

INNOVACIÓN EN LA INGENIERÍA DE REQUISITOS IMPULSADO POR LOS GRANDES MODELOS DE LENGUAJE

Leonardo Espinosa Arévalo¹

¹ Facultad de Matemáticas, Universidad Autónoma de Yucatán. Inn, Anillo Periférico Norte tablaje 13615, Chuburna de Hidalgo, 97119 Mérida, Yuc.

Autor de correspondencia: 1a24216469@alumnos.uady.mx

RESUMEN

Los grandes modelos de lenguaje (LLMs) están transformando la Ingeniería de Software al introducir nuevas formas de aprendizaje, automatización y aseguramiento de la calidad. Este ensayo analiza críticamente el impacto de los LLMs en los procesos de Ingeniería de Requisitos, destacando su papel como herramientas de innovación en la elicitación y verificación de requisitos. Se revisan estudios recientes que demuestran la integración de estos modelos en la generación de Historias de Usuario, la identificación de requisitos no funcionales y la evaluación de su calidad bajo marcos como QUS e INVEST. Asimismo, se abordan los dilemas éticos, técnicos y de dependencia tecnológica derivados de su uso. Los resultados sugieren que los LLMs deben concebirse como asistentes complementarios y no sustitutos del ingeniero de software, promoviendo una sinergia entre la inteligencia humana y la artificial.

Finalmente, se plantean líneas futuras orientadas a la especialización y adaptación de modelos para dominios específicos de la ingeniería de requisitos, con el fin de optimizar la calidad del software y los procesos ágiles de desarrollo.

Palabras clave: Innovación, Grandes Modelos de Lenguaje, Requisitos, Software.

INTRODUCCIÓN

Un modelo de lenguaje es un agente computacional diseñado originalmente para predecir la próxima palabra en una oración (Minaee, et al., 2025). Este tipo de inteligencia artificial, que ha escalado por sus capacidades conversacionales, ha impulsado el desarrollo de agentes más robustos. Los ahora llamados grandes modelos de lenguaje (LLMs por sus siglas en inglés) consisten en sistemas con miles de millones de parámetros entrenados con corpus masivos de texto utilizando un aprendizaje auto supervisado (Xin Zhao, et al., 2025). Este tipo de tecnología representa un gran avance en habilidades de procesamiento de lenguaje natural (PLN), las cuales ya están siendo implementadas en un gran número de industrias (Wang & Shao, 2025), así como en la investigación académica, donde se observa un aumento del estado del arte.

Por su parte, la ingeniería de software (IS) se define como un proceso sistemático y disciplinario para el diseño, desarrollo y mantenimiento de software de calidad (IEEE Standard Glossary of Software

Engineering Terminology, 1990). Dentro de la IS, la fase de Ingeniería de Requisitos (IR) es ampliamente reconocida como una de las más críticas y de mayor apalancamiento. Esto se debe a su énfasis, especialmente en metodologías ágiles, en identificar "el qué" y "el porqué" del sistema. La IR establece los lineamientos para las fases de diseño y desarrollo. El mal manejo de los requisitos puede llevar a una deuda técnica (Biazotto, Feitosa, Avgeriou, & Nakagawa, 2025). Asimismo, puede conducir a la pérdida significativa de calidad durante los procesos y del resultado final.

Históricamente, la IR ha sido un proceso manual, intensivo en trabajo y fundamentalmente conversacional (con el cliente o usuario final), propenso a errores humanos comunes, propios del lenguaje natural como la ambigüedad, la estimación imprecisa y las inconsistencias (Dalpiaz, van der Schalk, & Lucassen, 2018). Dentro de las metodologías ágiles, a través de artefactos como las Historias de Usuario (HU), se ha mejorado la flexibilidad y el enfoque en el valor para el cliente, ofreciendo un mayor enlace entre las necesidades del usuario y las especificaciones técnicas. Sin embargo, la baja calidad dentro de artefactos como las HU genera atrasos en los procesos. Es en este contexto de alta complejidad y alto impacto donde los LLMs emergen como una alternativa viable para este contexto.

El propósito de este ensayo es analizar críticamente el estado del arte e identificar cómo los grandes modelos de lenguaje (LLMs) están transformando la innovación en los procesos de Ingeniería de Requisitos, redefiniendo particularmente el papel del ingeniero de

software y del analista de negocio en las etapas tempranas del desarrollo. El texto se organiza en las siguientes secciones: un marco conceptual que introduce los fundamentos teóricos; un análisis del papel de los LLMs en la formación, automatización y aseguramiento de la calidad del software; la exposición de los principales riesgos, limitaciones y dilemas; una reflexión crítica sobre la dimensión innovadora de esta tecnología; y, finalmente, las conclusiones y líneas de trabajo futuro.

MARCO CONCEPTUAL

2.1. Los Grandes Modelos de Lenguaje.

Los grandes modelos de lenguaje (LLMs) forman parte de una rama de la inteligencia artificial conocida como procesamiento del lenguaje natural (Young, Hazarika, Poria, & Cambria, 2018). El PLN está enfocado en la realización de técnicas y herramientas que permiten a los sistemas computacionales interpretar, entender y replicar el lenguaje humano. Impulsados por la arquitectura Transformer (Vaswani, et al., 2017), los LLMs modernos (como los de las series GPT, Gemini o Llama) (Xin Zhao, et al., 2025) han demostrado capacidades notables, así como modelos BERT de clasificación. Más allá de su arquitectura específica (codificadores, decodificadores o una combinación), lo relevante para la Ingeniería de Requisitos son sus capacidades emergentes como: comprensión semántica; generación de lenguaje coherente y clasificación.

2.2. Innovación en la ingeniería de requisitos.

La innovación en el DS se refiere a la serie de ideas, herramientas, procesos o tecnologías novedosas destinadas a mejorar la eficiencia, calidad, agilidad o entrega de valor de los sistemas de software (Yatish, Jiarpakdee, Thongtanunam, & Tantithamthavorn, 2019).

Si bien la innovación puede ocurrir en cualquier etapa (codificación, pruebas, despliegue), específicamente en la IR tiene potencial en la implementación de nuevas herramientas, sobre todo impulsadas por el PLN. Ya que, la elicitación de requisitos, suele ser un proceso conversacional y de entendimiento técnico con un lenguaje amigable para los clientes. Es decir, suele suceder en un lenguaje natural. Un error detectado en etapas de desarrollo o de mantenimiento es mucho más costoso que en las etapas tempranas del desarrollo (Boehm & Basili, 2001). Por ello metodologías ágiles como SCRUM ofrecen el análisis de sus listas de requisitos (backlogs) (Iivari & Iivari, 2011), y determinar el valor de cada una de las HU y su calidad. Siendo sometidos a marcos de calidad como INVEST o Quality User Story (Lucassen, Dalpiaz, van der Werf, & Brinkkemper, 2016).

3. Los LLMs como herramienta de innovación en la ingeniería de requisitos.

Los LLMs están inmersos en numerosas áreas y de distintas maneras. Hoy en día representan un enfoque viable para la IS y específicamente para el área de elicitación de requisitos []. Podemos distinguir múltiples áreas, aunque el enfoque de

experimentación en el estado del arte apunta a tres ramas importantes.

3.1. LLMs como herramientas de aprendizaje.

Los esfuerzos de una inteligencia artificial general (IAG) aún son sometidos a evaluación. Actualmente, los modelos más conocidos (como GPT, Gemini, DeepSeek) aunque se trate de modelos robustos, aún dependen de la construcción correcta de un prompt y de las técnicas de reducción de alucinaciones (Huang, et al., 2025). Esto convierte al ingeniero de requisitos en el autor y guía principal del proceso, asistiéndose de las distintas herramientas. Actualmente, son conocidas como una herramienta ideal para la generación de resúmenes y el apoyo en el aprendizaje.

Un estudio reportó entrevistas cualitativas a profesores en áreas de IS, se ha percibido un aumento en la implementación de LLMs como herramientas de estudio y se muestran con una actitud positiva ante la inteligencia artificial. Sin embargo, persisten preocupaciones éticas, legales y de privacidad (Clear, et al., 2024).

Este mismo estudio ofrece un panorama del estado del arte evaluando los marcos de competencia y las exigencias del entendimiento de la inteligencia artificial (IA). Por ende entendemos dos puntos clave del aprendizaje impulsado por los LLMs:

- Aprendizaje de los lineamientos y de la teoría de la IS y de la IR. Es decir, utilizar estas tecnologías para aprender de los temas vinculados a los procesos de software.
- Aprendizaje de herramientas de IA. La industria solicita más ingenieros

capacitados en el manejo de LLMs (Gao, Hu, Gao, Xia, & Jin, 2024). abajo futuro.

El aprendizaje de procesos de IR impulsado por LLMs representa un aspecto crucial en las nuevas competencias. No obstante, es necesario determinar aspectos éticos y enfrentar la dependencia de estas tecnologías, ya que el conocimiento del ingeniero debe ser evidenciable (Clear, et al., 2024).

3.2. Generación de requisitos de software.

Tradicionalmente, la elicitación de requisitos implica un proceso de entrevistas, entendimiento con el cliente y culminar con la redacción de las funcionalidades y características que el usuario necesita en su sistema de software (Arora, Grundy, & Abdelrazek, 2024). Una mala interpretación de las necesidades lleva a que, en el mejor de los casos, se ofrezca un software de calidad que no resuelve el problema en cuestión (Saleem, Asim, Van Elst, & Dengel, 2025). Los procesos de elicitación funcionan en mayor grado con lenguaje natural. Es ahí donde los LLMs tienen relevancia ofreciendo:

- Generación de HU: A partir de transcripciones de entrevistas y prompts de especificaciones, puede ofrecer borradores de historias de usuario siguiendo estructuras y bajo el cumplimiento de ciertas características narrativas (Zadenoori, Dąbrowski, Alhoshan, Zhao, & Ferrari, 2025).
- Identificar requisitos no funcionales: Conociendo el dominio de negocio o las especificaciones del cliente y su giro o marca, pueden inferir opciones viables e identificar requisitos no

funcionales (Saleem, Asim, Van Elst, & Dengel, 2025).

- Propuestas de herramientas: mediante un prompt adecuado, identificar posibles preguntas y orientar a entrevistas más robustas y de mayor enriquecimiento para el problema.

Estudios demuestran que los LLMs en áreas como empresas de alto crecimiento suelen ser útiles. Sin embargo, aún requieren una supervisión humana y una integración que vaya más allá de una única tecnología, ya que algunas aún cuentan con deficiencias como alucinaciones, sesgos e incluso falta de privacidad para las necesidades de los clientes (Ojo, Hedayati, Jambharunkar, & Chimezie, 2024).

3.3. Verificación de la calidad impulsado por LLMs.

Estudios secundarios indican la implementación de modelos de aprendizaje profundo para clasificación, detección de vulnerabilidades y otras faltas, con base en un marco de calidad para los requisitos funcionales (Hendriana, Raharjo, & Nurfitriani, 2024). El enfoque mediante LLMs se vuelve crucial para la verificación de la calidad, un tema que permanece relevante y está en aumento. Actualmente, la producción de software adopta un enfoque ágil, es decir, entregas continuas y tempranas de valor en el sistema. Este ritmo de producción relega la documentación y el aseguramiento de la calidad.

Se realizó un análisis, mediante un estudio secundario basado en el marco de (Petersen, Vakkalanka, & Kuzniarz, 2015) sobre la verificación de la calidad de las historias de usuario, evidenciando un

aumento significativo en las investigaciones recientes, principalmente en los últimos años.

Existen diferentes propuestas para la verificación de la calidad, algunas reflejan el uso de modelos de clasificación como BERT (Borges Scoggin & Torres Marques-Neto, 2024), sin embargo, las propuestas no han sido dirigidas con marcos de calidad de HU o de requisitos funcionales, lo que lo vuelve incierta sus aplicaciones en producción. Otros estudios, ofrecen un panorama de la verificación de la calidad mediante modelos generativos como GPT y en el marco de Quality User Story. Sin embargo, estos estudios reportan alucinaciones y, aún, un campo de divergencia entre modelos generales, y la adaptación de los modelos a los campos específicos (Sharma & Tripathi, 2025).

La verificación de la calidad se vuelve crucial al tratarse de requisitos que comúnmente puedan ser interpretados por el equipo técnico experto y el cliente o usuario final. Esto hace que puedan arrastrar ambigüedades y faltas como atomicidad (que un requisito solo hable de una funcionalidad) o duplicidad y dependencia (un requisito obstruya la realización de otro) (Dalpiaz, van der Schalk, & Lucassen, 2018). El uso de LLMs como herramienta de apoyo en esta etapa relegada de verificación representa un filtro adicional para el aseguramiento de la documentación en el desarrollo de software.

4. Los dilemas de la implementación de LLMs.

Los estudios citados para las distintas capacidades y formas de implementación

reflejan que, si bien resulta una herramienta útil, en rápido crecimiento y, sobre todo, destinada a aligerar la carga de trabajo del ingeniero de software. Aún representa un cúmulo de cuestiones por resolver.

4.1. La ética en el uso de LLMs.

Los modelos de lenguaje forman parte de una ola tecnológica. Lo anterior contribuye a que su uso sea cada vez más cotidiano (Arora, Grundy, & Abdelrazek, 2024). Es por ello que se debe determinar las cuestiones éticas, profesionales y de educación, que implican.

Los LLMs no son una herramienta de reemplazo intelectual del ingeniero de software o del experto en cuestión de los requisitos. La forma de utilizarlo influye en las capacidades a adquirir (Clear, et al., 2024). El ser humano debe encargarse de dirigir e interpretar la información generada por los modelos de IA. Es por ello que no solo existe una cuestión ética en la construcción de los modelos, sino también en el usuario que los implementa en su marco de aprendizaje o de trabajo (Johnson & Hyland-Wood, 2024).

No obstante, aún se cuentan con limitantes para los que desarrollan los modelos, cuestiones como la privacidad hacen que existan reacciones adversas de la comunidad hacia estas tecnologías.

4.2. Limitantes en el caso de uso de elicitación.

Los modelos actuales aún cuentan con limitantes dentro del caso de uso de elicitación de requisitos. En particular, los modelos generativos presentan sesgos, falta de explicabilidad, alucinaciones,

errores semánticos y problemas de privacidad de datos (Clear, et al., 2024).

Del mismo modo, el uso de LLMs de propósito general genera la necesidad de la supervisión humana, así como la aplicación de enfoques híbridos entre múltiples tecnologías de la misma índole, cada una con mayores habilidades en algún área.

5.- Conclusiones y trabajos futuros.

Los grandes modelos de lenguaje representan un gran salto en la elicitación de requisitos: desde el aprendizaje y la capacitación del ingeniero de software, pasando por la asistencia en la creación de requisitos funcionales, hasta el aseguramiento de la calidad. Se puede inferir que, gracias a sus capacidades de procesamiento del lenguaje natural (PLN), estos modelos se están convirtiendo en la herramienta predilecta (Hendriana, Raharjo, & Nurfitriani, 2024). Tanto en el campo de la investigación, como en la industria, empieza a existir la generalización e implementación de dicha tecnología a los marcos de trabajo y de desarrollo de software ágil. Sin embargo, persisten desafíos significativos en torno a la implementación. Aspectos éticos como el abuso de estos modelos, la baja supervisión o la infracción de la privacidad. También aspectos de limitación tecnológica, tales como alucinaciones y sesgos. Es por eso que se recomienda implementarlos como herramientas de asistencia, no de sustitución de labores, ya que la intervención humana sigue siendo crucial, aún cuando los modelos aligeren la carga y ayuden en la realización del software de calidad en plazos ágiles.

En un panorama de investigación, algunos de los trabajos futuros son la especialización de modelos a los problemas que busca resolver el aseguramiento de la calidad de los documentos de software. Implementando ajustes finos a los modelos y, nuevamente, proporcionar modelos menos sesgados y orientados a la industria.

Finalmente, se plantea realizar un análisis comparativo y estadístico de los modelos generativos de propósito general, con el objetivo de identificar sus alcances, limitaciones y posibles aplicaciones en el ámbito de la ingeniería de requisitos. El propósito último de estas investigaciones es fomentar la mejora continua y la innovación en los procesos de desarrollo de software.

REFERENCIAS

- Arora, C., Grundy, J., & Abdelrazek, M. (2024). Advancing Requirements Engineering Through Generative AI: Assessing the Role of LLMs. *Generative AI for Effective Software Development*.
- Biazotto, J. P., Feitosa, D., Avgeriou, P., & Nakagawa, E. Y. (2025). Understanding practitioners' reasoning and requirements for efficient tool support in technical debt management. *Empirical Software Engineering*.
- Boehm, B., & Basili, V. R. (2001). Top 10 list [software development. *Computer*.
- Borges Scoggin, S., & Torres Marques-Neto, H. (2024). Identifying Valid User Stories Using BERT Pre-trained Natural Language Models. *Information Systems and Technologies*.
- Clear, T., Cajander, Å., Clear, A., McDermott, R., Daniels, M., Divitini, M., . . . Zhu, T. (2024). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *Working Group Reports on Innovation and Technology in Computer Science Education*, 34-64.

- Dalpiaz, F., van der Schalk, I., & Lucassen, G. (2018). Pinpointing Ambiguity and Incompleteness in Requirements Engineering via Information Visualization and NLP. *Requirements Engineering: Foundation for Software Quality*.
- Gao, C., Hu, X., Gao, S., Xia, X., & Jin, Z. (2024). The Current Challenges of Software Engineering in the Era of Large Language Models. *ACM Transactions on Software Engineering and Methodology*.
- Hendriana, A., Raharjo, T., & Nurfitriani, A. (2024). Approaches in Determining User Story Quality through Requirement Elicitation : A Systematic Literature Review. *Indonesian Journal of Computer Science*.
- Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., . . . Liu, T. (2025). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Transactions on Information Systems*, 1-55.
- IEEE Standard Glossary of Software Engineering Terminology. (1990). IEEE Std 610.12.
- Iivari, J., & Iivari, N. (2011). The relationship between organizational culture and the deployment of agile methods. *Information and Software Technology*.
- Johnson, S., & Hyland-Wood, D. (2024). A Primer on Large Language Models and their Limitations. *arXiv*.
- Kitchenham, B. A., Budgen, D., & Brereton, P. (2015). Evidence-Based Software Engineering and Systematic Reviews November 2015. Chapman & Hall/CRC.
- Lucassen, G., Dalpiaz, F., van der Werf, J. M., & Brinkkemper, S. (2016). Improving agile requirements: the Quality User Story framework and tool. *Requirements Engineering*.
- Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2025). Large Language Models: A Survey. *arXiv*.
- Ojo, S. O., Hedayati, H., Jambharunkar, T. B., & Chimezie, U. (2024). Leveraging Generative AI for Requirement Gathering in Software Startups. *Scientific Work for Computer Science*.
- Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*.
- Saleem, S., Asim, M. N., Van Elst, L., & Dengel, A. (2025). Generative language models potential for requirement engineering applications: insights into current strengths and limitations. *Complex & Intelligent Systems*.
- Sharma, A., & Tripathi, A. K. (2025). Evaluating user story quality with LLMs: a comparative study. *Journal of Intelligent Information Systems*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention Is All You Need.
- Wang, S., & Shao, Y. (2025). Integrative innovation of large language models in industries: technologies, applications, and challenges. *Data Science and Management*.
- Xin Zhao, W., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., . . . Li, Y. (2025). A Survey of Large Language Models. *arXiv*.
- Yatish, S., Jiarpakdee, J., Thongtanunam, P., & Tantithamthavorn, C. (2019). Mining Software Defects: Should We Consider Affected Releases? 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE).
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent Trends in Deep Learning Based Natural Language Processing. *IEEE Computational Intelligence Magazine*.
- Zadenoori, M. A., Dąbrowski, J., Alhoshan, W., Zhao, L., & Ferrari, A. (2025). Large Language Models (LLMs) for Requirements Engineering (RE): A Systematic Literature Review. *arXiv*.