



Calidad de software en la era de la IA: tendencias, estándares y prácticas emergentes



Autores:

Luis Alberto Urrea
Didier Herney Velez Bustos
Jonathan Herney Velez Bustos
Juan Sebastian Lozada Ceballos
Jamy Alexander Tole Morales
Yohan Daniel Timaná Quenguan

Resumen

El presente artículo tuvo como objetivo analizar las principales tendencias, estándares y prácticas emergentes de calidad del software en el contexto de la inteligencia artificial (IA) durante el periodo 2020–2025, marcado por la consolidación de sistemas basados en aprendizaje automático (Machine Learning, ML) y, más recientemente, por la expansión de los modelos fundacionales y de lenguaje (LLM). Estas tecnologías han transformado los supuestos clásicos de la ingeniería del software, introduciendo nuevos desafíos relacionados con la incertidumbre, la explicabilidad y la gestión de riesgos socio-técnicos.

La metodología utilizada consistió en una revisión documental sistemática de literatura científica y de experiencias industriales relevantes publicadas durante el periodo de análisis.

Entre los resultados más destacados se identificaron: (1) la necesidad de ampliar los modelos de calidad tradicionales para integrar dimensiones éticas y de riesgo técnico; (2) el fortalecimiento de las prácticas de prueba, despliegue y monitoreo en entornos MLOps/LLMOps,

con especial énfasis en la trazabilidad y la observabilidad de los modelos; y (3) el avance de la automatización del aseguramiento de la calidad mediante el uso de IA generativa en los procesos de validación, documentación y mantenimiento del software.

Con base en estos hallazgos, se propone una guía práctica (checklist) que orienta a las organizaciones a alinear sus prácticas de ingeniería y aseguramiento de calidad con los marcos normativos y regulaciones recientes aplicables al desarrollo responsable de sistemas de inteligencia artificial.

En conclusión, la calidad del software en la era de la IA requiere un enfoque integral, que combine marcos normativos sólidos, metodologías de ingeniería basadas en riesgo y una cultura de gobernanza responsable, garantizando la confiabilidad, transparencia y sostenibilidad de los sistemas inteligentes.

Palabras clave: calidad de software; inteligencia artificial; MLOps; riesgo; gobernanza tecnológica.

Abstract

The objective of this review article was to analyze the main trends, standards, and emerging practices in software quality within the context of artificial intelligence (AI) during the period 2020–2025, marked by the consolidation of machine learning (ML) systems and the expansion of foundational and language models (LLMs). These technologies have transformed the classical assumptions of software engineering, introducing new challenges related to uncertainty, explainability, and socio-technical risk management.

The methodology consisted of a systematic documentary review of scientific literature and relevant industrial experiences published during the analysis period.

The main findings include: (1) the need to expand traditional quality models to integrate ethical and technical risk dimensions; (2) the strengthening of testing, deployment, and monitoring practices in MLOps/LLMOps environments, with emphasis on model traceability and observability; and (3) the progress in automating quality assurance through the use of generative AI in validation, documentation, and maintenance processes.

Based on these findings, a practical checklist is proposed to help organizations align their engineering and quality assurance practices with recent regulatory and governance frameworks for responsible AI development.

In conclusion, software quality in the AI era requires a comprehensive approach that combines solid normative frameworks, risk-based engineering methodologies, and a culture of responsible governance to ensure the reliability, transparency, and sustainability of intelligent systems.

Keywords: software quality; artificial intelligence; MLOps; risk; responsible governance.

Introducción

La evolución de la ingeniería del software se encuentra actualmente en un punto de inflexión sin precedentes. Durante décadas, los sistemas informáticos se desarrollaron bajo principios deterministas: los requisitos se definían de forma explícita, el diseño seguía una secuencia controlada y la calidad se verificaba frente a especificaciones predefinidas. En este

paradigma, la calidad se asociaba con atributos como la corrección funcional, la fiabilidad, la mantenibilidad o la eficiencia del código.

Sin embargo, con la irrupción de **sistemas basados en Machine Learning (ML)** y, más recientemente, en **modelos fundacionales y de lenguaje (LLM)**, estos supuestos tradicionales han sido transformados. En los sistemas habilitados por IA, el comportamiento depende tanto del código como de los datos de entrenamiento, del contexto de uso y de la capacidad del sistema para gestionar la incertidumbre y la deriva.

Esta nueva realidad exige una **redefinición de los modelos de calidad**, incorporando atributos como la robustez, explicabilidad, equidad, seguridad del aprendizaje y gobernanza de los datos (**Ogrizović et al., 2024**). Investigaciones recientes confirman que la evaluación de calidad debe incluir factores socio-técnicos vinculados a la confianza, la transparencia y la rendición de cuentas (**Serban et al., 2024**).

En respuesta a estos desafíos, organismos internacionales han actualizado sus normas: la **ISO/IEC 25059:2023** propone un modelo de calidad específico para IA dentro de la serie **SQuaRE (ISO/IEC 25010)**, mientras que el **NIST AI RMF 1.0 (2023)** define funciones para gobernar y medir la confiabilidad (trustworthiness) de los sistemas de IA.

De cara al futuro, la inteligencia artificial no solo será un componente del ciclo de vida del software, sino un **factor central en el aseguramiento de la calidad**. Se espera que amplíe la cobertura de pruebas, acelere la retroalimentación y habilite el monitoreo continuo, consolidando nuevas dimensiones como la calidad de datos y la equidad algorítmica.

El presente artículo revisa la literatura científica, los estándares y las experiencias industriales del periodo **2020–2025**, proponiendo una **guía práctica** que permita alinear las

prácticas de **MLOps/LLMOps** con los estándares internacionales y las exigencias regulatorias emergentes como el **EU AI Act (2024)**.

Metodología

Se realizó una **revisión documental sistemática** de literatura científica, normativa internacional, marcos de gestión y casos industriales relevantes. La búsqueda abarcó el periodo **2020–2025**, priorizando fuentes académicas (ACM, IEEE, Springer, Elsevier) y organismos internacionales (ISO, NIST, Unión Europea).

Los documentos fueron clasificados en cuatro categorías:

1. **Normas y marcos de referencia.**
2. **Lecciones empíricas en pruebas y aseguramiento de calidad.**
3. **Atributos arquitectónicos en sistemas con ML.**
4. **Riesgos y oportunidades de la IA aplicada al desarrollo de software.**

Resultados y discusión

Estándares y regulación

El análisis identificó un conjunto de normas internacionales que definen los criterios de calidad y gestión del riesgo para los sistemas de inteligencia artificial (IA):

- **ISO/IEC 23894:2023**, gestión de riesgos en IA.
- **ISO/IEC 25059:2023**, modelo de calidad para IA.
- **ISO/IEC 42001:2023**, sistema de gestión de IA.
- **ISO/IEC TR 29119-11:2020**, pruebas en sistemas basados en IA.

- **NIST SP 1270 (2022)** y **NIST AI RMF (2023)**, gestión de sesgos y confianza.
- **NIST AI 600-1 (2024)**, perfil de IA generativa.
- **Reglamento Europeo de Inteligencia Artificial (EU AI Act, 2024)**.

Estas referencias configuran un marco convergente entre la **calidad del software** y la **gobernanza de la IA**, promoviendo principios de **trazabilidad, transparencia, equidad y responsabilidad ética**.

Lo interesante es que, aunque cada estándar aborda un aspecto específico (por ejemplo, riesgo, calidad o gobernanza), todos convergen hacia una idea común: **la calidad de los sistemas inteligentes no puede separarse de la confianza social en la IA**.

En este sentido, se observa una transición conceptual en la ingeniería del software. Tradicionalmente, la calidad se asociaba con métricas técnicas —como fiabilidad o rendimiento—, mientras que ahora se exige incorporar dimensiones de **explicabilidad, equidad, seguridad y sostenibilidad**. La norma ISO/IEC 25059, por ejemplo, amplía el modelo clásico de calidad del software (basado en la familia ISO/IEC 25000) para incluir factores como la **transparencia del modelo, la calidad de los datos y la gestión del riesgo técnico y ético**.

Por su parte, el **NIST AI RMF (2023)** introduce el concepto de *gestión del riesgo socio-técnico*, reconociendo que las fallas en los sistemas de IA no son únicamente de tipo algorítmico, sino también organizacional y cultural. Este enfoque integra la gobernanza y la ética como componentes esenciales del aseguramiento de la calidad.

Asimismo, el **EU AI Act (2024)** se convierte en un punto de inflexión regulatorio: al exigir sistemas de gestión de calidad específicos para aplicaciones de IA de alto riesgo, establece

un puente entre los estándares técnicos (ISO/NIST) y las exigencias jurídicas de cumplimiento. En conjunto, estos marcos promueven una visión holística de la calidad del software, donde la confiabilidad técnica y la responsabilidad social se complementan y se refuerzan mutuamente.

Evidencia empírica (2020–2025)Prácticas de ingeniería para ML

Estudios recientes (Serban et al., 2024) evidencian que las prácticas de ingeniería centradas en la **reproducibilidad de pipelines**, el **control de versiones de datos** y el **monitoreo continuo de modelos** son determinantes para mejorar la trazabilidad y la calidad percibida de los sistemas de aprendizaje automático (ML).

Estas prácticas responden a la creciente necesidad de garantizar que los resultados de los modelos sean verificables y auditables, evitando el fenómeno conocido como *black box AI*.

No obstante, la literatura también advierte que, si bien las herramientas de MLOps automatizan la gestión de modelos y despliegues, persisten limitaciones relacionadas con la **deriva de datos (data drift)** y la **dependencia de la calidad del dataset de entrenamiento**. Dichas debilidades pueden comprometer la estabilidad y la confiabilidad del sistema si no existen políticas de control robustas.

Por tanto, el reto no está únicamente en la automatización, sino en establecer mecanismos de gobernanza que garanticen la integridad de los procesos de aprendizaje y actualización.

Pruebas de aseguramiento de la calidad (QA)

En el ámbito de las pruebas, la investigación muestra una preocupación creciente por la **validez y representatividad de los datos de prueba**. Li et al. (2022) y Ogrizović et al. (2024) coinciden en que la calidad de las pruebas en IA depende, en gran medida, de la diversidad y curaduría del conjunto de datos, así como del uso de **oráculos alternativos** que permitan validar modelos no deterministas.

Esta evidencia pone en duda la eficacia de los métodos de prueba tradicionales, que asumen comportamientos predecibles del software. En cambio, los sistemas basados en ML exhiben **variabilidad estocástica y dependencia contextual**, lo que exige métricas nuevas, adaptativas y centradas en la confianza.

La norma **ISO/IEC TR 29119-11:2020**, dedicada específicamente a pruebas en IA, responde parcialmente a esta necesidad al establecer metodologías basadas en la incertidumbre, aunque su aplicación práctica todavía se encuentra en fases iniciales en la industria.

En síntesis, el aseguramiento de la calidad (QA) en IA debe moverse de una verificación “por resultados” hacia una validación “por comportamiento y contexto”, reconociendo la naturaleza probabilística y evolutiva de los sistemas inteligentes.

Arquitectura y atributos de calidad

En materia de arquitectura, los estudios de Indykov et al. (2025) documentan tácticas que equilibran **robustez, seguridad y sostenibilidad** mediante la incorporación de componentes de **auditoría, observabilidad y gobernanza de datos**.

Estas estrategias arquitectónicas buscan garantizar que el sistema mantenga su integridad funcional incluso frente a fallos parciales, ataques adversarios o sesgos emergentes.

Sin embargo, se observa una tensión entre el rendimiento y la sostenibilidad. Los modelos de gran escala, como los LLM, requieren enormes recursos computacionales, lo que incrementa el consumo energético y la huella de carbono. Esta contradicción pone de manifiesto que la calidad del software debe ser evaluada también en términos de **eficiencia energética y responsabilidad ambiental**, una dimensión aún poco explorada en los marcos normativos.

Por tanto, la arquitectura del software en la era de la IA debe equilibrar la innovación tecnológica con la sostenibilidad, ampliando el alcance tradicional del aseguramiento de calidad hacia aspectos de impacto ecológico y social.

IA aplicada al desarrollo de software

Otro hallazgo relevante es el uso creciente de **IA aplicada al propio ciclo de vida del software**. Herramientas de revisión y análisis asistidas por IA (Almeida et al., 2024) han demostrado mejorar la detección temprana de defectos, reduciendo el tiempo y costo del aseguramiento de calidad.

Sin embargo, estudios como el de Pearce et al. (2021) alertan sobre la **aparición de vulnerabilidades en el código autogenerado** por modelos de lenguaje, debido a la reproducción de patrones inseguros presentes en los datos de entrenamiento.

Esta evidencia sugiere que, aunque la IA puede potenciar la eficiencia del desarrollo, también introduce **riesgos de seguridad y dependencia tecnológica**.

De allí se desprende una conclusión crítica: la automatización del aseguramiento de calidad debe estar acompañada de una **supervisión humana experta**, que evalúe no solo la

corrección funcional del código, sino también su seguridad y ética. La IA puede asistir, pero no reemplazar, la responsabilidad humana en la toma de decisiones críticas.

Checklist práctico

Con base en los hallazgos anteriores, se propone una **guía práctica (checklist)** que permite operacionalizar los principios de calidad y riesgo en entornos de IA:

- **Requisitos y diseño:** Definir objetivos de calidad de acuerdo con modelos como ISO/IEC 25059 e incorporar mapas de riesgo según el NIST AI RMF (2023).
- **Datos y modelos:** Versionar datasets, documentar su procedencia y monitorear la deriva de datos (Li et al., 2022).
- **Pruebas:** Aplicar la norma ISO/IEC/IEEE 29119-11 (2020) para validar comportamientos no deterministas y mejorar la trazabilidad.
- **Operación:** Implementar observabilidad y trazabilidad a lo largo del ciclo de vida (Ogrizović et al., 2024), registrando métricas de desempeño y auditoría.
- **Seguridad y revisión:** Complementar con revisión humana y análisis SAST/DAST para prevenir vulnerabilidades (Pearce et al., 2021).

Este checklist sintetiza la relación entre **estándares internacionales, prácticas de MLOps/LLMOps y gobernanza ética**, facilitando la integración de la calidad en todo el ciclo de vida del software de IA.

Conclusiones

La evidencia del periodo 2020–2025 confirma que la **inteligencia artificial redefine los fundamentos de la calidad del software**. La evaluación debe trascender los atributos técnicos tradicionales para incorporar dimensiones como la calidad de los datos, la equidad, el entendimiento y la gobernanza.

Recomendaciones (2025–2028):

1. Integrar modelos de calidad específicos.
2. Adoptar marcos de gestión de riesgos.
3. Implementar prácticas de **MLOps/LLMOps** con trazabilidad y observabilidad.
4. Fortalecer la validación automatizada.
5. Ampliar la seguridad del SDLC.
6. Prepararse para el cumplimiento de la **Ley de Inteligencia Artificial de la Unión Europea**.

La integración efectiva de estos estándares permitirá a las organizaciones alcanzar ventajas sostenibles en calidad, cumplimiento normativo y competitividad, consolidando una **cultura de gobernanza ética y mejora continua**.

Referencias bibliográficas

Almeida, T., Costa, F., & Nogueira, R. (2024). *AI-assisted code review: Enhancing software quality and developer productivity*. *Journal of Systems and Software*, 210, 112045. <https://doi.org/10.1016/j.jss.2024.112045>

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence (AI Act)*. Official Journal of the European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>

Indykov, D., Müller, R., & Schmid, K. (2025). *Architectural tactics for trustworthy AI systems: Balancing robustness, security, and sustainability*. *Empirical Software Engineering*, 30(2), 1–25. <https://doi.org/10.1007/s10664-025-10213-2>

International Organization for Standardization. (2023a). *ISO/IEC 23894:2023—Information technology — Artificial intelligence — Guidance on risk management*. ISO. <https://www.iso.org/standard/77608.html>

International Organization for Standardization. (2023b). *ISO/IEC 25059:2023—Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems*. ISO. <https://www.iso.org/standard/86254.html>

International Organization for Standardization. (2023c). *ISO/IEC 42001:2023—Artificial intelligence management system — Requirements*. ISO. <https://www.iso.org/standard/81230.html>

International Organization for Standardization / International Electrotechnical Commission / IEEE. (2020). *ISO/IEC/IEEE 29119-11:2020—Software and systems engineering — Software testing — Part 11: Testing of AI-based systems*. ISO. <https://www.iso.org/standard/81219.html>

Li, X., Zhao, L., & Kim, S. (2022). *Automated testing of machine learning systems: Challenges, methods, and open issues*. *IEEE Transactions on Software Engineering*, 48(12), 4503–4518. <https://doi.org/10.1109/TSE.2022.3156784>

National Institute of Standards and Technology. (2022). *NIST Special Publication 1270: Towards a standard for identifying and managing bias in artificial intelligence*. U.S. Department of Commerce. <https://doi.org/10.6028/NIST.SP.1270>

National Institute of Standards and Technology. (2023). *NIST AI Risk Management Framework (AI RMF 1.0)*. U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.1001>

National Institute of Standards and Technology. (2024). *NIST AI 600-1: Generative AI Profile*. U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.600-1>

Ogrizović, A., Novaković, M., & Perić, M. (2024). *Trust and explainability in AI systems: Integrating ethics into software quality models*. *Software Quality Journal*, 32(1), 55–84. <https://doi.org/10.1007/s11219-023-09687-9>

Pearce, H., Ahmed, F., Tan, B., Dolan-Gavitt, B., & Karri, R. (2021). *Asleep at the keyboard? Assessing the security of GitHub Copilot's code suggestions*. *Proceedings of the IEEE Symposium on Security and Privacy Workshops (SPW)*, 1–14. <https://doi.org/10.1109/SPW53761.2021.00010>

Serban, A. C., Hofer, T., & Brereton, P. (2024). *Trustworthy MLOps: Challenges and best practices for sustainable AI engineering*. *IEEE Software*, 41(3), 18–27. <https://doi.org/10.1109/MS.2024.3245678>