Guillén, M., & Rodríguez, M. del P. (2019). Revisión de la literatura sobre la toma de decisiones éticas en organizaciones. *Información Tecnológica*, 30(3), 25-38. <https://doi.org/10.4067/s0718-07642019000300025>
- Eaton, S. E. (2025). Global trends in education: Artificial intelligence, postplagiarism, and future-focused learning for 2025 and beyond—2024-2025 Werklund Distinguished Research Lecture. *International Journal for Educational Integrity*, 21(1). <https://doi.org/10.1007/s40979-025-00187-6>
- Eke, D. (2024). Ethics and governance of neurotechnology in Africa: Lessons from AI. *JMIR Neurotechnology*, 3, e56665. <https://doi.org/10.2196/56665>
- Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, 21(5). <https://doi.org/10.2196/13216>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023018-9482-5>
- Francisco, M. (2023). Artificial intelligence for environmental security: National, international, human and ecological perspectives. *Current Opinion in Environmental Sustainability*, 61. <https://doi.org/10.1016/j.cosust.2022.101250>
- Friedrich, O., Seifert, J., & Schleidgen, S. (2021). AI-based self-tracking of the mind: Philosophical-ethical implications. *Psychiatrische Praxis*, 48, S42-S47. <https://doi.org/10.1055/a-1364-5068>
- Ghanem, S., Moraleja, M., Gravesande, D., & Rooney, J. (2025). Integrating health equity in artificial intelligence for public health in Canada: A rapid narrative review. *Frontiers in Public Health*, 13. <https://doi.org/10.3389/fpubh.2025.1524616>
- Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. *eBioMedicine*, 90, 104512. <https://doi.org/10.1016/j.ebiom.2023.104512>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6.ª ed.). McGraw-Hill.
- Holzinger, A., Zatloukal, K., & Müller, H. (2025). Is human oversight to AI systems still possible? *New Biotechnology*, 85, 59-62. <https://doi.org/10.1016/j.nbt.2024.12.003>

- Howard, A., & Borenstein, J. (2018). The ugly truth about ourselves and our robot creations: The problem of bias and social inequity. *Science and Engineering Ethics*, 24(5), 1521-1536. <https://doi.org/10.1007/S11948-017-9975-2/METRICS>
- Huang, R., Shi, L., Wu, Y., & Chen, T. (2024). Constructing content framework for artificial intelligence literacy instruction in China from a global perspective. *Documentation, Information and Knowledge*, 41(3), 27-37. <https://doi.org/10.13366/j.dik.2024.03.027>
- Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, 13(1). <https://doi.org/10.1186/s40504017-0050-1>
- Kamila, M. K., & Jasrotia, S. S. (2025). Ethical issues in the development of artificial intelligence: Recognizing the risks. *International Journal of Ethics and Systems*, 41(1), 45-63. <https://doi.org/10.1108/IJOES-05-2023-0107>
- Korteling, J. E. (Hans), Van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C., & Eikelboom, A. R. (2021). Human- versus artificial intelligence. *Frontiers in Artificial Intelligence*, 4. <https://doi.org/10.3389/frai.2021.622364>
- Kuhn, E., Fiske, A., Henningsen, P., & Buyx, A. (2021). Psychotherapy with an autonomous artificial intelligence: Ethical benefits and challenges. *Psychiatrische Praxis*, 48, S26-S30. <https://doi.org/10.1055/a-1369-2938>
- Ligthart, S., Ienca, M., Meynen, G., Molnar-Gabor, F., Andorno, R., Bublitz, C., Catley, P., Claydon, L., Douglas, T., Farahany, N., Fins, J. J., Goering, S., Haselager, P., Jotterand, F., Lavazza, A., McCay, A., Wajnerman Paz, A., Rainey, S., Ryberg, J., & Kellmeyer, P. (2023). Minding rights: Mapping ethical and legal foundations of 'neurorights'. *Cambridge Quarterly of Healthcare Ethics*, 32(4), 461-481. <https://doi.org/10.1017/s0963180123000245>
- Liu, J., Sandjaja, I., & Wunsch, D. C. (2024). AI trustworthy: Ethical challenges and strategies. En *Proceedings of the IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI 2024)*. <https://doi.org/10.1109/SOLI63266.2024.10956105>
- Plá Herrero, M. T. (2025). Neurorights: Legal relevance and regulation through comparative law. *Cuadernos de Derecho Transnacional*, 17(1), 631-653. <https://doi.org/10.20318/cdt.2025.9346>
- Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, 27(3), 491-497. <https://doi.org/10.1093/jamia/ocz192>
- Reep, M. J. (2024). Leveraging hybrid systems to forecast needs and analyze gaps for dual commercial and military use. En *2024 19th Annual System of Systems Engineering Conference (SoSE 2024)* (pp. 124-129). <https://doi.org/10.1109/SoSE62659.2024.10620960>

- Schwendicke, F., Samek, W., & Krois, J. (2020). Artificial intelligence in dentistry: Chances and challenges. *Journal of Dental Research*, 99(7), 769-774. <https://doi.org/10.1177/0022034520915714>
- Sobhi, N., Sadeghi-Bazargani, Y., Mirzaei, M., Abdollahi, M., Jafarizadeh, A., Pedrammehr, S., Alizadehsani, R., Tan, R. S., Islam, S. M. S., & Acharya, U. R. (2025). Artificial intelligence for early detection of diabetes mellitus complications via retinal imaging. *Journal of Diabetes and Metabolic Disorders*, 24(1). <https://doi.org/10.1007/s40200-025-01596-7>
- Tran, B. X., Vu, G. T., Ha, G. H., Vuong, Q. H., Ho, M. T., Vuong, T. T., La, V. P., Ho, M. T., Nghiem, K. C. P., Nguyen, H. L. T., Latkin, C. A., Tam, W. W. S., Cheung, N. M., Nguyen, H. K. T., Ho, C. S. H., & Ho, R. C. M. (2019). Global evolution of research in artificial intelligence in health and medicine: A bibliometric study. *Journal of Clinical Medicine*, 8(3). <https://doi.org/10.3390/jcm8030360>
- Wang, B., Rau, P. L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 42(9), 1324-1337. <https://doi.org/10.1080/0144929X.2022.2072768>
- Yuste, R., Goering, S., Y Arcas, B. A., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., Friesen, P., Gallant, J., Huggins, J. E., Illes, J., Kellmeyer, P., Klein, E., Marblestone, A., Mitchell, C., Parens, E., Pham, M., Rubel, A., Sadato, N., ... & Wolpaw, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature*, 551, 159-163. <https://doi.org/10.1038/551159a>