

Artículo de Investigación

# La evolución del procesamiento del lenguaje natural y su influencia en la inteligencia artificial: Una revisión y líneas de investigación futura

## The Evolution of Natural Language Processing and its Influence on Artificial Intelligence: A Review and Future Research Directions

Alberto Tomás Delso Vicente<sup>1</sup>: Universidad Rey Juan Carlos, España.

[alberto.delso@urjc.es](mailto:alberto.delso@urjc.es)

Marisol Carvajal Camperos: Universidad Rey Juan Carlos, España.

[marisol.carvajalc@urjc.es](mailto:marisol.carvajalc@urjc.es)

Daniel Ángel Corral De La Mata: Universidad Rey Juan Carlos, España.

[daniel.corral@urjc.es](mailto:daniel.corral@urjc.es)

Fecha de Recepción: 07/06/2024

Fecha de Aceptación: 26/07/2024

Fecha de Publicación: 10/12/2024

### Cómo citar el artículo

Delso Vicente, A. T., Carvajal Camperos, M. y Corral De La Mata, D. A. (2025). La evolución del procesamiento del lenguaje natural y su influencia en la inteligencia artificial: Una revisión y líneas de investigación futura [The Evolution of Natural Language Processing and its Influence on Artificial Intelligence: A Review and Future Research Directions]. *European Public & Social Innovation Review*, 10, 01-23. <https://doi.org/10.31637/epsir-2025-782>

### Resumen

**Introducción:** Este estudio revisa los desarrollos significativos en el procesamiento del lenguaje natural (PLN) y su impacto en la inteligencia artificial (IA), enfocándose en los avances en modelos de lenguaje, infraestructuras computacionales y la integración de métodos de aprendizaje automático. **Metodología:** Se realizó una revisión sistemática de la literatura utilizando las directrices PRISMA, centrada en artículos publicados entre 2022 y 2024. Se utilizó Web of Science, con términos de búsqueda como “procesamiento del lenguaje natural”, “PLN”. **Resultados:** La revisión destaca el papel crítico de los modelos de lenguaje avanzados

<sup>1</sup> Autor Correspondiente: Alberto Tomás Delso Vicente. Universidad Rey Juan Carlos (España).

como GPT-4, BERT y sus variantes en la mejora de la comprensión y generación del lenguaje natural, la importancia de infraestructuras de computación de alto rendimiento y el uso de técnicas de aprendizaje automático para optimizar tareas de PLN. **Discusión:** Los hallazgos confirman la relevancia de infraestructuras computacionales robustas y revelan nuevas perspectivas sobre la rápida evolución y adopción más amplia de técnicas de PLN en diversos sectores. **Conclusiones:** Es esencial continuar invirtiendo en infraestructuras computacionales y el desarrollo de modelos de lenguaje avanzados. La investigación futura debe ampliar el periodo de estudio, diversificar los idiomas, incluir literatura gris, realizar estudios longitudinales y explorar los desafíos de la ética y la privacidad en la implementación de técnicas de PLN.

**Palabras clave:** Procesamiento del Lenguaje Natural; Modelos de Lenguaje; Aprendizaje Automático; Inteligencia Artificial; Infraestructura Computacional; Revisión de literatura; Energía Renovables; Ciencia de Datos y Análisis.

### Abstract

**Introduction:** This study reviews the significant developments in natural language processing (NLP) and its impact on artificial intelligence (AI), focusing on advancements in language models, computational infrastructures, and the integration of machine learning methods. **Methodology:** A systematic literature review was conducted using the PRISMA guidelines, targeting articles from 2022 to 2024. Web of Science with search terms like “natural language processing”, “PNL”. **Results:** The review highlights the critical role of advanced language models such as GPT-4, BERT, and their variants in improving natural language understanding and generation, the importance of high-performance computing infrastructures, and the use of machine learning techniques to optimize NLP tasks. **Discussions:** The findings confirm the relevance of robust computational infrastructures and reveal new perspectives on the rapid evolution and broader adoption of NLP techniques across various sectors. **Conclusions:** Continued investment in computational infrastructures and the development of advanced language models is essential. Future research should expand the study period, diversify languages, include grey literature, conduct longitudinal studies, and explore the ethical and privacy challenges in implementing NLP techniques.

**Keywords:** Natural Language Processing; Language Models; Machine Learning; Artificial Intelligence; Computational Infrastructure; Literature review; Renewable energy; Data science and analytics.

## 1. Introducción

En la última década, el campo del procesamiento del lenguaje natural (PLN) ha experimentado un cambio significativo, impulsado no solo por avances en los modelos de lenguaje, sino también por transformaciones en las infraestructuras computacionales y un creciente interés en el aprendizaje automático (Giraldo Forero y Orozco Duque, 2023; Schaub, 2020). Este fenómeno ha capturado la atención tanto de investigadores como de profesionales, dado su potencial para revolucionar la manera en que se interactúa con la tecnología y se procesa el lenguaje humano (Bojórquez, 2021).

La industria del Procesamiento del Lenguaje Natural (PLN) ha experimentado cambios significativos debido a la integración de ciencias de datos y tecnologías avanzadas en el marketing digital y otros sectores (Plangger *et al.*, 2022; Shankar y Parsana, 2022). Este cambio ha creado un entorno donde la capacidad de las máquinas para comprender, interpretar y generar lenguaje humano de manera valiosa es crucial (Davenport y Ronanki, 2018). La adopción de tecnologías de PLN no solo promete mejorar la comunicación con las máquinas,

sino también permitir el análisis de grandes volúmenes de texto para obtener insights valiosos (Gill *et al.*, 2022).

La aplicación del PLN en el marketing digital ha revolucionado la manera en que las empresas interactúan con los consumidores. Herramientas avanzadas de PLN permiten analizar comentarios en redes sociales, reseñas de productos y otros textos generados por usuarios para comprender mejor las preferencias y comportamientos del consumidor (Ellis-Chadwick y Chaffey, 2022). Esto, a su vez, facilita la personalización de campañas de marketing y mejora la experiencia del cliente (Rust y Huang, 2014). Por otra parte, el PLN ha demostrado ser esencial en el desarrollo de asistentes virtuales y *chatbots*, que son cada vez más utilizados en el servicio al cliente. Estos sistemas pueden manejar consultas de clientes, proporcionar respuestas precisas y personalizadas, y aprender de las interacciones pasadas para mejorar continuamente su rendimiento (Adamopoulou y Moussiades, 2020).

Otra área crítica es el análisis de sentimientos, donde el PLN se utiliza para determinar las emociones detrás de los textos escritos por los usuarios. Esta capacidad es particularmente útil para monitorear la reputación de la marca y reaccionar rápidamente a posibles crisis (Liu, 2012).

El estudio del Procesamiento del Lenguaje Natural (PLN) es relevante debido a su impacto significativo en diversas industrias, como la salud, los servicios legales y el entretenimiento, lo que está redefiniendo cómo se gestionan y procesan los datos textuales (Ladeira, 2010; Hirschberg y Manning, 2015). Además, la infraestructura del PLN, que incluye tecnologías como modelos de lenguaje profundo, computación en la nube y hardware especializado como las unidades de procesamiento gráfico (GPU), ha permitido el procesamiento eficiente de datos textuales y la creación de modelos de lenguaje altamente precisos (Goyal *et al.*, 2018). La rápida adopción y evolución de estas tecnologías ha sido impulsada por la necesidad de analizar y generar lenguaje natural de manera más efectiva (Young *et al.*, 2018).

El procesamiento del lenguaje natural (PLN) se refiere a la capacidad de las máquinas para comprender, interpretar y generar lenguaje humano de manera que sea valiosa. Esta tecnología no solo está revolucionando la forma en que nos comunicamos con las máquinas, sino también cómo se analizan grandes volúmenes de texto para obtener insights valiosos (Gill *et al.*, 2022; Hirschberg y Manning, 2015). La infraestructura del PLN, que incluye modelos de lenguaje profundo, computación en la nube y hardware especializado como las unidades de procesamiento gráfico (GPU), ha permitido el procesamiento eficiente de datos textuales y la creación de modelos de lenguaje altamente precisos (Deng y Li, 2018). Estos avances han sido impulsados por la necesidad de analizar y generar lenguaje natural de manera más efectiva, beneficiando diversas industrias como la salud, los servicios legales y el entretenimiento (Ladeira, 2010; Young *et al.*, 2018).

El estudio del Procesamiento del Lenguaje Natural (PLN) es crucial debido a su impacto significativo en la industria y la transformación digital. La evolución de los ingresos y beneficios netos anuales de empresas clave en el desarrollo de tecnologías de PLN desde 1983 hasta 2023 ilustra cómo estas compañías han experimentado un crecimiento financiero notable. Este crecimiento refleja la creciente importancia económica del PLN y su papel central en el avance de la inteligencia artificial.

Se ha identificado una notable ausencia de estudios académicos que aborden de manera exhaustiva la evolución del Procesamiento del Lenguaje Natural (PLN) en los últimos tres años, especialmente en cuanto a infraestructuras computacionales, modelos de lenguaje y métodos de aprendizaje automático. Autores como Barman *et al.* (2023) y Rubino *et al.* (2017) han destacado la necesidad de investigar estos aspectos, señalando la falta de estudios que

integren estas áreas de manera cohesiva. Hirschberg y Manning (2015) también subrayan la importancia de una evaluación integral del progreso en PLN para comprender mejor su impacto y potencial en diversas industrias. Esta carencia en la literatura sugiere una oportunidad significativa para futuras investigaciones que proporcionen una visión más completa y actualizada de los desarrollos en PLN.

Este estudio se centrará en las siguientes preguntas de investigación:

1. ¿Cómo han evolucionado las infraestructuras computacionales para el PLN en los últimos tres años?
2. ¿Cuáles son los principales avances en los modelos de lenguaje y métodos de aprendizaje automático en PLN?
3. ¿Qué desafíos y oportunidades surgen de la integración de nuevas tecnologías en el PLN?

El objetivo principal de esta investigación es proporcionar una revisión sistemática de la evolución del PLN en los últimos tres años. Los objetivos específicos de esta investigación incluyen analizar las tendencias y avances en infraestructuras computacionales para el procesamiento del lenguaje natural (PLN), identificar y evaluar los principales modelos de lenguaje y métodos de aprendizaje automático, y examinar los desafíos y oportunidades que presentan estas tecnologías emergentes.

La originalidad de este artículo reside en el uso de gráficos visuales y matrices para ilustrar los avances en el Procesamiento del Lenguaje Natural (PLN) y su infraestructura de apoyo. La metodología adoptada es una revisión sistemática de la literatura (RSL) de los últimos cuarenta años, siguiendo el enfoque PRISMA, ampliamente reconocido y utilizado por autores como Moher *et al.* (2009) para estructurar revisiones sistemáticas de manera rigurosa y exhaustiva. La aplicación del método PRISMA garantizará una recopilación y análisis meticulosos de las investigaciones más relevantes en el campo del PLN. Se citarán los principales autores que han contribuido significativamente a este campo, proporcionando un análisis exhaustivo y actualizado de las tendencias y desarrollos. Posteriormente, se analizan los conceptos básicos del metaverso, las tecnologías inmersivas, la infraestructura digital, las energías renovables y la ciencia de datos y análisis.

El procesamiento del lenguaje natural, entendido como la capacidad de las máquinas para comprender y generar lenguaje humano, representa un avance significativo en la interacción con la tecnología (Giraldo Forero y Orozco Duque, 2023). Su desarrollo y adopción se basan en varias tecnologías clave, incluyendo los modelos de lenguaje profundo, la computación en la nube, y el hardware especializado (Cedeno-Moreno y Millan, 2023). Este marco teórico explora las principales áreas de estudio y desarrollo dentro del PLN, con un enfoque en la infraestructura computacional, los modelos de lenguaje y los métodos de aprendizaje automático (Schaub, 2020; Villamarín, 2024).

### **1.1. Modelos de Lenguaje**

Los modelos de lenguaje, como GPT-4 y BERT, son fundamentales para el Procesamiento del Lenguaje Natural (PLN) (Wei *et al.*, 2023; Wei, J., *et al.* 2023). Estos modelos permiten la creación de sistemas que pueden entender y generar lenguaje natural con una precisión sin precedentes (Radford *et al.*, 2019; Jing y Xu, 2019). Los avances en la arquitectura de redes neuronales, la mejora de los algoritmos de entrenamiento y el acceso a grandes volúmenes de

datos han permitido experiencias más realistas y útiles en aplicaciones de PLN (Devlin *et al.*, 2019; Brown *et al.*, 2020).

El avance en la resolución de los modelos, la reducción de la latencia y la mejora de los algoritmos de seguimiento han permitido experiencias más precisas y atractivas en aplicaciones de PLN (Huang *et al.*, 2024). Estos desarrollos han sido impulsados por la demanda de sistemas más inteligentes y la necesidad de herramientas más efectivas para el análisis y la generación de lenguaje (Nagda *et al.*, 2020). Además, la implementación de técnicas de aprendizaje profundo y transfer learning ha revolucionado el campo, permitiendo que los modelos de PLN sean aplicados en una variedad de contextos con resultados altamente eficientes (Vaswani *et al.*, 2017).

### **1.2. Infraestructura Computacional**

Esta colección de artículos explora diversos aspectos de la infraestructura y las aplicaciones del Procesamiento del Lenguaje Natural (PLN). Algunos estudios se centran en el desarrollo de herramientas y plataformas que facilitan la investigación y experimentación en PLN, como laboratorios en línea (López *et al.*, 2014) y plataformas de integración como InTiMe (Gómez, 2008). Otros trabajos discuten aplicaciones específicas del PLN, incluyendo el análisis de medios sociales y el apoyo a la investigación biomédica (Cedron *et al.*, 2018).

Se destaca la importancia de los recursos computacionales para las tareas de PLN y la necesidad de una representación estructurada del conocimiento (Martínez *et al.*, 1999). Además, se examina la naturaleza interdisciplinaria del PLN, incorporando conocimientos de ciencias de la computación, ciencias de la información y lingüística (Ladeira, 2010). La extracción de relaciones semánticas es identificada como una tarea clave dentro del PLN.

### **1.3. Energías Renovables**

A medida que el uso del PLN crece, también lo hace su consumo de energía. Las infraestructuras computacionales que soportan el PLN requieren una cantidad significativa de energía para funcionar, lo que plantea desafíos en términos de sostenibilidad ambiental (Ivanovski *et al.*, 2021). Integrar fuentes de energía renovable, como la solar y la eólica, en estas infraestructuras es esencial para mitigar su impacto ambiental. Las tecnologías de almacenamiento de energía, como las baterías de litio, están evolucionando rápidamente para satisfacer estas demandas crecientes de energía (Bolla *et al.*, 2010).

La adopción de energías renovables no solo reduce la huella de carbono del PLN, sino que también mejora la resiliencia de la infraestructura computacional al diversificar las fuentes de energía y reducir la dependencia de los combustibles fósiles. Además, la integración de tecnologías inteligentes de gestión de energía puede optimizar el uso de recursos y mejorar la eficiencia operativa (Song *et al.*, 2022).

### **1.4. Ciencia de Datos y Análisis**

La implementación de ciencia de datos y análisis avanzados está transformando el Procesamiento del Lenguaje Natural (PLN) (Sarker, 2021). Los datos generados a partir de interacciones en aplicaciones de PLN pueden ser utilizados para mejorar la personalización y la eficiencia de los servicios digitales. La inteligencia artificial y el aprendizaje automático permiten analizar grandes volúmenes de datos en tiempo real, proporcionando insights valiosos que pueden mejorar la experiencia del usuario y optimizar los recursos (Akter *et al.*, 2022).

Los algoritmos de aprendizaje automático pueden predecir patrones de comportamiento del usuario, personalizar experiencias y mejorar la seguridad en aplicaciones de PLN. Además, el análisis de big data puede identificar tendencias emergentes y facilitar la toma de decisiones informadas para el desarrollo de nuevas funcionalidades y servicios (Awan *et al.*, 2021; Shamim *et al.*, 2020).

### **1.5. Desafíos y Oportunidades**

A pesar de los avances tecnológicos, el Procesamiento del Lenguaje Natural (PLN) enfrenta varios desafíos. La estandarización de tecnologías y plataformas es crucial para garantizar la interoperabilidad y una experiencia de usuario consistente (Melluso *et al.*, 2022; Jurafsky y Martin, 2019). Además, existen preocupaciones sobre la privacidad y la seguridad de los datos, ya que el PLN implica la recolección y análisis de grandes cantidades de información personal (Weber, 2010).

Las oportunidades para el PLN son vastas. Desde la educación y la formación profesional hasta el entretenimiento y las redes sociales, el PLN tiene el potencial de transformar múltiples industrias (Cambria y White, 2014). La colaboración público-privada será esencial para abordar los desafíos y aprovechar las oportunidades que presenta esta nueva tecnología (Marr, 2021). El resto de este documento está organizado de la siguiente manera: La sección 2 presenta la metodología, la sección 3 los resultados, la sección 4 la discusión y por último, la sección 5 las conclusiones y las limitaciones y futuras líneas de investigación.

## **2. Metodología**

La elección de una revisión sistemática de la literatura (RSL) se justifica frente a otras metodologías debido a la necesidad de abordar un tema emergente y de gran actualidad, como es el procesamiento del lenguaje natural (PLN). Dado que este campo está en constante evolución y se espera que cambie significativamente el paradigma de interacción con la tecnología, es fundamental realizar un análisis exhaustivo y sistemático de la literatura existente. La RSL permite identificar, evaluar y sintetizar la evidencia de manera rigurosa, lo que resulta esencial para comprender las tendencias actuales y futuras del PLN.

### **2.1. Pasos de la Revisión Sistemática de la Literatura**

- i. Definición de Términos de Búsqueda: Los términos de búsqueda utilizados incluyen “Natural Language Processing” OR “NLP”.
- ii. Selección de Bases de Datos: Las bases de datos seleccionadas para la revisión incluyen Web of Science.
  - Resultados Iniciales: La búsqueda inicial arrojó un total de 6.111 artículos relevantes.
  - Criterios de Inclusión y Exclusión: Se incluyeron artículos publicados en los últimos cuarenta años (1983-2023), en inglés, y que aborden específicamente avances en PLN y sus aplicaciones. Se excluyeron artículos duplicados, aquellos que no estén disponibles en texto completo y estudios que no cumplan con los criterios de calidad establecidos.

## 2.2. Recogida de Datos

La recogida de datos se llevó a cabo entre el 1 de marzo y 30 de junio de 2024. Durante este período, se recopilaron y revisaron los artículos seleccionados para su análisis detallado.

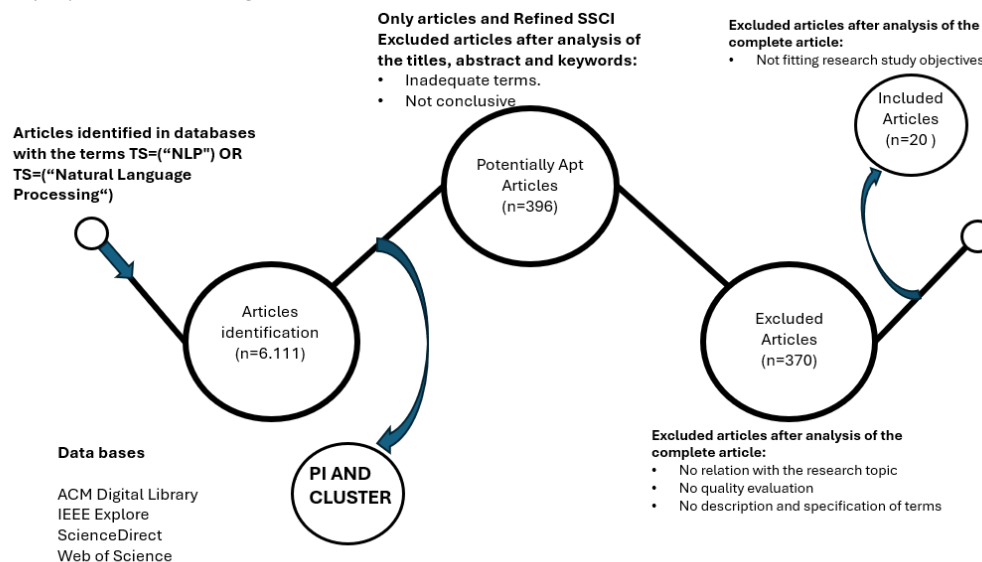
## 2.3. Proceso de filtrado

El proceso de filtrado se realizó siguiendo los pasos establecidos por el criterio PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). La Figura 1 señala el diagrama de flujo mediante la guía PRISMA, se detalla cada fase del filtrado:

1. **Identificación:** Se identificaron 5.111 artículos mediante las búsquedas en las bases de datos seleccionadas.
2. **Elegibilidad:** De los 5.111 artículos, excluyeron debido a que los términos no eran adecuados o no eran conclusivos, hasta que se evaluaron 396 artículos para verificar su pertinencia y calidad, resultando en la exclusión de 26 artículos que no cumplían con los criterios de inclusión, ya que no cumplían exactamente con los topics seleccionados o no cumplían con la calidad estipulada por los autores.
3. **Inclusión:** Finalmente, se incluyeron 370 artículos en la revisión sistemática. De los cuales, 10 fueron analizados en profundidad, según las citas, la metodología empleada y los usos. Por lo que se evaluaron en total 370, pero 10 se hizo un estudio recopilatorio en profundidad que queda plasmado en la tabla 2.

**Figura 1.**

Diagrama de flujo mediante la guía PRISMA



**Fuente:** Elaboración propia (2024).

La siguiente tabla resumen proporciona una visión clara y estructurada de los criterios y resultados de la revisión sistemática de la literatura:

**Tabla 1.**

*Resumen de la metodología*

PASO	DESCRIPCIÓN	RESULTADOS
DEFINICIÓN DE TÉRMINOS DE BÚSQUEDA	"Natural Language Processing" OR "NLP"	5.111 artículos identificados
BASES DE DATOS SELECCIONADAS	Web of Science	-
RESULTADOS INICIALES	Artículos relevantes encontrados	5924 artículos
ELEGIBILIDAD	Evaluación de pertinencia y calidad	396 artículos evaluados
INCLUSIÓN	Artículos incluidos en la RSL	370 artículos

**Fuente:** Elaboración propia (2024).

### 3. Resultados

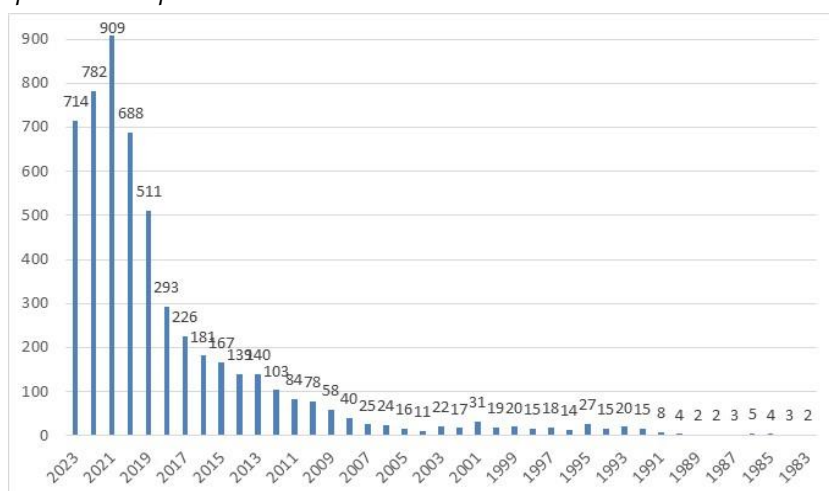
#### 3.1. Productividad

El análisis de los resultados obtenidos de la revisión sistemática de la literatura revela la creciente importancia del Procesamiento del Lenguaje Natural (PLN) en múltiples disciplinas. A continuación, se presenta un resumen detallado de las principales categorías de publicaciones, años de publicación, temas de citación, títulos de publicaciones y países que contribuyen significativamente a la investigación en PLN.

En la Figura 2, se muestra la productividad de publicación por año, la tendencia por año desde el 1983 hasta el 2023, muestra un incremento constante, alcanzando su punto máximo en 2021 con 909 artículos. Este aumento sostenido indica un interés creciente y continuo desarrollo en el campo del PLN. A pesar de que los datos de 2024 aún no están completos, ya se han registrado 376 publicaciones, lo que sugiere que la tendencia al alza se mantiene.

**Figura 2.**

*Productividad de publicación por año*



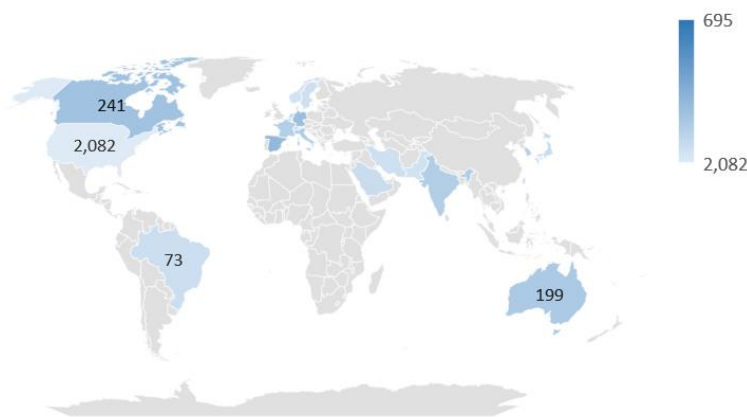
**Fuente:** Elaboración propia (2024).



En la Figura 3, muestra el mapa con la productividad por país. Estados Unidos lidera la producción académica en PLN con 2,082 artículos, seguido de China con 695 y Reino Unido con 456. Este liderazgo destaca la significativa contribución de estos países en el avance del PLN. Otros países como España, Alemania y Canadá también muestran una notable producción académica, subrayando la importancia global y la colaboración internacional en el desarrollo del PLN.

**Figura 3.**

*Productividad por país*



**Fuente:** Elaboración propia (2024).

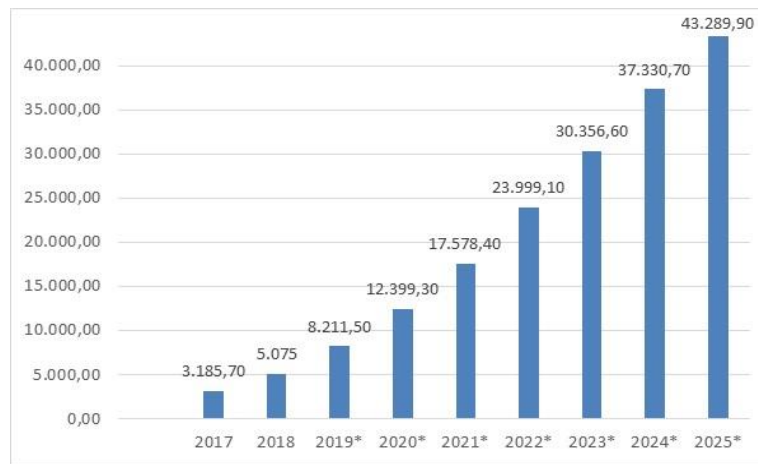
### 3.2. Mercado

El análisis de la literatura revela un crecimiento significativo en el desarrollo e implementación de tecnologías de PLN en los años 2014 al 2024. Modelos de lenguaje como GPT-4 y BERT han mejorado la comprensión y generación del lenguaje natural, proporcionando flexibilidad y escalabilidad necesarias para aplicaciones avanzadas.

La Figura 4, muestra los ingresos del mercado de procesamiento del lenguaje natural en todo el mundo incluyendo la proyección a 2025. Desde 2017, el mercado del procesamiento del lenguaje natural (PLN) ha crecido notable con ingresos que aumentaron en 3,185.70 millones de dólares estadounidenses. Este aumento se dio de manera significativa en los años posteriores, alcanzando 5,075 millones en 2018 y 8,211.50 millones en 2019. La tendencia alcista se mantuvo en 2020 con ingresos de 12,399.30 millones, y en 2021 se registraron 17,578.40 millones. Las proyecciones para los años siguientes indican que los ingresos seguirán creciendo, llegando a 23,999.10 millones en 2022, 30,356.60 millones en 2023, 37,330.70 millones en 2024 y finalmente 43,289.90 millones en 2025. Este crecimiento sostenido refleja la creciente adopción y demanda de tecnologías de PLN en diversas industrias y aplicaciones.

**Figura 4.**

*Ingresos del mercado de procesamiento del lenguaje natural en todo el mundo 2017-2025.  
(November 17, 2020).*



**Fuente:** Elaboración propia (2024).

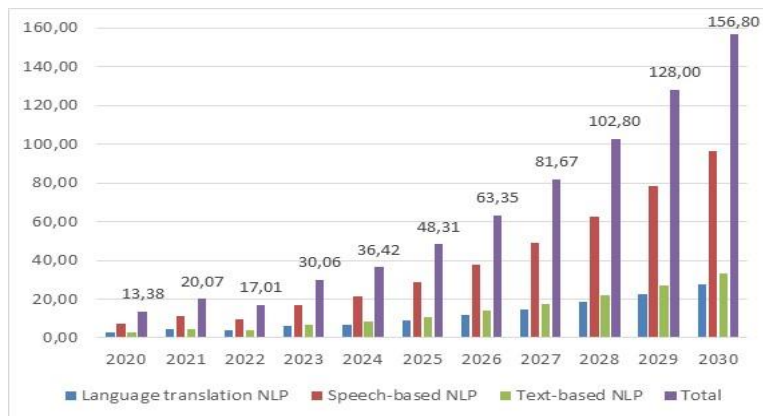
Además, la computación en la nube y las infraestructuras de alto rendimiento son cruciales para las aplicaciones avanzadas de PLN. La computación en la nube ofrece la flexibilidad y escalabilidad necesarias, mientras que las infraestructuras de alto rendimiento reducen la latencia al acercar el procesamiento de datos al lugar donde se generan, mejorando así la eficiencia y la experiencia del usuario.

Las tecnologías inmersivas, particularmente en el contexto del PLN, también han experimentado avances en términos de hardware y software. Las mejoras en la precisión de los modelos de lenguaje y la reducción de la latencia han permitido experiencias de lenguaje natural más realistas y útiles. Mientras, la integración de técnicas de aprendizaje automático ha generado un incremento en la adopción de tecnologías de PLN para una variedad de aplicaciones.

La Figura 5, la evolución del mercado del PLN desde 2020 hasta 2030. En 2020, el mercado se valoró en 13.38 mil millones de dólares estadounidenses, distribuidos entre traducción de lenguaje, PLN basado en voz y PLN basado en texto. En 2021, el mercado total ascendió a 20.07 mil millones. Aunque en 2022 hubo una ligera disminución a 17.01 mil millones, el mercado retomó su crecimiento en 2023 con 30.06 mil millones y continuó aumentando año tras año, alcanzando 48.31 mil millones en 2025 y 63.35 mil millones en 2026. Se proyecta que para 2030 el mercado alcanzará 156.80 mil millones de dólares estadounidenses. Este crecimiento está impulsado principalmente por las aplicaciones de PLN basadas en voz, seguidas de las de texto y las de traducción de lenguaje.

**Figura 5.**

*Comparación del tamaño del mercado en 2024*

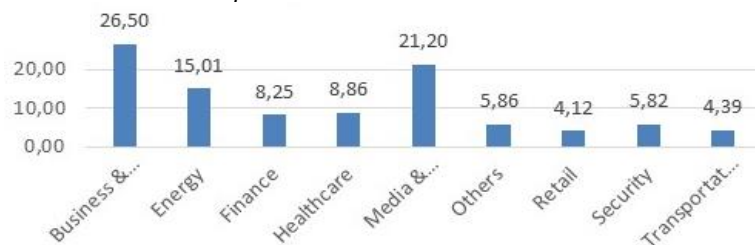


**Fuente:** Elaboración propia (2024).

La Figura 6, ilustra la distribución del mercado de Procesamiento de Lenguaje Natural (PLN) por industrias y sector en el año 2022. Se muestra la distribución del mercado de PLN por industrias en 2022, los servicios empresariales y legales lideran con una participación del 26.50%, lo que refleja la alta demanda de tecnologías de PLN en estos sectores. La industria de medios y entretenimiento ocupa el segundo lugar con un 21.20%, indicando un uso significativo de PLN en la creación y distribución de contenido. Otras industrias como la energía (15.01%), salud (8.86%), finanzas (8.25%) y seguridad (5.82%) también muestran una adopción considerable de tecnologías de PLN.

**Figura 6.**

*Comparación del tamaño del mercado por sectores en 2022*



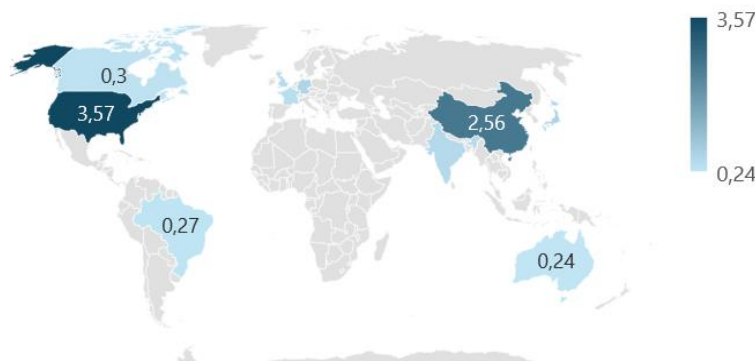
**Fuente:** Elaboración propia (2024).

En la Figura 7, se muestra el crecimiento en la última década del tamaño del mercado de procesamiento del lenguaje natural (PLN) en varios países. Estados Unidos lidera el mercado, con un incremento de 3.57 mil millones de USD en 2020 a 41.79 mil millones en 2030. Este crecimiento es reflejo del fuerte desarrollo y adopción de tecnologías de PLN en múltiples sectores de EE UU. China sigue de cerca con un crecimiento robusto, de 2.56 mil millones de USD en 2020 a 29.95 mil millones en 2030. Este incremento subraya el rápido avance tecnológico y la integración de PLN en diversas industrias chinas. Japón, aunque en menor escala, muestra un crecimiento notable, pasando de 0.59 mil millones de USD en 2020 a 6.96 mil millones en 2030, lo que resalta la importancia de esta tecnología en la innovación y desarrollo industrial del país. El Reino Unido y Alemania destacan en este análisis, con mercados que crecen de 0.43 mil millones a 5.06 mil millones de USD y de 0.57 mil millones a 6.71 mil millones de USD respectivamente para el año 2030. Estos países muestran una

adopción significativa de tecnologías de PLN en sectores clave como servicios legales y empresariales, así como en salud y servicios financieros. India presenta un crecimiento notable en el mercado de PLN, pasando de 0.47 mil millones de USD en 2020 a 5.56 mil millones en 2030, reflejando la creciente importancia de las tecnologías de PLN en el sector tecnológico y de comunicaciones. Francia y Brasil también experimentan un crecimiento significativo, con aumentos de 0.39 mil millones a 4.56 mil millones de USD y de 0.27 mil millones a 3.15 mil millones de USD respectivamente para el año 2030. Canadá y Australia muestran patrones similares de crecimiento, aumentando de 0.30 mil millones a 3.51 mil millones de USD y de 0.24 mil millones a 2.79 mil millones de USD respectivamente para 2030, subrayando la adopción creciente de tecnologías de PLN en múltiples sectores.

**Figura 7.**

*Mercado de PNL por país*



**Fuente:** Elaboración propia (2024).

La Tabla 2, basada en el análisis previo, resalta las principales contribuciones y aplicaciones académicas y prácticas de artículos claves en diversos campos.

**Tabla 2**

*Principales artículos y contribuciones en el campo de PNL*

AUTOR(ES)	JOURNAL	CITACIONES	METODOLOGÍA	DATOS	PRINCIPALES CONTRIBUCIONES	USOS
Clark <i>et al.</i> , (2020)	INFORMATION SYSTEMS RESEARCH	650	Masked language modeling (MLM), replaced token detection	Experimentos con modelos preentrenados en múltiples tareas de PLN	Proponen una tarea de preentrenamiento más eficiente que MLM, mejorando las representaciones contextuales	Desarrollo y mejora de modelos de PLN
Hannigan <i>et al.</i> , (2019)	JOURNAL OF MANUFACTURING TECHNOLOGY MANAGEMENT	650	Revisión sistemática y análisis de contenido	Análisis de 178 documentos utilizando procesamiento de lenguaje natural y análisis avanzado de texto	Identificación de principios de diseño y tendencias tecnológicas de la Industria 4.0; Propuesta de una hoja de ruta estratégica	Implementación de la Industria 4.0 en manufactura
Hartmaan <i>et al.</i> , (2019)	MANAGEMENT SCIENCE	438	Codificación de contenido, análisis de datos de redes sociales	Análisis de 106,316 mensajes de Facebook de 782 compañías	Encuentran que el contenido relacionado con la personalidad de la marca mejora el compromiso	Diseño de estrategias de contenido en redes sociales
Lee <i>et al.</i> , (2018)	JOURNAL OF MARKETING	330	Análisis textual automatizado	Revisión de interacciones textuales en diferentes contextos de marketing	Detallan cómo el análisis de texto puede generar conocimientos de marketing, unificando áreas	Generación de insights de marketing a partir de datos textuales
Li Y Xie, (2019)	JOURNAL OF MARKETING RESEARCH	292	Análisis empírico de contenido de imagen en redes sociales	Datos de Twitter e Instagram sobre aerolíneas y marcas de vehículos	Encuentran que la presencia de imágenes de alta calidad y profesionales incrementa el engagement	Mejora de la estrategia de contenido visual en redes sociales

Timoshenko Y Hauser , (2019)	MARKETING SCIENCE	218	aprendizaje automático para análisis cualitativo	análisis de contenido generado por usuarios (ugc) en productos de cuidado oral	demuestran que el ugc es valioso para identificar necesidades del cliente y que los métodos de aprendizaje automático mejoran la eficiencia	identificación de necesidades del cliente a partir de contenido generado por usuarios
Liu <i>et al.</i> , (2021)	JOURNAL OF BUSINESS RESEARCH	200	Análisis de big data en redes sociales	Análisis de 3.78 millones de tweets de 15 marcas de lujo en Twitter	Encuentran que el enfoque en entretenimiento, interacción y tendencias aumenta el compromiso del cliente	Diseño y gestión de marketing en redes sociales para marcas de lujo
Longoni Y Cian, (2022)	JOURNAL OF MARKETING	199	Análisis empírico de preferencias utilitarias y hedónicas	Estudios empíricos sobre recomendaciones de IA versus recomendaciones humanas	Identifican el efecto "word-of-machine", donde las preferencias utilitarias favorecen las recomendaciones de IA y las hedónicas no	Comprensión del impacto de las recomendaciones de IA en contextos utilitarios y hedónicos

**Fuente:** Elaboración propia (2024).

Los resultados de la revisión sistemática de la literatura destacan la creciente relevancia del Procesamiento del Lenguaje Natural (PLN) en múltiples disciplinas. Clark *et al.* (2020) propusieron una metodología de preentrenamiento más eficiente que el Masked Language Modeling (MLM), mejorando significativamente las representaciones contextuales en modelos de PLN, lo que es esencial para tareas como traducción automática y generación de texto. Hannigan *et al.* (2019) identificaron principios de diseño y tendencias tecnológicas cruciales para la Industria 4.0, ofreciendo una hoja de ruta estratégica para la implementación de tecnologías avanzadas de PLN en el sector manufacturero.

En el ámbito del marketing digital, Hartmaan *et al.* (2019) demostraron que la personalidad de marca en contenidos de redes sociales aumenta el compromiso del consumidor, proporcionando una base para el diseño de estrategias de contenido más efectivas. Lee *et al.* (2018) señalaron cómo el análisis textual automatizado puede unificar áreas fragmentadas del marketing, generando insights valiosos a partir de interacciones textuales. Li y Xie (2019) destacaron que las imágenes de alta calidad en redes sociales incrementan el compromiso, subrayando la importancia de la calidad visual en la estrategia de contenido digital.

Timoshenko y Hauser (2019) encontraron que el contenido generado por usuarios (UGC) es crucial para identificar necesidades del cliente, y que los métodos de aprendizaje automático pueden mejorar la eficiencia en este proceso. Liu *et al.* (2021) enfatizaron que un enfoque en entretenimiento, interacción y tendencias es clave para aumentar el compromiso del cliente en el marketing de marcas de lujo en redes sociales. Finalmente, Longoni y Cian (2022) exploraron el efecto "word-of-machine", destacando que las recomendaciones de IA son preferidas en contextos utilitarios, mientras que las recomendaciones humanas son más valoradas en contextos hedónicos, ofreciendo valiosos insights para la personalización de experiencias de usuario.

Estos estudios no solo subrayan la versatilidad del PLN en diversas aplicaciones, sino que también demuestran su capacidad para transformar prácticas empresariales y mejorar la interacción con los consumidores.

A pesar de estos avances, el desarrollo del PLN enfrenta varios desafíos. La falta de estándares unificados para tecnologías y plataformas ha sido un obstáculo significativo, afectando la interoperabilidad entre diferentes sistemas y dispositivos. La recopilación y análisis de grandes cantidades de datos personales plantean importantes preocupaciones sobre la privacidad y la seguridad. Así como la falta de regulaciones unificadas y la incertidumbre en la inversión requieren una mayor cooperación público-privada para ser superados.

## 4. Discusión

La interpretación de los resultados obtenidos a partir del análisis de la literatura sobre la evolución del PLN y su infraestructura resalta la relevancia de recientes avances tecnológicos. Estos hallazgos no solo validan tendencias de estudios anteriores, sino que también plantean una nueva perspectiva y retos para el futuro desarrollo del PLN.

### 4.1. Implicaciones de los Hallazgos

**Infraestructura Computacional:** El crecimiento significativo de las infraestructuras de alto rendimiento y la adopción de la computación en la nube han sido fundamentales para soportar las aplicaciones de PLN. Estos avances no solo mejoran la capacidad de procesamiento de datos y reducen la latencia, sino que también proporcionan la flexibilidad y escalabilidad necesarias para las aplicaciones de PLN. Estos hallazgos corroboran las teorías previas sobre la importancia de infraestructuras computacionales robustas para el desarrollo del PLN.

**Modelos de Lenguaje:** Las mejoras en la precisión de los modelos de lenguaje y la reducción de la latencia han permitido experiencias de PLN más realistas y útiles. Estos avances tecnológicos no solo han impulsado el análisis y la generación de lenguaje, sino que también han encontrado aplicaciones en la educación y la industria. Estos resultados amplían el cuerpo de conocimiento existente, mostrando que los modelos de lenguaje están evolucionando rápidamente y están siendo adoptados en una variedad de sectores más amplia de lo que se había anticipado previamente.

**Energías Renovables:** La integración de fuentes de energía renovable, como la solar y la eólica, en la infraestructura del PLN, ha demostrado ser una estrategia efectiva para reducir la huella de carbono. Estos hallazgos no solo son consistentes con estudios anteriores sobre la sostenibilidad de las infraestructuras computacionales, sino que también destacan la importancia de avanzar en las tecnologías de almacenamiento de energía para satisfacer las crecientes demandas del PLN. Este enfoque hacia la sostenibilidad es crucial para mitigar el impacto ambiental y mejorar la resiliencia energética de estas infraestructuras.

### 4.2. Relevancia Práctica y teórica

Los resultados de este estudio tienen implicaciones tanto prácticas como teóricas. En términos prácticos, destacan la necesidad de seguir invirtiendo en infraestructuras computacionales y modelos de lenguaje para el crecimiento del PLN, integrando energías renovables para garantizar la sostenibilidad. Desde una perspectiva teórica, los hallazgos amplían nuestra comprensión de cómo las tecnologías emergentes pueden ser utilizadas para crear experiencias digitales más inmersivas y eficientes. También subrayan la importancia de la interoperabilidad y la estandarización en el desarrollo de estas tecnologías.

## 5. Conclusiones

El presente estudio ha sintetizado los avances más significativos en la evolución del procesamiento del lenguaje natural, destacando la importancia de las infraestructuras computacionales, los modelos de lenguaje y la integración de energías renovables. Los hallazgos clave incluyen el crecimiento exponencial de las infraestructuras de alto rendimiento y la computación en la nube, los avances en modelos de lenguaje como GPT-4 y BERT, y la adopción de fuentes de energía renovable para mejorar la sostenibilidad del PLN.

Los resultados de este estudio aportan significativamente al cuerpo de conocimiento existente en varias maneras. En primer lugar, se ha demostrado la crucial importancia de las infraestructuras de alto rendimiento y la computación en la nube y *edge computing* en el soporte de aplicaciones de PLN. Estos avances facilitan experiencias inmersivas en tiempo real, ampliando las aplicaciones del PLN en diversos sectores. En segundo lugar, las mejoras en modelos de lenguaje no solo están revolucionando el análisis y la generación del lenguaje, sino también la educación y la formación profesional, abriendo nuevas oportunidades para el aprendizaje y el trabajo remoto. En tercer lugar, la integración de energías renovables en las infraestructuras del PLN es una estrategia esencial para reducir la huella de carbono, subrayando la necesidad de continuar avanzando en tecnologías de almacenamiento de energía para satisfacer las demandas futuras. La aplicación de algoritmos de aprendizaje automático y la inteligencia artificial en el PLN está optimizando los recursos y personalizando las experiencias de usuario, demostrando el poder transformador de la ciencia de datos en este ámbito.

Estos resultados no solo confirman tendencias observadas en investigaciones previas, sino que también introducen nuevas perspectivas y desafíos para el futuro desarrollo de este campo. Se ha dado respuesta a todas las preguntas de investigación planteadas:

- ¿Cuáles son las tendencias recientes en el desarrollo del PLN? Las tendencias recientes incluyen el crecimiento de las infraestructuras computacionales de alto rendimiento y la adopción de la computación en la nube. Estos avances permiten una mayor capacidad de procesamiento y reducen la latencia, lo que es crucial para las aplicaciones de PLN. Además, la integración de fuentes de energía renovable en las infraestructuras de PLN ha demostrado ser una estrategia efectiva para reducir la huella de carbono y mejorar la sostenibilidad.
- ¿Qué avances tecnológicos han sido más significativos en los últimos tres años? Los avances más significativos incluyen las mejoras en la precisión y eficiencia de los modelos de lenguaje como GPT-4 y BERT, así como la adopción de la computación en la nube y *edge computing*. Estos avances han permitido experiencias de PLN más realistas y útiles, ampliando su aplicación en la educación, la industria y otros sectores.
- ¿Cuáles son los principales desafíos que enfrenta el desarrollo del PLN? Los principales desafíos incluyen la estandarización de tecnologías y plataformas para garantizar la interoperabilidad, así
- como la privacidad y seguridad de los datos debido a la recolección y análisis de grandes cantidades de información personal. Además, la necesidad de avanzar en tecnologías de almacenamiento de energía es crucial para satisfacer las crecientes demandas del PLN de manera sostenible.

Se alcanzó el objetivo principal propuesto pues se realizó una revisión sistemática de la evolución del PLN en los últimos tres años. Este objetivo se ha cumplido al documentar los avances en infraestructuras computacionales, modelos de lenguaje y la integración de energías renovables, así como al identificar las tendencias, desafíos y oportunidades en el desarrollo del PLN. En cuanto a los objetivos específicos propuestos se cumplieron todos, se analizaron las tendencias y avances en infraestructuras digitales y se ha identificado la importancia de las infraestructuras de alto rendimiento y la computación en la nube para el soporte de aplicaciones de PLN, facilitando experiencias inmersivas en tiempo real.

Se evaluó el impacto de las tecnologías inmersivas en el desarrollo del PLN y se comprobó que las mejoras en los modelos de lenguaje no solo están revolucionando el análisis y la generación del lenguaje, sino también la educación y la formación profesional, abriendo nuevas oportunidades para el aprendizaje y el trabajo remoto. Se identificaron los desafíos y oportunidades en la integración de energías renovables en el PLN y se comprobó que la integración de energías renovables en las infraestructuras del PLN es una estrategia esencial para reducir la huella de carbono, subrayando la necesidad de continuar avanzando en tecnologías de almacenamiento de energía para satisfacer las demandas futuras.

Basado en los hallazgos del estudio, se proponen varias recomendaciones para la práctica y la política. Es crucial continuar invirtiendo en el desarrollo y la expansión de infraestructuras computacionales, incluyendo redes de alto rendimiento y tecnologías de computación en la nube y edge computing, para soportar el crecimiento del PLN. Además, los desarrolladores y planificadores de infraestructuras deben priorizar la integración de fuentes de energía renovable y tecnologías de almacenamiento de energía para garantizar la sostenibilidad a largo plazo del PLN. También es necesario desarrollar y adoptar estándares unificados para tecnologías y plataformas en el PLN para mejorar la interoperabilidad y la experiencia del usuario. Las políticas y regulaciones deben enfocarse en proteger la privacidad y la seguridad de los datos de los usuarios, asegurando que las innovaciones tecnológicas no comprometan los derechos de los individuos.

Para seguir avanzando en el conocimiento y desarrollo del PLN, se sugieren varias áreas para futuras investigaciones. Realizar estudios longitudinales que sigan la evolución de las tecnologías y aplicaciones del PLN proporcionará datos sobre tendencias y cambios continuos. Incluir estudios en múltiples idiomas y contextos culturales obtendrá una visión más global y diversa del desarrollo del PLN. Además, investigar innovaciones emergentes reportadas en literatura gris o no revisada por pares capturará prácticas industriales y desarrollos recientes que aún no han sido formalmente documentados. Es fundamental explorar el impacto social y económico del PLN en diferentes comunidades y sectores, evaluando tanto los beneficios como los posibles riesgos y desventajas. Finalmente, examinar las barreras y oportunidades para la estandarización de tecnologías y plataformas desarrollará soluciones que faciliten una mayor interoperabilidad y mejoren la experiencia del usuario.

### *5.1. Limitaciones del estudio*

A pesar de los hallazgos significativos, este estudio tiene varias limitaciones que deben ser consideradas. Primero, la revisión se limitó a artículos publicados entre 2022 y 2024, lo que puede excluir investigaciones relevantes anteriores que podrían ofrecer una perspectiva histórica más amplia. Además, la inclusión de solo estudios en inglés puede sesgar los hallazgos hacia investigaciones predominantemente de países anglófonos, excluyendo potencialmente valiosas contribuciones de otros contextos lingüísticos y culturales.

### *5.2. Áreas para Futuras Investigaciones*

Para superar estas limitaciones y explorar nuevas preguntas surgidas del estudio actual, se sugieren varias áreas para futuras investigaciones:

1. Ampliación del Período de Estudio: Incluir estudios de periodos anteriores para proporcionar una perspectiva más completa sobre la evolución del PLN y sus tecnologías asociadas.



2. Diversificación de Idiomas: Incluir estudios en múltiples idiomas para capturar una visión más global y diversa del desarrollo del PLN.
3. Investigación de Literatura Gris: Incorporar literatura gris y estudios no revisados por pares para captar innovaciones emergentes y prácticas industriales que aún no han sido formalmente documentadas en la literatura académica.
4. Estudios Longitudinales: Realizar estudios longitudinales que sigan el desarrollo del PLN y sus tecnologías en el tiempo, proporcionando datos sobre tendencias y cambios a largo plazo.
5. Evaluación de Impacto: Investigar más a fondo el impacto de la integración de energías renovables y las tecnologías de almacenamiento de energía en la sostenibilidad del PLN.
6. Interoperabilidad y Estandarización: Explorar las barreras y oportunidades para la estandarización de tecnologías y plataformas en el PLN, con el objetivo de mejorar la interoperabilidad y la experiencia del usuario.

## 6. Referencias

- Adamopoulou, E. y Moussiades, L. (2020). *An overview of chatbot technology*. En IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 373-383). Springer, Cham. [https://doi.org/10.1007/978-3-030-49186-4\\_31](https://doi.org/10.1007/978-3-030-49186-4_31)
- Akter, S., Michael, K., Uddin, M. R., McCarthy, G. y Rahman, M. (2022). Transforming business using digital innovations: The application of AI, blockchain, cloud and data analytics. *Annals of Operations Research*, 1-33. <https://doi.org/10.1007/s10479-020-03620-w>
- Awan, U., Shamim, S., Khan, Z., Zia, N. U., Shariq, S. M., y Khan, M. N. (2021). Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance. *Technological Forecasting and Social Change*, 168. <https://doi.org/10.1016/j.techfore.2021.120766>
- Barman, P., Dutta, L., Bordoloi, S., Kalita, A., Buragohain, P., Bharali, S. y Azzopardi, B. (2023). Renewable energy integration with electric vehicle technology: A review of the existing smart charging approaches. *Renewable and Sustainable Energy Reviews*, 183. <https://doi.org/10.1016/j.rser.2023.113518>
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., y Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, 84(1), 1-25. <https://doi.org/10.1177/0022242919873106>
- Bojórquez, D. M. (2021). De redes neuronales recurrentes a modelos de lenguaje: la evolución del pln en la generación de textos. *Publicación*, 4, octubre de 2021. <https://110.22201/dgtic.26832968e.2021.4.1>
- Bolla, R., Bruschi, R., Davoli, F. y Cucchietti, F. (2010). Energy efficiency in the future internet: a survey of existing approaches and trends in energy-aware fixed network infrastructures. *IEEE Communications Surveys and Tutorials*, 13(2), 223-244. <https://doi.org/10.1109/SURV.2011.071410.00073>

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P. y Amodei, D. (2020). *Language models are few-shot learners*. *Advances in Neural Information Processing Systems*, 33, 1877-1901. <https://arxiv.org/abs/2005.14165>
- Cambria, E. y White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational intelligence magazine*, 9(2), 48-57. <https://doi.org/10.1109/MCI.2014.2307227>
- Cedeno-Moreno, D. E. y Millan, A. (2023). Arquitectura de PLN aplicada al contexto de la salud mental. *I+ D Tecnológico*, 19(2), 24-29. <https://doi.org/10.33412/idt.v19.2.3770>
- Cedron, F., Carballal, A., Fernandez-Lozano, C., Munteanu, C. y Pazos, A. (2018). *Infrastructure to support biomedical applications*. <https://doi.org/10.3390/mol2net-04-05507>
- Clark, K., Luong, M. T., Le, Q. V. y Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv*, 2003.10555. <https://doi.org/10.48550/arXiv.2003.10555>
- Davenport, T. H. y Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108-116. <https://blockqai.com>
- Deng, L. y Liu, Y. (Eds.). (2018). *Deep learning in natural language processing*. Springer. <https://doi.org/10.1007/978-981-10-520-5>
- Devlin, J., Chang, M. W., Lee, K. y Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, arXiv:1810.04805. <https://doi.org/10.48550/arXiv.1810.04805>
- Ellis-Chadwick, F. y Chaffey, D. (2012). *Digital marketing: strategy, implementation and practice*. Pearson. <https://lontar.ui.ac.id/detail?id=20419965>
- García-Martínez, J. A., Herrera-Villalobos, G. y Fallas-Vargas, M. A. (2021). Aprender conectados: Un estudio sobre las redes personales de aprendizaje de estudiantes universitarios. *Educatio Siglo XXI*, 39(2), 41-60. <https://doi.org/10.6018/educatio.463821>
- Ghobakhloo, M. (2018). The future of manufacturing industry: a strategic roadmap toward Industry 4.0. *Journal of Manufacturing Technology Management*, 29(6), 910-936. <https://doi.org/10.1108/JMTM-02-2018-0057>
- Gill, S. S., Kumar, A., Singh, H., Singh, M., Kaur, K., Usman, M. y Buyya, R. (2022). Quantum computing: A taxonomy, systematic review and future directions. *Software: Practice and Experience*, 52(1), 66-114. <https://doi.org/10.1002/spe.3039>
- Gill, S. S., Xu, M., Ottaviani, C., Patros, P., Bahsoon, R., Shaghaghi, A., Golec, M., Stankovski, V., Wu, H., y Abraham, A. (2022). AI for next generation computing: Emerging trends and future directions. *Internet of Things*, 19, 100514. <https://doi.org/10.1016/j.iot.2022.100514>
- Giraldo Forero, A. F. y Orozco Duque, A. F. (2023). Evolución del procesamiento natural del lenguaje. *Tecnológicas*, 26(56). <https://doi.org/10.22430/22565337.2687>

- Gómez, J. M. (2008). InTiMe: plataforma de integración de recursos de PLN. *Procesamiento Del Lenguaje Natural*, 40.
- Goyal, P., Pandey, S. y Jain, K. (2018). *Deep learning for natural language processing*. Apress. <https://doi.org/10.1007/978-1-4842-3685-7>
- Hannigan, T. R., Haans, R. F., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S. y Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, 13(2), 586-632. <https://doi.org/10.5465/annals.2017.0099>
- Hartmann, J., Huppertz, J., Schamp, C. y Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20-38. <https://doi.org/10.1016/j.ijresmar.2018.09.009>
- Hirschberg, J. y Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261-266. <https://doi.org/10.1126/science.aaa8685>
- Huang, S., Dong, L., Wang, W., Hao, Y., Singhal, S., Ma, S., Lv, T., Cui, L., Mohammed, O. K. y Patra, B. (2024). *Language is not all you need: Aligning perception with language models*. *Advances in Neural Information Processing Systems*, 36. <https://proceedings.neurips.com>
- Ivanovski, K., Hailemariam, A. y Smyth, R. (2021). The effect of renewable and non-renewable energy consumption on economic growth: Non-parametric evidence. *Journal of Cleaner Production*, 286. <https://doi.org/10.1016/j.jclepro.2020.124956>
- Jing, K. y Xu, J. (2019). A survey on neural network language models. *ArXiv Preprint ArXiv:1906.03591*. <https://doi.org/10.48550/arXiv.1906.03591>
- Jurafsky, D. y Martin, J. H. (2019). *Speech and Language Processing* (3ª ed.). Prentice Hall. <https://web.stanford.edu/~jurafsky/slp3>
- Ladeira, A. P. (2010). *Processamento de linguagem natural: caracterização da produção científica dos pesquisadores brasileiros* [Tesis de doctorado]. <http://hdl.handle.net/1843/ECID-8B3Q6C>
- Lee, D., Hosanagar, K. y Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science*, 64(11), 5105-5131. <https://doi.org/10.1287/mnsc.2017.2902>
- Li, Y. y Xie, Y. (2020). Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research*, 57(1), 1-19. <https://doi.org/10.1177/0022243719881113>
- Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature. DOI: <https://doi.org/10.1007/978-3-031-02145-9>

- Liu, X., Shin, H. y Burns, A. C. (2021). Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. *Journal of Business Research*, 125, 815-826. <https://doi.org/10.1016/j.jbusres.2019.04.042>
- Longoni, C. y Cian, L. (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The “word-of-machine” effect. *Journal of Marketing*, 86(1), 91-108. <https://doi.org/10.1177/0022242920957347>
- López, J., Sánchez-Sánchez, C. y Villatoro-Tello, E. (2014). Laboratorio en línea para el procesamiento automático de documentos. *RCS*, 72, 1-10. [https://rcs.cic.ipn.mx/2014\\_72/RCS\\_72\\_2014.pdf](https://rcs.cic.ipn.mx/2014_72/RCS_72_2014.pdf)
- Marr, B. (2020). *Tech Trends in Practice: The 25 technologies that are driving the 4<sup>a</sup> Industrial Revolution*. John Wiley y Sons.
- Martínez, P., García-Serrano, A. y de Miguel Castaño, A. (1999). Estructuración del Conocimiento para la Interpretación de Textos y su Aplicación al Diseño de Esquemas Conceptuales de Bases de Datos. *Inteligencia Artificial*, 3(8), 36-58.
- Melluso, N., Grangel-González, I. y Fantoni, G. (2022). Enhancing industry 4.0 standards interoperability via knowledge graphs with natural language processing. *Computers in Industry*, 140. <https://doi.org/10.1016/j.compind.2022.103676>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. y PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264-269. <https://doi.org/10.1016/j.ijisu.2010.02.007>
- Nagda, K., Mukherjee, A., Shah, M., Mulchandani, P. y Kurup, L. (2020). *Ascent of pre-trained state-of-the-art language models*. Advanced Computing Technologies and Applications: Proceedings of 2nd International Conference on Advanced Computing Technologies and Applications – ICACTA 2020, 269–280. [https://doi.org/10.1007/978-981-15-3242-9\\_26](https://doi.org/10.1007/978-981-15-3242-9_26)
- Plangger, K., Grewal, D., de Ruyter, K. y Tucker, C. (2022). The future of digital technologies in marketing: A conceptual framework and an overview. *Journal of the Academy of Marketing Science*, 50(6), 1125-1134. <https://doi.org/10.1007/s11747-022-00906-2>
- Radford, A., Metz, L. y Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint*, arXiv:1511.06434. <https://doi.org/10.48550/arXiv.1511.06434>
- Rubino, L., Capasso, C. y Veneri, O. (2017). Review on plug-in electric vehicle charging architectures integrated with distributed energy sources for sustainable mobility. *Applied Energy*, 207, 438-464. <https://doi.org/10.1016/j.apenergy.2017.06.097>
- Rust, R. T. y Huang, M. H. (2014). The service revolution and the transformation of marketing science. *Marketing Science*, 33(2), 206-221. <https://doi.org/10.1287/mksc.2013.0836>
- Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, 2(5), 377. <https://doi.org/10.1007/s42979-021-00765-8>

- Schaub, L. P. (2020). La industria del lenguaje en la era del dato. *Ábaco*, 103, 82-89. <https://www.jstor.org/stable/10.2307/27135841>
- Shamim, S., Zeng, J., Khan, Z. y Zia, N. U. (2020). Big data analytics capability and decision-making performance in emerging market firms: The role of contractual and relational governance mechanisms. *Technological Forecasting and Social Change*, 161. <https://doi.org/10.1016/j.techfore.2020.120315>
- Shankar, V. y Parsana, S. (2022). An overview and empirical comparison of natural language processing (NLP) models and an introduction to and empirical application of autoencoder models in marketing. *Journal of the Academy of Marketing Science*, 50(6), 1324-1350. <https://doi.org/10.1007/s11747-022-00840-3>
- Song, L., Hu, X., Zhang, G., Spachos, P., Plataniotis, K. N. y Wu, H. (2022). Networking systems of AI: On the convergence of computing and communications. *IEEE Internet of Things Journal*, 9(20), 20352-20381. <https://doi.org/10.1109/IIOT.2022.3172270>
- Timoshenko, A. y Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, 38(1), 1-20. <https://doi.org/10.1287/mksc.2018.1123>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gómez, A. N. y Polosukhin, I. (2017). Attention Is All You Need. arXiv preprint arXiv:1706.03762.
- Villamarín, A. T. (2024). Big data en ciencias sociales. Una introducción a la automatización de análisis de datos de texto mediante procesamiento de lenguaje natural y aprendizaje automático. *Revista CENTRA de Ciencias Sociales*, 3(1). <https://doi.org/10.54790/rccs.51>
- Weber, R. H. (2010). Internet of Things–New security and privacy challenges. *Computer Law y Security Review*, 26(1), 23-30. <https://doi.org/10.1016/j.clsr.2009.11.008>
- Wei, C., Wang, Y.-C., Wang, B. y Kuo, C.-C. J. (2023). *An overview on language models: Recent developments and outlook*. ArXiv Preprint ArXiv:2303.05759. <https://doi.org/10.48550/arXiv.2303.05759>
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S. y Fedus, W. (2022). *Emergent abilities of large language models*. arXiv preprint arXiv:2206.07682. <https://doi.org/10.48550/arXiv.2206.07682>
- Young, T., Hazarika, D., Poria, S. y Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3), 55-75. <https://doi.org/10.1109/MCI.2018.2840738>

## CONTRIBUCIONES DE AUTORES/AS, FINANCIACIÓN

**Conceptualización:** Delso Vicente, Alberto; Carvajal Camperos, Marisol; Corral De La Mata Daniel Ángel. **Análisis formal:** Delso Vicente, Alberto; Carvajal Camperos, Marisol; Corral De La Mata Daniel Ángel. **Redacción-Preparación del borrador original:** Delso Vicente, Alberto; Carvajal Camperos, Marisol; Corral De La Mata Daniel Ángel. **Redacción-Revisión y Edición:** Delso Vicente, Alberto; Carvajal Camperos, Marisol; Corral De La Mata Daniel Ángel. **Supervisión:** Delso Vicente, Alberto; Carvajal Camperos, Marisol. **Nombres Administración de proyectos:** Delso Vicente, Alberto; Carvajal Camperos, Marisol; Corral De La Mata Daniel Ángel. **Todos los/as autores/as han leído y aceptado la versión publicada del manuscrito:** Delso Vicente, Alberto; Carvajal Camperos, Marisol; Corral De La Mata Daniel Ángel. **Financiación:** Esta investigación no financiamiento externo. **Conflicto de intereses:** no hay conflicto de interés.

### AUTORES:

#### **Alberto Tomás Delso Vicente**

Universidad Rey Juan Carlos, España.

Experiencia en gestión de operaciones con enfoque en la optimización de recursos y la eficiencia operativa. Habilidades analíticas, de gestión y comunicación, así como experiencia en gestión de proyectos, liderando proyectos exitosos internacionales y cumpliendo con regulaciones industriales. Comprometido con la innovación. Estudiante de Doctorado, con un MBA en Business Administration, además de un Grado en Ingeniería Química. Ha trabajado como Ingeniero Logístico y Mantenimiento de cisternas en empresas internacionales. La experiencia académica se complementa con el rol de profesor y coordinador en la Universidad Rey Juan Carlos, donde se han desarrollado habilidades en gestión educativa y liderazgo. Destaca en áreas como análisis de datos, CRM, ERP, y marketing digital. Además, cuenta con competencias avanzadas en idiomas: Inglés, Italiano y alemán.

[alberto.delso@urjc.es](mailto:alberto.delso@urjc.es)

**Orcid ID:** <https://orcid.org/0009-0000-6410-1132>

#### **Marisol Carvajal Camperos**

Universidad Rey Juan Carlos, España.

PhD en Administración y Dirección de Empresas por la Universidad Complutense de Madrid, cum laude y mención internacional. Ingeniera Industrial homologada en España, MBA por el Instituto de Estudios Bursátiles (IEB) de Madrid, y especialista en Recursos Humanos y Gerencia de Empresas por universidades en Venezuela y Colombia. Profesora y conferenciante internacional, ha trabajado en Westfield Business School, CEIPA Business School y la Universidad Rey Juan Carlos (URJC). Ha publicado artículos en revistas JCR y Scopus, y es revisora de la revista por pares ciego. Ha sido Vicedecana, Decana y Vicerrectora en Euro-EAD Madrid, y es Directora de la Maestría en Gestión de Talento Humano en EIG-UIDES. Ha desempeñado puesto de alta dirección en empresas de América Latina y en España.

[marisol.carvajalc@urjc.es](mailto:marisol.carvajalc@urjc.es)

**Orcid ID:** <https://orcid.org/0000-0001-9639-4136>

**Scopus:** <https://www.scopus.com/authid/detail.uri?authorId=57226771620>

### **Daniel Ángel Corral De La Mata**

PhD en Administración y Dirección de Empresas por la Universidad Rey Juan Calos (URJC), con calificación cum laude. Experiencia avalada en más de 30 años de actividad en el sector financiero, desempeñando funciones de Dirección Comercial tanto a nivel de red como de Dirección Territorial. Experiencia en puestos de responsabilidad a nivel de Servicios Centrales, en departamentos tales como "Marketing", "Canales Alternativos", "Medios de Pago", "Control de Gestión y Desarrollo de Negocio", "Internet y Banca online" y "Asset Management". Profesor de Universidad de la Rey Juan Carlos (URJC). Ha publicado artículos en revistas JCR y Scopus. Miembro del equipo de investigación: MARPRISO y del grupo de Innovación docente consolidado en inteligencia de datos, sistemas de la información y nuevas tendencias.  
[daniel.corral@urjc.es](mailto:daniel.corral@urjc.es)

**Orcid ID:** <https://orcid.org/0000-0002-1375-0092>

**ResearcherID:** <https://www.webofscience.com/wos/author/record/HKF-7223-2023>