

Entorno para la Evaluación de la Adecuación Funcional en Sistemas IA

Jesús Oviedo Lama^{1,2}, Moisés Rodríguez Monje^{1,2}, Mario Piattini Velthuis²

¹AQCLab Software Quality, Puertollano, España

²Instituto de Tecnologías y Sistemas de Información y Escuela Superior de Informática, Universidad de Castilla-La Mancha Ciudad Real, España

Abstract. *Artificial Intelligence is currently a fundamental part of the digital transformation of organizations and the appearance of applications with AI algorithms means that their impact on everyday activities is increasing every day. In this situation, there is a growing need for these AI systems, as software systems, to have the necessary quality parameters to guarantee their use. This article presents a methodological and technological environment for the measurement and evaluation of Functional Suitability in AI Systems. The environment proposes a set of metrics and quality properties aligned with the new ISO/IEC 25059 standard. It also has an assessment methodology aligned with ISO/IEC 25040. On the other hand, a set of automatic tools to facilitate the evaluation of the quality of Functional in IA Systems.*

Resumen. *En la actualidad la Inteligencia Artificial es una parte fundamental de la transformación digital de las organizaciones y la aparición de aplicaciones con algoritmos de IA hace que su incidencia en las actividades cotidianas sea cada día mayor. Ante esta situación, existe una necesidad de que estos Sistemas IA, como sistemas software que son, cuenten con los parámetros de calidad necesarios para garantizar su uso. En este artículo se presenta un entorno metodológico y tecnológico para la medición y la evaluación de la Adecuación Funcional en Sistemas IA. El entorno propone un conjunto de métricas y propiedades de calidad alineadas con el nuevo estándar ISO/IEC 25059. También cuenta con una metodología de evaluación alineada con la norma ISO/IEC 25040. Por otro lado, se ha desarrollado un conjunto de herramientas automáticas para facilitar la evaluación de la Adecuación Funcional en Sistemas IA.*

1. Introducción

Históricamente las comunidades de Ingeniería del Software y de Inteligencia Artificial han permanecido poco comunicadas, teniendo sus propias terminologías, métodos y desafíos. A nivel de I+D, es más habitual que los investigadores se centren en la “**IA para la IS**” es decir cómo utilizar la Inteligencia Artificial para mejorar algunos aspectos de las Ingeniería del Software, como pueden ser los requisitos, la estimación, pruebas, etc. [Romero et al. 2023], que en usar la Ingeniería del Software para mejorar la IA, lo que se ha llamado la “**IS para la IA**”. Sin embargo, en los últimos tres años ha aumentado las investigaciones sobre técnicas de Ingeniería del Software para construir sistemas inteligentes con una calidad y productividad adecuadas, esta nueva área de investigación como indica el SEI se podría denominar “Ingeniería de la Inteligencia Artificial ¹”, que se

¹<https://www.sei.cmu.edu/our-work/artificial-intelligence-engineering/>

fundamenta en los principios establecidos por Horneman [Horneman et al. 2019], al que podemos denominar el padre de esta nueva área, al considerar que los sistemas de IA son sistemas intensivos, y que, por lo tanto:

- “Los principios establecidos para diseñar y desplegar sistemas de software de calidad que cumplan los objetivos de su misión a tiempo se deben aplicar a los sistemas de ingeniería de IA”.
- “Los equipos deben esforzarse por entregar la funcionalidad a tiempo y con calidad, diseñar para los requisitos arquitectónicos importantes (como la seguridad, la usabilidad, la fiabilidad, el rendimiento y la escalabilidad) y planificar el mantenimiento del sistema durante toda su vida útil”.

Si bien, a pesar de que “La inteligencia artificial (IA) y el aprendizaje automático (ML) son omnipresentes en el panorama actual de la informática, todavía existe una falta de experiencia en ingeniería de software y mejores prácticas en este campo” [Van Oort et al. 2021]. Los desarrolladores de sistemas intensivos en IA son conscientes de la necesidad de organizar el desarrollo de una forma sistemática aplicando métodos y herramientas de ingeniería, pero también reconocen que los métodos y herramientas tradicionales de ingeniería del software no son adecuados ni suficientes por sí mismos y deben adaptarse y ampliarse [Lavazza and Morasca 2021].

Por todo ello, en la nueva área de la Ingeniería de la Inteligencia Artificial, han surgido retos relacionados con la definición y la garantía de los atributos de comportamiento y calidad de estos sistemas y aplicaciones, que nos han impulsado a trabajar en el desarrollo de un entorno que permita la Evaluación de la Adecuación Funcional en los Sistemas IA. Por tanto, el objetivo principal del presente artículo es presentar los componentes de una primera versión de dicho entorno de evaluación, que serán utilizados próximamente de manera práctica para su validación y mejora. Para ello, el resto del artículo se estructura de la siguiente manera: en el apartado 2 se presenta el nuevo estándar para la evaluación de la calidad en los Sistemas IA como producto. En el apartado 3 se describe el modelo de calidad que se ha definido para la evaluación de la Adecuación Funcional en los Sistemas IA. En el apartado 4 se presenta cómo se lleva a cabo el proceso de evaluación del producto. En el apartado 5 se expone el conjunto de herramientas empleadas para llevar a cabo la evaluación y presentación de los resultados. Y finalmente, el apartado 6 presenta las conclusiones obtenidas con este trabajo y las líneas futuras de trabajo e investigación.

2. Nuevo estándar ISO/IEC 25059 para la Evaluación de la Calidad en los Sistemas IA

El aseguramiento de la calidad de los sistemas basados en la IA es un área emergente que no ha sido bien explorada aún y que requiere la colaboración entre las comunidades de investigación de la Ingeniería del Software (SE) y de la Inteligencia Artificial (IA). Recientemente, como ha sucedido en otros casos (sostenibilidad de productos software [Calero et al. 2021], servicios cloud [Navas 2016], sistemas de e-learning [Rahman 2019], etc.), ha sido necesario adaptar el modelo de calidad definido en la ISO/IEC 25010 para el producto software a las características especiales de la Inteligencia Artificial, surgiendo así la nueva norma ISO/IEC 25059 [ISO/IEC 2023b] conforme al marco de las normas ISO/IEC 25000. Este nuevo estándar para la “calidad para sistemas basados en IA” propone un nuevo modelo de calidad donde se recogen las características de calidad de producto que debe cumplir cualquier Sistema de IA, basándose en las

Adecuación Funcional <ul style="list-style-type: none"> • Completitud Funcional • Corrección Funcional • Pertinencia Funcional • Adaptabilidad Funcional 	Seguridad <ul style="list-style-type: none"> • Confidencialidad • Integridad • No Repudio • Responsabilidad • Autenticidad • Intervenibilidad 	Usabilidad <ul style="list-style-type: none"> • Intelegibilidad • Aprendizaje • Operabilidad • Protección a Errores • Atraktividad • Accesibilidad • Capacidad de Control • Transparencia 	Fiabilidad <ul style="list-style-type: none"> • Madurez • Disponibilidad • Tolerancia a Fallos • Capacidad de Recuperación • Robustez
Rendimiento <ul style="list-style-type: none"> • Comportamiento Temporal • Utilización de Recurso • Capacidad 	Mantenibilidad <ul style="list-style-type: none"> • Modularidad • Reusabilidad • Analizabilidad • Capacidad de Ser Modificado • Capacidad de Ser Probado 	Portabilidad <ul style="list-style-type: none"> • Adaptabilidad • Facilidad de Instalación • Intercambiable 	Compatibilidad <ul style="list-style-type: none"> • Coexistencia • Interoperabilidad

Figura 1. Modelo de calidad para sistemas IA ISO/IEC 25059 (en negrita la subcaracterísticas añadidas o modificada)

propuestas por la ISO/IEC 25010, pero presentando ciertos cambios y adaptaciones, en algunas de las características y subcaracterísticas [Oviedo et al. 2023]. En la Figura 1 se pueden ver las características y subcaracterísticas propuestas para el modelo de calidad de producto de sistema de inteligencia artificial, donde se ha resaltado en negrita las Subcaracterísticas que han sido añadidas o modificadas.

Si bien, este nuevo estándar supone un punto de partida importante, la propuesta realizada por ISO/IEC 25059 es alto nivel y está incompleta, siendo necesario para poder ser aplicable en la industria, definir un conjunto de métricas y umbrales que nos permitan determinar los valores de calidad de cada una de las subcaracterísticas y características definidas. En los siguientes apartados se describe el trabajo realizado para disponer de cada uno de los elementos necesarios para llevar a cabo la evaluación de la Adecuación Funcional para los Sistemas IA.

3. Modelo de Calidad de Adecuación Funcional para Sistemas IA

El modelo de adecuación funcional definido para los sistemas de IA está basado en el estándar ISO/IEC 25059, siendo la adecuación funcional una de las características que se han visto modificadas en este nuevo estándar. Estos cambios han motivado la necesidad de definir un nuevo modelo para abordar la calidad de los Sistemas IA, no pudiendo emplear el modelo de calidad definido por AQCLab [Verdugo et al. 2024] para la evaluación de la adecuación funcional en los productos software [Rodríguez et al. 2016], basado en la ISO/IEC 25010, siendo necesario abordar el desarrollo de un nuevo modelo que pueda evaluar la adecuación funcional para los sistemas de IA.

Este nuevo modelo propuesto en el estándar ISO/IEC 25059 está basado en el estándar ISO/IEC 25010 y presenta el mismo enfoque jerárquico donde existen diferentes niveles. En los niveles superiores del modelo se encuentran la **Característica** y **Subcaracterísticas de Calidad** que son definidas por el estándar. Sin embargo, el modelo está

incompleto y ha sido necesario concretar los atributos de calidad para los niveles inferiores para hacer operativo el modelo. Así se ha definido un conjunto de Medidas de Calidad y Elementos de Medida que se han denominado en este modelo como **Propiedades de Calidad y Métricas Base y Derivadas**, las cuales son los elementos más concretos del modelo.

En el nivel inferior del modelo está formado por el conjunto de Métricas Base y Métricas Derivadas, las cuales son los elementos más concretos al obtenerse como resultado de las mediciones realizadas sobre el Sistema de IA y las métricas derivadas se obtienen a partir de las métricas base mediante la aplicación de alguna operación (suma, multiplicación, etc.) entre ellas.

A partir de estas métricas derivadas se calcula el siguiente nivel superior, las Propiedades de Calidad, tomando valor entre 0 y 100, las cuales nos permiten determinar el valor de las Subcaracterísticas de calidad, que también toman un valor entre 0 y 100. Por último, siguiendo el modelo jerárquico, se obtendría el valor de la Adecuación Funcional, Característica de calidad, tomando un valor discreto entre 1 y 5.

La Adecuación Funcional en el estándar ISO/IEC 25059 se define como el grado con el que un Sistema de IA proporciona funciones que satisfacen necesidades declaradas e implícitas cuando se usa con las condiciones especificadas y es evaluada tomando como base la evaluación de sus Subcaracterísticas (Figura 1) empleando una función por perfiles como la presentada en [Rodríguez et al. 2016].

3.1. Evaluación de las Subcaracterísticas de Adecuación Funcional

El modelo calidad que se ha definido ha tomado como base las subcaracterísticas definida en el estándar ISO/IEC 25059, que son las siguientes:

- **Completitud Funcional:** Esta subcaracterística no ha sido modificada respecto al modelo de calidad ISO/IEC 25010, y por tanto la definición realizada en el modelo de la evaluación de la calidad del producto software por AQCLAB [Rodríguez et al. 2016] es válida para la evaluación de la Completitud Funcional de los Sistemas IA. Si bien, esta subcaracterística se define como el grado en el que el conjunto de funciones implementadas en el producto o sistema cubren todos los requisitos funcionales y objetivos de los usuarios.
- **Corrección Funcional:** Esta Subcaracterística ha sido modificada en el estándar ISO/IEC 25059, esta modificación viene justificada porque en los Sistemas IA la corrección es un concepto más amplio que en el software tradicional [Lavazza and Morasca 2021], debido a que en los sistemas de IA los resultados correctos no se producen en la totalidad de los casos, por lo que también se espera un porcentaje de resultados incorrectos. Por ello, en la evaluación de la corrección funcional se debe tener en cuenta los resultados correctos e incorrectos.
- **Pertinencia Funcional:** Esta subcaracterística no ha sido modificada en el nuevo estándar ISO/IEC 25059. Por ello, la definición realizada en el modelo de la evaluación de la calidad del producto software por AQCLAB [Rodríguez et al. 2016] es válida para evaluar la Pertinencia Funcional de los Sistemas de IA. Así, esta subcaracterística se define cómo el grado en el que las funciones del producto o sistema facilitan el cumplimiento de las tareas y objetivos especificados.
- **Adaptabilidad Funcional:** Es la nueva subcaracterística definida en el estándar ISO/IEC 25059 para la Adecuación Funcional. Esta subcaracterística se define

como el grado en que un sistema de IA puede realizar nuevas predicciones en base a nuevo conocimiento adquirido. Esta subcaracterística presenta una diferencia significativa respecto a otras subcaracterísticas, ya que un aumento de su valor no implica una mejora de la calidad del sistema, es decir, no se puede interpretar utilizando una escala lineal. Por ello, en este modelo, se ha realizado una interpretación de la Adaptabilidad Funcional para permitir la linealidad, como en el resto de subcaracterísticas, y sólo se tiene en cuenta si la Adaptabilidad Funcional del sistema de IA mejora los resultados de éste. Así, se define la adaptabilidad funcional como la capacidad que tiene el sistema IA para mejorar sus resultados en base de su experiencia, es decir en base de nuevas entradas y/o resultados previos.

La evaluación de cada una de las subcaracterísticas se realiza a partir de un conjunto de Propiedades de Calidad. Si bien, en este artículo sólo se van a presentar las Propiedades de Calidad de las Subcaracterísticas nuevas o modificadas, ya que el resto de propiedades no presentan cambios respecto el modelo de evaluación de la adecuación funcional para el producto software de AQCLAB [Rodríguez et al. 2016].

3.2. Propiedad para evaluar la Subcaracterística Corrección Funcional

La Propiedad **Exactitud Funcional** evalúa si los requisitos que han sido implementados son exactos, es decir si su tasa de éxito es igual o superior a la tasa de éxito establecida para dicho requisito. Para conocer si un requisito es exacto es necesario conocer el porcentaje de ejecuciones correctas e incorrectas, teniendo en cuenta los falsos positivos y los falsos negativos [ISO/IEC 2023b] y que la proporción de ejecuciones correctas sea superior al umbral establecido para dicho requisito para considerar que sus resultados son exactos. Para calcular el valor de esta Propiedad es necesario conocer el número de requisitos implementados y el número de requisitos exactos.

Para determinar si un requisito es exacto se debe ejecutar varias veces cada uno de los casos de prueba asociados al requisito. El número de ejecuciones se determina mediante la formula que se muestra en la Figura 2, donde los valores asignados son:

- Los valores de p y q es 0,5 al desconocer la probabilidad del hecho (ejecución correcta o ejecución incorrecta).
- El valor de e se corresponde con el error que se quiera aceptar: 1 %, 5 %, 10 %, etc.
- El valor de Z ha sido obtenido a partir de las tablas de distribución normal, en función de los errores seleccionados. Se ha tomado una distribución normal, porque el número de veces que se puede ejecutar una funcionalidad no está limitado. Además, el resultado de la ejecución puede tratarse como una distribución nominal, el resultado de la funcionalidad puede ser correcto o incorrecto.

$$n = \frac{Z_{\sigma/2}^2 * p * q}{e^2}$$

Figura 2. Ecuación para determinar el tamaño de la muestra

Tras ejecutar el caso de prueba el número de veces determinado se obtendrá el número de ejecuciones que han sido correctas, considerando los verdaderos positivos y los verdaderos negativos. Si el porcentaje de resultados correctos en todos los casos de

prueba trazados con el requisito es igual o superior a la tasa de éxito establecida en la especificación del requisito entonces se podrá considerar que el requisito es exacto.

3.3. Propiedad para evaluar la Subcaracterística Adaptabilidad Funcional

La Propiedad **Aprendizaje Adecuado** examina si la utilización de nuevas entradas de datos genera mejores resultados por parte del modelo de IA, es decir si se reduce la probabilidad de resultados incorrectos. Para el análisis de esta propiedad es necesario conocer el número de requisitos que han mejorado su tasa de resultados correctos y el número de requisitos implementados. Para determinar si un requisito ha mejorado sus resultados correctos, se debe comprobar, tras la evaluación de la corrección funcional, si el porcentaje de resultados correctos del requisito es superior al umbral de corrección definido en la especificación del requisito entonces se podrá considerar que la tasa de resultados correctos del requisito ha mejorado.

4. Proceso de Evaluación

Para abordar la evaluación de la adecuación funcional de los sistemas IA se ha definido un proceso basado en el estándar ISO/IEC 25040 [ISO/IEC 2023a]. Sin embargo, existen ciertas matizaciones en el procedimiento definido con respecto al estándar. Concretamente, en el caso del proceso de la evaluación de la adecuación funcional de los sistemas de inteligencia artificial el objetivo es siempre el mismo y por ello se utiliza un módulo predefinido, lo que supone que las actividades iniciales del proceso, donde se establece el propósito de la evaluación y se define el módulo de evaluación a emplear requieran menos profundidad que la establecida en la norma ISO/IEC 25040. Lo mismo sucede en la actividad de planificación de la evaluación, ya que no es necesario identificar las tareas a realizar dado que la evaluación de la adecuación funcional se compone siempre de las mismas tareas. El resto de las actividades (ejecución y conclusión de la evaluación) se realizan en base a lo establecido en dicha norma. En la Figura 3 se muestran las actividades y tareas que componen este proceso de evaluación.

5. Herramientas de Evaluación

El entorno software para la medición y evaluación de la adecuación funcional en sistemas de IA implementado (en adelante entorno AQCLab) se encuentra compuesto fundamentalmente por cuatro componentes:

- **Herramienta de Medición** se trata de una herramienta desarrollada en JAVA para obtener métricas relacionadas con la evaluación de la adecuación funcional de los Sistemas de IA. Permite registrar información relacionada con los requisitos y casos de prueba del sistema genera un fichero XML con los resultados de las métricas, que es utilizado como entrada por la herramienta de evaluación.
- **Herramienta de Evaluación** se encarga de medir y evaluar la adecuación funcional de los sistemas IA, se compone de dos componentes, un componente estándar de evaluación y un componente específico para evaluar la adecuación funcional de los sistemas IA. Este componente específico incluye el modelo para evaluar la adecuación funcional implementado en XML y un programa implementado en Python que realiza los cálculos necesarios para llevar a cabo la evaluación a partir de los resultados de la herramienta de medición.

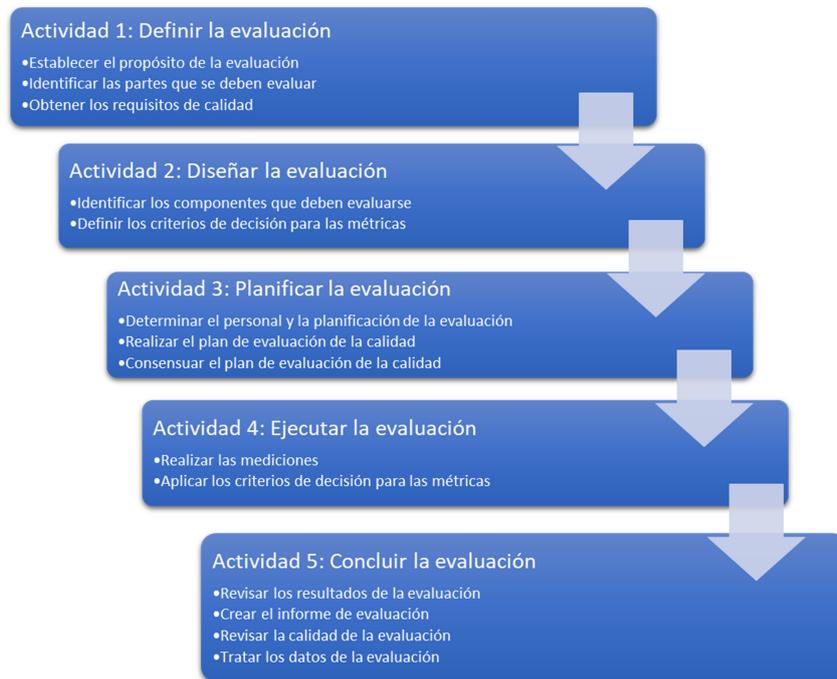


Figura 3. Proceso de evaluación de la adecuación funcional en sistemas IA

- **Herramienta de Visualización.** se trata de una herramienta formada por: un entorno de visualización proporcionado por AQCLab en PowerBI, para la creación de diferentes informes interactivos y una ampliación específica dedicada la visualización y análisis de datos de la adecuación funciona en Sistemas IA.
- **Base de Datos** el entorno se compone de una base de datos gestionada mediante PostgreSQL, donde se almacenan los resultados obtenidos tras la evaluación de la adecuación funcional de los sistemas de IA.

6. Conclusiones y Trabajo Futuro

El auge de la Inteligencia Artificial en todas las facetas de la vida cotidiana hace necesario, al igual que en el software, el empleo de técnicas de ingeniería del software, como es la Calidad, para el desarrollo de estos sistemas. Si bien, los métodos y herramientas tradicionales de la ingeniería del software deben ser adaptados a las particularidades de los sistemas de IA. Por ello, en este artículo se ha presentado un nuevo entorno para la evaluación de la adecuación funcional en sistemas IA, adaptado a las particularidades de esta nueva tecnología.

En este nuevo entorno de evaluación se ha definido un nuevo modelo de calidad alineado al nuevo estándar ISO/IEC 25059 donde se han analizado nuevas métricas con las que abordar la evaluación de los sistemas IA. También ha sido necesario adaptar el proceso de evaluación definido en la ISO/IEC 25040 a las particularidades de las actividades que se deben realizar en la evaluación de la adecuación funcional de un sistema de IA. Por último, se ha desarrollado una nueva herramienta de medición para obtener las nuevas métricas y se han realizado adaptaciones en las herramientas de evaluación y visualización de AQCLab para realizar la evaluación y la presentación de los resultados

Por tanto, una vez desarrollado este entorno de evaluación se nos plantea como línea principal de trabajo futuro el estudio de la aplicación práctica de dicho entorno, a través de evaluaciones de sistemas de IA. Además, esta aplicación práctica permitirá realizar la validación del entorno asegurando que los valores obtenidos en las evaluaciones presentan una correlación con la Adecuación Funcional de los Sistemas de IA.

7. Agradecimientos

Esta investigación ha contado con el apoyo de ADAGIO (SBPLY/21/180501/000061), financiado por la Junta de Comunidades de Castilla La Mancha, Unión Europea, Ministerio de Hacienda y Función Pública y FEDER (Programa Operativo de Castilla-La Mancha 2021-2027), UNION (2022-GRIN-34110), Ayudas para la realización de proyectos de investigación aplicada, en el marco del Plan Propio de Investigación, cofinanciadas en un 85 % por FEDER y CASIA financiado por la Junta de Comunidades de Castilla La Mancha.

Referencias

- Calero, C., Moraga, M. Á., and Piattini, M. (2021). *Software Sustainability*. Springer.
- Horneman, A., Mellinger, A., and Ozkaya, I. (2019). Ai engineering: 11 foundational practices-recommendations for decision makers from experts in software engineering, cybersecurity, and applied artificial intelligence white paper dm19-0624, 06.06.
- ISO/IEC (2023a). Iso/iec 25040:2023. In *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Evaluation process*.
- ISO/IEC (2023b). Iso/iec 25059:2023. In *Software engineering - Systems and software Quality Requirements and Evaluations (Square) - Quality model for AI systems*.
- Lavazza, L. and Morasca, S. (2021). Understanding and modeling ai-intensive system development. In *2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN)*, pages 55–61. IEEE.
- Navas, R. (2016). Modelo de calidad para servicios cloud.
- Oviedo, J., Márquez, R., Rodríguez, M., and Piattini, M. (2023). Calidad en los sistemas ia: adaptación de modelos de procesos y productos. *Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos*.
- Rahman, A. (2019). Quality consideration for e-learning system based on iso/iec 25000 quality standard.
- Rodríguez, M., Oviedo, J. R., and Piattini, M. (2016). Evaluation of software product functional suitability: A case study. *Software Quality Professional*, 18(3).
- Romero, J., Medina-Bulo, I., and Chicano, F. (2023). *Optimising the Software Development Process with Artificial Intelligence*. Springer.
- Van Oort, B., Cruz, L., Aniche, M., and Van Deursen, A. (2021). The prevalence of code smells in machine learning projects. In *2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN)*, pages 1–8. IEEE.
- Verdugo, J., Oviedo, J., Rodríguez, M., and Piattini, M. (2024). Connecting research and practice for software product quality evaluation and certification: a software laboratory's 25-year journey. *IEEE Software*.